Contribution à une cartographie de la chimiodiversité du vivant

Guillaume Marti

Habilitation à Diriger des Recherches

Présentée et soutenue le 30 avril 2024 par : Guillaume Marti

Contribution à une cartographie de la chimiodiversité du vivant

JURY

Guillaume Tcherkez	Professeur d'Université
Didier Stien `	Directeur de Recherche
Pierre Petriacq	Maitre de Conférences
Jean Charles Portais	Professeur d'Université
Veronique Eparvier	Directrice de Recherche
Fabien Jourdan	Directeur de Recherche

École doctorale et spécialité :
SEVAB : Sciences Ecologiques, Vétérinaires, *Agronomiques et Bioingénieuries*

Unité de Recherche :
Laboratoire de Recherche en Sciences Végétales (UMR 5546)

Direction des Etudes et de la Vie de l'Etudiant
3e cycle sciences

CURRICULUM VITAE

Candidat:

ECOLE DOCTORALE DE RATTACHEMENT : CEVAB

Nom :Marti Prénom :Guillaume
Nom d'usage : Marti
Né le :12/02/1978
Fonctions :MCU
Doctorat obtenu le 5 Mai 2009 à Paris (MNHN)
Mail : guillaume.marti@univ-tlse3.fr
Adresse : Faculté de Pharmacie, 35 chemin des maraichers, 31400 Toulouse

I. PARCOURS UNIVERSITAIRE

2006 - 2009 *Thèse de doctorat du Museum National d'Histoire Naturelle, Paris-France*
Laboratoire d'accueil: Institut de Chimie des Substances Naturelles.
Encadrants: Drs Marc Litaudon et Christian Moretti
Titre : Etude chimique et Biologique des espèces de la famille des Clusiaceae de Guyane Française
Soutenance : 5 Mai 2009

2004 - 2005 *Master de Recherche, Valorisation chimique de la biodiversité végétale Université des Antilles et de la Guyane, France.*

1999 - 2001 *Licence en chimie et biologie des plantes tropicales, Université de La Réunion, France.*

II. PARCOURS PROFESSIONNEL

J'ai intégré en septembre 2014 le département de pharmacie de l'**Université Toulouse 3**-Paul Sabatier en tant que **Maitre de conférences universitaire**. Mes activités de recherche ont débuté au sein de l'UMR152-Pharmadev (IRD/UPS), elles concernaient l'utilisation d'approches innovantes en chimie analytique pour la découverte de substances naturelles bioactives. En 2020, j'ai intégré le Laboratoire de Recherche en Sciences Végétales (UMR-5546, CNRS/UPS) afin de de réorienter mes activités de recherche en biologie végétale et plus particulièrement sur l'étude et la valorisation de la chimiodiversité du vivant. Ainsi j'ai pris la **direction de l'équipe et de la plateforme Metatoul-AgromiX.** J'ai donc pu intégrer également le réseau de plateformes toulousaines faisant partie du GIS de la Génotoul ainsi que le bureau exécutif de l'infrastructure nationale MetaboHub.
En sus de mes activités de recherche, j'effectue un service d'enseignement de 210 HETD en moyenne depuis ma prise de poste à l'université dans les domaines de la pharmacognosie, la chimie analytique, le contrôle qualité, et la botanique. Depuis 2021, j'interviens également dans plusieurs UEs de niveau **Master** en **métabolomique appliqué** auprès de la faculté des sciences et ingénierie (FSI) de l'Université Toulouse 3.

2020-présent	***Direction plateforme** Metabolomique: Metatoul-AgromiX (LRSV, UMR5546, CNRS/UT3).*
2018	*Obtention de la prime d'encadrement doctoral et de recherche (PEDR).*
2017-2021	*Chercheur associé à l'équipe IRIS de l'Institut de Recherche en Informatique de Toulouse (IRIT, UMR 5505, CNRS/UT3).*
2015-2020	***Membre élu de la commission recherche** de l'Université Toulouse 3-collège B*
2014-2020	*Activité de recherche au sein du laboratoire Pharmadev (Equipe du Pr. Nicolas Fabre).*
2009-2014	***Postdoctorat** en métabolique à l'Université de pharmacie de Genève (Suisse) au sein du laboratoire du Pr Jean-Luc Wolfender.*
2006-2009	***Thèse de doctorat** du Museum Nationale d'Histoire Naturelle de Paris en Phytochimie.*
	Laboratoire: Institut de Chimie des Substances Naturelles (ICSN-CNRS), équipe du professeure Françoise Guéritte.
	Co-encadrants: Drs Christian Moretti (DR-IRD), Marc Litaudon (IR-CNRS).
	Sujet: « Études chimiques et biologiques de Clusiaceae de la forêt équatoriale guyanaise ».
2014-présent	***Enseignement** en Pharmacognosie au sein de l'UFR Pharmacie-Université Toulouse 3.*
2010-2014	*Enseignant en Pharmacognosie à l'Université de pharmacie de Genève (Suisse) 90 HETD/an.*

III. ACTIVITES D'ENSEIGNEMENT

J'effectue en moyenne 210 HETD par an dans le service de Botanique-Mycologie-Pharmacognosie du département pharmacie depuis la rentrée 2014 ainsi que pour la FSI depuis 2020. Ces activités sont résumées dans le tableau 1.

Mes activités se concentrent sur l'enseignement pratique de la pharmacognosie (123 HETD Licence pro et DFGSP1 et 2) ainsi que la botanique (35 HETD DFGSP2 et DFASP1-officine). Dans le cadre des travaux pratiques, **j'ai modifié sensiblement les protocoles** inspirés de la pharmacopée européenne afin de les adapter aux objectifs pédagogiques (illustration des voies d'accès aux substances naturelles actives et contrôle de leurs qualités). J'ai ajouté une séquence préliminaire d'explication du protocole d'un point de vue fondamental en début de séance accompagnée d'une session finale de comparaison statistique des résultats de chaque groupe. Ces ajouts pédagogiques ont été très favorablement accueillis par les étudiants et les résultats des contrôles continus se sont sensiblement améliorés.

J'ai par ailleurs **la charge du Jardin botanique de la faculté de pharmacie**. À ce titre, j'encadre un technicien des espaces verts qui assure les travaux manuels quotidiens indispensables à ce type structure. À mon arrivée, nous avons hérité d'un jardin qui n'était plus entretenu depuis plusieurs années. Aujourd'hui ce jardin est propre et contient une serre à semis, un cabanon et rassemble une soixantaine d'espèces (toxiques, médicinales et ornementales). Des collaborations avec le Museum de Toulouse nous permettent d'augmenter notre collection chaque année. Ce jardin est un outil pédagogique important dont notre équipe se sert pour les 2ème, 3ème et 4ème années de pharmacie (TP, TD et CM), principalement pour la filière officine. Cet outil pédagogique est en constante évolution. Depuis la rentrée 2022, nous avons mis en place des parcours spécifiques (plantes toxiques, médicinales, aromatiques…) sous forme d'étiquettes pédagogiques rassemblant les informations essentielles à l'identification de chaque plante et de ses principes actifs accompagnés de liens sous forme de QR code vers l'outil « smartjardin » et « telabotanica ». L'objectif étant de faciliter les révisions par les étudiants en

autonomie et de pouvoir acquérir une compréhension intégrée entre familles/espèce botaniques et les grandes classes d'actifs chimiques associées.
J'effectue par ailleurs plusieurs cours magistraux (10 HETD) dans le cadre d'UEs liées à la phytochimie et la pharmacognosie en troisième et quatrième année de pharmacie.

En dehors de mon service à l'UFR de pharmacie, j'interviens depuis deux ans au sein de quatre Masters incluant 180 étudiants : le M2 « Functional and Structural Biology », le M1 « Biologie santé », M1 « Biotechnologies des plantes » et enfin le M1 « Omiques au service de la physiopathologie ». Ces interventions, en anglais, consistent en la création de nouveaux modules d'apprentissage par analyse chimique de matrices complexes, analyses statistiques multivariées associées et par des outils d'interprétations de données.
En concertation avec l'équipe pédagogique du M2-SFB nous avons décidé de faire **nos cours magistraux** en **pédagogie inversée**. Sur un CM de 3h, un des chapitres doit être construit et présenté par un binôme d'étudiant supervisé par l'enseignant. Cette approche en anglais s'est avérée très efficace tant pour les apprentis enseignants que pour les élèves.
Lorsque c'est possible, **je privilégie des travaux dirigés à partir de données issues de la plateforme Metatoul-AgromiX** afin d'assoir les connaissances théoriques sur des cas pratiques en incluant une visite des locaux. Cela permet aux étudiants d'appréhender de manière concrète le travail d'analyse chimique, de la production des données brutes à leur contextualisation. En tout, plus d'une cinquantaine d'étudiants par ans se familiarise avec la plateforme et ses outils.
J'ai aussi, depuis 2021, **créé un cours pour l'école des docteurs de l'Université fédérale de Toulouse** (l'unité régionale de formation à l'information scientifique et technique d'Occitanie). Cette formation de trois heures en anglais intitulé « Chemical database handling: find the needle in the haystack » s'adresse aux doctorants, postdoctorants et chercheurs utilisant les bases de données ouvertes en chimie. J'ai élaboré ce cours de manière à favoriser au maximum l'interactivité avec l'auditoire en alternant connaissances fondamentales et exercices réalisés en salle informatisée.

Tableau 1: enseignements effectués pour l'année 2020-21/21-22

Année	Niveau	Diplôme	Intitulé	Type de formation (1)	Nature (2)	Effectifs	Volume horaire annuel
4eme	Master1	Master	Functional Biology and Ecology	initiale	CM	24	2
5eme	Master 2	Master	Structural and Functional Biology	initiale	CM/TP	15	7
Bac+6	Doctorat et +	aucun	Database Handling	continue	CM	15	4.5
4eme	Master1	Master	Biosanté	initiale	CM	80	8
4eme	Master1	Master	Biotechnologie	initiale	CM	80	8
3eme	Licence 3	Licence	UE « Valorisation du végétal »	initiale	CM	15	2
2eme	DFGSP2	Pharmacie	Technique GdB	initiale	TP	148	15
2eme	DFGSP2	Pharmacie	Biodiversité	initiale	TP	148	33
3eme	**DFGSP3**	**Pharmacie**	**Contrôle qualité**	**initiale**	**TP**	**148**	**69**
3eme	**DFGSP3**	**Pharmacie**	**TP coordonnées**	**initiale**	**TP**	**148**	**36**
3eme	DFGSP3	Pharmacie	Projets tutorés	initiale	encadrement	3	3
3eme	Licence	Licence pro-dermo	Phytothérapie	professionnelle	TP	20	6
4eme	DFASP1	Pharmacie	Plantes médicinales	initiale	CM+TP	24	6+30
4eme	DFASP1	Pharmacie	UE sportif	initiale	CM	24	1.5

5eme	DFASP2	Pharmacie	intoxication	initiale	TP	24	20
6eme	DFASP2	Pharmacie	Thèse exercice	initiale	encadrement	2	6

(1) formation initiale / continue, professionnelle, présentielle / à distance
(2) cours magistraux, TP, TD, encadrement de travaux de fin d'étude et de stages
(**Gras**) Responsable d'enseignement

IV. ACTIVITÉS DE RECHERCHE

J'ai, à ce jour, 44 publications scientifiques et 9 communications orales ayant fait l'objet d'acte de colloques (h-index = 22, Scholar). La figure 1 suivante offre une vision synthétique au cours du temps de mon activité de publication par « catégorie » : articles à comité de lecture en collaboration, **en premier auteur, *en correspondant**, les conférences invitées, ainsi que les posters). Les numéros correspondent à ceux décrits dans la section bibliographie en annexe. Ces articles sont disponibles en ligne sur le portail Hal et orcid (https://orcid.org/0000-0002-6321-9005).

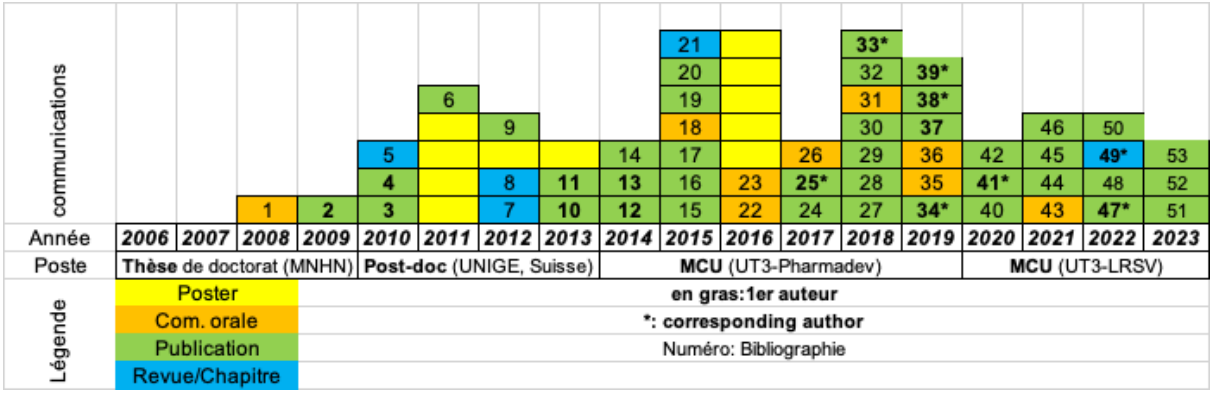


Figure 1 : Synthèse des communications scientifiques : Les numéros correspondent à ceux de la partie publication ci-dessous

Publications à comité de lecture

2. Marti, G.; Eparvier, V.; Moretti, C.; Susplugas, S.; Prado, S.; Grellier, P.; Retailleau, P.; Guéritte, F.; Litaudon, M. Antiplasmodial Benzophenones from the Trunk Latex of Moronobea Coccinea (Clusiaceae). Phytochemistry 2009, 70 (1), 75–85. https://doi.org/10.1016/j.phytochem.2008.10.005.

3. Marti, G.; Eparvier, V.; Litaudon, M.; Grellier, P.; Guéritte, F. A New Xanthone from the Bark Extract of Rheedia Acuminata and Antiplasmodial Activity of Its Major Compounds. Molecules 2010, 15 (10), 7106–7114. https://doi.org/10.3390/molecules15107106.

4. Marti, G.; Eparvier, V.; Moretti, C.; Prado, S.; Grellier, P.; Hue, N.; Thoison, O.; Delpech, B.; Guéritte, F.; Litaudon, M. Antiplasmodial Benzophenone Derivatives from the Root Barks of Symphonia Globulifera (Clusiaceae). Phytochemistry 2010, 71 (8–9), 964–974. https://doi.org/10.1016/j.phytochem.2010.03.008.

6. Glauser, G.; Marti, G.; Villard, N.; Doyen, G. A.; Wolfender, J.-L.; Turlings, T. C. J.; Erb, M. Induction and Detoxification of Maize 1,4-Benzoxazin-3-Ones by Insect Herbivores: Defense Induction and Detoxification in Maize. The Plant Journal 2011, 68 (5), 901–911. https://doi.org/10.1111/j.1365-313X.2011.04740.x.

9. Robert, C. A. M.; Veyrat, N.; Glauser, G.; Marti, G.; Doyen, G. R.; Villard, N.; Gaillard, M. D. P.; Köllner, T. G.; Giron, D.; Body, M.; Babst, B. A.; Ferrieri, R. A.; Turlings, T. C. J.; Erb, M. A

Specialist Root Herbivore Exploits Defensive Metabolites to Locate Nutritious Tissues: A Root Herbivore Exploits Plant Defences. Ecology Letters 2012, 15 (1), 55–64. https://doi.org/10.1111/j.1461-0248.2011.01708.x.

10. Marti, G.; Eparvier, V.; Morleo, B.; Ven, J.; Apel, C.; Bodo, B.; Amand, S.; Dumontet, V.; Lozach, O.; Meijer, L.; Guéritte, F.; Litaudon, M. Natural Aristolactams and Aporphine Alkaloids as Inhibitors of CDK1/Cyclin B and DYRK1A. Molecules 2013, 18 (3), 3018–3027. https://doi.org/10.3390/molecules18033018.

11. Marti, G.; Erb, M.; Boccard, J.; Glauser, G.; Doyen, G. R.; Villard, N.; Robert, C. a M.; Turlings, T. C. J.; Rudaz, S.; Wolfender, J.-L. Metabolomics Reveals Herbivore-Induced Metabolites of Resistance and Susceptibility in Maize Leaves and Roots. Plant, Cell & Environment 2013, 36 (3), 621–639. https://doi.org/10.1111/pce.12002.

12. Marti, G.; Boccard, J.; Mehl, F.; Debrus, B.; Marcourt, L.; Merle, P.; Delort, E.; Baroux, L.; Sommer, H.; Rudaz, S.; Wolfender, J.-L. Comprehensive Profiling and Marker Identification in Non-Volatile Citrus Oil Residues by Mass Spectrometry and Nuclear Magnetic Resonance. Food Chemistry 2014, 150, 235–245. https://doi.org/10.1016/j.foodchem.2013.10.103.

13. Marti, G.; Schnee, S.; Andrey, Y.; Simoes-Pires, C.; Carrupt, P.-A.; Wolfender, J.-L.; Gindro, K. Study of Leaf Metabolome Modifications Induced by UV-C Radiations in Representative Vitis, Cissus and Cannabis Species by LC-MS Based Metabolomics and Antioxidant Assays. Molecules 2014, 19 (9), 14004–14021. https://doi.org/10.3390/molecules190914004.

14. Mehl, F.; Marti, G.; Boccard, J.; Debrus, B.; Merle, P.; Delort, E.; Baroux, L.; Raymo, V.; Velazco, M. I.; Sommer, H.; Wolfender, J.-L.; Rudaz, S. Differentiation of Lemon Essential Oil Based on Volatile and Non-Volatile Fractions with Various Analytical Techniques: A Metabolomic Approach. Food Chemistry 2014, 143, 325–335. https://doi.org/10.1016/j.foodchem.2013.07.125.

15. Chawech, R.; Jarraya, R.; Girardi, C.; Vansteelandt, M.; Marti, G.; Nasri, I.; Racaud-Sultan, C.; Fabre, N. Cucurbitacins from the Leaves of Citrullus Colocynthis (L.) Schrad. Molecules 2015, 20 (10), 18001–18015. https://doi.org/10.3390/molecules201018001.

16. Erb, M.; Robert, C. A. M.; Marti, G.; Lu, J.; Doyen, G.; Villard, N.; Barrière, Y.; French, B. W.; Wolfender, J.-L.; Turlings, T.; Gershenzon, J. A Physiological and Behavioral Mechanism for Leaf-Herbivore Induced Systemic Root Resistance. Plant Physiology 2015, pp.00759.2015. https://doi.org/10.1104/pp.15.00759.

17. Gutjahr, C.; Sawers, R. J. H.; Marti, G.; Andrés-Hernández, L.; Yang, S.-Y.; Casieri, L.; Angliker, H.; Oakeley, E. J.; Wolfender, J.-L.; Abreu-Goodger, C.; Paszkowski, U. Transcriptome Diversity among Rice Root Types during Asymbiosis and Interaction with Arbuscular Mycorrhizal Fungi. PNAS 2015, 201504142. https://doi.org/10.1073/pnas.1504142112.

19. Mehl, F.; Marti, G.; Merle, P.; Delort, E.; Baroux, L.; Sommer, H.; Wolfender, J.-L.; Rudaz, S.; Boccard, J. Integrating Metabolomic Data from Multiple Analytical Platforms for a Comprehensive Characterisation of Lemon Essential Oils: Lemon Oil Characterisation by Multiblock Metabolomic Analysis. Flavour and Fragrance Journal 2015, 30 (2), 131–138. https://doi.org/10.1002/ffj.3230.

20. Queiroz, M. M. F.; Marti, G.; Queiroz, E. F.; Marcourt, L.; Castro-Gamboa, I.; Bolzani, V. S.; Wolfender, J.-L. LC–MS/MS Quantitative Determination of Tetrapterys Mucronata Alkaloids, a Plant Occasionally Used in Ayahuasca Preparation. Phytochem. Anal. 2015, 26 (3), 183–188. https://doi.org/10.1002/pca.2548.

24. Biloa, B. M.; Ho, R.; Marti, G.; Lannang, A. M.; Wolfender, J.-L.; Hostettmann, K. A Rapid Determination and Quantification of Three Biologically Active Polyisoprenylated Benzophenones Using Liquid Chromatography-Tandem Mass Spectrometry (MRM) Method in Five Garcinia Species from Cameroon. NATURAL PRODUCT COMMUNICATIONS 2017, 12 (12), 1893–1896.

25. Chervin, J.; Perio, P.; Martins-Froment, N.; Pharkeovilay, C.; Reybier, K.; Nepveu, F.; Fabre, N.; Talou, T.; Bonzon-Ponnet, V.; Marti, G. Dereplication of Natural Products from Complex Extracts

by Regression Analysis and Molecular Networking: Case Study of Redox-Active Compounds from Viola Alba Subsp. Dehnhardtii. Metabolomics 2017, 13 (8). https://doi.org/10.1007/s11306-017-1227-6.

27. Cano, L.; Cerapio, J. P.; Ruiz, E.; Marchio, A.; Turlin, B.; Casavilca, S.; Taxa, L.; Marti, G.; Deharo, E.; Pineau, P.; Bertani, S. Liver Clear Cell Foci and Viral Infection Are Associated with Non-Cirrhotic, Non-Fibrolamellar Hepatocellular Carcinoma in Young Patients from South America. Sci Rep 2018, 8 (1), 9945. https://doi.org/10.1038/s41598-018-28286-0.

28. Chassagne, F.; Haddad, M.; Amiel, A.; Phakeovilay, C.; Manithip, C.; Bourdy, G.; Deharo, E.; Marti, G. A Metabolomic Approach to Identify Anti-Hepatocarcinogenic Compounds from Plants Used Traditionally in the Treatment of Liver Diseases. Fitoterapia 2018, 127, 226–236. https://doi.org/10.1016/j.fitote.2018.02.021.

29. Diop, F.; Vial, T.; Ferraris, P.; Wichit, S.; Bengue, M.; Hamel, R.; Talignani, L.; Liegeois, F.; Pompon, J.; Yssel, H.; Marti, G.; Missé, D. Zika Virus Infection Modulates the Metabolomic Profile of Microglial Cells. PLOS ONE 2018, 13 (10), e0206093. https://doi.org/10.1371/journal.pone.0206093.

30. Haidara, M.; Haddad, M.; Denou, A.; Marti, G.; Bourgeade-Delmas, S.; Sanogo, R.; Bourdy, G.; Aubouy, A. In Vivo Validation of Anti-Malarial Activity of Crude Extracts of Terminalia Macroptera, a Malian Medicinal Plant. Malaria Journal 2018, 17 (1). https://doi.org/10.1186/s12936-018-2223-7.

32. Villaret, J.; Marti, G.; Dubois, F.; Reybier, K.; Gaudre, N.; Haddad, M.; Valentin, A. Adaptation of a Microbead Assay for the Easy Evaluation of Traditional Anti-Sickling Medicines: Application to DREPANOSTAT and FACA. Pharmaceutical Biology 2018, 56 (1), 385–392. https://doi.org/10.1080/13880209.2018.1501585.

33. Chassagne, F.; Cabanac, G.; Hubert, G.; David, B.; Marti, G. The Landscape of Natural Product Diversity and Their Pharmacological Relevance from a Focus on the Dictionary of Natural Products®. Phytochemistry Reviews 2019, 18 (3), 601–622. https://doi.org/10.1007/s11101-019-09606-2.

34. Chervin, J.; Talou, T.; Audonnet, M.; Dumas, B.; Camborde, L.; Esquerré-Tugayé, M.-T.; Roux, C.; Cabanac, G.; Marti, G. Deciphering the Phylogeny of Violets Based on Multiplexed Genetic and Metabolomic Approaches. Phytochemistry 2019, 163, 99–110. https://doi.org/10.1016/j.phytochem.2019.04.001.

37. Marti, G.; Joulia, P.; Amiel, A.; Fabre, B.; David, B.; Fabre, N.; Fiorini-Puybaret, C. Comparison of the Phytochemical Composition of Serenoa Repens Extracts by a Multiplexed Metabolomic Approach. Molecules 2019, 24 (12), 2208. https://doi.org/10.3390/molecules24122208.

38. Phakeovilay, C.; Bourgeade-Delmas, S.; Perio, P.; Valentin, A.; Chassagne, F.; Deharo, E.; Reybier, K.; Marti, G. Antileishmanial Compounds Isolated from Psidium Guajava L. Using a Metabolomic Approach. Molecules 2019, 24 (24), 4536. https://doi.org/10.3390/molecules24244536.

39. Vial, T.; Tan, W.-L.; Wong Wei Xiang, B.; Missé, D.; Deharo, E.; Marti, G.; Pompon, J. Dengue Virus Reduces AGPAT1 Expression to Alter Phospholipids and Enhance Infection in Aedes Aegypti. PLoS Pathog 2019, 15 (12), e1008199. https://doi.org/10.1371/journal.ppat.1008199.

40. Dienes-Nagy, Á.; Marti, G.; Breant, L.; Lorenzini, F.; Fuchsmann, P.; Baumgartner, D.; Zufferey, V.; Spring, J.-L.; Gindro, K.; Viret, O.; Wolfender, J.-L.; Rösti, J. Identification of Putative Chemical Markers in White Wine (Chasselas) Related to Nitrogen Deficiencies in Vineyards. OENO One 2020, 54 (3), 583–599. https://doi.org/10.20870/oeno-one.2020.54.3.3285.

41. Fraisier-Vannier, O.; Chervin, J.; Cabanac, G.; Puech, V.; Fournier, S.; Durand, V.; Amiel, A.; André, O.; Benamar, O. A.; Dumas, B.; Tsugawa, H.; Marti, G. MS-CleanR: A Feature-Filtering Workflow for Untargeted LC–MS Based Metabolomics. Anal. Chem. 2020, 92 (14), 9971–9981. https://doi.org/10.1021/acs.analchem.0c01594.

42. Vial, T.; Tan, W.-L.; Deharo, E.; Missé, D.; Marti, G.; Pompon, J. Mosquito Metabolomics Reveal That Dengue Virus Replication Requires Phospholipid Reconfiguration via the Remodeling Cycle. Proc Natl Acad Sci USA 2020, 202015095. https://doi.org/10.1073/pnas.2015095117.

44. Vásquez-Ocmín, P. G.; Gadea, A.; Cojean, S.; Marti, G.; Pomel, S.; Van Baelen, A.-C.; Ruiz-Vásquez, L.; Ruiz Mesia, W.; Figadère, B.; Ruiz Mesia, L.; Maciuk, A. Metabolomic Approach of the Antiprotozoal Activity of Medicinal Piper Species Used in Peruvian Amazon. Journal of Ethnopharmacology 2021, 264, 113262. https://doi.org/10.1016/j.jep.2020.113262.

45. Vásquez-Ocmín, P. G.; Marti, G.; Bonhomme, M.; Mathis, F.; Fournier, S.; Bertani, S.; Maciuk, A. Cannabinoids vs. Whole Metabolome: Relevance of Cannabinomics in Analyzing Cannabis Varieties. Analytica Chimica Acta 2021, 339020. https://doi.org/10.1016/j.aca.2021.339020.

46. Vial, T.; Marti, G.; Missé, D.; Pompon, J. Lipid Interactions Between Flaviviruses and Mosquito Vectors. Front. Physiol. 2021, 12, 763195. https://doi.org/10.3389/fphys.2021.763195.

(1)

47. Chervin, J.; Romeo-Oliván, A.; Fournier, S.; Puech-Pages, V.; Dumas, B.; Jacques, A.; Marti, G. Modification of Early Response of Vitis Vinifera to Pathogens Relating to Esca Disease and Biocontrol Agent Vintec® Revealed By Untargeted Metabolomics on Woody Tissues. *Front. Microbiol.* **2022**, *13*, 835463. https://doi.org/10.3389/fmicb.2022.835463.

48. Zhour, H.; Bray, F.; Dandache, I.; Marti, G.; Flament, S.; Perez, A.; Lis, M.; Cabrera-Bosquet, L.; Perez, T.; Fizames, C.; Baudoin, E.; Madani, I.; Zein, L. E.; Véry, A.-A.; Rolando, C.; Sentenac, H.; Chokr, A.; Peltier, J.-B. Wild Wheat Rhizosphere-Associated Plant Growth-Promoting Bacteria Exudates: Effect on Root Development in Modern Wheat and Composition. Int. J. Mol. Sci. 2022, 24.

50. Espichán, F.; Rojas, R.; Quispe, F.; Cabanac, G.; Marti, G. Metabolomic Characterization of 5 Native Peruvian Chili Peppers (Capsicum Spp.) as a Tool for Species Discrimination. Food Chemistry 2022, 386, 132704. https://doi.org/10.1016/j.foodchem.2022.132704.

51. Vásquez-Ocmín, P. G.; Marti, G.; Gadea, A.; Cabanac, G.; Vásquez-Briones, J. A.; Casavilca-Zambrano, S.; Ponts, N.; Jargeat, P.; Haddad, M.; Bertani, S. Metabotyping of Andean Pseudocereals and Characterization of Emerging Mycotoxins. Food Chemistry 2023, 407, 135134. https://doi.org/10.1016/j.foodchem.2022.135134.

52. Danneels, B.; Blignaut, M.; Marti, G.; Sieber, S.; Vandamme, P.; Meyer, M.; Carlier, A. Cyclitol Metabolism Is a Central Feature of Burkholderia Leaf Symbionts. Environmental Microbiology 2023, 25 (2), 454–472. https://doi.org/10.1111/1462-2920.16292.

53. Vásquez-Ocmín, P. G.; Cojean, S.; Roumy, V.; Marti, G.; Pomel, S.; Gadea, A.; Leblanc, K.; Dennemont, I.; Ruiz-Vásquez, L.; Cotrina, H. R.; Mesia, W. R.; Bertani, S.; Mesia, L. R.; Maciuk, A. Deciphering Anti-Infectious Compounds from Peruvian Medicinal Cordoncillos Extract Library through Multiplexed Assays and Chemical Profiling. 2023. https://doi.org/10.1101/2022.11.15.516654.

Revues/Chapitres

5. Wolfender, J.-L.; **Marti, G**.; Ferreira Queiroz, E. Advances in Techniques for Profiling Crude Extracts and for the Rapid Identification of Natural Products: Dereplication, Quality Control and Metabolomics; 2010; Vol. 14.

7. Gindro, K.; Alonso-Villaverde, V.; Viret, O.; Spring, J.-L.; **Marti, G**.; Wolfender, J.-L.; Pezet, R. Stilbenes: Biomarkers of Grapevine Resistance to Disease of High Relevance for Agronomy, Oenology and Human Health. In Plant Defence: Biological Control; Mérillon, J. M., Ramawat, K.

G., Eds.; Springer Netherlands: Dordrecht, 2012; pp 25–54. https://doi.org/10.1007/978-94-007-1933-0_2

8. **Marti, G**.; Erb, M.; Rudaz, S.; Turlings, T.; Wolfender, J.-L. Search for Low-Molecular-Weight Biomarkers in Plant Tissues and Seeds Using Metabolomics: Tools, Strategies, and Applications. In Seed Development: OMICS Technologies toward Improvement of Seed Quality and Crop Yield; Agrawal, G. K., Rakwal, R., Eds.; Springer Netherlands: Dordrecht, 2012; pp 305–341.

21. Wolfender, J.-L.; **Marti, G**.; Thomas, A.; Bertrand, S. Current Approaches and Challenges for the Metabolite Profiling of Complex Natural Extracts. Journal of Chromatography A 2015, 1382, 136–164. https://doi.org/10.1016/j.chroma.2014.10.091.

49. Vial, T.; Phakeovilay, C.; Watanabe, S.; Chan, K. W. K.; Peng, M.; Deharo, E.; Chassagne, F.; Vasudevan, S. G.; Marti, G. Chapter 13 - Antiviral Potential of Medicinal Plants: A Case Study with Guava Tree against Dengue Virus Using a Metabolomic Approach. In Medicinal Plants as Anti-Infectives; Chassagne, F., Ed.; Academic Press, 2022; pp 439–458. https://doi.org/10.1016/B978-0-323-90999-0.00010-0.

Communications orales

1. Marti, G. MS/MS of Natural Polyprenylated Benzophenone Derivatives; Athen, 2008.

18. Marti, G. Approche Holistique d'une Fleur de Rond-Point: La Violette de Toulouse. In 7 ème journé e de la spectrométrie de masse en Midi-Pyrénées.; Toulouse, 2015.

22. Marti, G. Embezzlement. Dengue Virus Manipulation of Mosquito Metabolome. In Invited speaker at Duke-NUS; Singapore, 2016.

23. Marti, G. LC-MS Based Metabolomic Approach Applied to Bioactive Molecules Identification: Case Study of a Comprehensive Chemodiversity Screening of Viola Genus. In RBB-14; Ghent, Belgium, 2016.

26. Marti, G. Développement d'approches Holistiques Pour l'identification Précoce de Composes Bioactifs Au Sein de Matrices Complexes. In 33eme journées scientifiques de l'AFERO; Toulouse, 2017.

31. Marti, G. LCMS Base Metabolomics: A Fake News Machine? In conférence invité; Pisa, Italie, 2018.

35. Marti, G. Exploring the Dark Matter of Chemodiversity Using Metabolomics. In The Impact of Biodiversity on Human Health The Role of Biodiversity in Disease Ecology and Evolution; Atlanta US, 2019.

36. Marti, G. Presentation Plateforme Metatoul-Metabolites Végétaux. In Congrés Biocontrol; Bordeaux, 2019.

43. Marti, G. MEtaboHub 2.0: WP4: Extand Metabolome Coverage; Nantes, 2021.

Posters

Marti G. Projet VIOLA TOLOSA : Biodiversité des fragrances de violettes et valorisation d'un patrimoine végétal-Journée de la violette-Serre Municipale de Toulouse-2015

Poinso A., Vicendo P., Couderc F., Fabre N., **Marti G**., Vansteelandt M., Haddad M., Cabanillas B., Ong-Meang V. SOD inhibitors research from endophytic fungi extracts. 9 Th Joint Natural Products Conference 2016, July 24-27th. Copenhagen (Denmark).

Chervin J., **Marti G**., Perio P., Martins N., Reybier K., Nepveu F., Fabre N., Vilarem G., Talou T.. A comprehensive chemodiversity screening of *viola* genus in Midi-Pyrenees by a holistic approach; Montpellier - Colloque Recherche Chimie et Procédés du Végétal, janvier 2016.

Chervin J., **Marti G**., Perio P., Martins N., Nepveu F., Fabre N., Pharkeovilay C., Vilarem G., Talou T. Metabolomics as a tool for biorefinery: case study of a comprehensive chemotaxonomy screening of violets. - A Ghent - RRB-12: The 12th International Conference on Renewable Resources & Biorefineries, mai-juin 2016.

Chervin J., **Marti G**., Perio P., Martins N., Reybier K., Nepveu F., Fabre N., Vilarem G., Talou T. LC-MS based metabolomic approach applied to bioactive molecules identification: case study of a comprehensive chemodiversity screening of viola genus. 9th Joint Natural Products Conference 2016, July 24-27 th Copenhagen (Denmark).

Chervin J., Dumas B., Camborde L., Esquerre M.-T., Fabre N., Marti G., Talou T. Chemotaxonomy *vs* Genotaxonomy of *Viola genus*: Can native volatile compounds of flowers be genuine tracers for violets differentiation? – Nice - 47th International Symposium on Essential Oils, September 2016.

Queiroz, M.F., E.F. Queiroz, **G. Marti**, L. Marcourt, I.G. Castro, V.S. Bolzani, and J.L. Wolfender, Isolation and quantification of alkaloids from *Tetrapterys mucronata* – a plant used in the ayahuasca preparation. Planta Med, 2013. 79(13): p. PI86. 61st International Congress and Annual Meeting of the Society for Medicinal Plant and Natural Product Research; Münster, Germany, 1st – 5th September 2013

G. Marti., V. Zufferey, K. Gindro, O. Viret, and J.L. Wolfender, A comprehensive metabolomic study of wine from the Vaud Switzerland vineyard. Planta Med, 2012. 78(11): p. PJ67. 8th Joint Meeting of AFERP, ASP, GA, PSE and SIF; New York City, July 28 – August 1, 2012

G. Marti., J. Boccard, S. Rudaz, K. Gindro and J-L. Wolfender (2011). Study of UV irradiation effects on *Vitis* leaves at the metabolome level by UHPLC-TOFMS. Swiss Chemical Society Fall meeting; Lausanne, CHIMIA 2011, 65, No. 7/8 461

Stefanou, A., S. Bertrand, J. Boccard, N. Lemonakis, **G. Marti**, S. Rudaz, S. Kostidis, V. Gikas, A.L. Skaltsounis, K. Gindro, M. Halabalaki, and J.L. Wolfender, NMR- and UHPLC-MS-based metabolomics for the discrimination of different resistant *Vitis vinifera* L. cultivar woods. Planta Med, 2011. 77(12): p. PL76. 59th International Congress and Annual Meeting of the Society for Medicinal Plant and Natural Product Research; Antalya, Turkey, 4.–9. September 2011

Mehl, F., **G. Marti**, J. Boccard, L. Marcourt, B. Debrus, P. Merle, E. Delort, L. Baroux, H. Sommer, J-L Wolfender, S. Rudaz (2011). Analyse Multi-tableaux de données UHPLC-QTOF-MS, RMN et FT-MIR, des Huiles Essentielles de citron. Chimiométrie 2011-Aix Marseille III

Biloa Messi, B., **G. Marti**, R. Ho, K. Ndjoko Ioset, A. Meli Lannang, K. Hostettmann, and J. Wolfender, A rapid simultaneous quantification of five biologically active polyisoprenylated benzophenones using liquid chromatography-tandem mass spectrometry (MRM) method in two *Garcinia* species from Cameroon. Planta Med, 2010. 76(12): p. P551. Tannin Conference (Presymposium) and 58th International Congress and Annual Meeting of the Society for Medicinal Plant and Natural Product Research; Berlin, Germany, 29th August – 2d September 2010

V. ACTIVITÉS D'ENCADREMENT :

À ce jour (Mars 2023), j'ai co-encadré 4 doctorants. J'ai par ailleurs supervisé l'accueil d'un étudiant péruvien dans le cadre d'un séjour d'un an durant sa thèse de doctorat. À ces encadrements s'ajoutent les CDDs que j'ai recrutés au travers de projets de recherche financés et qui étaient sous ma responsabilité. Aujourd'hui je co-dirige une thèse débutée en octobre 2021 ayant obtenu une dérogation d'HDR à titre individuel. Mon habilitation définitive devant être acquise au second semestre 2023.

La figure 2 et les tableaux 2 suivants offrent une vision synthétique de mon activité d'encadrement.

Figure 2 : détail des encadrements hors stage (CDD, doctorant, post-doc)

V.1 Thèses soutenues :

- Justine Chervin « Valorisation chimique de la violette en Occitanie »
  Diplôme préparé : doctorat en phytochimie et métabolomique de l'UT3
  Financement : Bourse CIFRE
  Durée de l'encadrement : 3 ans (octobre 2015 – octobre 2018)
  Co-direction
  Part d'encadrement : 50 % (co-dirigé avec T. Talou, INP ENSIACET)

- Phimpha Paboriboune « Approche holistique pour la mise en évidence de marqueurs diagnostics, dynamiques et pronostics des hépatites B, C, et du VIH sur plasma de patients Laotiens »
  Diplôme préparé : doctorat en Biologie clinique de l'UT3
  Financement : Bourse Campus France
  Durée de l'encadrement : 3 mois/ an sur 3 ans (octobre 2015 – octobre 2018)
  Rôle : Co-encadrement
  Part d'encadrement : 25 % (directeur : E. Deharo, PHARMADEV)

- Chiobouaphong Phakeavolay « Étude des changements REDOX chez Leishmaniose par des approches métabolomiques»
  Diplôme préparé : Doctorat en chimie de l'UT3
  Financement : bourse de la Fondation Pierre Fabre
  Durée de l'encadrement : 3 ans (octobre 2016 – octobre 2019)
  Rôle : Co-direction
  Part d'encadrement : 25 % (directrice : K. Reybier, PHARMADEV)

- Fabio Espichan « Approche métabolomique pour le contrôle qualité de piments péruviens »
  Diplôme préparé : Doctorat de l'université Cayetano Ereidia (Pérou)
  Financement : Gouvernement péruvien
  Durée de l'encadrement : 1an/3 ans (Avril 2017- Avril 2018)
  Rôle : Supervision d'un séjour d'apprentissage
  Part d'encadrement : 15 % (directeur : Rosario Rojas Duran)
  Publications communes : En révision

- *Thomas Vial* « Dengue virus diverts the mosquito phospholipid metabolism for replication»
  Diplôme préparé : Doctorat de l'université Toulouse 3
  Financement : Bourse DukeNUS
  Durée de l'encadrement : 2017-2020
  Rôle : Supervision scientifique
  Part d'encadrement : 25 % (directeur : E. Deharo et J. Pompon)

V.2 Thèse en cours :

- Solweig Hennechart « Chemomaps : approche deep learning pour appréhender la chimiodiversité du vivant »
  Diplôme préparé : Doctorat de l'université Toulouse 3
  Financement : Bourse CNRS 80PRIME
  Durée de l'encadrement : 36 mois (Octobre 2021- Septembre 2024)
  Rôle : Co-direction
  Part d'encadrement : 50 % (co-direction: Guillaume Cabanac, MCU UT3)

Tableau 2: détail des encadrements

Personnes	Poste	Titre	financement	durée	role
Justine Chervin	Doctorante	Valorisation chimique de la violette en Occitanie	CIFRE	3 ans 2015-2018	co-encadrant
Phimpha Paboriboune	Doctorante	Mise en évidence de marqueurs diagnostics, dynamiques et pronostics des hépatites B, C, et du VIH sur plasma de patients Laotiens	Campus france	3 ans 2015-2018	co-encadrant
Chiobouaphong Phakeavolay	Doctorante	Étude des changements REDOX chez Leishmaniose par des approches métabolomiques	Pierre-Fabre fondation	3 ans 2016-2019	co-encadrant
Fabio Espichan	Doctorant	Approches metabolomiques pour le contrôle qualité des piments péruviens	Gouvernement péruvien	1 an 2017-2018	Acceuil et formation
Aurélien Amiel	IE	Approche métabolomique appliquée à des problématiques industrielles	Contrat Pierre-Fabre	1 an 2016	Direction scientifique
Ophélie Fraisier-Vannier	Post-Doc	Approche métabolomique appliquée à des problématiques industrielles	Contrat Pierre-Fabre	1 an 2019-2020	Direction scientifique
Thomas Contini	IE	Étude des steviol glicosides chez la Stevia	ANR	1,5 ans 2020-2021	co-direction
S. Hennechart	Doctorante	Projet ChemoMaps	Bourse CNRS 80PRIME	3 ans 2021-2024	co-encadrant
A. Dablanc	IE	Projet MTH 2.0	ANR	1,5 ans 2023-2024	Direction scientifique
A. Pirat	IE	Contrat plateforme	Projet Region	1 an 2022	Direction scientifique
P Vasquez	Post-Doc	Contrat plateforme	Projet PIA3	2 ans 2022-2023	Direction scientifique
Julian Janer	Stage BTS anabiotech	Stage BTS	Aucun	4 mois 2021	Direction scientifique
Anaïs Apolinario	Stage BTS anabiotech	Stage BTS	Aucun	4 mois 2022	Direction scientifique
Victoria Dickinson	M1 Biosanté	Stage M1	Aucun	3 mois 2023	Direction scientifique

VI - AUTRES

VI.1 Responsabilités collectives et d'intérêt général

- **Membre élu de la commission recherche** de l'université Paul-Sabatier (2015-2020).
- **Membre de la commission BIATSS du département de pharmacie** (2022-aujourd'hui).
- **Responsable scientifique et administratif de la plateforme analytique Metatoul-AgromiX** hébergée au LRSV (UMR 5546).
- **Responsable du WP4 pour l'ANR MTH 2.0** (2.5M sur 5 ans)
- **Membre du bureau de direction de Metatoul**: La plateforme Metatoul-AgromiX fait partie de la genotoul de Toulouse qui est un rassemblement de plateformes de la région Occitanie.
- **Membre du conseil Scientifique du LRSV**
- **Membre du bureau exécutif de l'infrastructure nationale MétaboHub.**
- **Expert pour l'ANRT** depuis 2020
- **Membre de jurys de recrutements** (IGR BAP A, UT3, Septembre 2020 et MCU, Paris Saclay, Mai 2020)

VI.2 Financements obtenus :

- Contrat Pierre-Fabre (2016-2019): Un contrat de collaboration industriel avec l'entreprise Pierre-Fabre a été initié en 2016 sous la direction de Nicolas Fabre (Pr, UPS) et ma co-direction. Ce contrat d'un montant de 100 KE par an s'est prolongé jusqu'en 2019 et a permis de financer l'embauche d'un IE (Aurélien Amiel) et d'une postdoctorante (Ophélie Fraisier-Vannier). La thématique de recherche sur l'application des approches métabolomiques à des problématiques industrielles a été valorisée sous forme d'une co-publication (22). Ce type de contrat est très exigeant étant donné les attentes forcément concrètes en termes de résultats des partenaires industriels, mais permet faire des pontages très valorisants entre la recherche académique et appliquée.
- Projet Région Occitanie « Viola tolosa » (2017-2020): Cette bourse CIFRE m'a permis de réaliser mon premier co-encadrement de thèse (Justine Chervin). L'ensemble de ces travaux ont abouti à plusieurs communications sous forme d'article et de congrès.
- Bourse CNRS 80PRIME (2021-2024): J'ai obtenu en 2021 une bourse CNRS sur l'enveloppe « 80PRIME ». Le projet s'intéresse à la cartographie de la chimiodiversité du vivant à l'aide d'études de données massives et d'approches de machine learning. Le recrutement d'un doctorant est donc en cours pour un démarrage effectif en septembre 2021.
- ANR MTH 2.0 (2021-2024): je suis impliqué dans ce projet ANR d'une infrastructure national (MetaboHub, https://www.metabohub.fr/) dont fait partie Metatoul en tant que responsable d'un work package (Bases de données et améliorations des annotations en métabolomique). Ce projet financera un postdoctorant sur 18 mois qui est en cours de recrutement.

- Bourse PIA Metex+ (2021-2024): Dans le cadre de MetaboHub, nous avons obtenu le financement de plusieurs équipements de spectrométrie de masse (4.5 M) qui viendront agrandir le parc existant. Le plateau AgromiX que j'anime bénéficiera de l'acquisition d'un spectromètre à mobilité ionique (500 KE).
- ANR CREZY (2023-2027) : collaboration avec Dr Cyril Saintenac (DR, INRAE Clermont Ferrand).

Tableau 2: récapitulatif des financements obtenus

Financeur	Dates	Projet	Montants gérés (KE)	ETP
Pierre Fabre medicament	2016-2019	Contrat privée	300	0.5
Région Occitanie	2017-2020	Viola tolosa	250	0.5
Région Occitanie	2019-2022	Compevia	150	0.1
CNRS-80PRIME	2021-2024	Chemomaps	120	0.5
ANR	2021-2024	MTH 2.0	120	0.2
PIA4	2019-2023	Solstice	300	0.1
PIA4	2021-2024	MeteX+	500	0.1
ANR	2023-2027	Crezy	40	0.1
Prestation de services de la plateforme	/an	académiques et industrielles	100 à 150	0.3

VI.3 Responsabilités scientifiques et administratives :

- J'effectue en moyenne **quatre relectures d'article par an** pour les journaux suivant : Plant, Cell & Environment, Analytica Chemica Acta, PLOS one, Molecules, Scientific Reports, Metabolomics, Molecules, Journal of Agricultural and Food Chemistry et Journal of Ethnopharmacology, Food Chemistry, Phytochemistry…
- J'ai été **responsable de l'animation scientifique** de l'UMR Pharmadev (80 personnes, non-permanents inclus) de 2015 à 2017. Ceci implique l'animation mensuelle de séminaire interne ou invité ainsi que l'organisation des journées scientifiques de l'UMR.
- **Membre élu de la commission recherche** de l'université Paul-Sabatier (2015-2020). En tant que membre élu de la commission recherche de l'université Toulouse 3 (UT3), j'ai pu découvrir les arcanes décisionnels de la politique d'une grande université. Cette responsabilité implique deux réunions mensuelles (commission recherche et conseil académique) précédées d'une réunion de préparation en amont (cellule opérationnelle).
- **Responsable scientifique et administratif de la plateforme analytique Metatoul-AgromiX** hébergée au LRSV (UMR 5546). J'anime depuis janvier 2020 une équipe de recherche et prestations analytique. Ce nouveau métier est passionnant, mais présente des challenges importants, notamment au niveau de la gestion des ressources humaines et de l'environnement matériel. La politique du plateau est de réalisé des prestations pour les partenaires

académiques et industriels et de développer ses propres outils pour la communauté internationale de métabolomiciens. Cet équilibre permet de valoriser le savoir-faire du plateau et de chacun de ses partenaires.

- **Responsable du WP4 pour l'ANR MTH 2.0** (2.5M sur 5 ans) en vue de développer des outils d'aide à l'annotation des spectres de masse de matrices complexes. Dans ce cadre, j'anime un groupe d'une quarantaine de personnes dispersées sur l'ensemble du territoire afin d'améliorer les synergies de compétences complémentaires en vue de réaliser les objectifs que nous nous sommes fixés.
- **Membre du comité de gestion et du comité stratégique de la plateforme Metatoul**: La plateforme Metatoul-AgromiX fait partie de la genotoul de Toulouse qui est un rassemblement de plateformes de la région Occitanie. J'effectue des missions d'animation, de gestion de la qualité (les plateformes sont toutes en ISO9001), de coordinations scientifiques au sein de ce consortium et prends part activement au pilotage scientifique et stratégique de la plateforme Metatoul.
- **Membre du comité scientifique de Laboratoire de Recherche en Sciences végétales** : Le CS du LRSV se réuni une fois par mois afin d'orient les politiques scientifiques de l'unité (recrutement, orientation, stratégie à moyen et long terme).

Signature du candidat

Activités de recherche : Cartographie de la chimiodiversité du vivant

1 Table des matières

Sommaire des abbréviations

LC-MS : Chromatographie liquide couplée à un spectromètre de masse
RMN : Rainsonnace Magnétique Nucleaire
CCM : Chromatographie sur couche mince
MS/MS : Spectrométrie de masse en tandem
MPLC : Chromatographie liquide Moyenne Pression
UHPLC : Ultra High Pressure Liquid Chromatography
CAD : Corona Aerosol Detector
PDA : PhotoDiode Array detector
SPE: Solid Phase Extraction
ACP : Analyse en Composante Principale
PLS-DA : Analyse discriminante par projection sur structure latente
HMDB : Human Metabolome Database
CMNPD : Comprehensive Marine Natural Product Database
DNP: Dictionnary of Natural Products
CAS: Chemical Abstract Service
EMBL: European Molecular Biology Laboratory
ADMET: absorption, distribution, metabolism, excretion and toxicity
PNUE: Programme des Nations unis pour l'Envirronnement
PPAPs : benzoylphloroglucinols polyisoprénylés polycycliques
DMAPP : Dimethyl-allyl pyrophosphate
SMILES : simplified molecular input line entry system
InChI :International Chemical Identifier

2 Résumé

Depuis les premières purifications de substances naturelles bioactives, des centaines de milliers de composés ont été isolés issus de l'ensemble de la biodiversité. Ce sont essentiellement des méthodes réductionnistes orientées par des essais bioguidés qui ont conduit à la valorisation de plusieurs milliers de molécules. Cependant, on a pu observer ces dernières années un essoufflement de ce type d'approche au profit des méthodes holistiques. En effet, ces dernières permettent d'accélérer la recherche de nouvelles entités chimiques, mais donnent aussi des clefs d'interprétation de systèmes biologiques complexes permettant de modéliser et manipuler le vivant. À ce titre, la métabolomique joue un rôle majeur en tant que produit final de la cascade des -omiques au sein d'un organisme.
Ces deux approches, se nourrissent l'une de l'autre : la métabolomique s'appuie largement sur les données spectrales issues des travaux d'isolements de substances naturelles alors que la phytochimie s'est approprié l'outil métabolomique afin d'accélérer ses découvertes. L'exploitation des signaux spectrométriques et particulièrement ceux issus de la spectrométrie de masse dans le cadre de l'analyse de mélanges complexes est un des outils les plus utilisés aujourd'hui. Cependant, malgré des avancées majeures, plusieurs écueils limites ses potentialités telles que la déconvolution des signaux, leurs quantifications ou encore l'absence de bases de données chimiques exhaustives et biosourcées. C'est à ces points particuliers qu'essaiera de répondre une première partie du projet de recherche avant d'aborder la thématique de l'architecture de la chimie du vivant et de sa cartographie en couplant des approches analytiques innovantes à des modèles d'inférences.

3 Introduction

L'exploration de la biodiversité est l'un des domaines les plus prolifiques de la biologie tant sur le plan de la compréhension des écosystèmes complexes que sur les enjeux liés à son exploitation. Depuis les premières purifications de substances actives telles que la morphine isolée du pavot en 1817 [1] ou la strychnine purifiée de la noix vomique en 1819 [2], plusieurs centaines de milliers de substances naturelles ont été identifiées. Cette ruée vers la chimiodiversité du vivant s'explique par les nombreux succès thérapeutiques qui ont jalonné le développement de cette discipline [3–5]. Son histoire est autant liée au développement de nouvelles stratégies de prospection qu'à l'évolution des législations en vigueur.

Par exemple, l'article premier de la **Convention sur la Diversité Biologique (CBD)** implique à la fois une gestion raisonnée sur le long terme des écosystèmes et pose un cadre au transfert marchand des ressources génétiques. Ce dernier aspect a été renforcé par le **protocole de Nagoya** qui vise à « améliorer l'accès aux ressources génétiques » notamment au niveau juridique et à « permettre un partage plus juste et équitable des avantages » via des mécanismes de compensations multilatéraux (PNUE, 2010).

Par ailleurs, plusieurs stratégies de bioprospection ont progressivement vu le jour pour sélectionner les taxons d'intérêts et découvrir des entités chimiques à fort potentiel de valorisation :

- En se basant sur les **savoirs ethnobotaniques** des communautés indigènes afin d'orienter la sélection de plantes médicinales, nourricières ou d'intérêt cosmétique.
- Le développement du **criblage à haut débit** à partir des années 1980 qui grâce à la miniaturisation des essais biologiques *in vitro* a conduit aux premières campagnes de fouille à l'aveugle de la diversité du vivant.
- Les approches se basant sur la **chimiotaxonomie** qui explorent les taxons proches d'un organisme d'intérêts.
- Les récents développements de l'**écologie chimique** qui ont conduit à des stratégies de bioprospection d'interactions chimiques biotiques et abiotiques.
- Et enfin, les **approches computationnelles** basées sur la prédiction des activités biologiques *in silico* permettant d'orienter la sélection de motifs chimiques d'intérêts [6].

Ces stratégies de criblage partagent une approche réductionniste du vivant ou la biodiversité est considérée comme un réservoir d'entités chimiques dont une fraction peut être valorisée pour le bien-être de l'homme.
Le bilan de ces années de bioprospection depuis les années 1980 est relativement mitigé. Malgré quelques succès [7–9], les résultats tant sur le plan des découvertes que des partages équitables des retombées ont été limités. Aujourd'hui, la plupart des entreprises pharmaceutiques/cosmétiques se sont retirées des programmes de bioprospections. *In fine* pour 17 entreprises impliquées dans les années 1980-90, seules 6 avaient des programmes actifs en 2010 [10].
Cette situation peut s'expliquer par plusieurs facteurs :
- Les limitations du modèle simplifié « enzyme-ligand » (clé-serrure) utilisé dans les programmes de criblage à haut débit (on estime à 1 hit sur 1 million de composés tester).
- Les problèmes d'identification binomiale et de réapprovisionnement dans des écosystèmes sous pression anthropique croissante.
- Les faibles retombées économiques ayant découlées de ces programmes et les risques associés avec pour conséquence un net durcissement de l'accès aux bioressources des pays à forte biodiversité.

Aujourd'hui, suite au désintéressement des « big-pharma », les programmes de bioprospections sont essentiellement supportés par le milieu académique [10].

Cet échec souligne les limites de ces approches réductionnistes dont l'ethnocentrisme a occulté la complexité du vivant et la place des sociétés humaines dans leurs environnements. Un changement de paradigme survient dans les années 2000 avec l'émergence des premiers séquençages complets de génomes [11], [12]. Il s'agit donc de capturer la totalité d'un ensemble moléculaire complexe issu d'un organisme (ici le support du code génétique) et de le rendre accessible à la communauté. Ces **approches holistiques** auront pour objectif de cartographier le vivant dans sa globalité et sa complexité. Contrairement aux méthodes réductionnistes, ces approches sont sans *a priori* et aboutissent à la génération de nouvelles hypothèses issues de l'analyse des données. **Les « -omiques »** se sont propagées dans toutes les sphères des sciences du vivant sous plusieurs avatars liés à la classe de composés moléculaires ou macromoléculaires ciblés (génomique, protéomique, transcriptomique…).

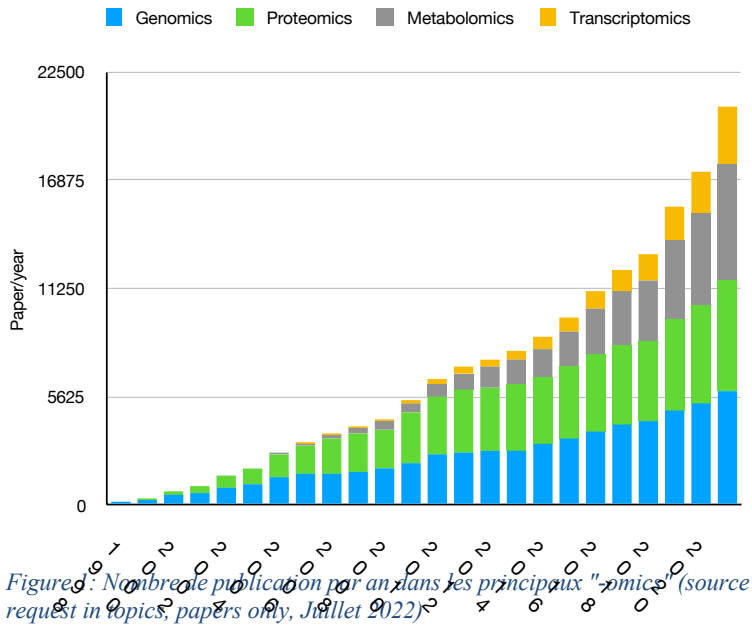


Figure 1 : Nombre de publication par an dans les principaux "-omics" (source WOS, request in topics, papers only, Juillet 2022)

La **génomique** et la **protéomique** se sont développées de manière concomitante et ont joué le rôle d'amorce aux développements des autres -omiques, que ce soit grâce aux avancées des méthodes d'hybridations pour l'exploration du génome et du transcriptome ou bien par l'essor des outils de spectrométrie de masse en protéomique dont la métabolomique a largement bénéficié. Aujourd'hui, plus de 20 000 articles sont publiés chaque année en utilisant l'une ou plusieurs de ces approches -omiques (Figure 1).
La phytochimie tel qu'elle était pratiquée jusque dans les années 2010 a été fortement impactée par l'essor de la **métabolomique**. D'un autre côté, les bases de données nécessaires à l'annotation des signaux en spectrométrie de masse ou résonance magnétique nucléaire se sont largement nourries des résultats de la phytochimie.

Le projet présenté ici se propose d'explorer la rencontre entre la phytochimie et la métabolomique et leurs articulations et possibles synergies.
Qu'en est-il du bilan de ce siècle et demi de recherche en chimie des substances naturelles ? Comment quantifier et qualifier la chimiodiversité du vivant par ces approches ? Peut-on tirer des informations sur son évolution, sa répartition et sa diversité ? Quel est l'apport des récents développements en chimie analytique pour qualifier et quantifier le métabolome ?
Ce manuscrit s'attachera à donner des éléments de réponses à ces questions en s'appuyant sur mon expérience scientifique dans une première partie puis en développant des perspectives de recherches à même d'apporter des solutions pertinentes.

J’aborderai dans une première partie l’articulation entre une approche réductionniste et son évolution vers des approches holistiques et leurs applications à l’interface entre la métabolomique et la phytochimie. Ce cheminement sera illustré au travers d’une sélection de mes publications (uniquement premier auteur ou auteur correspondant, numéro de référence en gras dans le texte) en abordant dans un premier temps mes travaux de thèse sur la purification de substances naturelles antipaludiques issus de la biodiversité guyanaise (Chapitre 3). Cet exemple sera approfondi afin d’introduire les notions d’empreintes chimiques et de déréplication illustrées par l’utilisation de différentes méthodes de détection (CCM, LC-MS, RMN). L’intérêt de ces approches couplées aux analyses statistiques multivariées dans le cadre de recherche de substances naturelles bioactives conclura cette partie (Chapitre 4). La problématique spécifique de la déréplication en spectrométrie de masse sera détaillée avant d’introduire l’intérêt de ces empreintes LC-MS dans la classification taxonomique des espèces (Chapitre 5). Cet exemple servira de support au projet de recherche : la cartographie de la chimiodiversité du vivant. La faisabilité de ce projet sera démontrée à l’aide de l’analyse des catalogues chimiques de substances naturelles puis illustrée au travers de modèles d’inférences permettant de relier les structures chimiques à leurs origines taxonomiques potentielles. Ces modèles seront corrigés à l’aide de données expérimentales. Dans ce cadre, le développement de méthodes d’acquisition d’empreintes chimiques et de traitements spécifiques sera également abordé (Chapitre 6).

4 Un exemple d'approche réductionniste pour la purification de substances naturelles

Le plateau des Guyanes (Guyane, Surinam, Guyana, Est du Venezuela) est un milieu de très forte biodiversité, floristiquement très riche, bien distinct du reste de l'Amazonie. La forêt guyanaise couvre 80 000 km^2 , soit 95 % du territoire à l'exception d'une mince frange côtière. C'est une forêt dense, sempervirente, ombrophile, de type tropical humide. Elle représente 1 % de la superficie de l'ensemble forestier amazonien. Ce milieu forestier se caractérise par une apparente homogénéité du couvert végétal qui cache de fortes variations spatiales de structures et de composition (forêts sur sols hydromorphes, sur schistes, sur quartzites, sur conglomérats, sur roches éruptives basiques, sur roches cristallines, sur cuirasses latéritiques ; forêts submontagnardes à nuages ; forêts basses sèches, cambrousses …). Une diversité importante de tous les organismes vivants y est présente, notamment les végétaux ligneux. Il est recensé plus de 5 625 espèces de végétaux vasculaires, dont 1 200 espèces d'arbres (contre 136 en France métropolitaine), plus de 150 espèces de mammifères, 132 espèces de reptiles, 109 espèces d'amphibiens, 199 espèces d'oiseaux et plus de 400 000 espèces d'insectes (World Resources Institute, 2003). La richesse floristique de ces forêts est renforcée par la faible capacité d'adaptation de certaines espèces, ce qui explique la modestie de leurs aires de répartition et le taux d'endémisme relativement élevé pour ces forêts [13].

4.1 Un site de bioprospection : la piste Saint-Élie

Situé à quelques kilomètres de Sinnamary, ce dispositif a été mis en place entre 1999 et 2002 par l'équipe de l'unité S84 Biodival de l'IRD, en Guyane. Durant ces trois années de prospections et d'identifications botaniques, les 7 hectares de forêt répartis le long d'un sentier (« la piste Saint-Élie) ont pu être caractérisés. Tous les individus ayant un diamètre supérieur ou égal à 10 cm, ont été répertoriés selon la méthode des quadras. Sur cette parcelle nous avons mesuré qu'un hectare de forêt primaire peut contenir plus de 200 espèces d'arbres (dbh> 10 cm de diamètre).

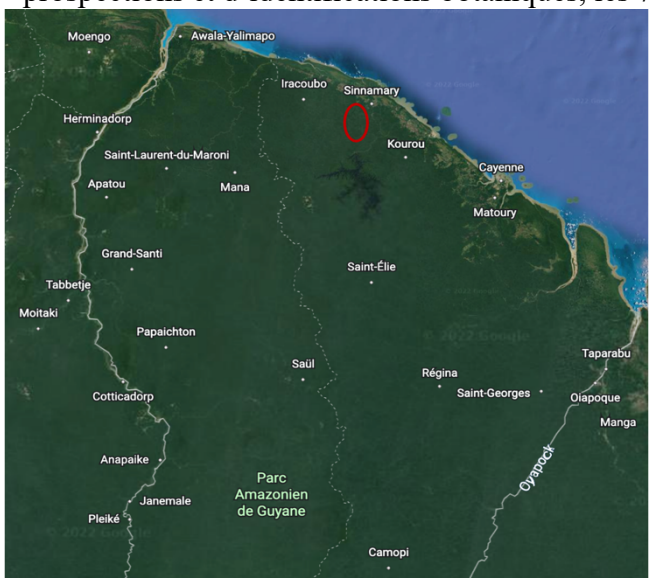


Figure 2: vue satellite de la Guyane française, l'emplacement du dispositif de la piste Saint-Élie est entouré de rouge

Nous avons compté près de 400 espèces pour 7 hectares contigus, avec seulement 9 espèces dominantes (>100 individus) et plus de 200 espèces représentées par un seul individu. Ce dispositif a été utilisé pour assurer la collecte de nombreux échantillons notamment pour des études phytochimiques du genre *Siparuna* (Monimiaceae) et des espèces de la famille des Clusiaceae.

4.2 Illustration d'une approche réductionniste : Étude phytochimique des arbres de la famille des Clusiaceae du plateau des Guyanes

Mes travaux de thèses se sont focalisés sur l'étude des potentialités phytothérapeutiques de quelques représentants de la famille des Clusiaceae récoltés sur le dispositif de la piste Saint-Élie décrit plus haut. Ces travaux ont été effectués à l'Institut de Chimie des Substances

Naturelles dans l'équipe de la Directrice de Recherche Françoise Guéritte sous la co-direction des Drs Marc Litaudon et Christian Moretti.

L'objectif de cette étude était d'identifier des substances naturelles actives sur l'agent du paludisme, *Plasmodium falciparum*, au travers d'une approche réductionniste largement utilisée dans le domaine. Les principales étapes sont résumées ci-dessous (Figure 4) :

1) **Sélection des taxons d'intérêt** selon des critères prédéfinis qui peuvent par exemple se baser sur des études ethnopharmacologiques ou des critères chimiotaxonomiques. Des essais *in vitro* sur la cible biologique d'intérêt permettront de choisir les extraits les plus pertinents pour la suite de l'étude.
2) **Fractionnement bioguidé** : Un fractionnement grossier est réalisé, suivi de nouveaux essais biologiques afin de focaliser la purification sur la fraction la plus active.
3) **Purification des composés actifs** : Un fractionnement plus fin permettra d'isoler les composés de la fraction. Leurs puretés et leurs élucidations structurales seront réalisées par des méthodes analytiques et spectroscopiques (MS et RMN essentiellement).
4) **Essais biologiques sur substances purifiées** : Les essais biologiques *in vitro* et *in vivo* seront réalisés sur les produits purs afin d'en déterminer leurs mécanismes d'action.

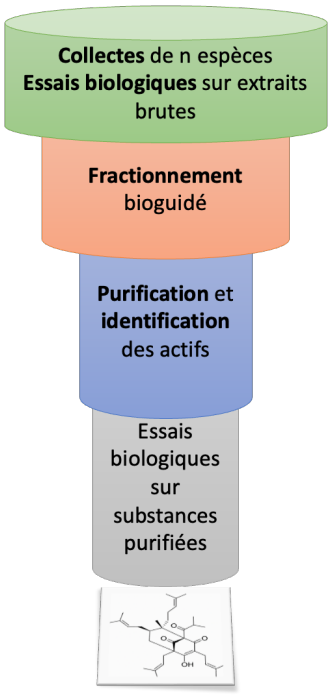


Figure 3:approche réductionniste de purification et d'identification d'actif par bioguidage

Cette approche en « entonnoir » permet de cibler les composés actifs en se basant sur les résultats des essais biologiques successifs. Ainsi, les composés isolés à la fin de ce processus itératif long seront largement dépendant de l'essai utilisés (souche, conditions, reproductibilité…). Des essais supplémentaires seront nécessaires afin d'élucider le mécanisme d'action des composés purifiés (cible protéique, cytotoxicité, activité *in vivo* …).

4.3 Sélection des taxons et des extraits d'intérêts

Dans le cadre de cette étude, nous nous sommes focalisés sur la famille des Clusiaceae tant pour le nombre de représentants de cette famille sur le dispositif de la piste Saint-Élie que pour les différents usages tirés des études ethnobotaniques réalisées par l'équipe [13].
La sélection des espèces candidates s'est basée sur plusieurs critères (Tableau 5):

- Identification, disponibilité et accessibilité de la ressource végétale : c'est un critère déterminant dans ce genre d'étude sachant que l'identification botanique des espèces en forêt tropicale humide nécessite l'avis de spécialistes et la comparaison avec des herbiers de références. De nombreux genres en forêt guyanaise ne sont pas identifiés au niveau des espèces. La disponibilité et l'accessibilité de la ressource sont aussi un critère essentiel, notamment pour pouvoir assurer des collectes multiples sachant qu'un milieu forestier présente une dynamique importante au cours du temps. Le dispositif Saint-Élie décrit plus haut permet de remplir ces critères de sélection en assurant une traçabilité du matériel végétal *in situ*. L'ensemble des spécimens récoltés ont fait l'objet d'un dépôt à l'Herbier de Guyane (https://herbier-guyane.ird.fr/).
- **La bibliographie phytochimique disponible** : qui doit être peu documentée afin de maximiser la découverte de nouveaux squelettes. La recherche bibliographique a été menée sur Chemical Abstract Service (interrogé en 2006).

- **Activité biologique sur *Plasmodium falciparum*** (souche chloroquinorésistante FcB1) : Les extraits réalisés à l'aide de solvants organiques doivent présenter une activité marquée afin d'être sélectionnée. Ces essais ont été réalisés à l'Institut Pasteur de Cayenne par le Dr Éric Deharo selon le protocole de Desjardins et al. [14].

Tableau 1: Critères de sélections des espèces de la famille des Clusiaceae

FAMILLES	Espèce	Utilisation traditionnelle[38]	Biblio CAS	Résultat bio P.f in vitro
CLUSIACEAE	*Caraipa densifolia*[1]	-	1	x
	Platonia insignis[1]	Fruit comestible ("bacuri")	12	+++
	Rheedia acuminata[1]	relaxant musculaire	2	x
	Rheedia benthamiana[1]	relaxant musculaire	5	x
	Symphonia globulifera[1]	Dermatoses	49	-
	Symphonia sp.[1]	-	0	+++
	Moronobea coccinea[1,3]	Dermatoses	0	+++
	Tovomita sp[1]	-	6	+
	Mahurea palustris[1]	Poison	1	-
	Vismia macrophylla[1]	Dermatoses	6	x

1: collection piste saint-élie, 3: Saul, x: non testé, -: pas d'activité, +: activité modérée, +++: activité importante

L'inventaire des arbres de la famille des Clusiaceae présents sur la piste de saint Élie a permis de répertorier 10 espèces représentées par au moins 2 individus de 20 à 40 mètres de hauteur. L'étude bibliographique sur Chemical Abstract Service (interrogé en 2006) ainsi que les résultats des essais sur *P. falciparum* ont **abouti à la sélection de 2 espèces : *Moronobea coccinea*** et ***Symphonia* sp**. Cette dernière a été identifiée comme proche de *Symphonia globulifera* sans toutefois pouvoir lui être totalement affilié.
Ces premiers résultats ont été confortés par une seconde série d'essais biologique réalisée sur l'ensemble des espèces et plusieurs organes afin d'affiner la sélection des extraits méritant une poursuite de leurs études phytochimiques. En plus de leur activité sur *P. falciparum,* les extraits ne devaient pas présenter de cytotoxicité.
Ainsi **les extraits du latex de *M. coccinea* et les écorces de racines de *S.* aff *globulifera* et les écorces de tronc de *R. acuminata* ont été retenues** pour une étude phytochimique biodirigée**.**

4.3.1 Fractionnement bioguidé

À titre d'exemple, l'extrait méthanolique de latex de *M. coccinea* a été dans un premier temps fractionné par MPLC (chromatographie liquide moyenne pression) pour donner 9 fractions sur colonne en phase inverse selon un gradient sur 115 minutes acétonitrile/Eau. Les fractions ont été testées sur *P. falciparum* à une seule concentration (10 µg/mL). Seules les fractions ayant montré les plus fortes activités antiplasmodiales ont été retenues pour une séparation des composés sur HPLC-semi-préparative (fraction 5 & 6, figure 4). Chacun

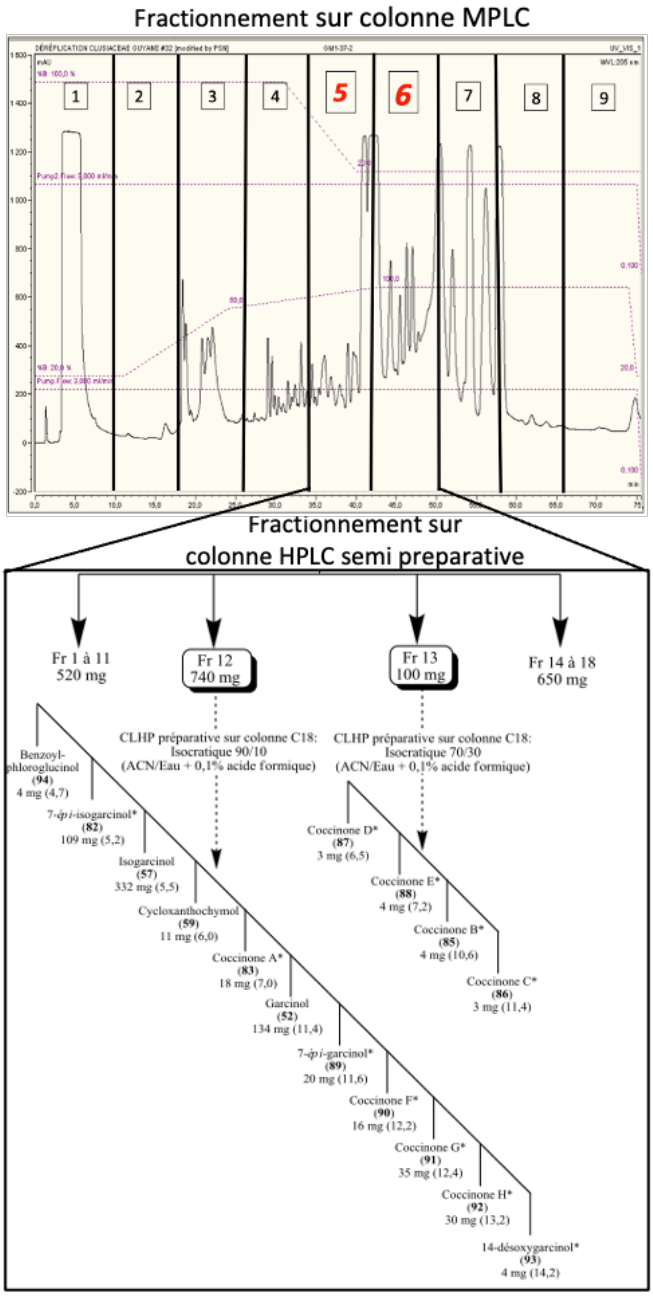


Figure 4: Schéma simplifié du fractionnement bioguidé suivi pour le latex de M. coccinea

des pics détectés (détection UV, À = 220 nm) en HPLC semi-préparative ont été collectés séparément et plusieurs injections successives ont été nécessaires pour accumuler suffisamment de matière afin de pouvoir effectuer une élucidation structurale et des essais biologiques avec les produits purifiés.

Cette approche réductionniste a été appliquée aux trois extraits sélectionnés pour leurs activités contre *P. flaciparum*. **Ces travaux ont conduit à la purification et l'identification de 31 composés** que nous avons par la suite caractérisés **[15,16]**.

4.3.2 Purification et élucidation structurale des composés isolés

Chacun des composés a fait l'objet d'analyses spectroscopiques et spectrométriques afin d'en déduire sa structure et éventuellement sa stéréochimie. L'élucidation structurale de substances naturelles est un processus complexe qui implique l'utilisation de plusieurs outils et techniques permettant d'accéder à des informations parcellaires sur la molécule pure analysée. L'ensemble de ces données spectrales constituent un puzzle dont la résolution permettra d'appréhender le squelette final du composé. Ce processus implique plusieurs étapes successives résumées ci-après (figure 5).

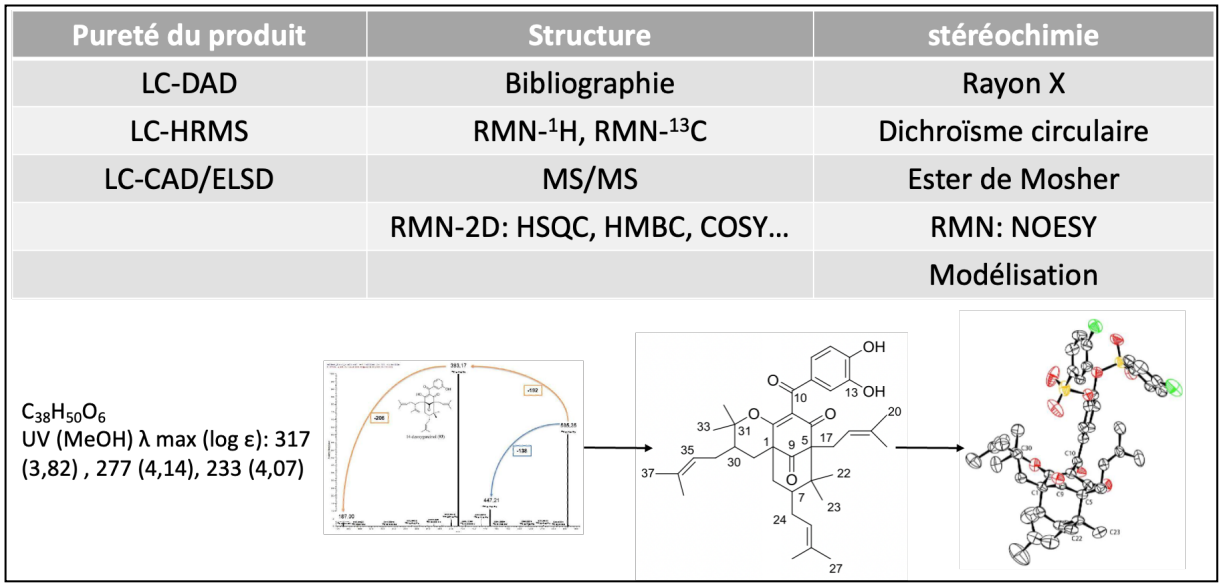


Figure 5: outils utilisés pour l'élucidation structurale avec l'exemple de l'isogarcinol purifié du latex de M. coccinea

L'un des points clés est une étude exhaustive de la bibliographie phytochimique de la famille ou du genre afin d'appréhender rapidement les squelettes possibles par comparaisons des signaux expérimentaux aux données de la bibliographie. Dans le cas de la famille des Clusiaceae, les études phytochimiques démontraient une majorité de benzoylphloroglucinols polyisoprénylés polycycliques (PPAPs) issus des voies de biosynthèse du shikimate (des polyacétates constituant le noyau benzoylphloroglucinol) suivie de prénylations successives. Un benzoylphloroglucinol trisubstitué se trouve dans une conformation favorable à l'attaque d'une unité DMAPP pour former un carbocation tertiaire favorisant la cyclisation en trois types de noyaux PPAPs (figure 6).

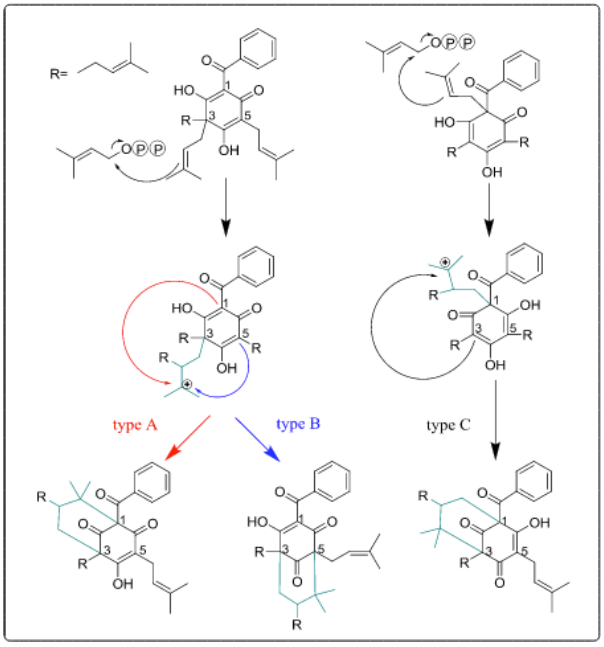


Figure 6: Schéma représentant l'étape finale de formation des différents noyaux PPAPs

Les 31 composés purifiés à l'étape précédente ont été identifiés. Parmi ces derniers, 20 présentaient des motifs terpéniques originaux autour des squelettes PPAPs de type B **[15–17]**. Ces modifications sont principalement dues au positionnement d'une chaine monoterpénique en C3 ou C5 possédant des isoméries hétérogènes, une cyclisation en C2 pour former un tetrahydropyrane ou tetrahydrofurane et l'acylation en C1 par un groupement pyrocatéchol ou phénol en position meta (Figure 7).

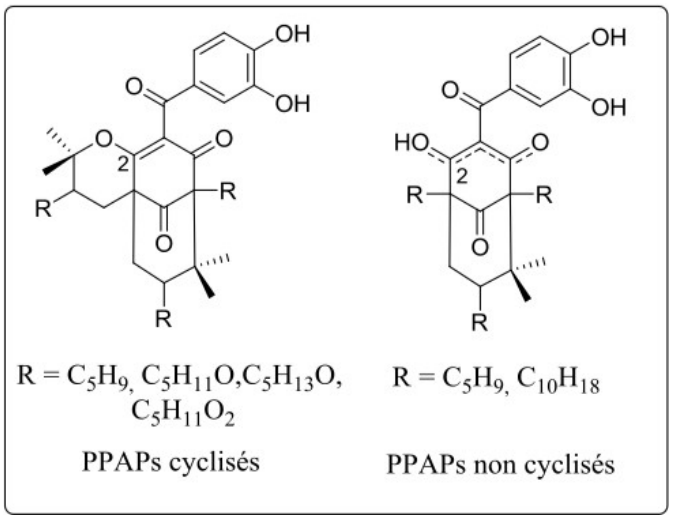


Figure 7: exemple de squelettes isolés des espèces de la famille des Clusiaceae. Pour plus de détails, se référer aux références citées 14 à 16

4.3.3 Essais biologiques sur les composés purifiés

Afin de confirmer l'activité biologique enregistrée préalablement dans les extraits de départs, des essais biologiques ont été réalisés sur la souche plasmodiale chloroquinorésistante FcB1. L'évaluation de la cytotoxicité des composés a été réalisée sur la lignée cellulaire MRC5 (à partir de fibroblastes de poumon fœtal) dans le laboratoire du professeur Philippe Grellier au MNHN de Paris. Les résultats sont présentés sur la figure 8.

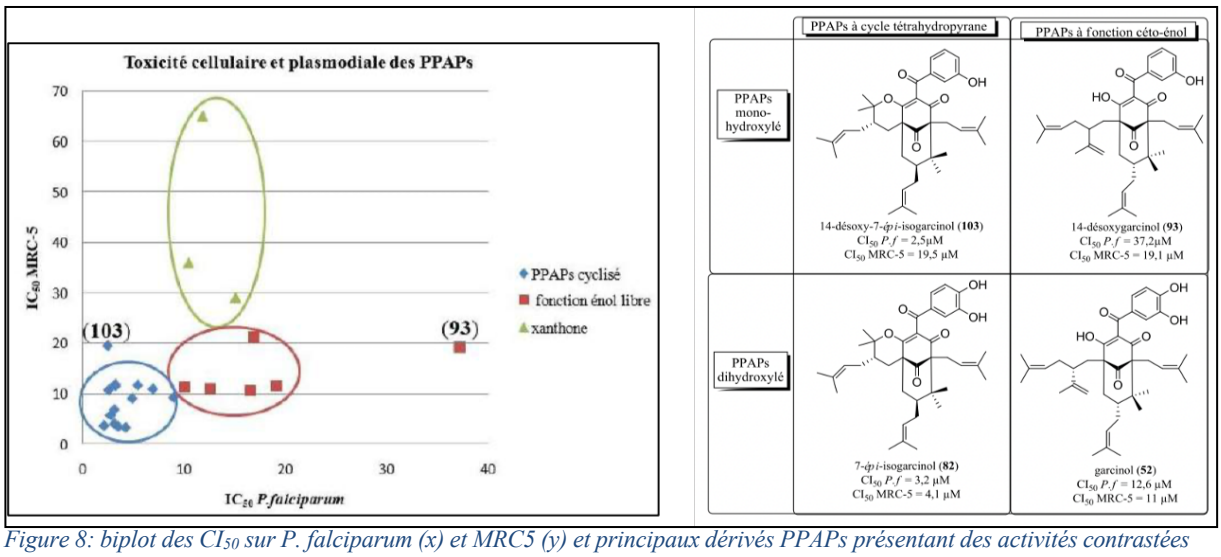


Figure 8: biplot des CI_{50} sur P. falciparum (x) et MRC5 (y) et principaux dérivés PPAPs présentant des activités contrastées

Cette représentation permet de dégager des relations structure/activité selon trois groupes de composés en fonction de leur toxicité cellulaire et plasmodiale :

- Les xanthones qui apparaissent comme les moins cytotoxiques avec une activité antiplasmodiale comprise entre 10 et 20 µM, ce qui correspond aux données de la littérature.
- Les PPAPs possédant une fonction céto-énol libre qui sont les moins actifs avec une CI_{50} comprise entre 10 et 20 µM (e.g. garcinol-52). Le dérivé **93** se distingue de ce groupe par une moindre activité (37,2 µM). La présence du groupement phénol/catéchol joue donc un rôle important dans l'activité antiplasmodiale de ces composés.
- Les PPAPs cyclisées sur l'oxygène en C-2 ou C-4 qui possèdent une activité plus marquée avec des CI_{50} inférieures à 10 µM. Ceci confirme l'importance de l'hétérocycle en position 2 pour l'activité antiplasmodiale de ces composés. Le dérivé 103 se distingue des autres PPAPs cyclisés par une cytotoxicité relative moins importante du fait de l'acylation par un cycle phénolique et un hétérocycle en position C2.

4.3.4 Bilan quantitatif

Une valorisation potentielle de produits purifiés se heurte généralement à un écueil important en chimie des substances naturelles : le bilan de masse de purification des produits est généralement très faible, les composés solubles représentant une fraction minoritaire des matrices végétales. Par exemple dans le cas des composés purifiés du latex de *M. coccinea*, le composé majoritaire atteint 10 % de rendement (isogarcinol) et le second 5 % (garcinol). Les autres composés isolés sont compris entre 1 et 0,1 % de la masse de départ. Or ce sont ces composés minoritaires qui renferment les squelettes aux motifs terpéniques originaux tels que ceux de la série des coccinones (figure 10, haut).

On peut observer une distribution similaire pour l'investigation de l'espèce *S.* aff. *Globulifera*. Dans ce cas le produit majoritaire est le 7-*épi*-isogarcinol qui est dix à cent fois plus concentré que les autres produits isolés de cette matrice (Figure 10, bas). Les rendements d'extractions observés sont aussi plus faibles que ceux du latex.

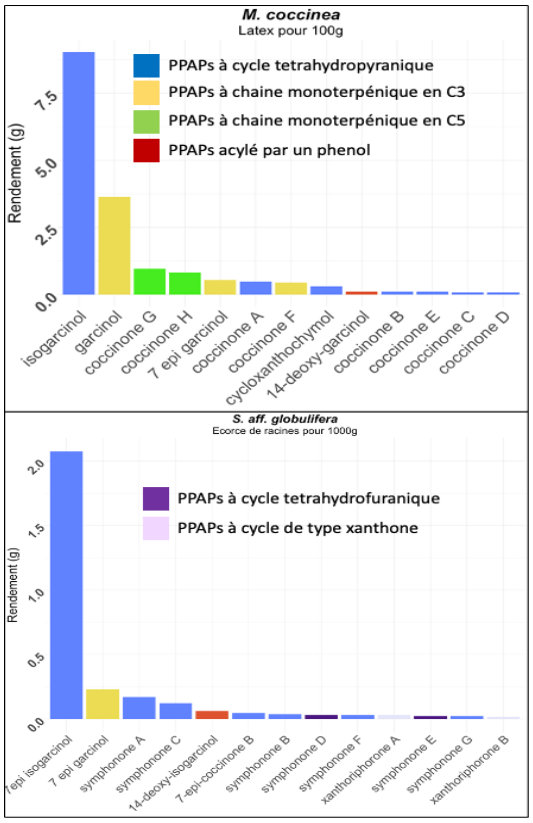


Figure 10: bilan de masse des composés purifiés à partir du latex de M. coccinea (haut) et des écorces de racine de S. aff. Globulifera (bas).

Le schéma d'une plasticité enzymatique conduisant à des séries de sous-produits minoritaires à partir de quelques synthons majoritaires a été observé dans ces deux exemples. Cette distribution est fréquemment retrouvée dans les extraits de plantes, comme en témoigne l'utilisation intensive des réseaux de similarité spectrale dans la littérature qui permettent d'agréger des spectres de fragmentations MS/MS et de refléter ainsi la présence de motifs structuraux très proches. Un exemple peut être trouvé ici :**[18]**.

4.3.5 Bilan de l'approche réductionniste pour l'isolement de substances naturelles :

Les travaux résumés en partie dans ce paragraphe ont nécessité 3 ans de travaux et ont bénéficié de la richesse de l'écosystème de recherche de l'Institut de Chimie des Substances naturelles (ICSN-CNRS). Les étapes les plus consommatrices de ressources étant les multiples fractionnements et purifications afin d'isoler les composés du reste des matrices étudiées. Le rassemblement du puzzle des différents spectres permettant d'inférer la structure bi- et tridimensionnelle des composés purifiés est aussi une étape très importante en termes de temps humain, le tout étant jalonné par les essais biologiques permettant d'orienter les efforts d'isolement (figure 9).

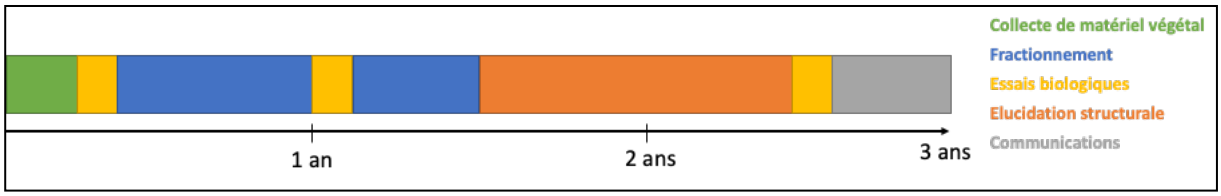


Figure 9: diagramme de gant simplifié des travaux de purifications par bioguidage sur les extraits de la famille des Clusiaceae

L'approche réductionniste déployée au cours de ces travaux a conduit à la purification de plusieurs dizaines de produits au **cout d'un temps de manipulation conséquent** et d'une mobilisation importante de plusieurs technologies parfois très couteuses. De plus, les résultats restent **incertains en termes de débouchés**. Dans le cadre de cette étude, les PPAPs isolés ont montré un indice de sélectivité *in vitro* compris entre 20 et 1 ce qui apparait insuffisant pour une poursuite de l'étude sur la cible *P. falciparum* souche FcB1. Les rendements de purifications ont été faibles à très faibles, ce qui limite une valorisation potentielle des structures sans le relais de la synthèse ou d'outils biotechnologiques. Bien que ces travaux aient permis d'améliorer les connaissances sur la phytochimie de ces espèces tropicales, leurs valorisations sont restées cantonnées aux études académiques.

Les produits majoritaires étaient déjà identifiés dans la littérature et donc reconnaissables par des méthodes spectroscopiques sans purification préalable. L'utilisation de ces outils peut permettre d'identifier les composés connus et de focaliser les investigations phytochimiques sur les motifs originaux. Ce genre d'approche, dite déréplicative, est une des plus utilisés en métabolomique appliqué aux substances naturelles.

5 La déréplication par empreintes spectrométriques

La déréplication consiste à identifier l'empreinte spectrométrique caractéristique d'un composé au sein d'une matrice complexe. Cette approche nécessite l'acquisition de spectres très informatifs, généralement à l'aide d'instrument de spectroscopie en haute résolution (RMN, MS haute résolution, MS en tandem…), afin d'améliorer la qualité des identifications. Étant donnée la complexité des mélanges, les composés sont généralement séparés par des méthodes chromatographiques en amont des mesures spectrométriques. L'analyse d'un mélange complexe produit une matrice de 2 à n dimensions dont les données chromatographiques et spectrales seront caractéristiques de chaque composé pour un système d'analyse donné (figure 11 : RT × intensité × *m/z* × MS/MS).

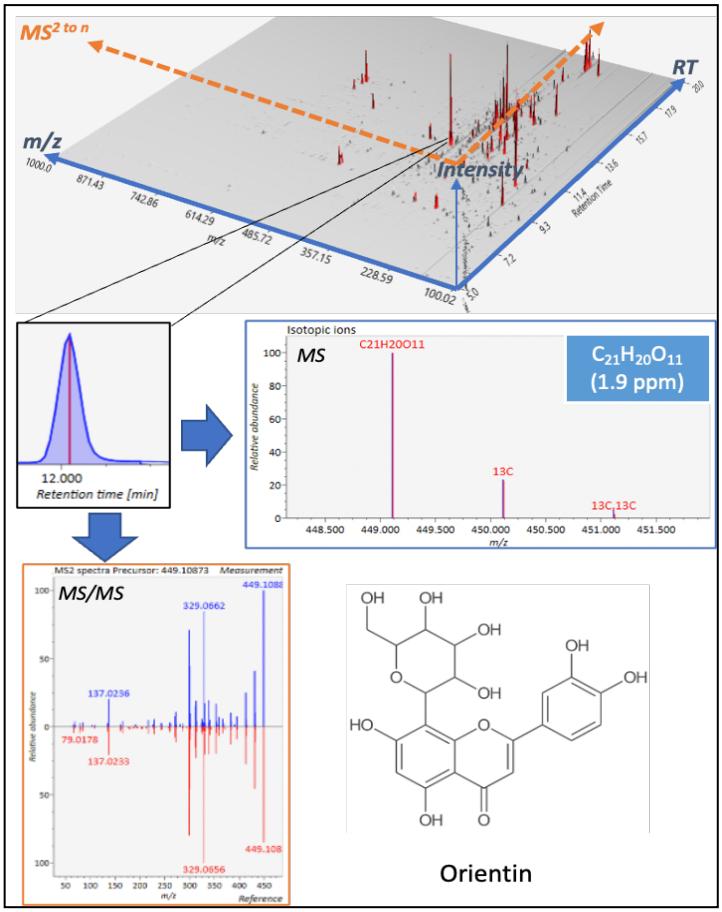


Figure 11: Déréplication par LCMS-MS/MS de l'orientine dans un mélange complexe

5.1 Un exemple de déréplication par chromatographie sur couche mince : étude des espèces du genre *Siparuna* du plateau des Guyanes

Dans le cadre d'une étude sur les inhibiteurs de protéines kinase DIRK1A et CDK1/Cyclin B, le dispositif de la piste Saint-Élie a été utilisé pour collecter 6 espèces du genre *Siparuna* (S. *guianensis, S. poepiggi, S. descipiens, S. cristata, S. pachyantha et S. cuspidata*) de la famille des Monimiaceae afin d'exploiter leurs variabilités phytochimiques. Cette famille est connue pour ses dérivés d'alcaloïdes aporphiniques. Ainsi, lors de cette étude, 11 alcaloïdes ont été isolés à l'aide du système chromatotron (CCM préparative par centrifugation). Puis testés pour leurs inhibitions sur des protéines kinases. La figure 13 montre les traces des profils chimiques de *totum* alcaloïdique effectués par chromatographie sur couche mince (CCM silice, parcoure de 20 cm, $CHCl_3$/MeOH 90/10, dépôt 20 ug). Les capacités d'identification en CCM sont très limitées sans l'aide de standards purifiés. Ici, la déréplication s'est effectuée à l'aide de l'élution des standards isolés et comparaison de leurs rapports frontaux respectifs (non montré sur la figure 13). Les classes chimiques peuvent être identifiées par des réactifs spécifiques complétés par la mesure de l'absorbance UV/Vis. Ici, le réactif de Dragendorff a été utilisé pour révéler spécifiquement les alcaloïdes. L'importance des taches est proportionnelle à la quantité de composés, les concentrations relatives de chacun des constituants peuvent donc être appréhendées.

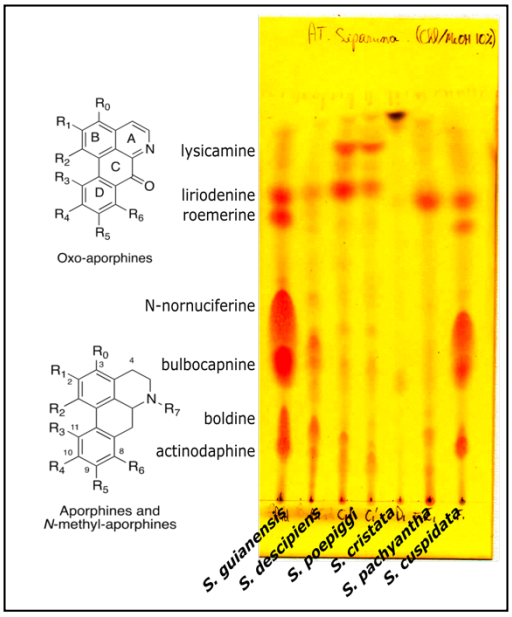


Figure 13: Profils CCM d'extraits d'alcaloïdes totaux des espèces du genre Siparuna

L'approche CCM est encore largement utilisée notamment en contrôle qualité ou pour le suivi des profils de fractionnement. En effet son coût très limité et sa facilité de mise en œuvre constituent ses principaux atouts. Cependant, en termes d'annotation, cette approche se cantonne aux grandes classes chimiques en l'absence de standards analytiques de références.

Il est remarquable de noter des différences significatives entre les profils alcaloïdiques qui semblent caractéristiques de chaque espèce. Ces variations sont autant quantitatives (richesse des profils pour *S. guianensis* et *S. cuspidata*) que qualitatives, même pour des espèces peu différenciées (figure 13). En effet *S. guianensis* et *S. poepiggi* ont un phénotype très proche qui se distingue uniquement au niveau de la taille des sépales des inflorescences (figure 14). La figure 13 montre que l'oxo-aporphine lysicamine est largement majoritaire chez *S. poepiggi* alors que *S. guianensis* est plus riches en dérivés *N*-methyl-aporphines (roemerine, boldine).

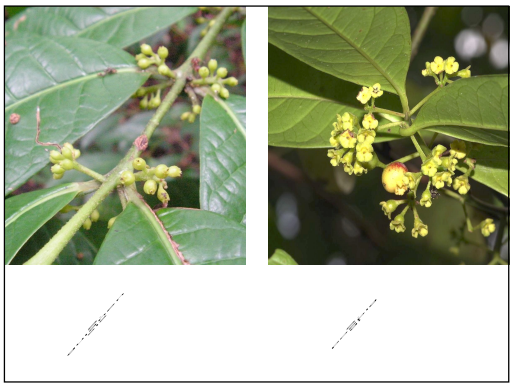


Figure 14: Photographies des inflorescences de S. guianensis *et* S. poepiggi

La liriodenine et la lysicamine se sont révélées être les meilleurs inhibiteurs des kinases de cette série (respectivement, IC_{50} de 2, 1 et 3,4 uM) **[19]**.

5.2 Un exemple de déréplication par spectrométrie de masse en tandem des extraits de la famille des Clusiaceae.

Les produits décrits au paragraphe 3.3.1 ont fait l'objet d'une analyse de fragmentation par spectrométrie de masse en tandem. L'analyse des spectres MS/MS a permis de dégager des modèles de fragmentation pour cette classe de composés (figure 15).

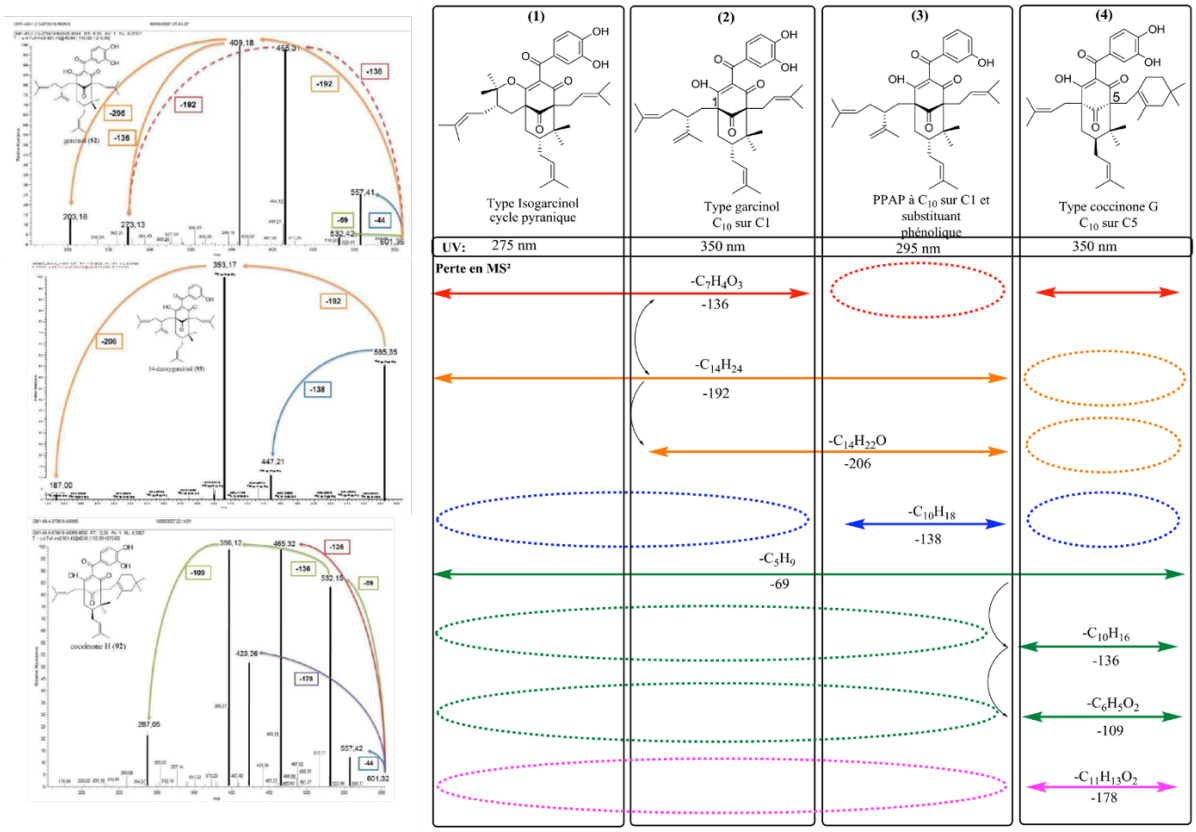


Figure 15: spectre de fragmentation MS/MS (gauche) et tableau des pertes de neutres observées pour chaque type de PPAPs (flèches) ou leur absence (ronds pointillés) ainsi que leur maximum d'absorption en spectre UV.

Par cette étude nous avons pu mettre en évidence que les pertes de masse sur charge observées en MS/MS sont étroitement liées aux motifs retrouvés autour du noyau phloroglucinol. Chaque grande classe de PPAPs peut donc être déduite à l'aide des spectres de fragmentation et du maximum d'absorption en UV. L'exploitation de ces deux informations orthogonales augmente la pertinence de l'annotation. L'acquisition des profils LC-UV-MS-MS/MS peut donc être suffisante pour appliquer cette approche de déréplication à un extrait ou une fraction. **L'analyse fine des chromatogrammes obtenus permet l'annotation de plusieurs pics chromatographiques sans passer par les étapes de purifications** (figure 16).

On comprend ici les potentialités de ce type d'approche. À titre d'exemple, la comparaison des profils de l'extrait de latex et de feuilles de *M. coccinea* met en évidence une richesse bien plus importante dans le latex, notamment dans les dérivés PPAPs à cycle tétrahydropyranique qui ont démontré une bonne activité *in vitro* contre *P. falciparum*. Un autre intérêt peut être **la mise en évidence de schémas de fragmentations originaux**, comme chez les écorces de racines de *S.* aff. *globulifera* qui ont permis d'orienter les travaux d'isolement conduisant à la purification de la série des xanthoriphorones **[16]**.

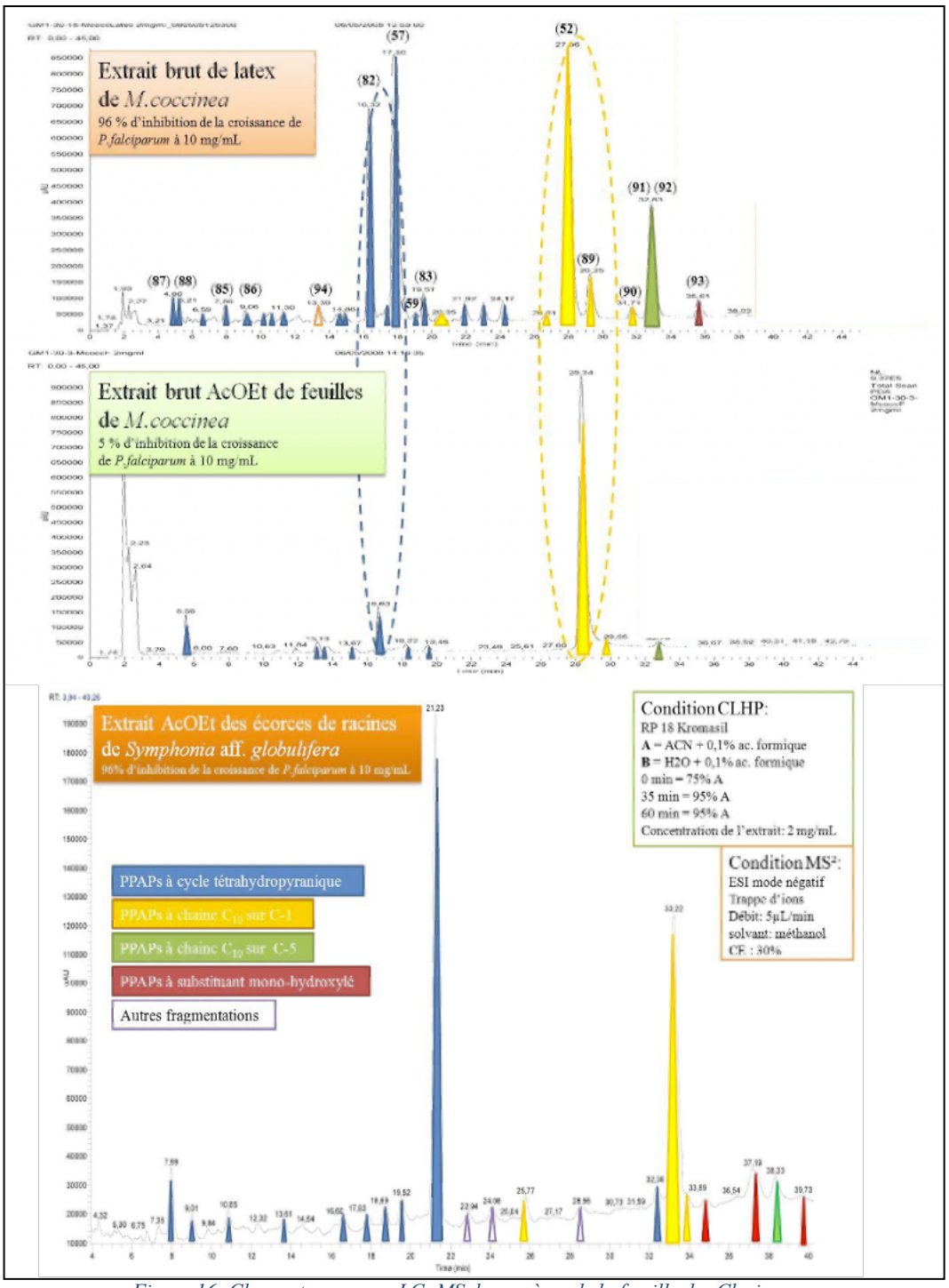


Figure 16: Chromatogrammes LC -MS des espèces de la famille des Clusiaceae

Cette méthode de déréplication couplée à la comparaison des profils chimiques acquis dans les mêmes conditions (même instrument et même concentration d'extrait) ouvre la voie aux approches métabolomiques. Des analyses statistiques multivariées sont nécessaires afin de mettre en évidence les différences les plus significatives entre les profils chimiques. Un exemple basique de ces approches peut être trouvé dans les travaux de thèse de Kevin Cottet (Université Paris Descartes) sur l'analyse des empreintes de *Symphonia globulifera* qui s'est inspiré des résultats détaillés dans ce paragraphe. Ses travaux ont utilisé une approche de déréplication similaire afin de comparer les empreintes chimiques des organes de cette espèce [20], puis de rechercher des chimiomarqueurs discriminant l'espèce d'origine africaine et guyanaise [21].

5.3 Un exemple de déréplication par Résonnance Magnétique Nucléaire : Caractérisation de l'origine géographique des sous-produits du genre *Citrus*

Dans le cadre d'un projet de collaboration entre l'université de Genève et l'entreprise de fragrance Firmenich visant à établir des méthodes d'authentification d'origine géographique des huiles essentielles des fruits du genre *Citrus*, la partie non volatile (CPLOR : Cold Pressed Lemon Oil Residue) obtenue à partir de lots industriels provenant d'Argentine et d'Italie a été analysée par RMN 1D et 2D. Les CPLOR sont riches en dérivés de types coumarines et flavonoïdes. Les empreintes en RMN du proton ont été utilisées pour les analyses multivariées afin d'identifier des signaux caractéristiques pouvant prédire l'origine des échantillons.

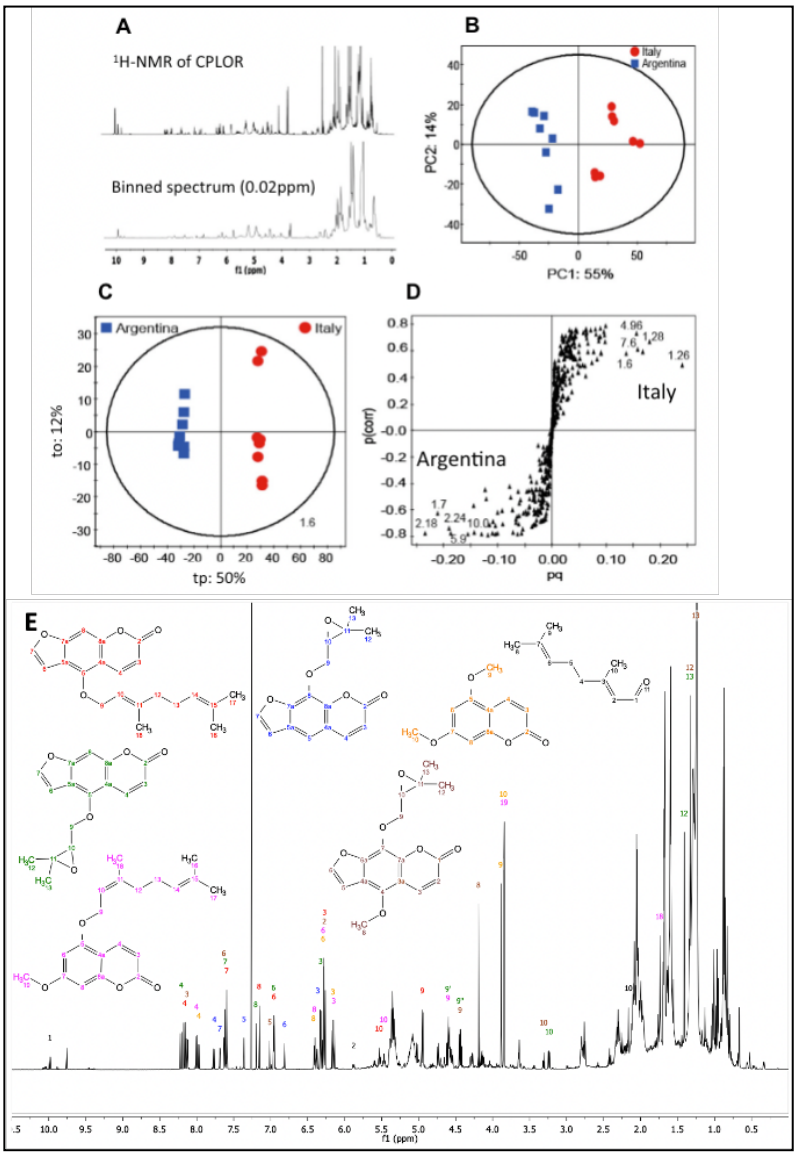


Figure 17: Empreintes en RMN 1H des fractions non volatile d'huiles essentielles de Citrus *sp.; A: RMN ^{1}H brut et en bucket; B: ACP échelle pareto; C: OPLS-DA, D: S-plot du modèle OPLS-DA ; E : Attribution des signaux sur le spectre RMN ^{1}H*

Le traitement des spectres RMN du proton passe par une étape d'alignement et de normalisation suivi d'un « bucketing » ou découpage du spectre en portion de 0.02 ppm dans ce cas (Figure 17A). Ces étapes préliminaires sont indispensables aux analyses multivariées qui débutent par une analyse en composante principale (ACP) afin d'appréhender la distribution des échantillons dans l'espace de variance du jeu de données (Figure 17B). Les échantillons provenant d'Italie et d'Argentine sont clairement séparés sur l'axe 1 (55 % de la variance exprimé) témoignant

d'une proportion importante de « bucket » différenciant les 2 sites. L'axe 2 de l'ACP (14 % de l'ACP) sépare les échantillons indépendants reflétant la variabilité au sein des lots. L'extraction des buckets les plus significatifs s'est effectuée ici à l'aide d'une analyse discriminante orthogonale partielle des moindres carrés (OPLS-DA, figure 17C). Un S-plot a été utilisé pour projeter les buckets en fonction de leurs importances dans le modèle OPLS-DA. Ainsi, les buckets les plus extrêmes seront les plus pertinents pour séparer les sites géographiques (Figure 17D)

L'identification des composés par RMN est un processus complexe, notamment pour les mélanges. Dans le cas de cette étude, la comparaison avec des spectres de standards analytiques et l'acquisition de spectres en 2D (HMBC, HSQS, COSY) ont permis d'identifier 7 structures majoritaires dans les fractions analysées, principalement des dérivés coumariniques (Figure 17E). **La richesse des signaux enregistrés par la spectroscopie RMN permet une identification précise des composés d'un mélange. L'autre avantage sera d'utiliser l'intensité des signaux des protons afin d'évaluer les quantités respectives des composés dans les extraits.** Dans le cadre de cette étude, la méthode PULCON a été utilisée afin de quantifier les signaux caractéristiques de chaque composé [22]. Les résultats démontrent des quantités significativement plus importantes de citroptène et géranial dans les échantillons provenant d'Argentine alors que les échantillons italiens seront plus riches en bergamotine et 5-géranyloxy-7-methoxycoumarine (Figure 18). Ces analyses ont été complétées par des profilages LC-MS et une méthode d'hétérospectroscopie a été élaborée afin de corréler les signaux de ces deux méthodes d'acquisition. Les détails ne seront pas abordés dans le cadre de ce devoir, mais peuvent être trouvés dans la référence **[23]**.

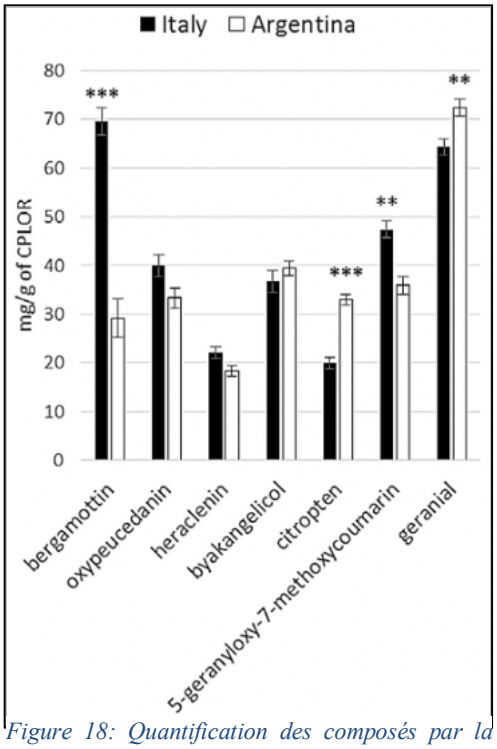


Figure 18: Quantification des composés par la méthode PULCON au sein des CPLOR provenant d'Italie et d'Argentine

5.4 Approche métabolomique par LC-MS : détection des produits antioxydants chez la violette de Toulouse par corrélation des profils chimiques et des essais biologiques.

Dans le cadre du projet région « Viola Tolosa » sur la violette de Toulouse (*Viola alba* subsp *dehnhardtii*) et plus particulièrement sur la détection de produits à activité antioxydante, une approche par métabolomique a été envisagée. Il s'agit ici d'exploiter la richesse des profils chimiques associée à des modèles statistiques de corrélations sur les activités antioxydantes mesurées à partir des extraits puis des fractions. Cette stratégie permet de réduire drastiquement les étapes de fractionnements /suivies d'essais bioguidés grâce à la mise en évidence précoce de composés fortement corrélés à l'activité biologique. **L'exploitation de cette liste de suspects oriente les travaux de purifications et accélère le processus d'identification d'actifs au regard des stratégies réductionnistes habituellement utilisées dans ce cas.**
La variabilité des profils chimiques est appréhendée en deux étapes successives qui servent à alimenter le modèle statistique de corrélation : i) variabilité des répliquas biologiques, ii) fractionnement grossier sur cartouche SPE en 5 à 6 fractions de polarité décroissante (Figure 19A). Les fractions de chaque individu sont profilées par LC-MS et évaluées pour l'activité biologique d'intérêt. La matrice de pics est injectée dans le modèle (variables x) afin d'être corrélée à l'activité biologique de chaque fraction (y = IC_{50}). La robustesse du modèle est

fortement influencée par le nombre de profils LC-MS et la qualité des mesures biologiques pour chaque fraction (Figure 19B). Le résultat de ces modèles est un classement des variables en fonction de leur corrélation avec l'activité biologique mesurée. Les variables fortement corrélées sont ciblées prioritairement pour les travaux de déréplication puis de purification (Figure 19C). Enfin, les composés purifiés sont de nouveau testés afin de confirmer qu'ils sont porteurs de l'activité biologique (Figure 19D).

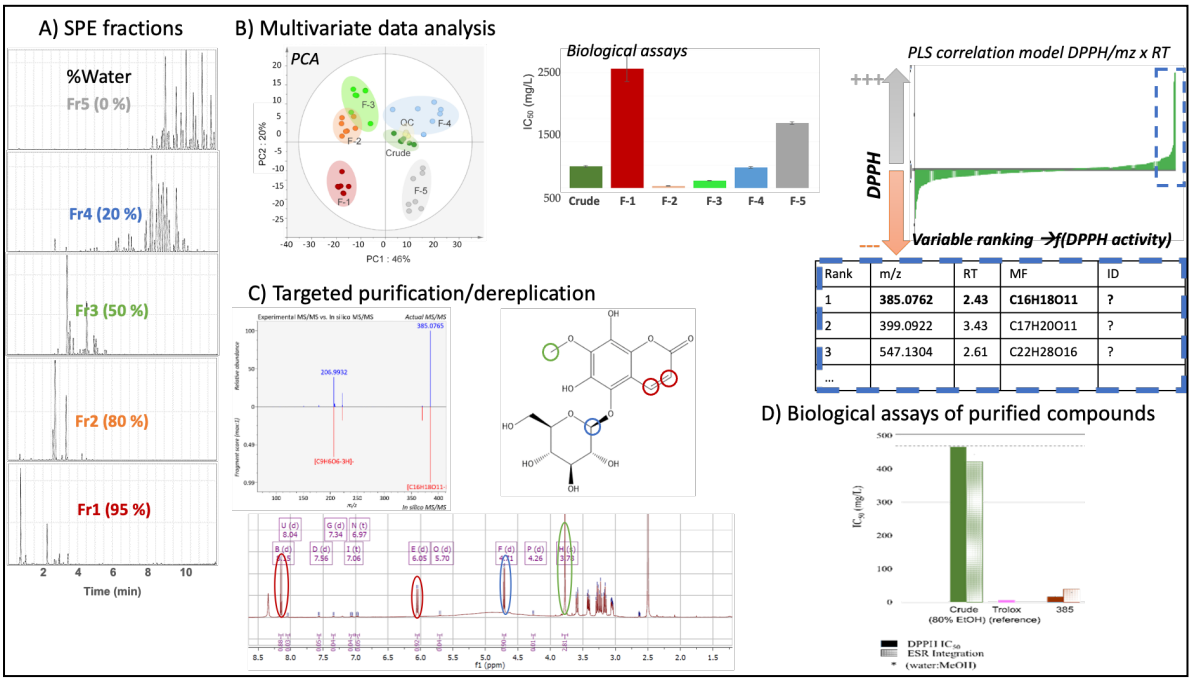


Figure 19: Stratégie de purification ciblée à l'aide d'une approche métabolomique

Cette stratégie permet (1) de réduire considérablement le temps nécessaire au processus d'isolement d'actifs d'extraits complexes, mais aussi (2) d'obtenir une connaissance plus globale de la diversité chimique de l'extrait. L'exploitation de ces données peut s'avérer utile dans le cadre de valorisation directe du mélange qui est un domaine en plein essor, notamment en parapharmacie. Dans le cadre de cette étude, le profil chimique de la violette de Toulouse à révéler une richesse importante en flavonoïdes glycosilés ainsi qu'en dérivés coumariniques. Ces derniers se sont révélés les plus actifs sur le test DPPH (Diphenylpicrylhydrazyle) utilisé ici **[18]**.

5.5 Les approches déréplicatives : intérêts et limitations

L'objectif final de la déréplication sera l'identification de l'ensemble des signatures spectrales que l'on peut détecter dans un système d'acquisition [24]. L'identification d'un composé s'effectuera par la comparaison de ses données chromatographiques et/ou spectrométriques avec un standard de référence. L'une des limites principales de cette approche sera la disponibilité des standards analytiques qui ne représentent qu'une fraction de la chimiodiversité du vivant (Cf paragraphe 6.5). En l'absence de standards permettant de comparer les signaux, le terme d'annotation sera utilisé. Une annotation implique une note de confiance ou de précision en fonction des données spectrales interprétables permettant d'inférer une identification. Celle-ci sera donc une **hypothèse basée sur un faisceau d'indices concordants** rassemblant idéalement plusieurs mesures spectrométriques orthogonales. Dans tous les cas, contrairement à l'identification, l'annotation implique 2 à n possibilités pour une signature spectrale.

Ce système d'inférence par faisceaux de preuves a conduit à de nombreuses ambiguïtés dans les résultats des publications utilisant ces approches de déréplication. Un formalisme a été proposé par plusieurs équipes de recherches permettant d'apprécier le niveau de confiance pour

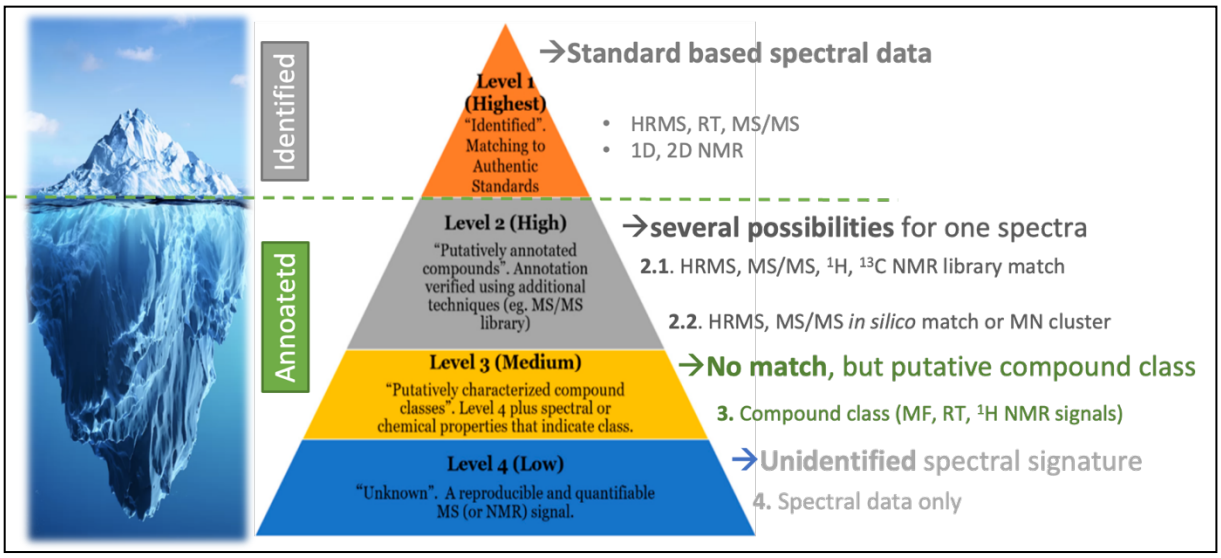


Figure 12: Niveau d'annotation communément admis dans la littérature pour la déréplication par spectrométrie de masse et/ou RMN d'après les références 19 à 21.

une annotation donnée [25–27] (figure 12). L'enjeu ici sera de pouvoir comparer les méthodes d'acquisition et les résultats d'annotation qui en découlent sur des matrices similaires (plasmas issus de cohortes cliniques, plantes modèles…). Bien que l'adoption de cette échelle ait rencontré un consensus relativement large au sein de la communauté, elle fait encore l'objet de nombreux débats et activités de recherche portés principalement par la Metabolomics Society.

Les exemples cités ci-dessus illustrent la pluralité des approches possibles pour acquérir des empreintes chimiques de matrices complexes. L'outil analytique sera choisi en fonction de l'objectif de l'étude, car **aucune méthode à l'heure actuelle ne permet d'obtenir une cartographie complète du métabolome à la fois qualitative et quantitative.** Une inspection rapide du site de dépôt de données métabolomiques « Metabolights » (https://www.ebi.ac.uk/metabolights/, consulté le 08/09/22) montre que 85% des études déposées sont réalisées par spectrométrie de masse et 15% par RMN.
Le tableau ci-dessous résume les avantages et inconvénients des approches abordées dans cette partie.

	Reductionist	Holistic		
Detection	Multiple	CCM	NMR	MS
Metabolome coverage	Few compounds	Tiny	Medium	Large
Annotation support	Puzzle of experimental data	AS	AS/Dbase/InSilico	AS/Dbase/InSilico
Annotation quality	Very good	Poor	Good	Medium
Quantification	According to purification yield	Overview	Good	Bad
Bioactivity linkage	Driving force	Bioautography	Statistical analysis	Statistical analysis
Biochemical knowledge	Poor	Poor	Medium	Good
Valorization	Poor	Quality control	Quality control/system biology	Quality control/system biology
Time consuming	Long	Short		

Tableau 2: Avantages et inconvénients des différentes approches illustrées dans cette partie (AS : Analytical Standard)

Les approches réductionnistes appliquées en chimie des substances naturelles ont conduit à la purification de dizaine de milliers de composés généralement orientée sur une activité biologique d'intérêt. C'est un processus long et couteux en termes de manipulation et de ressource (Cf. 3.3.5). L'apport des approches holistiques peut accélérer ce processus comme démontré dans l'exemple du paragraphe 4.4.
Concernant les approches sans a priori, bien que la RMN présente de nombreux avantages, notamment en termes d'identification et de stabilité des mesures, les possibilités de couverture du métabolome sont réduites du fait de sa sensibilité limitée et des difficultés de couplage aux méthodes chromatographiques. La spectrométrie de masse couplée à la chromatographie gazeuse ou liquide constitue l'une des méthodes les plus sensibles actuellement autorisant l'accès à un très grand nombre de métabolites en une injection. **L'interopérabilité des méthodes d'acquisition, les incertitudes sur les annotations et les possibilités quantitatives constituent cependant des écueils importants au développement de ces approches.**
On peut distinguer deux facteurs d'évolution majeurs ces dernières années :
- la montée en gamme constante des instruments LC-MS favorisants principalement la résolution, la fréquence d'acquisition des spectres et la sensibilité. Ces innovations sont essentiellement portées par les industriels.
- le développement des outils de traitement des signaux principalement axé sur la déconvolution des spectres et leurs annotations particulièrement développées dans le monde académique. Ce dernier volet sera abordé en détail dans les paragraphes suivants.

6 L'annotation par spectrométrie de masse : un outil pour répertorier la diversité chimique du vivant

Une des conséquences remarquables de la complexité du métabolome sera la multiplication des méthodes d'acquisition des empreintes chimiques tant du point de vue des approches séparatives (colonne, gradient, solvant…) que des détecteurs de masses. Pour ces derniers, il existe une importante variété de méthodes d'acquisitions (data dependant/independant analysis, source d'ionisation, mobilité ionique…) auxquelles s'ajoute la dégénérescence des signaux que l'on rencontre au cours de la nébulisation et de l'ionisation.

L'appréhension de cette chimiodiversité par **les approches de chimie analytique s'avère donc particulièrement composite**, **limitant l'interopérabilité et la standardisation des méthodes d'acquisition, des traitements des chromatogrammes ainsi que des annotations.** De fait, de nombreuses recherches du milieu académique tentent de normaliser les processus de traitements des données de métabolomiques.

6.1 MS-CleanR : une tentative de traitement de la dégénérescence des signaux en spectrométrie de masse.

Le développement de cet outil a été réalisé d'après un constat simple : il n'est pas possible avec les méthodes actuelles d'appréhender le nombre de métabolites réellement détectés par spectrométrie de masse dans un extrait complexe. En effet, la dégénérescence observée des signaux en ionisation électrospray complexifie la détection et l'annotation sans ambiguïté des composés d'un mélange. L'ionisation d'un composé va générer plusieurs signaux détectables en spectrométrie de masse résultant des adduits, des pertes de neutres, des dimères, ou hétéromères ainsi que des différents sites d'ionisation possible pour un composé donné. Cette dégénérescence étant très dépendante de la structure du composé ainsi que de la méthode analytique et de l'instrumentation, elle est très difficile à inférer. Plusieurs méthodes ont été proposées afin de détecter les liens entre des ions apparentés à un même composé tel que, pour les plus récents, CliqueMS [28], Ion Identity Network [29] ou encore MS-Dial-PCE (Peak Character Estimation) couplé à MS-CleanR **[30]**. Dans tous les cas il va s'agir de détecter des liens de causalités entre les ions dans une fenêtre très courte de temps de rétention afin d'identifier l'ion pseudomoléculaire le plus représentatif. Ce dernier sera retenu pour être utilisé à l'annotation du composé (figure 20).

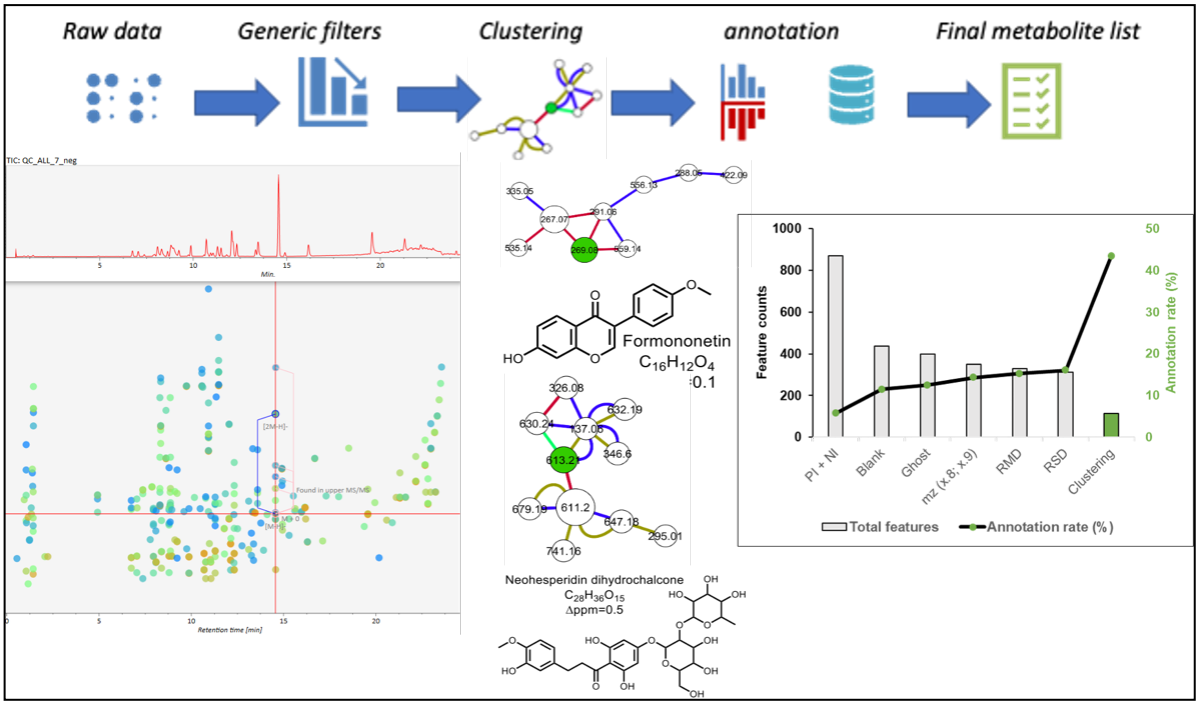


Figure 20: Processus de traitement MS-CleanR pour la sélection des ions et l'annotation basé sur le MS/MS

MS-CleanR exploite les données générées par le logiciel de traitement des spectres LC-MS MS-Dial [31]. Dans un premier temps une série de filtres sont appliqués afin de supprimer des ions sans liens avec la détection d'un métabolite : ions provenant du solvant de dilution, du bruit de fond instrumental, ions présentant un défaut de masse anormal ou des variations importantes d'intensités entre les répliquas biologiques ou d'injection. L'ensemble de ces filtres pouvant être paramétré afin de régler la profondeur du nettoyage. Environ 55% des ions détectés sont retirés par cette première approche (figure 20 graphe en barres). Dans un second temps, un algorithme de clustering des ions apparentés dans une fenêtre de temps de rétention prédéfini (typiquement 0.025 minutes en UHPLC) sera déployé sur l'ensemble des ions restants. Ces regroupements seront basés sur ceux définis par l'algorithme « peak character estimation » calculé au cours de la déconvolution MS-Dial ainsi que par les corrélations de Pearson entre les répliquas et une liste exhaustive d'adduits disponible dans MS-CleanR. Pour chaque cluster, un seul ion sera retenu en se basant sur des règles de sélections itératives tenant compte de

l'intensité de chaque ion du cluster, le degré de leurs relations ainsi que de la prédominance des adduits détectés. Cette étape aura pour objectif d'éviter la redondance d'un même métabolite sous plusieurs formes dans la liste de pic finale. Ainsi, dans l'exemple de la figure 21, pour 50 standards injectés en LC-MS en mode ESI positif et négatif, 850 paires m/z × RT ont été détectées. L'application de MS-CleanR a permis de réduire cette liste à 100 paires sans perte d'ions liés à l'un des standards.

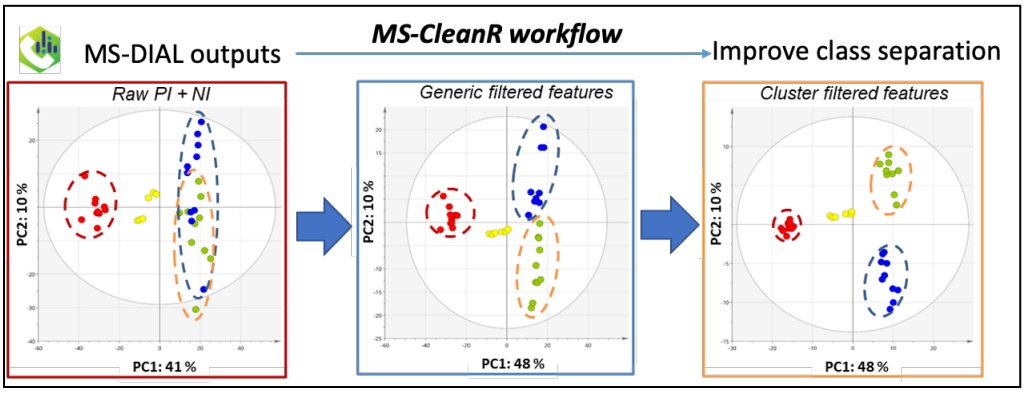


*Figure 21: Résultats des analyses en composante principale du traitement par MS-CleanR d'un jeu de donnée issu de trois conditions de cultures de plantules d'*Arabidopsis thaliana

Ce traitement des empreintes LC-MS aura pour conséquences de réduire significativement le temps nécessaire à l'annotation des spectres, particulièrement lors de l'utilisation des logiciels d'inférence spectrale *in silico* généralement très gourmands en ressource computationnelle. Le second avantage étant une diminution significative du bruit analytique et des covariables venant du même métabolite dans le jeu de données obtenu (figure 22). La réduction de la colinéarité du tableau améliorera les regroupements des classes d'échantillons facilitant ainsi l'interprétation des modèles statistiques.

Enfin, les bases de catalogues chimiques servant de support aux annotations *in silico* peuvent être priorisées afin de prendre en considération les données bibliographiques de la bioressource [32]. Ainsi dans l'exemple de la figure 22, la base de données du catalogue chimique du genre *Arabidopsis* a été priorisée si plusieurs matches étaient possibles entre la base du genre, de la famille (Brassicaceae) puis des bases toute substance naturelle confondue (générique). Un effet remarquable du clustering sera d'augmenter sensiblement les matchs d'annotation sur le catalogue du genre. Ce résultat conforte la pertinence de la sélection des ions qui apparaissent dans 50% des cas ici liés aux composés déjà identifiés dans le genre *Arabidopsis*.

La publication de MS-CleanR a été citée 35 fois depuis 2020 et le trafic sur le dépôt du code (https://github.com/eMetaboHUB/MS-CleanR) est toujours relativement important (une quinzaine de visites par semaine). La version en cours de développement de MS-Dial 5 intègre dorénavant MS-CleanR.

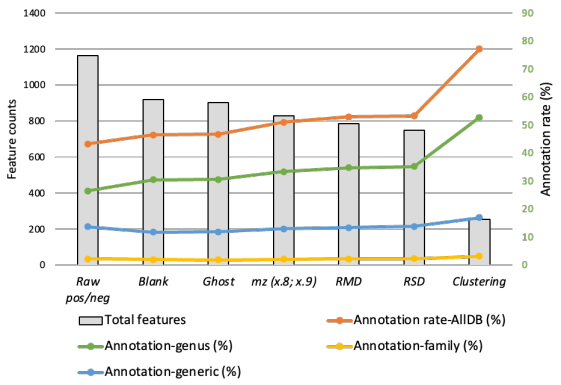


Figure 22: Niveau d'annotation en % en fonction du catalogue chimique utilisé sur le jeu de donnée totale pendant les étapes de filtrations de MS-CleanR

6.2 L'empreinte métabolomique comme outil de classification : exemple de la collection de violettes des serres de Toulouse.

Au cours d'une collaboration avec l'INP-ENSIACET, les serres municipales de Toulouse ainsi que les Parfums Berdoues (projet Région Occitanie « *Viola Tolosa* ») nous nous sommes intéressés à la classification de la collection des serres municipales. Cette collection provenant de diverses sources (dons, collaborations…), seulement 58% avaient fait l'objet d'une identification rigoureuse. L'approche métabolomique détaillée ci-dessus a été testée sur un jeu d'échantillons comprenant les extraits foliaires de l'ensemble du catalogue de violettes (soit une centaine de cultivars). Les empreintes chimiques ont servi à la construction de modèles statistiques de prédiction de l'espèce de violette (figure 24).

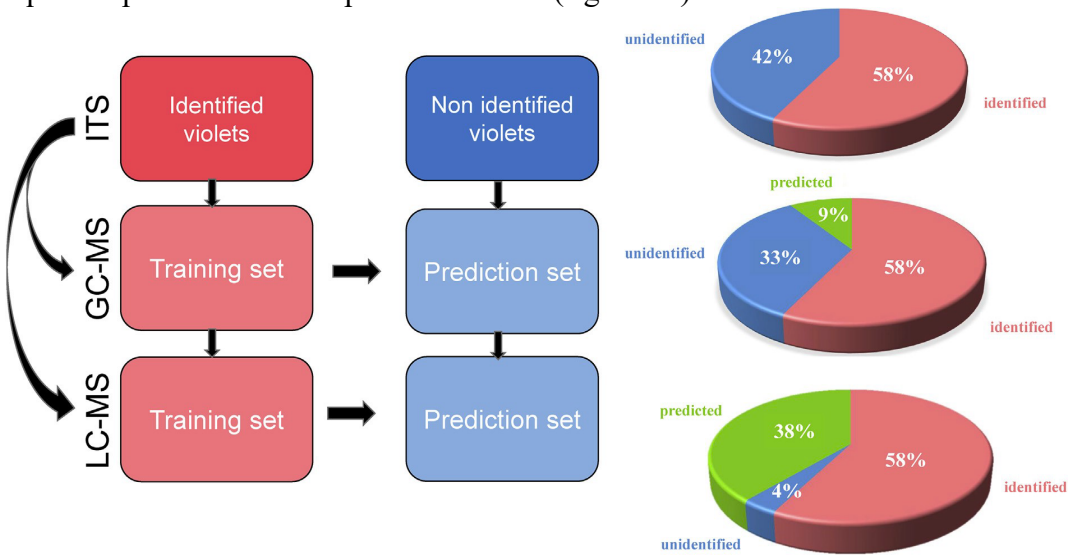


Figure 24:Stratégie de prédiction statistique des espèces de violettes à partir des empreintes chimiques GC- et LC-MS

Plus de trois cents composés ont été retenus après traitement MS-CleanR et ont été annotés soit *i)* à l'aide d'une base de données interne comprenant les spectres LC-MS/MS dans les mêmes conditions analytiques et *ii*) par modélisation du spectre MS/MS *in silico* en interrogeant les catalogues chimiques à partir du genre *Viola* et de la famille des Violaceae. Certains biomarqueurs d'intérêts ont été confirmés à l'aide de leurs spectres d'absorption UV. Cette matrice a servi de support pour l'élaboration de modèle OPLS (orthogonale projection to latent structure) discriminant l'espèce. Le jeu d'entrainement s'est basé sur les violettes clairement identifiées par le séquençage de leurs ITS1 et ITS2 soit 80 individus. Le modèle OPLS a permis de classer 26 individus sur les 30 restant à identifier (figure 24).

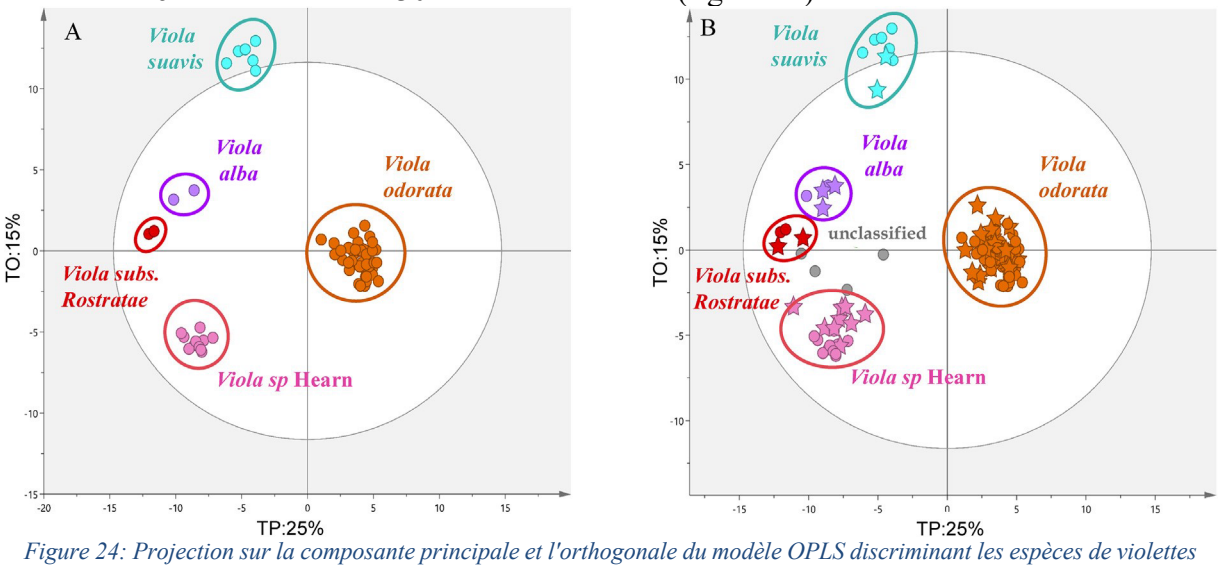


Figure 24: Projection sur la composante principale et l'orthogonale du modèle OPLS discriminant les espèces de violettes basées sur les empreintes chimiques LC-MS (droite : modèle avec les échantillons d'entrainement ; gauche : modèle comprenant la projection des échantillons inconnus)

Quelques marqueurs chimiques de la famille des flavonoïdes ainsi que des polypeptides cycliques (cyclotides) ont pu être identifiés du fait de leur surabondance dans une espèce en particulier **[33]**.
Bien que cette étude ait été réalisée à partir de plants issus de conditions de culture contrôlées, cet exemple démontre la possibilité d'utiliser **les empreintes chimiques complexes comme outils de classification des espèces**. D'après les résultats obtenus lors de l'étude par RMN du proton des CPORs de *Citrus* sp. (paragraphe 4.3) où les essences ont été différentiés selon leurs localités, il apparait qu'une partie du métabolome reflète également les conditions environnementales. On peut suspecter qu'au vu du nombre important de variables détectées par ces approches, **les modèles statistiques extraient aisément des sous-ensembles fortement corrélés à la question scientifique.** En dehors des exemples des parties 4 et 5, de nombreuses autres études témoignent de l'utilisation des modèles basés sur des empreintes chimiques pour discriminer les conditions d'études telles que le stress biotique [34] ou abiotique [35], classification chimiotaxonomique [36], stade de développements [37], recherche de bioactifs [38]...
La dynamique du métabolome reflète à la fois les interactions environnementales tout en restant suffisamment stable pour servir d'empreinte moléculaire spécifique, au moins en conditions contrôlées. C'est sans doute cette plasticité qui est à l'origine du succès de la métabolomique, mais aussi de son plus grand défaut. La littérature regorge de nombreuses mises en avant de sous-ensembles de métabolites fortement corrélés à une question de recherche, mais cela reste généralement insuffisant pour saisir l'ensemble du contexte biologique de l'étude. Les inférences dégagées par ces modèles doivent être confirmées par des expériences de biologie fonctionnelle parfois longue et couteuse ou simplement impossible à réaliser du fait du manque de connaissance des systèmes biologiques.
La contextualisation des données omiques en générale et métabolomiques en particulier reste aujourd'hui un défi majeur pour la communauté scientifique. Des approches spécifiques telles que la fluxomique [39] ou des méthodes d'explorations du métabolisme comme metexplore [40] ont permis des avancées significatives, mais se limitent à des organismes ou des contextes d'étude particulière. C'est pourquoi l'un des besoins majeurs nécessaires au développement de la métabolomique dans le cadre de l'étude de systèmes biologiques sera la mise en place d'une cartographie complète et documentée par les empreintes chimiques de la chimiodiversité du vivant, comme cela se fait dans les autres approches holistiques.

7 Projet de recherche : Cartographie de la chimiodiversité du vivant

La caractérisation du catalogue chimique du vivant a largement bénéficié des innovations de la chimie analytique et des outils spectroscopiques. Cependant, malgré plus d'un siècle de découverte de produits naturels, **la genèse du métabolite au métabolome s'est construite tardivement au regard des autres « omiques ».**

Ce décalage peut s'expliquer par l'absence de canonisation du langage biochimique contrairement au code génétique et protéique s'appuyant sur quatre bases nucléiques en 21 acides aminés conduisant aux millions de protéines connues ou prédites (Sur 230M de protéines connues, 150M sont prédites d'après uniprot statisitcs-Décembre 2022). Ce langage redondant et systémique au sein du vivant assure une grande stabilité de l'ensemble. La dissection de cette grammaire génétique a permis un essor fulgurant des outils d'exploration du vivant depuis les années 1960 : séquençage complet des génomes (>500 K espèces décrites sur GenBank), metagénomique, RNAseq single cell, protéomique bottom-up et top-down, modélisation 3D des protéines par apprentissage machine…cette unicité a sans doute favorisé la structuration de ces disciplines scientifiques autour de socles communs admis par l'ensemble de la communauté (standardisation des nomenclatures, base de données, méthodes d'analyses…). Elle a aussi catalysé la naissance d'une vision unifiée du vivant dont celle du métabolisme malgré une complexité apparente importante.

En effet le métabolome est le produit des interactions du complexe génome-transcriptome-protéome, de l'environnement et des biomes associés. Bien que les voies métaboliques des briques biochimiques élémentaires soient bien caractérisées, le puzzle métabolique conduisant à la diversité moléculaire du vivant reste encore peu déchiffré, particulièrement pour le métabolisme secondaire. Ce sont essentiellement des approches de traçage isotopique qui ont permis d'élucider les voies de biosynthèse de quelques substances naturelles d'intérêts [41]. Le site KEGG rassemble 12 000 réactions métaboliques impliquées dans la biosynthèse de 18 000 produits (https://www.genome.jp/kegg/pathway.html visité en décembre 2022), soit une fraction de la chimiodiversité répertoriée, estimée à plus d'un million de composés [42]. Pour la majorité des composés connue, des classifications manuelles ou automatisées ont été proposées : par exemple, le dictionnaire des produits naturels édité par CRCpress (DNP version 30.1, CRCpress-2022) distingue une quinzaine de classes chimiques différentes pour plus de 1000 sous-classes qui caractérisent les 300 000 composés de son catalogue. Des tentatives d'algorithme automatisé sont aussi disponibles telles que Classyfire [43] ,NPclassifier [44] ou encore BioNavi-NP [45] qui permettent d'inférer les principales voies métaboliques impliquées dans la biogenèse des composés. Ces tentatives multiples d'ontologies chimiques démontrent un manque de consensus de la part de la communauté scientifique autour de la structuration de ce langage biochimique.

La chimiodiversité du vivant est estimée à plusieurs millions de composés dont seule une fraction est répertoriée au travers de plusieurs ressources hétérogènes. Ce projet de recherche se propose de cartographier cette diversité chimique à l'aide d'approches holistiques en reliant plusieurs couches d'informations orthogonales par l'intermédiaire de modèles d'apprentissage machine. Le projet s'appuiera sur une base de données relationnelles comprenant plus d'un million de composés d'origine naturelle, accompagnés de leurs métadonnées, dont l'origine biologique, la classification chimique et les activités biologiques. Des modèles d'inférence seront développés afin de construire un réseau phylo-moléculaire. Ces inférences seront croisées aux modèles phylogénétiques existants puis validées à l'aide d'acquisitions métabolomiques par spectrométrie de masse sur des taxons d'intérêts. Ainsi, cette approche holistique permettra **d'étendre la couverture du métabolome du vivant, de caractériser les systèmes évolutifs moléculaires et de prioriser la recherche de composés bioactifs sur des taxons d'intérêts.**

Ce projet s'articule autour de trois axes de développements étroitement liés : *i)* la concaténation des catalogues chimiques issus du vivant en une base de données relationnelles *ii)* La mise en place d'un processus d'acquisition de données analytiques multiplexées *iii)* la confrontation des données expérimentales en LC-MS non ciblée acquises sur des taxons clefs par rapport aux modèles d'inférences.

Ces outils d'ingénieries seront déployés afin de cartographier la chimiodiversité du vivant, non pas dans un contexte expérimental de comparaisons du métabolome d'un organisme à un stimulus, mais à une échelle systématique en traitant les empreintes chimiques comme une signature caractéristique de l'espèce. Il s'agira donc ici de consolider les liens entre l'origine biologique des métabolites et leur répartition structurelle au sein du vivant par une approche théorique et expérimentale.

Cette répartition reflète-t-elle la systématique phylogénétique ? Peut-on établir des liens entre des métabolites spécialisés et leur environnement ? La notion de biomarqueur taxonomique est-elle valable à l'échelle d'une espèce ou à des niveaux trophiques plus élevés (genre, famille, phylum) ? Peut-on modéliser la chimie du vivant et prédire l'origine d'un métabolite ? Les inférences de ces modèles peuvent-elles être utilisées pour aider à la priorisation de recherche de molécules d'intérêts dans un contexte d'érosion de la biodiversité ?

7.1 Panorama des catalogues chimiques

Le paysage des catalogues chimiques de substances naturelles est pour le moins fragmenté avec plus d'une cinquantaine de bases de données répertoriées en 2021 [46,47]. Plus de la moitié ne sont pas maintenues ou mises à jour et seule une dizaine comportent plus de 10 000 composés répertoriés. Parmi ces dernières, certaines sont commerciales (DNP-CRCpress, SciFinder, Reaxys…) ou en accès libre (HMDB, Coconut, Lotus, KnapSack, NPAtlas…). Chacune de ces bases se focalise sur des modèles biologiques (humain pour HMDB, microorganismes pour NPatlas, marin pour CMNPD…) ou sur des classes chimiques telles que Lipidmaps ou phenolDB. Bien que ces couvertures chimiques soient tout à fait pertinentes pour les communautés scientifiques concernées, cette dispersion ne facilite pas leurs interrogations dans le cadre d'études d'interactions complexes. Toutes ces bases ont en commun d'avoir à minima une à plusieurs notations chimiques en caractères (SMILES, InChI ou InChIKey). Une extraction des bases effectuée en 2019 a permis de répertorier 530K structures uniques pour un taux de recouvrement inférieur à 5% entre les bases focalisées sur les modèles biologiques et les catalogues de substances naturelles (figure 25). À cette dispersion s'ajoute une grande hétérogénéité des métadonnées associées à chaque composé. En effet certaines renseigneront l'origine biologique (Lotus), les références bibliographiques (KnapsSacK), des données de compartimentations cellulaires et des voies de biosynthèse (Kegg, HMDB) ou encore les activités biologiques (EMBL). En l'état il est difficile d'évaluer l'étendue de ces catalogues tant du point de vue de leur diversité chimique que de leurs couvertures phylogéniques. C'est pourtant un enjeu essentiel pour l'annotation et la contextualisation des empreintes métabolomiques ou pour orienter les politiques

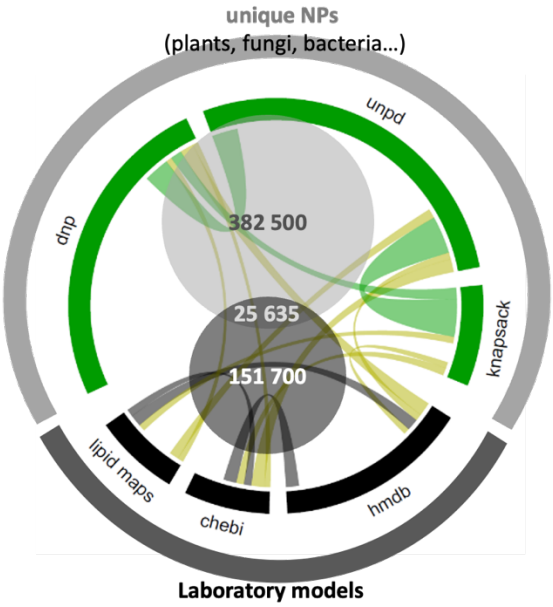


Figure 25: Couverture des bases de données utilisées en métabolomique pour l'annotation in silico. La largeur des ponts entre les bases est relative au nombre de composés communs.

de recherches d'actifs vers des taxons d'intérêts. Une première version d'une base de données unifiée appelée Pharmakon servira de support au reste du projet.

La diversité des composés rencontrés dans ces bases de données peut être hiérarchisée selon des ontologies manuelles (DNP) ou automatisées telles que proposées par Classyfire ou NPclassifier [43,44]. Les méthodes automatisées sont basées sur les empreintes binarisées des composés afin d'en extraire les motifs structuraux caractéristiques d'une classe. La méthode manuelle du DNP rassemble 11 superclasses réparties en 1000 sous classes. C'est une ontologie manuelle réalisée par des experts du domaine qui tient compte de la structure, mais aussi des voies métaboliques et de l'origine biologique des composés. Cette ontologie reflète donc la diversité des motifs ou sous structures répertoriées dans le catalogue DNP. Une analyse temporelle focalisée sur l'année d'isolement de la première structure à l'origine d'une nouvelle sous-classe chimique présente des tendances communes, quelle que soit la classe des composés (Figure 26). En effet, si les premières structures isolées au 19e siècle correspondent à des alcaloïdes ou des acides aromatiques simples (quinine, acide salicylique…), le véritable essor de la phytochimie a eu lieu à partir des années 1950. Cela peut s'expliquer par l'essor des méthodes chromatographiques, mais aussi par l'émergence de nouveaux détecteurs ou l'investigation de nouvelles matrices (polycétides des micro-organismes par exemple). Il apparait aussi clairement que depuis les années 2000, très peu de nouvelles classes chimiques ont été caractérisées.

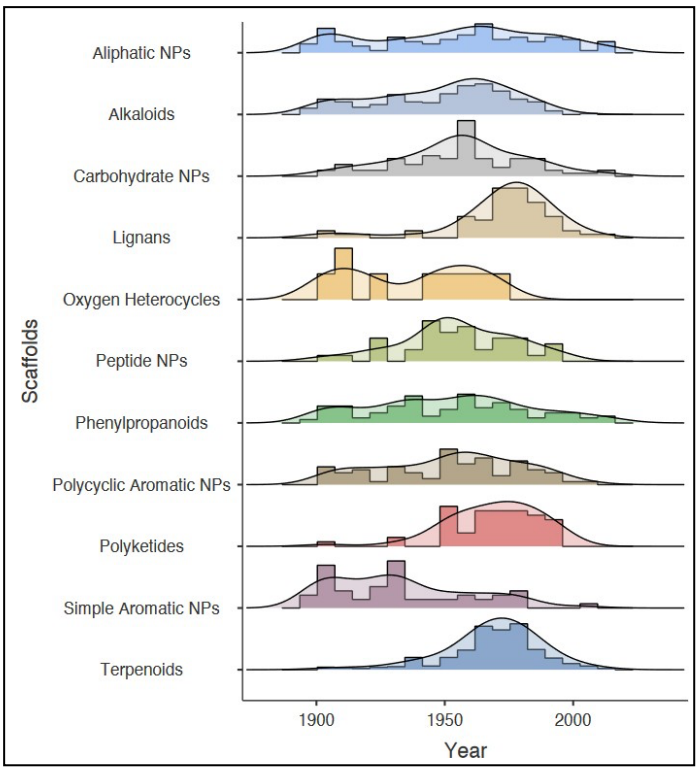


Figure 26: analyse temporelle de l'émergence de nouveaux motifs chimiques d'après l'ontologie du DNP

Ces données tendent à démontrer que **la plupart des motifs structuraux de base ont été répertoriés.** Ce résultat peut être lié aux limites des outils analytiques actuels ainsi qu'à la faible couverture des espèces explorées en phytochimie par rapport à l'ensemble de la biodiversité. *A contrario*, les études récentes de fouilles de génomes orientent vers une chimiodiversité encore largement méconnue qui semble s'exprimer uniquement dans des conditions particulières. On parle ici de plusieurs millions de composés potentiels, notamment chez les plantes, les micro-organismes et les insectes [38]. Il est donc difficile d'évaluer précisément la diversité structurelle des substances naturelles et si cette diversité potentielle concerne de nouvelles classes chimiques ou un ornement de structures existantes réaliser par des machineries enzymatiques ubiquistes (cytochromes, transférases..).

L'analyse du catalogue DNP (v2020) qui comprend 380K composés montre une dégénérescence importante du nombre de composés dérivés d'une structure unique, de l'ordre de 3 pour 1 (figure 27) pour 20K formules brutes uniques. Si l'on prend en compte les distances de tanimoto (>0.7) ce catalogue se résume à 2700 groupes de composés. Enfin, en ne retenant que les squelettes de bases calculés par la méthode de Murcko, nous pouvons distinguer un peu moins de 2000 groupes de composés.

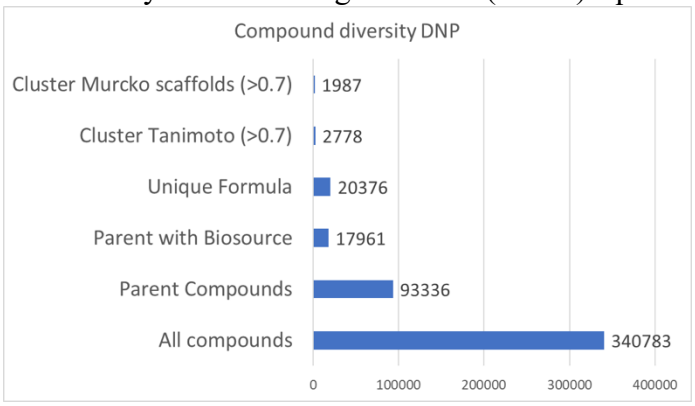


Figure 27: Analyse macroscopique du catalogue DNP

Cet exemple illustre la dégénérescence importante des structures chimiques biosynthétisés dans le vivant et appuie **l'hypothèse d'un nombre fini de squelettes de base à l'origine de la chimiodiversité.**

Une étude préliminaire de répartition des classes de composés entre les règnes du vivant a révélé des tendances très nettes, telles que la présence de flavonoïdes caractéristiques des plantes ou des polykétides que l'on trouve principalement chez les microorganismes (figure 28) **[5]**. L'analyse macroscopique de cette structuration de la diversité chimique au sein des taxons a été effectuée à l'aide des classes structurales et sous-classes telle que définie à l'aide d'une classification chimique automatisée. Cette approche présente l'avantage de donner un aperçu qualitatif général pour chaque niveau de classification. Cependant les ontologies chimiques restent très généralistes et peu adaptées aux métabolites secondaires.

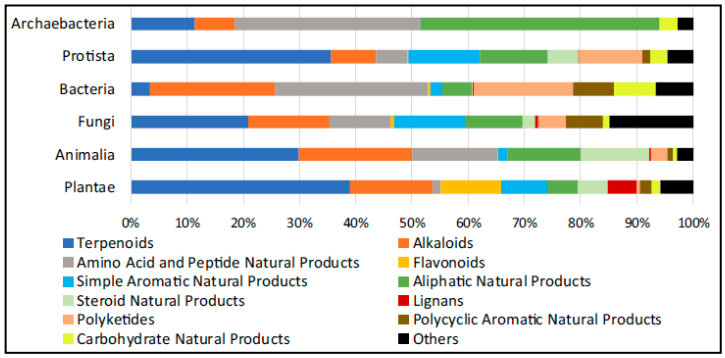


Figure 28: Répartition des classes structurales au sein des règnes du vivant (d'après référence 5)

Afin de pallier à cette limite, il est nécessaire d'analyser de manière plus fine les squelettes chimiques en 2 dimensions appartenant aux taxons référencés. **L'objectif ici sera de construire un réseau phylo-moléculaire** reliant les taxons à partir de modèles d'apprentissage machine.

7.2 Tentative de modèles d'inférences à partir de catalogues chimiques

Plusieurs méthodes d'empreintes moléculaires vectorisées (Molecular Fingerprint) ont été publiées dans le but de faciliter la fouille de données de librairies de composés [48]. Ces méthodes de vectorisation seront testées sur un groupe de structures représentatives avant d'être appliquées à l'ensemble de la base. L'obtention de ces empreintes vectorisées permettra de les relier de manière très fine aux autres métadonnées, notamment les entrées taxonomiques à l'aide de modèles d'inférences. Ces modèles peuvent servir à prédire des propriétés chimiques indispensables à la chimie médicinale tel que l'ADMET [49] ou être à la base d'ontologie chimique automatisée (e.g. classyfire). Ici nous nous intéressons aux liens possibles entre les composés et leurs origines taxonomiques. Il a été récemment démontré qu'il était possible de prédire l'origine d'un composé à partir de son empreinte binarisée [50]. Le modèle développé dans l'étude de l'équipe de Reymond classe les composés d'origine fongique, bactérienne ou végétale avec une précision supérieure à 80%.

Afin d'évaluer la faisabilité de la construction d'un modèle prédictif de l'origine botanique d'un composé, un sous-ensemble de 18K structures issus du DNP a été utilisé. Toutes ces molécules ayant des métadonnées relatives à leurs origines botaniques allant du phylum à l'espèce. Pour cet essai, l'origine du phylum a été modélisée. Le modèle développé ici atteint une précision moyenne de 73% pour les 19 phylums que comprend le jeu de donnée, allant de 55% pour les composés d'origine vertébrés à 100% pour le groupe des plantes apparentées aux fougères (Fernallies). Il est intéressant de noter l'inexactitude du modèle soulignant ainsi qu'une fraction de composés sont partagés entre plusieurs groupes (Figure 29). Par exemple le phylum des éponges partage un certain nombre de caractères chimiques commun avec les hemicordates (vers marins), les mollusques et les marchantias (liveworts). Les champignons sont aussi dispersés entre les lichens, les bactéries et les plantes vertes ce qui souligne des mécanismes symbiotiques (lichens) ou plus généralement des systèmes complexes de type holobiontes.

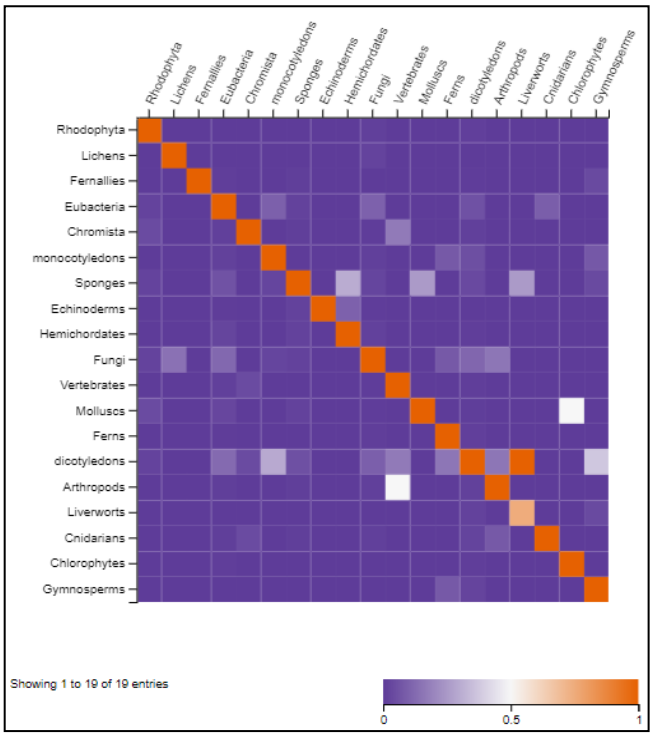


Figure 29: carte de corrélations issues d'un modèle d'inférence sur l'origine phylétique des composés chimiques

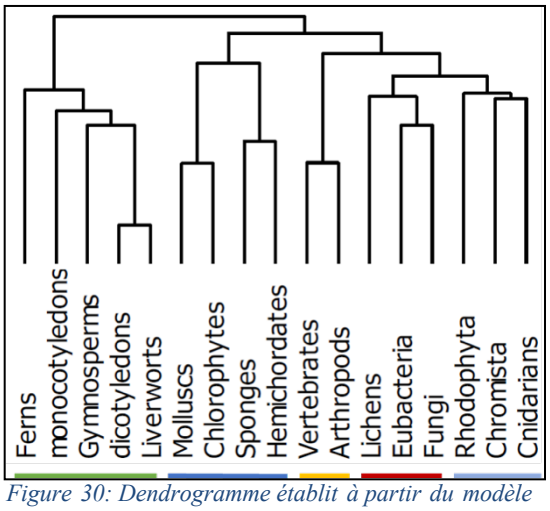


Figure 30: Dendrogramme établit à partir du modèle d'inférence des phylums

Un dendrogramme construit à partir des données du modèle permet de mettre en évidence les liens entre les phylums (Figure 30). Il est remarquable de noter que la branche verte constitue un groupe homogène séparé des autres phylums. Les lichens, bactéries et champignons sont aussi groupés dans un sous-embranchement rassemblant notamment les algues rouges et le plancton. Les autres organismes marins sont aussi séparés du reste (mollusques, éponges...).
Bien que cet arbre phylo-moléculaire ne reflète pas celui pouvant être obtenu avec les données phylogénétiques, **cette approche souligne une architecture sous-jacente** au vu de ces résultats préliminaires. L'exploration de cette chimiodiversité à l'aide de modèles d'inférences s'attachera à mettre en évidence le continuum structurel au sein de la chimie du vivant tout en contrastant les composés spécifiques à certains clades. Afin d'atteindre cet objectif, une base de données relationnelle sera construite à partir des dépôts disponibles en libre accès (cité plus haut) ainsi que des bases payantes telles que le DNP. Les modèles seront entrainés en se basant sur les liens disponibles entre structures et origines biologiques puis étendus à l'ensemble du catalogue ainsi construit en se basant sur des scores de prédictions.
Cette première partie du projet de recherche s'attachera donc à :

- **Unifier les bases de données structurales existantes** sous une même architecture pour faciliter la fouille de données.

- Appliquer des méthodes d'empreintes moléculaires vectorisées afin de **calculer des clusters de motifs structuraux communs aux taxons répertoriés**.
- Utiliser cette **cartographie moléculaire du vivant afin de détecter des marqueurs métaboliques émergeant au niveau des points clés de l'évolution**.
- Tester **des outils d'inférence** pour la prédiction des motifs chimiques dans les taxons encore peu travaillés.
- Évaluer ces hypothèses à l'aide **d'analyses LC-MS en métabolomique non ciblée** sur un panel d'espèces préalablement sélectionné (espèces modèles et espèces non travaillées).

7.3 Les limites de la métabolomique non ciblée par LC-MS

Le second axe de ce projet s'intéresse à un sous-ensemble des approches métabolomiques à savoir les empreintes chimiques globales ou non ciblées. Ces empreintes consistent à capturer l'ensemble de la diversité chimique d'un extrait complexe sans a priori. Les annotations issues de ces approches permettent ainsi de dégager des tendances spécifiques au contexte de l'étude comme illustré dans les exemples précédents.
Cependant de nombreuses limites entravent le développement de ces outils et leurs intégrations dans les autres sciences -omiques :

- **La qualité de l'annotation des signaux spectrométriques** : que ce soit en RMN ou en LC-MS, l'annotation des signaux est une étape déterminante dans le processus d'analyse. Le résultat aura un impact majeur sur les interprétations possibles. La qualité des annotations dépendra de la qualité des données spectrométriques et des bases de données expérimentales (comportant des signaux de références pour chaque composé) ou sous forme de catalogue chimique utilisées pour identifier les signatures spectrales.
- **Le manque d'organismes référencés** dans les bases de données : on en dénombre quelques centaines pour les réseaux métaboliques.
- **La dispersion importante des ressources** tant au niveau du traitement des chromatogrammes que des outils environnant l'annotation (bases de données spectrales et catalogues chimiques, outils *in silico*…).
- **L'absence de quantification des métabolites** : contrairement à la RMN du proton, la LC-MS, bien que plus sensible, ne peut pas quantifier la quantité de chaque métabolite détecté à moins d'avoir qualifié la réponse instrumentale pour chacun des composés de la matrice. Cette limitation est un écueil important notamment pour la mise en place des modèles statistiques qui se limitent en l'état à comparer la réponse instrumentale (intensité ou aire de pic normalisé) d'un signal entre les échantillons du jeu de données. L'accès à des données semi-quantitatives en LC-MS permettrait d'obtenir des modèles plus proches de la réalité biologique.
- **L'imbrication importante des -omiques** oriente vers des expériences multiplexées capables d'acquérir des données sur l'ensemble de la cascade -ome. Ceci implique une évolution des pratiques vers des combinaisons d'approches capable d'augmenter la couverture du métabolome.
- La nécessité de **développer des outils d'aide à l'interprétation** des données tels que MetExplore [40].

Il s'agira ici de répondre à ces écueils en proposant le développement d'outils nouveaux ou la combinaison d'outils existant sous licence ouverte.

7.4 Les bases de données pour la LC-MS

Toutes les méthodes -omiques font appel à des bases de données afin de transformer les signaux instrumentaux en séquence interprétable (gènes, transcrits, ontologies, protéines, métabolites…). Le développement des -omiques est donc étroitement lié à l'essor d'internet et des langages permettant le stockage digital des informations. Par exemple, la base Chemical Abstract Services a commencé à être informatisée dans les années 80 en même temps que l'apparition des codes SMILES ouvrants la voie à un décodage computationnel des structures chimiques. Les bases de données d'agrégation telles que PubChem ont vu le jour dans les années 2000 (PubChem est fondé en 2004 par le National Institute of Health).
Concernant la métabolomique, il existe deux types de bases de données utilisés pour les annotations :

- Les catalogues de composés chimiques comprenant un ensemble de métabolites liés à un domaine d'étude particulier tels que les substances naturelles (Dictionnary of Natural Products, Knapsack, NPatlas), l'humain (Human Metabolome DataBase), l'exposome (T3DB) etc... (Cf. paragraphe 6.1). Les catalogues sont utilisés pour l'annotation *in silico* : les identifiants interprétables par les algorithmes (SMILES, InChI) sont utilisés en entrée par des algorithmes de modélisation des spectres MS, MS/MS ou RMN. Ces spectres modélisés sont confrontés aux spectres acquis expérimentalement générant plusieurs possibilités classées selon un score de similarité pour un spectre donné. On parlera ici d'un niveau d'annotation 3 (figure 12).

- Les bases de données spectrales comprenant les données de composés chimiques purifiés et leurs signaux spectrométriques correspondant en HRMS, MS/MS ou RMN. L'interrogation des bases de données spectrales se base uniquement sur un score de similarité entre les spectres de la base et celui de l'échantillon. Là encore, plusieurs possibilités seront remontées et classées par degrés de similitude. On admet toutefois que les composés ayant des similarités spectrales supérieures à 90% sont supposés être un match de niveau 2. Les bases de données spectrales construites dans les mêmes conditions que les analyses permettent d'atteindre une annotation de niveau 1 (figure 12).

Bien que ces approches soient incontournables au vu du nombre de variables potentiellement identifiables dans les extraits complexes, les résultats de ce système d'annotation par hiérarchisation des structures chimiques possibles seront très dépendants de la base de données utilisée. La qualité des données spectrales et l'exhaustivité de la base utilisée auront un impact direct sur les annotations. Concernant l'utilisation des catalogues structuraux, il sera aussi très difficile d'établir un score au seuil duquel une annotation sera pertinente, particulièrement dans le cas des annotations *in silico*. En effet, les résultats des challenges CASMI ou de comparaisons de ces outils montrent que la structure réelle du composé est classée première dans 50% des cas et dans les 10 premières dans 95% des cas [51,52]. Les annotations à la volée peuvent aussi masquer des composés inconnus qui seront annotés de manière erronée. C'est pourquoi les métabolites sélectionnés par les modèles statistiques font généralement l'objet d'une inspection manuelle afin d'en confirmer l'identité. La base de données GNPS (Global Natural Products Social Molecular Network, [29]) a proposé à la communauté une méthode basée sur les proximités des spectres de fragmentations MS/MS afin d'identifier des composés inconnus dont le spectre est proche d'un standard dument identifié. Cette approche ayant fait ses preuves en peptidomique donnera cependant des résultats contrastés en annotation des produits naturels en fonction de la qualité des spectres de fragmentations ainsi que des classes chimiques rencontrées.

7.5 Concaténation des bases de données expérimentales en spectrométrie de masse

Les bases de données expérimentales réalisées à l'aide des empreintes MS et MS/MS de standards analytiques sont disponibles sous plusieurs formats : JSON, SQL, msp, mgf, ces deux derniers étant les plus utilisés. Ces formats texte rassemblent pour chaque composé un ensemble de métadonnées allant des identifiants alphanumériques des composés (SMILES, InChI, InChiKey) aux conditions expérimentales d'acquisitions. Ici encore, la qualité et la quantité des métadonnées renseignées vont grandement varier selon la base considérée. Une seconde partie concerne la masse exacte, éventuellement l'adduit et les pics de masse fragments et leurs intensités respectives. Il y a généralement très peu de curations sur ces données d'origine expérimentale. Peu de ressources sont disponibles en libre accès (GNPS, RIKEN, MONA, Massbank) alors que d'autres bases sont payantes (mzcloud, Metlin, NIST). Une dernière catégorie de spectres est générée *in silico* notamment à l'aide de lipidblast pour les composés lipidiques [53] ou CFM-ID pour les autres classes chimiques [51]. Même si plus d'un million de spectres sont disponibles en libre accès, la moitié sont d'origine computationnelle. En ce qui concerne le nombre de composés, seuls 25% sont des spectres expérimentaux, soit environ 50 000 composés uniques (figure 31A). Ce chiffre peut paraitre infime comparé au nombre de composés présents dans les catalogues chimiques (plus de 500 000), mais s'explique par la difficulté à isoler des produits naturels ainsi que par le faible taux de dépôts des spectres lors des publications associées aux purifications de nouveaux composés.

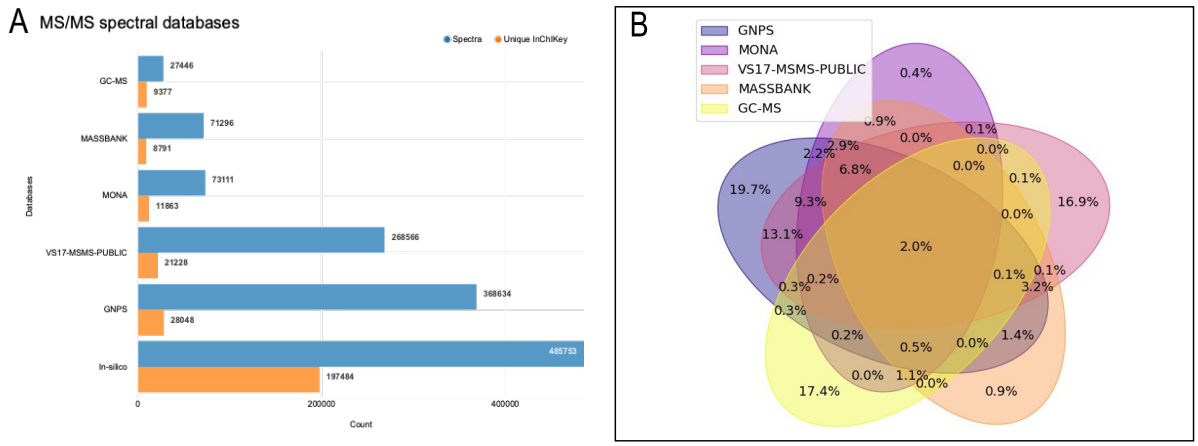


Figure 31: Données spectrales et nombres de composés uniques disponibles dans les bases de données ouvertes (A); Diagramme de ven résultant de la concaténation des espèces chimiques entre bases de données (B)

On observe une dégénérescence importante du nombre de spectres par composé. Cela peut s'expliquer par l'acquisition de plusieurs empreintes différentes pour un même composé en fonction de l'instrument, de l'énergie de collisions ou du type de cellule de collision considéré. On peut aussi constater un certain nombre de doublons (<5%) ce qui souligne le manque de curation des données notamment chez les bases qui agrègent plusieurs ressources (GNPS et VS17).

Chaque base possède en moyenne 15% de composés uniques (Figure 31B) ce qui atteste de l'intérêt d'agréger ces ressources en vus d'obtenir une couverture de spectres expérimentaux les plus larges possibles.

Il s'agira donc ici de concaténer ces bases de données tout en harmonisant les métadonnées qui y sont associées telles que les identifiants des composés, le mode d'acquisition…l'objectif sera d'augmenter sensiblement le nombre d'annotations de niveau 2.

Un premier essai réalisé sur les bases disponibles en libre accès fin 2022 a permis de mesurer le gain potentiel d'une telle approche. Les matchs retenus au cours de cet essai ont une similarité supérieure à 80%. En comparant les matchs de niveau 1 issus de la base interne au laboratoire avec le nombre de matchs considérés de niveau 2 issus de l'ensemble des bases de données

concaténées, le gain est très significatif. Cependant, ce résultat reste encore à étayer notamment du fait de manque de curation des spectres, ce qui peut donner de nombreux faux positif. Un second écueil à cette démarche étant le manque de données orthogonales aux spectres MS/MS tels que la prise en compte des adduits possibles, du temps de rétention ou encore des spectres d'absorption UV permettant potentiellement d'affiner ces annotations. Un travail est en cours dans le cadre du projet France 2030, MétaboHub 2.0 afin de consolider ces premiers résultats encourageants.

7.6 Extension de la couverture du métabolome par orthogonalisation des approches analytiques en LC-MS

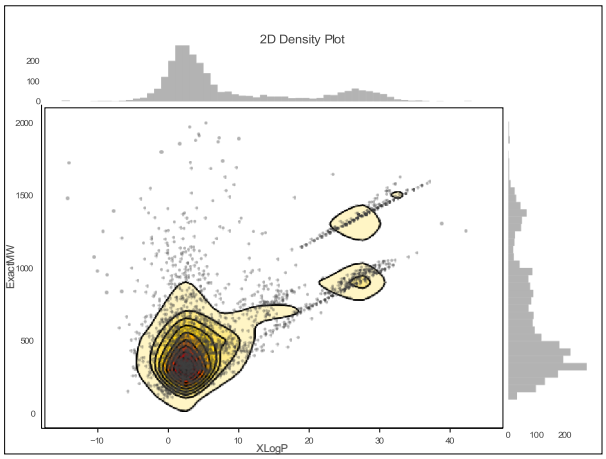


Figure 32 : Nuage de points entre le XlogP (x) et la masse exact (y) de 2500 composés naturels pris aléatoirement dans une banque de 300 000 composés. Les histogrammes représentent la densité des points pour chaque axe.

L'hétérogénéité des structures rencontrées dans les substances naturelles varie de 50 à 2000 daltons et sur une large gamme de polarité. La spectrométrie de masse couvre l'ensemble de la gamme rencontré, mais les variations des propriétés physico-chimiques impliquent la création de méthode de séparation spécifique.

On peut grossièrement séparer ces méthodes en trois catégories en fonction du coefficient de partition des composés à analyser (logP) : pour les logP négatifs, des méthodes de séparation basées sur la chromatographie d'interaction hydrophile (HILIC) seront envisagées [54]. Pour les logP compris entre 0 et 15, des méthodes en phase inverse vont être élaborées. Au-delà, tel que pour la séparation des triglycérides compris entre un logP de 20 à 40 (figure 32) des outils chromatographiques de type lipidomique seront privilégiés [55].

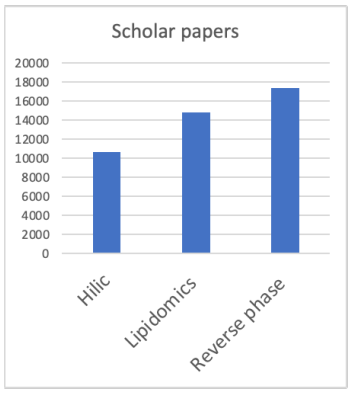


Figure 33: occurence des papiers contenant le mot "Metabolomique and LCMS and Hilic "or" Lipidomics "or" Reverse phase

La plupart des études de métabolites secondaires par LC-MS s'effectuent en chromatographie de phase inverse. C'est en effet le meilleur compromis qui englobe une majeure partie des composés biosynthétisés par les substances naturelles (Figure 33). Cependant, si l'intérêt se porte sur des aspects de biologie des systèmes, c'est bien l'ensemble du métabolome qui doit être appréhendé. Afin d'atteindre cet objectif, la méthode d'extraction doit d'abord être optimisée afin de couvrir une large gamme de polarité de composés. Dans un second temps, des méthodes chromatographiques orthogonales, adaptées à la polarité des extraits peuvent être mises en place. Enfin, des outils de concaténation des données issues de ces analyses multiplexées peuvent être exploités afin de limiter la covariance des données tout en maximisant les corrélations entre les méthodes et la question biologique.

Un premier essai d'orthogonalisation de méthode analytique a été effectué sur une matrice de bois de vigne dans le cadre d'un projet en collaboration avec l'EI de Purpan. Après application d'une extraction biphasique [56] à partir d'un seul échantillon, la phase lipophile a été profilée par une approche LC-MS de type lipidomique et la phase polaire a été séparée selon un mode en phase inverse classique (C18) et une seconde méthode utilisant une colonne Hilic.

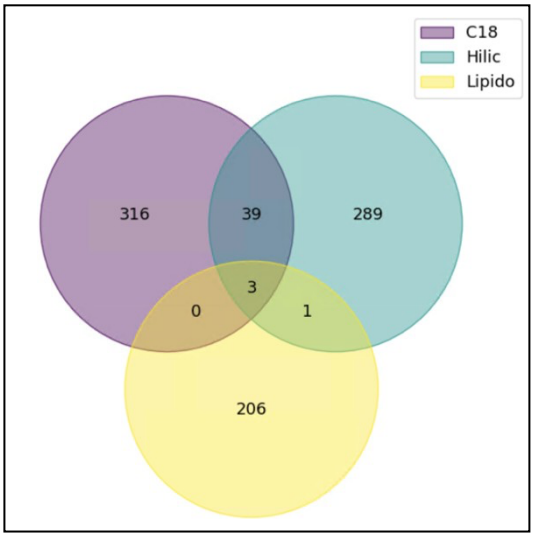


Figure 34: Diagramme de Venn basé sur les annotations communes trouvé entre les trois modes d'acquisition pour un même échantillon

La combinaison de ces méthodes séparatives confirme l'orthogonalisation des informations qu'elles apportent. Cette complémentarité est particulièrement importante entre l'approche par séparation Hilic et lipidomic avec très peu de redondance de composés observés dans ce cas. La méthode C18 et hilic ayant une quinzaine de pour cent de composés communs (Figure 34). La seconde étape de cet essai préliminaire a consisté en la concaténation de ces données afin d'établir des modèles statistiques pertinents pour répondre à la question biologique. Dans ce cas, une analyse multiblocs a été effectuée afin d'identifier la signature métabolomique potentielle apportée par chaque méthode analytique et l'effet de leur additivité [57].

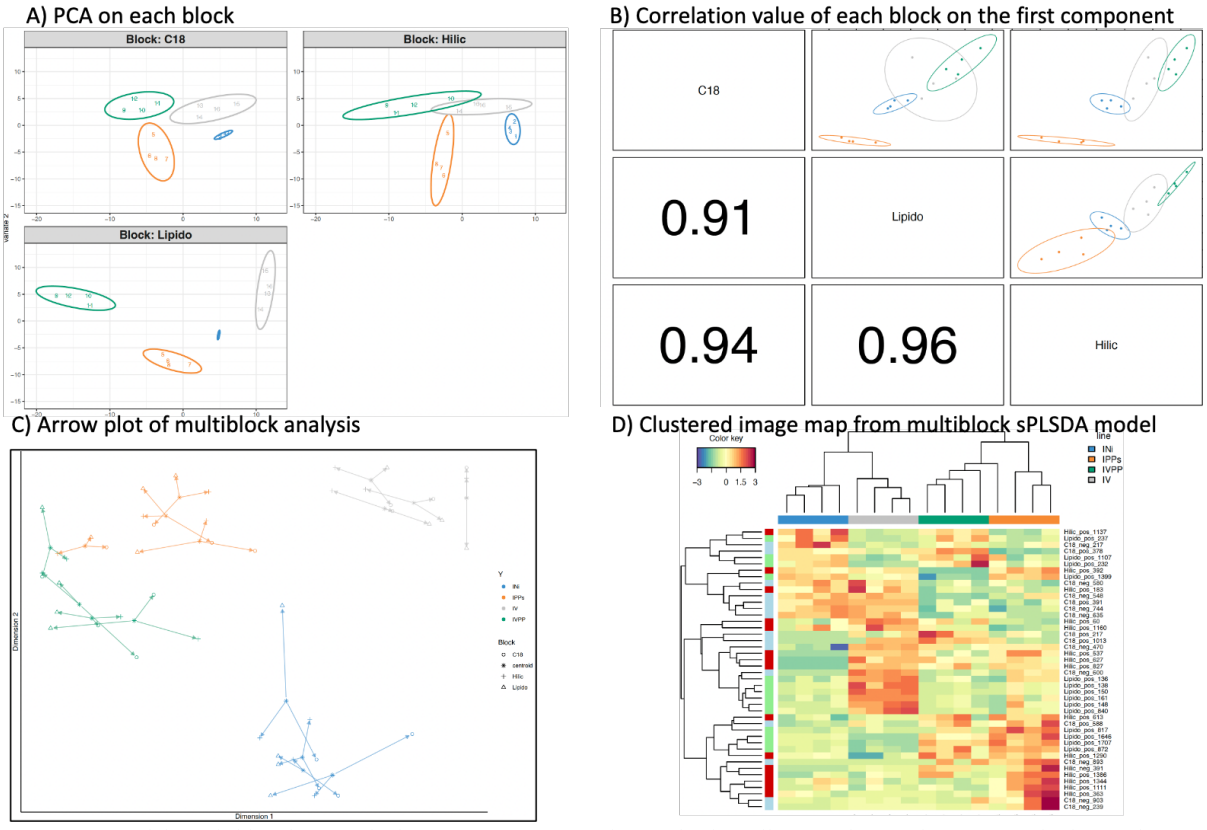


Figure 35: Modèle multibloc d'un jeu de données présentant 4 classes d'échantillons analysés selon trois méthodes analytiques orthogonales.

La figure 35-A montre une analyse en composante principale (2 premières composantes) pour chaque bloc de données. On observe une bonne séparation des 4 classes d'échantillon pour l'analyse lipidomique et C18. Afin de vérifier si les informations apportées par les 3 modes chromatographiques apportent des informations complémentaires au modèle discriminant, la

corrélation entre les blocs sur la première composante du modèle est mesurée (Figure 35-B). Elle est ici supérieure à 0.9 donc la combinaison des méthodes semble très pertinente pour la construction du modèle. Ces corrélations sont confirmées sur la figure 35-C qui permet d'appréhender le barycentre de chaque échantillon et sa projection dans chacun des blocs correspondant aux méthodes analytiques. Les classes d'échantillons restent bien séparées dans ce modèle. Enfin un modèle sPLS-DA sur l'ensemble du jeu de données permet de sélectionner les variables les plus discriminantes de chacun des blocs conduisant à la séparation des classes d'échantillon.
Cet exemple illustre la plus-value apportée par cette approche en termes de couverture du métabolome et des modèles statistiques associés pour rentrer dans la question biologique d'intérêts.

7.7 Métabolomique globale semi-quantitative par multiplexage de la détection

Comme souligné plus haut, la détection par spectrométrie de masse souffre d'un défaut important relatif au mécanisme de détection des composés. L'intensité du signal enregistré par un spectromètre de masse sera très dépendante de la stabilité d'un composé à supporter une charge en milieu gazeux. L'intensité des signaux qui en résulte sera donc variable en fonction de l'entité chimique détectée et très influencée par les effets matrice se produisant dans la source d'ionisation. Cet écueil a plusieurs conséquences, la principale étant l'impossibilité d'estimer la quantité réelle d'un composé donné dans une matrice complexe sans calibration par standard interne ou externe de la réponse du spectromètre de masse pour ce composé. En corolaire, les méthodes statistiques pouvant être appliquées aux données LC-MS se limitent à comparer l'intensité d'un signal relatif préalablement normalisé entre des échantillons sans pouvoir exploiter les données quantitatives.

La figure 36 montre le résultat d'un multiplexage de la détection pour un extrait de plante avec une détection PDA (200-500 nm), une détection en mode ESI positif ainsi qu'une détection avec le détecteur universel CAD. La détection PDA est aussi dépendante des chromophores présents dans les composés élués en sortie de colonne et sera donc soumise aux mêmes limitations que la spectrométrie de masse. Le CAD est classé dans les détecteurs dits « universel » à savoir que l'intensité du signal sera proportionnelle à la quantité de matière détectée. Le chromatogramme LC-CAD qui en résulte permettra donc de déduire à partir des pics majoritaires les composés les plus concentrés au sein d'une matrice complexe. Dans cet exemple on peut constater que les signaux détectés en électrospray sont complètement décorrélés des chromatogrammes PDA et CAD avec des pics de forte intensité compris entre 5 et 12 minutes. Le détecteur universel indique dans ce cas que les composés les plus concentrés sont élués entre 5 et 7 minutes. Il est intéressant de noter que la détection PDA est en accord avec celle du CAD sauf pour la partie apolaire du chromatogramme, au-delà de 9 minutes où le PDA ne détecte plus aucun signal.

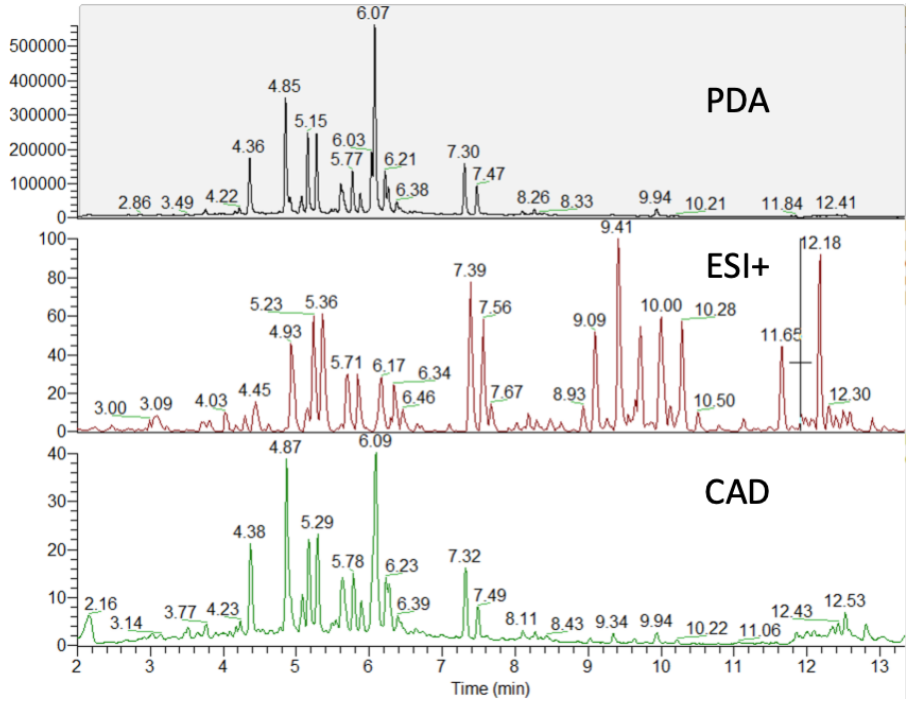


Figure 36: Superposition d'un chromatogramme d'un extrait de plante avec une triple détection: Photodiode Array (PDA), ionisation électrospray en mode positif (ESI+), et décharge corona par aérosol chargé (CAD)

L'axe de développement proposé ici consistera à exploiter les données de l'ensemble de ces détecteurs. Les données issues du CAD seront extraites et déconvoluées en une liste d'air de pics puis alignées sur les données spectrométriques afin de corriger les intensités relatives normalisées en données semi-quantitatives. De même, les données du détecteur PDA seront exploitées pour l'annotation des composés ayant des chromophores caractéristiques (e.g. flavonoides). Le résultat attendu sera une liste de pics détectés quantifiés à l'aide du CAD et identifier par spectrométrie de masse haute résolution. Cette méthode sera très influencée par la résolution chromatographique du système utilisé, une attention particulière sera donc apportée à cette partie au cours du développement de cette approche.

7.8 Stratégie d'implémentation du projet à moyen terme

L'appréhension de la chimiodiversité du vivant par les stratégies réductionnistes a permis d'isoler plusieurs centaines de milliers de composés. L'analyse globale du résultat de plus d'un siècle de recherche démontre que la majeure partie des grandes classes chimiques ont été répertoriées. Au regard de ces connaissances, les approches holistiques telles que la métabolomique sont capables d'annoter un grand nombre de composés au sein d'un extrait complexe en se basant sur ces savoirs. La combinaison de ces approches devrait permettre de donner une impulsion novatrice en chimie des substances naturelles sur plusieurs niveaux : la découverte de motifs structuraux non répertoriés pour un extrait donné, une amélioration de la contextualisation de la problématique biologique (comparaison d'extraits dans le cadre d'une problématique biologique) et enfin une vision d'ensemble de l'architecture de la chimiodiversité au sein de la taxonomie du vivant.

La stratégie qui sera déployée s'articule autour d'un axe de développement de chimie analytique (paragraphe 6.6 et 6.7) et de concaténation de bases de données ouvertes (paragraphe 6.4 et 6.5). Le plan de financement (Figure 39) pour les prochaines années s'articule autour du projet d'infrastructure nationale MetaboHub soutenue par France2030 et plus particulièrement dans l'axe « extension de la couverture du métabolome » qui comprend la création de bases de données et le développement d'outils analytiques innovants. Les ressources propres et RH

seront assurés par des demandes au fil de l'eau tel que les bourses 80PRIME du CNRS (projet Chemomaps en cours) ainsi que des appels d'offres soutenues par l'ANR. Ce projet étant très

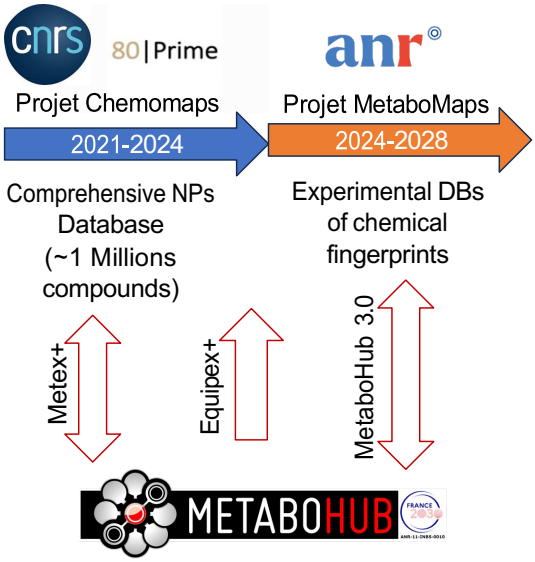


Figure 39: Plan de financement à moyen terme

consommateur en ressources informatiques tels que des serveurs de données ou de calculs, une collaboration étroite avec l'Institut de Recherche en Informatique de Toulouse et plus particulièrement l'équipe du Professeur G. Cabanac a été initiée depuis plusieurs années et sera renforcée au cours de ce projet.

Les deux axes de développement seront amalgamés à l'aide de la construction de modèles d'inférences (paragraphe 6.2, Figure 40). Ces modèles seront construits en se basant sur la chimie répertoriée pour chaque espèce puis confrontés aux données LC-MS issus de modèles de laboratoire ou d'espèce charnière présentant des traits évolutifs remarquables. Cette approche théorique et expérimentale permettra la génération de données à même de répondre à plusieurs questions scientifiques :

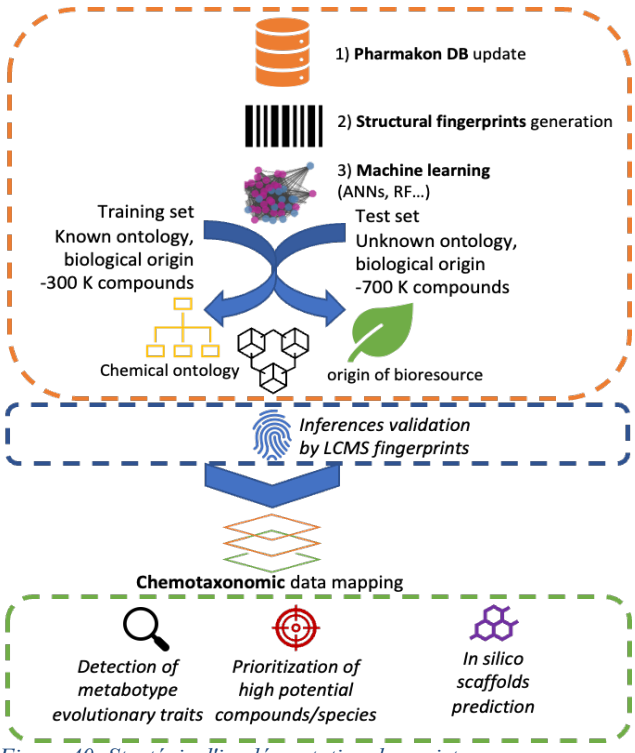


Figure 40: Stratégie d'implémentation du projet

- La détection de motifs chimiques fortement liés à des mécanismes évolutifs au sein du vivant.
- La priorisation de l'exploration de bioressource à fort potentiel de valorisation
- La prédiction de motifs structuraux susceptibles d'être retrouvés au sein d'un taxon.
- La création d'une base de données ouverte rassemblant les structures publiées ainsi que les composés détectés par des approches métabolomiques.

En conclusion, ce projet vise à modéliser la chimie du vivant dans son ensemble afin d'en comprendre l'architecture et d'en établir finement sa cartographie. Plus largement, le décryptage de ce langage biochimique servira de socle commun afin de relier les différentes strates des -omiques.

Ce projet sera susceptible d'aider la communauté scientifique en chimie des substances naturelles, en évolution et plus largement en biologie des systèmes.

(Biobliographie en rouge : Travaux personnels, auteurs souligné : encadrement)

8 Bibliographie

1. Warolin, C. La Pharmacopée Opiacée En France Des Origines Au XIXe Siècle. *Rev. Hist. Pharm.* **2010**, *97*, 81–90.
2. Pelletier PP; Caventou JB Note Sur Un Nouvel Alkalai. *Ann. Chim. Phys.* **1818**, *8*, 323–324.
3. Newman, D.J.; Cragg, G.M. Natural Products as Sources of New Drugs over the Nearly Four Decades from 01/1981 to 09/2019. *J. Nat. Prod.* **2020**, *83*, 770–803, doi:10.1021/acs.jnatprod.9b01285.
4. Baker, D.D.; Chu, M.; Oza, U.; Rajgarhia, V. The Value of Natural Products to Future Pharmaceutical Discovery. *Nat. Prod. Rep.* **2007**, *24*, 1225, doi:10.1039/b602241n.
5. Chassagne, F.; Cabanac, G.; Hubert, G.; David, B.; Marti, G. The Landscape of Natural Product Diversity and Their Pharmacological Relevance from a Focus on the Dictionary of Natural Products®. *Phytochem. Rev.* **2019**, *18*, 601–622, doi:10.1007/s11101-019-09606-2.
6. Atanasov, A.G.; Waltenberger, B.; Pferschy-Wenzig, E.-M.; Linder, T.; Wawrosch, C.; Uhrin, P.; Temml, V.; Wang, L.; Schwaiger, S.; Heiss, E.H.; et al. Discovery and Resupply of Pharmacologically Active Plant-Derived Natural Products: A Review. *Biotechnol. Adv.* **2015**, *33*, 1582–1614, doi:10.1016/j.biotechadv.2015.08.001.
7. Cushnie, T.; Cushnie, B.; Echeverría, J.; Fowsantear, W.; Thammawat, S.; Dodgson, J.L.; Law, S.; Clow, S.M. Bioprospecting for Antibacterial Drugs: A Multidisciplinary Perspective on Natural Product Source Material, Bioassay Selection and Avoidable Pitfalls. *Pharm. Res.* **2020**, *37*, 1–24.
8. Svenson, J. MabCent: Arctic Marine Bioprospecting in Norway. *Phytochem. Rev. Proc. Phytochem. Soc. Eur.* **2013**, *12*, 567–578, doi:10.1007/s11101-012-9239-3.
9. Rocha, J.; Peixe, L.; Gomes, N.C.M.; Calado, R. Cnidarians as a Source of New Marine Bioactive Compounds—An Overview of the Last Decade and Future Steps for Bioprospecting. *Mar. Drugs* **2011**, *9*, 1860–1886, doi:10.3390/md9101860.
10. David, B.; Wolfender, J.-L.; Dias, D.A. The Pharmaceutical Industry and Natural Products: Historical Status and New Trends. *Phytochem. Rev.* **2015**, *14*, 299–315, doi:10.1007/s11101-014-9367-z.
11. Fleischmann Robert D.; Adams Mark D.; White Owen; Clayton Rebecca A.; Kirkness Ewen F.; Kerlavage Anthony R.; Bult Carol J.; Tomb Jean-Francois; Dougherty Brian A.; Merrick Joseph M.; et al. Whole-Genome Random Sequencing and Assembly of Haemophilus Influenzae Rd. *Science* **1995**, *269*, 496–512, doi:10.1126/science.7542800.
12. Venter J. Craig; Adams Mark D.; Myers Eugene W.; Li Peter W.; Mural Richard J.; Sutton Granger G.; Smith Hamilton O.; Yandell Mark; Evans Cheryl A.; Holt Robert A.; et al. The Sequence of the Human Genome. *Science* **2001**, *291*, 1304–1351, doi:10.1126/science.1058040.
13. *Pharmacopées traditionnelles en Guyane: Créoles, Wayãpi, Palikur*; Grenand, P., Ed.; Ed. entièrement rev. et complétée.; IRD Éditions, Institut de recherche pour le développement: Paris, 2004; ISBN 978-2-7099-1545-8.
14. HAYNES, J.D.; DIGGS, C.L.; HINES, F.A.; DESJARDINS, R.E. Culture of Human Malaria Parasites Plasmodium Falciparum. *Nature* **1976**, *263*, 767–769, doi:10.1038/263767a0.
15. Marti, G.; Eparvier, V.; Moretti, C.; Susplugas, S.; Prado, S.; Grellier, P.; Retailleau, P.; Guéritte, F.; Litaudon, M. Antiplasmodial Benzophenones from the Trunk Latex of Moronobea Coccinea (Clusiaceae). *Phytochemistry* **2009**, *70*, 75–85, doi:10.1016/j.phytochem.2008.10.005.

16. Marti, G.; Eparvier, V.; Moretti, C.; Prado, S.; Grellier, P.; Hue, N.; Thoison, O.; Delpech, B.; Guéritte, F.; Litaudon, M. Antiplasmodial Benzophenone Derivatives from the Root Barks of Symphonia Globulifera (Clusiaceae). *Phytochemistry* **2010**, *71*, 964–974, doi:10.1016/j.phytochem.2010.03.008.
17. Marti, G.; Eparvier, V.; Litaudon, M.; Grellier, P.; Guéritte, F. A New Xanthone from the Bark Extract of Rheedia Acuminata and Antiplasmodial Activity of Its Major Compounds. *Molecules* **2010**, *15*, 7106–7114, doi:10.3390/molecules15107106.
18. Chervin, J.; Perio, P.; Martins-Froment, N.; Pharkeovilay, C.; Reybier, K.; Nepveu, F.; Fabre, N.; Talou, T.; Bonzon-Ponnet, V.; Marti, G. Dereplication of Natural Products from Complex Extracts by Regression Analysis and Molecular Networking: Case Study of Redox-Active Compounds from Viola Alba Subsp. Dehnhardtii. *Metabolomics* **2017**, *13*, doi:10.1007/s11306-017-1227-6.
19. Marti, G.; Eparvier, V.; Morleo, B.; Ven, J.; Apel, C.; Bodo, B.; Amand, S.; Dumontet, V.; Lozach, O.; Meijer, L.; et al. Natural Aristolactams and Aporphine Alkaloids as Inhibitors of CDK1/Cyclin B and DYRK1A. *Molecules* **2013**, *18*, 3018–3027, doi:10.3390/molecules18033018.
20. Cottet, K.; Genta-Jouve, G.; Fromentin, Y.; Odonne, G.; Duplais, C.; Laprévote, O.; Michel, S.; Lallemand, M.-C. Comparative LC–MS-Based Metabolite Profiling of the Ancient Tropical Rainforest Tree Symphonia Globulifera. *Phytochemistry* **2014**, *108*, 102–108, doi:10.1016/j.phytochem.2014.09.009.
21. Cottet, K.; Kouloura, E.; Kritsanida, M.; Wansi, J.-D.; Odonne, G.; Michel, S.; Halabalaki, M.; Lallemand, M.-C. Comparative Metabolomic Study between African and Amazonian Symphonia Globulifera by Tandem LC–HRMS. *Phytochem. Lett.* **2017**, *20*, 309–315, doi:10.1016/j.phytol.2017.01.012.
22. Wider, G.; Dreier, L. Measuring Protein Concentrations by NMR Spectroscopy. *J. Am. Chem. Soc.* **2006**, *128*, 2571–2576, doi:10.1021/ja055336t.
23. Marti, G.; Boccard, J.; Mehl, F.; Debrus, B.; Marcourt, L.; Merle, P.; Delort, E.; Baroux, L.; Sommer, H.; Rudaz, S.; et al. Comprehensive Profiling and Marker Identification in Non-Volatile Citrus Oil Residues by Mass Spectrometry and Nuclear Magnetic Resonance. *Food Chem.* **2014**, *150*, 235–245, doi:10.1016/j.foodchem.2013.10.103.
24. Wolfender, J.-L.; Marti, G.; Thomas, A.; Bertrand, S. Current Approaches and Challenges for the Metabolite Profiling of Complex Natural Extracts. *J. Chromatogr. A* **2015**, *1382*, 136–164, doi:10.1016/j.chroma.2014.10.091.
25. Schymanski, E.L.; Jeon, J.; Gulde, R.; Fenner, K.; Ruff, M.; Singer, H.P.; Hollender, J. Identifying Small Molecules via High Resolution Mass Spectrometry: Communicating Confidence. *Environ. Sci. Technol.* **2014**, *48*, 2097–2098, doi:10.1021/es5002105.
26. Tsugawa, H.; Nakabayashi, R.; Mori, T.; Yamada, Y.; Takahashi, M.; Rai, A.; Sugiyama, R.; Yamamoto, H.; Nakaya, T.; Yamazaki, M.; et al. A Cheminformatics Approach to Characterize Metabolomes in Stable-Isotope-Labeled Organisms. *Nat. Methods* **2019**, *16*, 295–298, doi:10.1038/s41592-019-0358-2.
27. Creek, D.J.; Dunn, W.B.; Fiehn, O.; Griffin, J.L.; Hall, R.D.; Lei, Z.; Mistrik, R.; Neumann, S.; Schymanski, E.L.; Sumner, L.W.; et al. Metabolite Identification: Are You Sure? And How Do Your Peers Gauge Your Confidence? *Metabolomics* **2014**, *10*, 350–353, doi:10.1007/s11306-014-0656-8.
28. Senan, O.; Aguilar-Mogas, A.; Navarro, M.; Capellades, J.; Noon, L.; Burks, D.; Yanes, O.; Guimerà, R.; Sales-Pardo, M. CliqueMS: A Computational Tool for Annotating in-Source Metabolite Ions from LC-MS Untargeted Metabolomics Data Based on a Coelution Similarity Network. *Bioinformatics* **2019**, *35*, 4089–4097, doi:10.1093/bioinformatics/btz207.
29. Schmid, R.; Petras, D.; Nothias, L.-F.; Wang, M.; Aron, A.T.; Jagels, A.; Tsugawa, H.; Rainer, J.; Garcia-Aloy, M.; Dührkop, K.; et al. Ion Identity Molecular Networking for Mass Spectrometry-Based Metabolomics in the GNPS Environment. *Nat. Commun.* **2021**, *12*, 3832, doi:10.1038/s41467-021-23953-9.
30. Fraisier-Vannier, O.; Chervin, J.; Cabanac, G.; Puech, V.; Fournier, S.; Durand, V.;

Amiel, A.; André, O.; Benamar, O.A.; Dumas, B.; et al. MS-CleanR: A Feature-Filtering Workflow for Untargeted LC–MS Based Metabolomics. *Anal. Chem.* **2020**, *92*, 9971–9981, doi:10.1021/acs.analchem.0c01594.
31. Tsugawa, H.; Cajka, T.; Kind, T.; Ma, Y.; Higgins, B.; Ikeda, K.; Kanazawa, M.; VanderGheynst, J.; Fiehn, O.; Arita, M. MS-DIAL: Data-Independent MS/MS Deconvolution for Comprehensive Metabolome Analysis. *Nat. Methods* **2015**, *12*, 523–526, doi:10.1038/nmeth.3393.
32. Rutz, A.; Dounoue-Kubo, M.; Ollivier, S.; Bisson, J.; Bagheri, M.; Saesong, T.; Ebrahimi, S.N.; Ingkaninan, K.; Wolfender, J.-L.; Allard, P.-M. Taxonomically Informed Scoring Enhances Confidence in Natural Products Annotation. *Front. Plant Sci.* **2019**, *10*, 1329, doi:10.3389/fpls.2019.01329.
33. Chervin, J.; Talou, T.; Audonnet, M.; Dumas, B.; Camborde, L.; Esquerré-Tugayé, M.-T.; Roux, C.; Cabanac, G.; Marti, G. Deciphering the Phylogeny of Violets Based on Multiplexed Genetic and Metabolomic Approaches. *Phytochemistry* **2019**, *163*, 99–110, doi:10.1016/j.phytochem.2019.04.001.
34. Marti, G.; Erb, M.; Boccard, J.; Glauser, G.; Doyen, G.R.; Villard, N.; Robert, C. a M.; Turlings, T.C.J.; Rudaz, S.; Wolfender, J.-L. Metabolomics Reveals Herbivore-Induced Metabolites of Resistance and Susceptibility in Maize Leaves and Roots. *Plant Cell Environ.* **2013**, *36*, 621–639, doi:10.1111/pce.12002.
35. Yadav, B.; Kaur, V.; Narayan, O.P.; Yadav, S.K.; Kumar, A.; Wankhede, D.P. Integrated Omics Approaches for Flax Improvement under Abiotic and Biotic Stress: Current Status and Future Prospects. *Front. Plant Sci.* **2022**, *13*, 931275, doi:10.3389/fpls.2022.931275.
36. Abdullah-Zawawi, M.-R.; Govender, N.; Karim, M.B.; Altaf-Ul-Amin, Md.; Kanaya, S.; Mohamed-Hussein, Z.-A. Chemoinformatics-Driven Classification of Angiosperms Using Sulfur-Containing Compounds and Machine Learning Algorithm. *Plant Methods* **2022**, *18*, 118, doi:10.1186/s13007-022-00951-6.
37. Jang, Y.K.; Jung, E.S.; Lee, H.-A.; Choi, D.; Lee, C.H. Metabolomic Characterization of Hot Pepper (*Capsicum Annuum* "CM334") during Fruit Development. *J. Agric. Food Chem.* **2015**, *63*, 9452–9460, doi:10.1021/acs.jafc.5b03873.
38. Medema, M.H.; de Rond, T.; Moore, B.S. Mining Genomes to Illuminate the Specialized Chemistry of Life. *Nat. Rev. Genet.* **2021**, *22*, 553–571, doi:10.1038/s41576-021-00363-7.
39. Heux, S.; Bergès, C.; Millard, P.; Portais, J.-C.; Letisse, F. Recent Advances in High-Throughput 13C-Fluxomics. *Curr. Opin. Biotechnol.* **2017**, *43*, 104–109.
40. Cottret, L.; Frainay, C.; Chazalviel, M.; Cabanettes, F.; Gloaguen, Y.; Camenen, E.; Merlet, B.; Heux, S.; Portais, J.-C.; Poupin, N.; et al. MetExplore: Collaborative Edition and Exploration of Metabolic Networks. *Nucleic Acids Res.* **2018**, *46*, W495–W502, doi:10.1093/nar/gky301.
41. Dewick, P.M. *Medicinal Natural Products: A Biosynthetic Approach*; 3rd edition.; Wiley, A John Wiley and Sons, Ltd., Publication: Chichester, West Sussex, United Kingdom, 2009; ISBN 978-0-470-74168-9.
42. Wang, S.; Alseekh, S.; Fernie, A.R.; Luo, J. The Structure and Function of Major Plant Metabolite Modifications. *Mol. Plant* **2019**, *12*, 899–919, doi:10.1016/j.molp.2019.06.001.
43. Djoumbou Feunang, Y.; Eisner, R.; Knox, C.; Chepelev, L.; Hastings, J.; Owen, G.; Fahy, E.; Steinbeck, C.; Subramanian, S.; Bolton, E.; et al. ClassyFire: Automated Chemical Classification with a Comprehensive, Computable Taxonomy. *J. Cheminformatics* **2016**, *8*, doi:10.1186/s13321-016-0174-y.
44. Kim, H.W.; Wang, M.; Leber, C.A.; Nothias, L.-F.; Reher, R.; Kang, K.B.; van der Hooft, J.J.J.; Dorrestein, P.C.; Gerwick, W.H.; Cottrell, G.W. NPClassifier: A Deep Neural Network-Based Structural Classification Tool for Natural Products. *J. Nat. Prod.* **2021**, *84*, 2795–2807, doi:10.1021/acs.jnatprod.1c00399.
45. Zheng, S.; Zeng, T.; Li, C.; Chen, B.; Coley, C.W.; Yang, Y.; Wu, R. Deep Learning Driven Biosynthetic Pathways Navigation for Natural Products with BioNavi-NP. *Nat.*

Commun. **2022**, *13*, 3342, doi:10.1038/s41467-022-30970-9.
46. Misra, B.B. New Software Tools, Databases, and Resources in Metabolomics: Updates from 2020. *Metabolomics* **2021**, *17*, 49, doi:10.1007/s11306-021-01796-1.
47. Sorokina, M.; Steinbeck, C. Review on Natural Products Databases: Where to Find Data in 2020. *J. Cheminformatics* **2020**, *12*, doi:10.1186/s13321-020-00424-9.
48. Willighagen, E.L.; Mayfield, J.W.; Alvarsson, J.; Berg, A.; Carlsson, L.; Jeliazkova, N.; Kuhn, S.; Pluskal, T.; Rojas-Chertó, M.; Spjuth, O.; et al. The Chemistry Development Kit (CDK) v2.0: Atom Typing, Depiction, Molecular Formulas, and Substructure Searching. *J. Cheminformatics* **2017**, *9*, 33, doi:10.1186/s13321-017-0220-4.
49. Yang, H.; Lou, C.; Sun, L.; Li, J.; Cai, Y.; Wang, Z.; Li, W.; Liu, G.; Tang, Y. AdmetSAR 2.0: Web-Service for Prediction and Optimization of Chemical ADMET Properties. *Bioinformatics* **2019**, *35*, 1067–1069, doi:10.1093/bioinformatics/bty707.
50. Capecchi, A.; Reymond, J.-L. Classifying Natural Products from Plants, Fungi or Bacteria Using the COCONUT Database and Machine Learning. *J. Cheminformatics* **2021**, *13*, 82, doi:10.1186/s13321-021-00559-3.
51. Wang, F.; Liigand, J.; Tian, S.; Arndt, D.; Greiner, R.; Wishart, D.S. CFM-ID 4.0: More Accurate ESI-MS/MS Spectral Prediction and Compound Identification. *Anal. Chem.* **2021**, *93*, 11692–11700, doi:10.1021/acs.analchem.1c01465.
52. Tsugawa, H.; Kind, T.; Nakabayashi, R.; Yukihira, D.; Tanaka, W.; Cajka, T.; Saito, K.; Fiehn, O.; Arita, M. Hydrogen Rearrangement Rules: Computational MS/MS Fragmentation and Structure Elucidation Using MS-FINDER Software. *Anal. Chem.* **2016**, *88*, 7946–7958, doi:10.1021/acs.analchem.6b00770.
53. Kind, T.; Okazaki, Y.; Saito, K.; Fiehn, O. LipidBlast Templates As Flexible Tools for Creating New In-Silico Tandem Mass Spectral Libraries. *Anal. Chem.* **2014**, *86*, 11024–11027, doi:10.1021/ac502511a.
54. Tang, D.-Q.; Zou, L.; Yin, X.-X.; Ong, C.N. HILIC-MS for Metabolomics: An Attractive and Complementary Approach to RPLC-MS: HILIC-MS FOR METABOLOMICS. *Mass Spectrom. Rev.* **2016**, *35*, 574–600, doi:10.1002/mas.21445.
55. Yang, K.; Han, X. Lipidomics: Techniques, Applications, and Outcomes Related to Biomedical Sciences. *Trends Biochem. Sci.* **2016**, doi:10.1016/j.tibs.2016.08.010.
56. Salem, M.A.; Jüppner, J.; Bajdzienko, K.; Giavalisco, P. Protocol: A Fast, Comprehensive and Reproducible One-Step Extraction Method for the Rapid Preparation of Polar and Semi-Polar Metabolites, Lipids, Proteins, Starch and Cell Wall Polymers from a Single Sample. *Plant Methods* **2016**, *12*, doi:10.1186/s13007-016-0146-2.
57. Singh, A.; Shannon, C.P.; Gautier, B.; Rohart, F.; Vacher, M.; Tebbutt, S.J.; Lê Cao, K.-A. DIABLO: An Integrative Approach for Identifying Key Molecular Drivers from Multi-Omics Assays. *Bioinformatics* **2019**, *35*, 3055–3062, doi:10.1093/bioinformatics/bty1054.

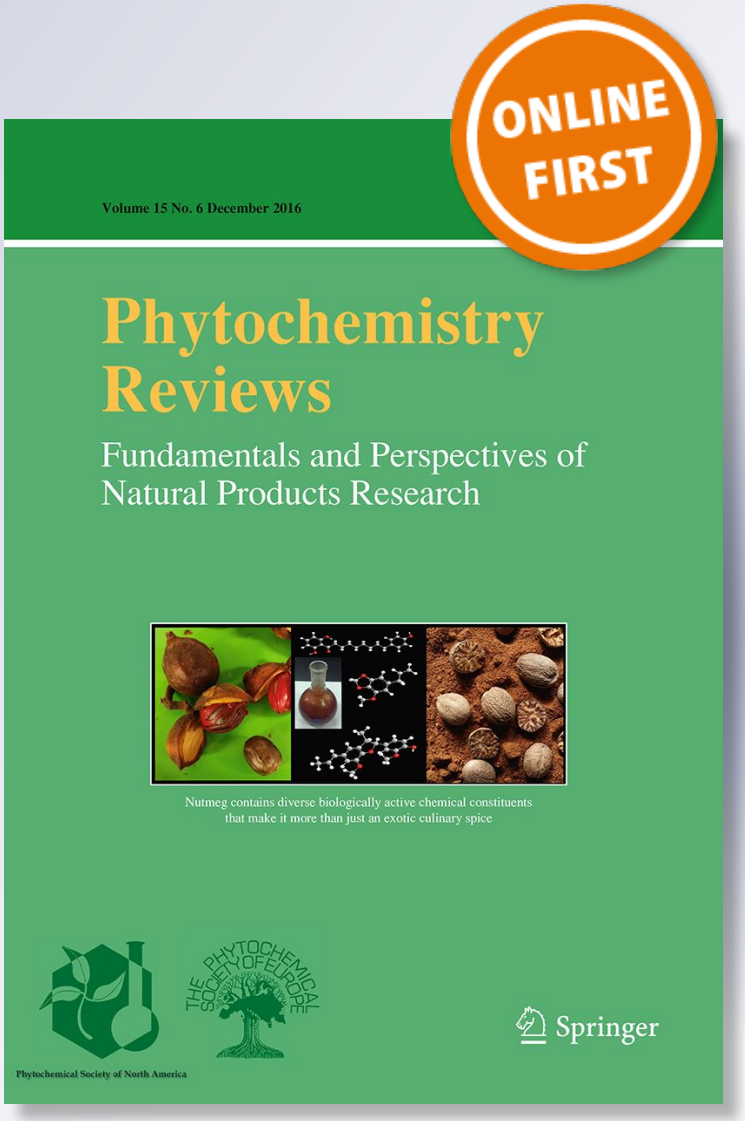
The landscape of natural product diversity and their pharmacological relevance from a focus on the Dictionary of Natural Products ®

François Chassagne, Guillaume Cabanac, Gilles Hubert, Bruno David & Guillaume Marti

Phytochemistry Reviews
Fundamentals and Perspectives of Natural Products Research

ISSN 1568-7767

Phytochem Rev
DOI 10.1007/s11101-019-09606-2



Springer

Phytochem Rev
https://doi.org/10.1007/s11101-019-09606-2

The landscape of natural product diversity and their pharmacological relevance from a focus on the *Dictionary of Natural Products*®

François Chassagne · Guillaume Cabanac · Gilles Hubert · Bruno David · Guillaume Marti

Received: 5 February 2019 / Accepted: 17 April 2019

Abstract Nature is considered a prolific source of diverse biologically active chemotypes. While most reviews have focused on the characteristics of the chemical backbones of natural products (NPs), few have tried to provide an overview of their origins (the living organisms in which they are produced), chemical classes, and biological activities. This review discusses the current knowledge on NP diversity by focusing on the *Dictionary of Natural Products*® (DNP). We datamined the 300,000 NPs covered by the DNP to reveal relevant, albeit dormant, knowledge about NP diversity. This holistic picture of NPs allows us to discuss the most abundant biological sources of NPs investigated in relation to their chemical features and biological activities. In a nutshell, a large part of NPs originated from plants (67%), especially from the Compositae and Leguminosae families. Among all kingdoms, NPs isolated from *Streptomyces* spp. were largely represented, while terpenoids and alkaloids were the two most represented chemical classes. Out of all NPs documented, only 3882 were reported to be bioactive (1163 from plants and 1006 from bacteria), with antibacterial, antibiotics, and antineoplastic agents being the most frequent therapeutic classes. In this paper, we also address the advantages and limitations of NP research from a pharmaceutical industry perspective. This work will provide useful insights and guidance to researchers involved in drug discovery from NPs.

Keywords Biological activity · Drug discovery · Genetic resources · Pharmaceutical industry · Plants

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11101-019-09606-2) contains supplementary material, which is available to authorized users.

F. Chassagne (✉)
Center for the Study of Human Health, Emory University, 615 Michael Street, Whitehead Building, Room 111, Atlanta, GA 30322, USA
e-mail: francois.chassagne@emory.edu

F. Chassagne · G. Cabanac · G. Hubert · G. Marti
UMR 5505, CNRS, Computer Science Department, Informatics Research Institute of Toulouse (IRIT), Université Toulouse 3 – Paul Sabatier, 118, Route de Narbonne, 31062 Toulouse, France

B. David
Green Mission Pierre Fabre, Institut de Recherche Pierre Fabre, 3, Avenue Hubert Curien, BP 13562, 31562 Toulouse, France

G. Marti
PharmaDev, UMR 152 IRD, Université Toulouse 3 – Paul Sabatier, 35 Chemin des Maraîchers, 31400 Toulouse, France

Published online: 24 April 2019

Abbreviations

ACE	Angiotensin converting enzyme
CAS	Chemical abstracts services
DNP	Dictionary of Natural Products®
HTS	High throughput screening
NP	Natural product
PSK	Polysaccharide krestin
PSP	Polysaccharopeptide

Introduction

Natural products (NPs) have been used in various branches of traditional medicine for millennia. The oldest written records of NPs as medicine date as far back as 4600 years ago (Dias et al. 2012; Cragg and Newman 2013). Due to their chemical diversity, their structural complexity, and their biological selectivity, NPs are considered to be a great source of inspiration for the development of potential novel drugs (Clardy and Walsh 2004; Atanasov et al. 2015). From 1981 to 2014, about 51% of all new approved drugs were of (or derived from) natural sources, amounting to 65% of all antibacterial compounds and 73% of all anticancer compounds (Newman and Cragg 2016). It is expected that newly discovered natural compounds will continue to play an important role in drug discovery (Baker et al. 2007; Harvey et al. 2015).

Natural product databases have been developed to assist with in silico and in vitro screening in drug discovery. A number of NP libraries are already available and include general databases such as *SuperNatural 2* (≈ 326,000 NP) (Banerjee et al. 2015) and the *Universal Natural Products Database* (≈ 229,000 compounds) (Chen et al. 2017b). Specialized databases have also been designed, including those focusing on indigenous medicines such as *AfroDb* for African medicinal plants (Ntie-Kang et al. 2013), *NuBBE* for Brazilian biodiversity (Valli et al. 2013), *iSMART*, and *TCMID* for traditional Chinese medicine (Chang et al. 2011; Xue et al. 2013); and those focusing on specific types of NPs such as *HIT* for herbal ingredients and their targets (Ye et al. 2011), *SuperToxic* for toxic compounds (Schmidt et al. 2009), *NPACT* for anticancer NPs (Mangal et al. 2013), and *MarinLit* for marine NPs (Chen et al. 2017b).

The *Dictionary of Natural Products*® (CRC Press, v. 27.1) (DNP) is a compilation of all known compounds derived from natural sources and can be considered as one of the most comprehensive libraries of NPs available to date (Quinn et al. 2008; Gaudêncio and Pereira 2015; Chen et al. 2017b). The latest version of the DNP (v. 27.1, at the time of writing) provides data for nearly 300,000 natural compounds, and it contains information on the chemical, physical, and biological properties of compounds; along with their systematic and common names, literature references, molecular structures, and natural sources for purification (including family, genus, and species). Due to its rich content, the DNP can be considered as a body of knowledge for NPs and can be used to guide investigations in NP-based drug discovery.

Since its development, several studies have used and explored the content of the DNP. One of the first comprehensive reports was published by Henkel et al. (1999) who investigated the differences between the structural properties of NPs found in the DNP and those of synthetic substances; then Whittle et al. (2003) evaluated various similarity measures for screening molecular fingerprints from the DNP; later Koch et al. (2005) introduced a structural classification of NPs to chart biologically relevant chemical space. With the same objective, Rosén et al. (2009) compared the chemical space of bioactive medicinal chemistry compounds from the WOMBAT database to that of products from the DNP; in 2008, Quinn et al. (2008) used the DNP to develop a library of NPs that exhibit drug-like properties; also Kong et al. (2011) performed a historical analysis on the structural novelty of products from the DNP; more recently, Pascolutti et al. (2015) identified fragment-sized NPs from the DNP to capture the structural diversity of nature. These studies mainly focused on the characteristics of the chemical backbones of the NP contained in this database, but none of them strove to provide an overview of their origins (the living organisms in which they are produced), chemical classes, and biological activities. To the best of our knowledge, only Bérdy (2005) summarized compounds isolated from natural sources, but his review was limited to bioactive compounds from microbial sources.

In this review, we unveil the current knowledge on NP diversity based on a descriptive study of the DNP (v. 27.1). This includes a characterization of all compounds (primary and secondary metabolites) from the DNP in relation to their natural sources (all types of organisms are considered), chemical classes, and

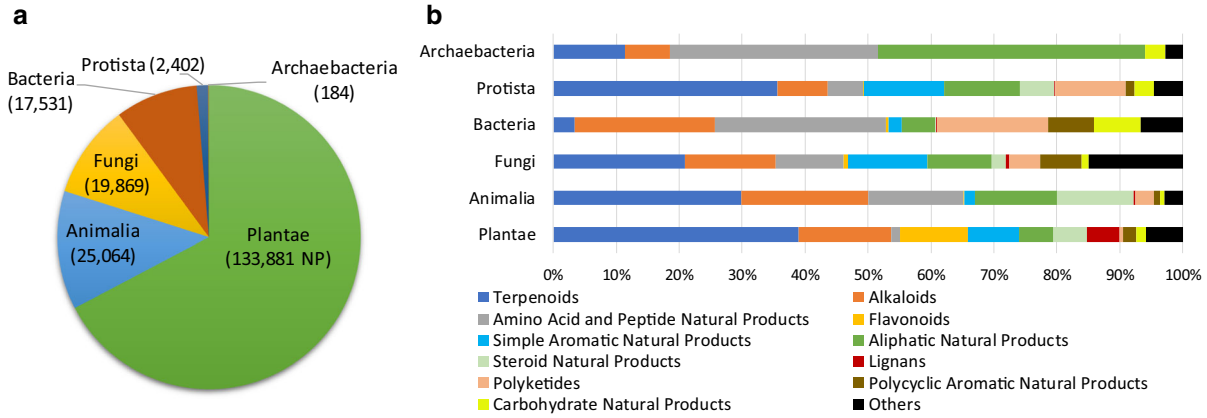


Fig. 1 Overview of the Dictionary of Natural Products database, **a** distribution of natural products per kingdom of life; **b** distribution of the main chemical classes of natural products in each kingdom of life

biological activities (NPs without bioactivities are also reviewed). Natural products with biological activities annotated in the DNP are defined as "compounds with established activity and being used as drugs or under investigation for drug use" (Taylor & Francis group, pers. comm.). Finally, we also discuss the potential application of NPs in the pharmaceutical industry.

Defining the number of NPs already identified

A key issue in NP research is the estimation of the number of known NPs. First, Bérdy (2005) estimated the approximate number of known NPs to be close to one million. Later, the same author proposed that the number of published natural compounds was actually between 300,000 and 600,000 (Bérdy 2012). However, little information was provided on the methodology used to infer these figures. Other authors based their estimations on the number of compounds tabulated in the NP databases (Blunt et al. 2012). At present, in the general NP databases, the number of compounds described amounts to 326,000 NPs in *Super Natural 2*, 293,798 NPs in the DNP, 283,000 NPs in the *Chemical Abstract Services (CAS) registry*, and 220,000 NPs in *Reaxys* (Gaudêncio and Pereira 2015; Chen et al. 2017b). By comparing the number of unique NPs from a wide range of commercial and freely available virtual databases, Chen et al. (2017b) estimated a figure close to 250,000 NPs. Since the authors did not include *SuperNatural 2* and *Reaxys* in their analysis, the number of known NPs is likely to be higher. Furthermore, the number of NPs could even be larger if we consider the molecular fossils produced from natural products (Falk and Wolkenstein 2017).

In the DNP, out of the 194,977 NPs with information on organism classification, most of the NPs were from the Plantae kingdom (133,881 NPs, 67.3%), while Animalia ranked second (25,064 NPs, 12.6%), followed by Fungi (19,869 NPs, 9.99%) and Eubacteria (17,531, 8.81%) (Fig. 1a). Regarding the chemistry of NPs, terpenoids and alkaloids groups were the two most represented chemical classes in these four kingdoms, and they represented more than half of all compounds isolated from the Plantae kingdom (Fig. 1b). In contrast, some chemical classes did not have a cross distribution in the different kingdoms of life. This is the case of steroid NPs which were seldomly represented in the Bacteria kingdom. Likewise, flavonoids and lignans were almost exclusively found in Plantae kingdom, while polyketides were mainly found in Bacteria, Fungi, and Protista kingdoms, which emphasize two distinct crossroads among the acetate pathway.

In the following sections, we provide a detailed analysis of the chemodiversity found in the most represented families and species of each kingdom with a special focus on their pharmacologically active compounds.

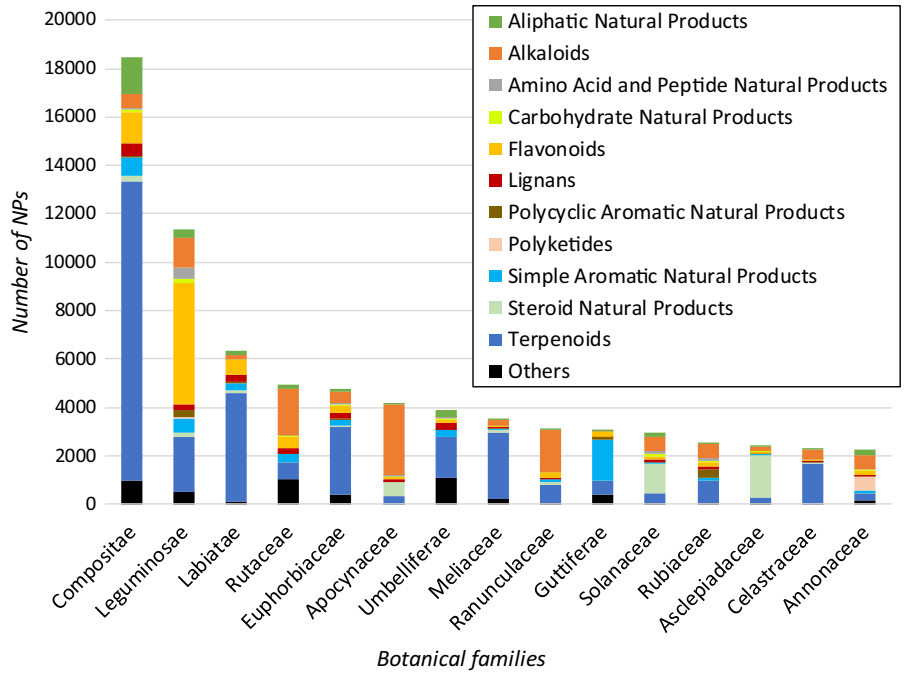


Fig. 2 The top 15 botanical families containing natural products and the distribution of the different chemical classes

NPs from the Plantae kingdom

A large percentage of the NPs (67.3%) recorded in the DNP are from the Plantae kingdom. This is consistent with Bérdy (2012) who estimated that approximately 70% of NP are of plant origin. From a historical perspective, this is not surprising since medicinal herbs were the first natural substances to be studied. It started with the isolation of morphine (alkaloid) in 1817 by a German pharmacist named Friedrich Sertuerner (Sertuerner, 1817). It gained momentum during the nineteenth century with the discovery of a wide range of natural compounds of plant origin (e.g., aspirin, atropine, caffeine, cocaine, colchicine, quinine, nicotine, and strychnine). Then, the discovery of penicillin in 1928 by Alexander Fleming and its purification by Howard Florey in 1938 led to an era of natural products discovery from microbial sources (David et al. 2015; Atanasov et al. 2015; Bernardini et al. 2017).

One of the distinctive features of the plant compounds compiled in the DNP is the high proportion of terpenoids, which are especially pervasive in the Gymnosperms group (57% of the total NPs) and in the Dicotyledons group (40%) (Fig. 1b). Terpenoids are known as the largest and the most diverse class of NPs, and they play various roles in plant development and growth, as well as in communication and defence (Langenheim 1994; Gershenzon and Dudareva 2007). In the DNP, one third of terpenoid NPs have antineoplastic-related activities including, for instance (Huang et al. 2012):

- Limonene, a monoterpene from *Citrus limon* (L.) Osbeck.
- Tanshinone IIA, a diterpene from *Salvia miltiorrhiza* Bunge.
- Celastrol, a triterpene from *Tripterygium wilfordii* Hook. f.
- Lycopene, a tetraterpene from *Lycopersicon esculentum* Mill.

Most of the NPs isolated from the Plantae kingdom are from the dicotyledons group (Angiosperms clade) (83.7%), followed by the monocotyledons group (8.1%), and the Gymnosperms clade (3%). Liverworts, ferns, fern allies, and mosses represent only a small part of the NP recorded (3.2%), just as rhodophyta and chlorophytes groups (2%). Furthermore, three botanical families (Compositae, Leguminosae, and Labiatae) host about one quarter of the total

compounds from the Plantae kingdom (Fig. 2). With over 32,700 species recorded, Compositae (also known as Asteraceae) is the largest family of flowering plants worldwide, while Leguminosae (or Fabaceae) is the third-largest with more than 20,800 species (Roskov et al. 2018).

In the Compositae family, the *Artemisia* genus, one of the largest genera in this family ($\approx$ 500 species), was the most represented in the DNP (1299 NP) (Suppl. Mat. S1A). The antimalarial compound, artemisinin, isolated from the traditional Chinese medicinal herb *Artemisia annua* L. in 1971 led to the development of various derivatives (e.g., artemether and sodium artesunate), and it has also attracted renewed interest into sesquiterpenes which are well represented in this genus (Kayser et al. 2003; Atanasov et al. 2015). For example, arglabin isolated from *A. glabella* Kar. and Kir. is used in Kazakhstan for cancer chemotherapy, and santonin produced by *A. cina* Berg ex Poljakov was widely used in the past as an anthelminthic agent (Ivanescu et al. 2015).

According to the DNP, about half of the compounds found in the Leguminosae family are flavonoids. Indeed, a wide range of flavonoids are found in this botanical family including quercetin, kaempferol, and their derivatives (Wink 2013). Some of which are currently available or under investigation as drugs, such as genistein from *Glycine max* (L.) Merr. which is in clinical trials for use as an angiogenesis inhibitor (Veitch 2010; Russo et al. 2016). Other compounds isolated from the Leguminosae family that are of major importance in human health include the oleanane triterpenoids from *Glycyrrhiza glabra* L. (e.g., glycyrrhizin used as anti-inflammatory agent), polycyclic aromatic NP from *Cassia angustifolia* M.Vahl (i.e., sennosides used as laxatives), and cyclotryptamine alkaloids from *Physostigma venenosum* Balf. (i.e., physostigmine used to treat neurological diseases). Altogether, the Leguminosae family is known as the most drug-prolific botanical family with 44 drugs either approved or in clinical trials in 2011 (Zhu et al. 2011).

The Labiatae family (or Lamiaceae) is the third largest botanical group, and according to the DNP a large number of the compounds (71%) isolated from this family are terpenoids. These results could be explained by the numerous reports on volatile oils (mainly monoterpenes and sesquiterpenes) found in genera of economic importance (e.g., *Lavandula*, *Mentha* and *Thymus*) (Wu et al. 2012). *Salvia* species account for almost 20% of the overall NPs isolated from Labiatae, with *Salvia miltiorrhiza* Bunge being the most represented. *S. miltiorrhiza* (Danshen in Chinese) is a very popular traditional Chinese herb used to treat several conditions including cardiovascular, cerebrovascular, and hyperlipidaemia diseases (Wu et al. 2012). Lipophilic diterpenoids (i.e., tanshinones) and hydrophilic phenolic acids (i.e., salvianolic acids) are the main bioactive constituents of this species. These compounds have been shown to possess various pharmacological effects ranging from anticancer properties to effects on cardiovascular diseases (Su et al. 2015). More than thirty clinical trials of *S. miltiorrhiza* and its tanshinones have been undertaken to justify its use in stroke patients, angina, and other ischemic conditions (Adams et al. 2006; Tan et al. 2018). Moreover, Dantonic, a Chinese botanical drug containing *S. miltiorrhiza* extracts, has gone through Phase III clinical trials (completed in 2016) by the U.S. Food and Drug Administration (FDA) for the treatment of angina pectoris and cardiovascular diseases (Chao et al. 2017).

Of the most cited genera in the DNP, the *Euphorbia* genus ranks first in the Plantae kingdom and fourth amongst all kingdoms. With about 2160 known species, *Euphorbia* is the largest genus in the Euphorbiaceae family and the third-largest amongst the flowering plants (Jassbi 2006; Ernst et al. 2015). In the DNP, approximately two thirds of the total NPs (1008 compounds) isolated from this genus are diterpenoids. The diterpenoids are the most studied chemical class from *Euphorbia* species, especially the polycyclic diterpenoids (e.g., jatrophane, ingenane, tigliane, and lathyrane) (Shi et al. 2008). Indeed, these compounds are taxonomic markers of the Euphorbiaceae family (they occur only in Thymelaeaceae and Euphorbiaceae families), and are lead compounds in drug discovery from NPs (Vasas and Hohmann 2014). One diterpenoid compound (i.e., ingenol mebutate isolated from *E. peplus*) was approved by the FDA and the European Medicines Agency in 2012 for the treatment of actinic keratosis, and it is currently used in clinical practice (Cantisani et al. 2013).

Interestingly, *Tripterygium wilfordii* Hook. f. (Celastraceae) was the Plantae species that exhibited the highest number of NPs in the DNP (315) (Table 1). This liana has been used in traditional Chinese medicine for more than 2000 years for the treatment

Table 1 Top 15 natural sources (at the species level) of natural products

Biological sources (species)	Kingdom of life	Number and frequency of NP		Chemical class[a]		Biological activities[b]			
		N	%	Top 1	%	*N*	Top 1	%	Examples
Escherichia coli	Bacteria	487	2.2	AA	79.4	12	Antineop.	29.4	Asparaginase
Streptomyces hygroscopicus	Bacteria	353	1.6	Polyk	58.1	44	Antibact.	70.5	Nigericin, rapamycin
Tripterygium wilfordii	Plantae	315	1.4	Terp	71.1	7	Antineop.	36.4	Celastrol, triptolide
Azadirachta indica	Plantae	278	1.3	Terp	84.8	6	Insectic.	25.0	Azadirachtin
Helianthus annuus	Plantae	263	1.2	Terp	45.7	0	NA		
Bacillus subtilis	Bacteria	252	1.1	AA	61.7	14	Antibact.	35.0	Bacitracin, rhizocticin
Nicotiana tabacum	Plantae	248	1.1	Terp	22.9	1	Gang. block.	50.0	Nicotine
Aspergillus terreus	Fungi	244	1.1	Alk	20.8	9	Antibact.	33.3	Patulin
Ganoderma lucidum	Fungi	239	1.1	Terp	86.9	1	Immuno.	100	Ling zhi-8
Gynostemma pentaphyllum	Plantae	239	1.1	Terp	95.4	1	Antineop.	100	Protopanaxadiol
Camellia sinensis	Plantae	236	1.1	Terp	37.1	2	Antifung.	50.0	Octanoic acid
Panax ginseng	Plantae	221	1.0	Terp	71.6	4	Antineop.	100	Ginsenosides
Saccharomyces cerevisiae	Fungi	220	1.0	AA	83.6	4	Antianaem.	20.0	Sargramostim
Lyngbya majuscula	Bacteria	220	1.0	AA	38.2	2	Antifung.	33.3	Laxaphycin B
Sorangium cellulosum	Bacteria	214	1.0	Polyk	48.7	2	Antineop.	100	Epothilones

NA not applicable

[a]Most represented chemical class in each species with its percentage. *AA* amino acid and peptide, *Alk* alkaloids, *Polyk* polyketides, *Terp* terpenoids

[b]Number (N) of NP (for each species) with established activity and being used as drugs or under investigation for drug use, and most represented biological activities (with their percentage) in each species. *Antianaem.* antianaemic agents, *Antibact.* antibacterial agents and antibiotics, *Antifung.* antifungal agents, *Antineop.* antineoplastics agents, *Gang. block.* ganglion blocking agents, *Immuno.* immunomodulators, *Insectic.* insecticides

of arthritis, muscle, and skeletal injury, as well as skin diseases (Chen 2001). In the 1960s, formulations of *T. wilfordii* began to be used in Chinese allopathic medicine to treat patients with inflammatory lesions caused by leprosy, and patients with rheumatoid arthritis (Tao and Lipsky 2000). Since then, numerous clinical trials have been conducted for the treatment of inflammatory diseases, and over 300 compounds have been identified from this plant, many of which displaying immunosuppressive and anti-inflammatory effects (Brinker et al. 2007). Its most abundant metabolites are diterpenoids including triptolide, tripdiolide, and triptonide (Goldbach-Mansky 2009). Triptolide analogues are currently in Phase II clinical trials for the treatment of rheumatoid arthritis and cancer (Chen et al. 2018).

NPs from the Bacteria kingdom

According to the DNP, the number of NP from the Bacteria kingdom is less than that from the other kingdoms of life (only 9% of the total compounds from living organisms). However, the percentage of bioactive compounds from the Bacteria kingdom is three to five times higher than that from the other kingdoms of life (Suppl. Mat. S1B). Bérdy (2012) estimated that approximately 8% of NPs were from bacterial origin (actinobacteria and unicellular bacteria), but 47% of these compounds exhibited some kind of biological activity. Interestingly, the latter figure was lower in plants (7%) and in animal-derived compounds (3%).

Bioactive compounds from bacterial origin mainly display antibiotic and antibacterial properties (67% of bioactive NPs from the Bacteria kingdom in the DNP) (Fig. 3). Regarding the number of FDA approved

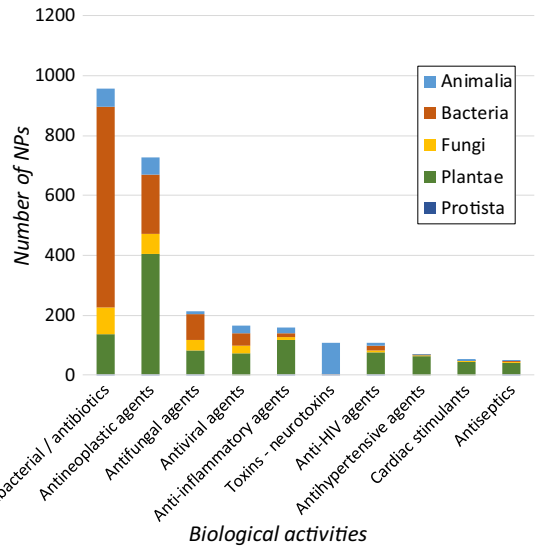


Fig. 3 The top 10 activities of natural products and their distribution by kingdom of life

drugs from natural sources, Patridge et al. (2016) reported that 51% of antibacterial agents were from bacterial origin.

Many NPs from bacteria belong to the amino acid and peptide class (Fig. 1). In the DNP, 27% of NPs from bacterial origin were amino acids and peptides, and cyclic-, oligo- and polypeptides were the main represented chemical subclass with 1147 NPs. Moreover, the amino acid and peptide group (236 NPs) and the polyketide group (248 NPs) combined amount to approximately half of the bioactive NPs from bacterial sources.

Among all living organisms, the Streptomycetaceae family and *Streptomyces* species rank top as producers of NPs (7951 NPs in total from *Streptomyces*) (Table 2). Bérdy (2005) estimated that 34% of metabolites from microbial sources were from this genus. In our study, we found that this number is even higher and reaches 45%. Indeed, with the discovery of streptothricin in 1942 and streptomycin in 1943 at the beginning of the antibiotic era, attention turned to *Streptomyces* species. Pharmaceutical companies engaged in large efforts to discover NPs by developing research programs based on microbial fermentation, and by focusing mainly on antibacterial and antifungal targets (Clardy et al. 2006; Baker et al. 2007; Katz and Baltz 2016). In the 1950s and 1960s, about 70–80% of antibiotics were discovered from *Streptomyces* species and this percentage was still high at the beginning of the 1990s ($\approx$ 40–50%) (Bérdy 2005). Several important compounds belonging to various antibiotic classes were discovered from *Streptomyces* species including β-lactams (cephamycin and carbapenems); aminoglycosides (neomycin and kanamycin); macrolides (tylosin and spiramycin); peptides (actinomycin); polyenes (candicidin, amphotericin B and nystatin); and tetracyclines (tetracycline, chlortetracycline and oxytetracycline) (Katz and Baltz 2016). Besides the discovery of antibiotics, a wide range of compounds related to various therapeutic classes have been isolated from this genus. Let us mention anthelminthic and antiparasitic drugs (e.g., ivermectins), anti-tumour products (e.g., bleomycin and doxorubicin), anti-obesity agents (e.g., lipstatin), immunosuppressive agents (e.g., rapamycin) and herbicides (e.g., bialaphos). Altogether, Watve et al. (2001) estimated that this genus is capable of producing approximately 100,000 antimicrobial compounds, of which only 1–3% have been discovered so far.

More generally, microbial organisms from the Actinomycetales order represent the largest group of bioactive bacterial compounds. According to the DNP 9261 compounds have been isolated from these species, including the largest genus *Streptomyces*, but also rare actinomycetes such as *Micromonospora* (436 total NPs and 84 bioactive NPs), *Nocardia* (312 and 32) and *Actinomadura* (267 and 27) species. The discovery of rare actinomycetes increased significantly between the 1970s and the 1990s, and during this period about 30% of new antibiotics stemmed from these species (Bérdy 2005). Besides the fact that they are difficult to isolate and cultivate, one of the distinctive features of the rare actinomycetes is their production of diverse, unique, and complex compounds with excellent antibacterial potency and usually low toxicity (Lazzarini et al. 2000; Kurtböke 2012). Glycopeptide and orthosomycin antibiotics are produced almost exclusively by rare actinomycetales (Bérdy 2005). Some of the successful antibacterial agents currently on the market include gentamicin isolated from *Micromonospora purpurea*, erythromycin produced by *Saccharopolyspora erythraea*, rifamycins by *Amycolatopsis rifamycinica* and vancomycin by *Amycolatopsis orientalis* (Tiwari and Gupta 2012). Moreover, the rare actinomycetes are highly drug-prolific, with the Pseudonocardiaceae family (including *Amycolatopsis* and *Saccharopolyspora* species) being the most productive with

Table 2 Top 15 natural sources (at the genus level) of natural products

Biological sources (genus)	Kingdom of life	Number and frequency of NP		Chemical class[a]		Biological activities[b]			
		N	%	Top 1	%	*N*	Top 1	%	Examples
Streptomyces	Bacteria	7951	3.4	Polyk	24.5	597	Antibact.	63.6	Amphotericin B, avermectin, deferoxamine, doxorubicin, nystatin, tetracycline
Aspergillus	Fungi	2374	1.0	Alk	25.1	67	Antibact.	25.7	Fumagillin, lovastatin
Penicillium	Fungi	2104	0.9	Alk	21.7	31	Antibact.	25.5	Benzylpenicillin, griseofulvin
Euphorbia	Plantae	1434	0.6	Terp	76.5	8	Antivir.	28.5	Ingenol
Artemisia	Plantae	1299	0.5	Terp	66.5	21	Antimal.	11.1	Artemisinin
Salvia	Plantae	1166	0.5	Terp	79.4	9	Antineop.	21.4	Salvinorin A, tanshinones
Pseudomonas	Bacteria	952	0.4	AA	37.5	40	Antibact.	45.1	Aeruginoic acid, mupirocin
Aconitum	Plantae	948	0.4	Alk	92.8	7	Anaesth. g.	12.5	Aconitine
Isodon	Plantae	895	0.4	Terp	96.2	4	Antineop.	57.1	Oridonin, ponicidin
Garcinia	Plantae	892	0.4	Aro	68.9	7	Antivir.	27.2	Gambogic acid
Senecio	Plantae	865	0.4	Terp	68.6	6	Antispasm.	28.5	Platyphylline
Laurencia	Plantae	851	0.4	Terp	64.3	0	NA		
Bacillus	Bacteria	849	0.4	AA	53.4	39	Antibact.	47.5	Bacitracin, spergualin
Piper	Plantae	847	0.4	Alk	30.8	17	Antipyr.	13.2	Piperine
Taxus	Plantae	797	0.3	Terp	68.4	3	Antineop.	16.7	Paclitaxel

NA not applicable

[a]Most represented chemical class in each genus with its percentage. *AA* amino acid and peptide, *Alk* alkaloids, *Aro* simple aromatic NP, *Polyk* polyketides, *Terp* terpenoids

[b]Number (N) of NP (for each genus) with established activity and being used as drugs or under investigation for drug use, and most represented biological activity (with its percentage) for each genus. *Anaesth. g.* general anaesthetics, *Antibact.* antibacterial agents and antibiotics, *Antimal.* antimalarial agents, *Antineop.* antineoplastic agents, *Antipyr.* antipyretic agents, *Antispasm.* antispasmodic agents, *Antivir.* antiviral agents

76 approved drugs in 2011 (Zhu et al. 2011). Nowadays, rare actinomycetes can be isolated from diverse environments (i.e., marine sources, soil samples, plant materials, and extreme environments) and intensively studied in antibiotic discovery programs (Jose and Jebakumar 2013; Dhakal et al. 2017).

Finally, other eubacteriales species including *Pseudomonas*, *Bacillus* and *Escherichia* species are among the highest producers of NPs. Bérdy (2005) already stressed the importance of the *Bacillus* and *Pseudomonas* species as the most prolific producers of bioactive metabolites from the unicellular bacteria group. Due to their ubiquity in the environment (i.e., soil, water, plants, and animals), *Pseudomonas* species are known to synthetize a wide range of metabolites (Gross and Loper 2009). In particular, *Pseudomonas* is one of the most common genera associated with plants as it contains epiphytic, endophytic, and pathogenic species (Strobel et al. 2004). They have been found to produce phytotoxic compounds as well as antimicrobial agents involved in the biocontrol of plant pathogens (Haas and Défago 2005). For example, *P. viridiflava*, associated with the leaves of grass species, produces ecomycins—a family of lipopeptides active against *Cryptococcus neoformans* and *Candida albicans* (Miller et al. 1998). *Bacillus* species produce mainly peptide antibiotics such as lantibiotics which contain the uncommon thioether amino acids lanthionine and ß-methyllanthionine (Stein 2005). Various lantibiotics have been isolated from *Bacillus* species including subtilin from *B. subtilis*, haloduracin from *B. halodurans*, lichenicidin from *B.*

licheniformis, and cerecidins from *B. cereus* (Lawton et al. 2007; Dischinger et al. 2009; Wang et al. 2014). Since the 1970s, *Escherichia coli* have been widely used as a model for the development of DNA recombinant technology (Katz and Baltz 2016). Nowadays, it is the most employed host for the expression of drug candidates and small molecules (Atanasov et al. 2015). Thus, metabolites from *E. coli* have been widely studied and this may explain the high number of NPs from this species in the DNP (Reed and Palsson 2003).

NPs from the Fungi kingdom

In the DNP, approximately 20,000 compounds were reported to belong to the Fungi kingdom, and two divisions (Ascomycota and Basidiomycota) account for almost 90% of the total fungal products. The Fungi kingdom is thought to contain the largest number of microbial metabolites.

Compared to the other kingdoms, fungal products are structurally less complex and have smaller molecular weights (Bérdy 2005). A wide range of chemical classes are represented among fungi, including terpenoids (e.g., paclitaxel from *Taxomyces*), alkaloids (e.g., ergot alkaloids from *Claviceps*), simple aromatic NPs (e.g., griseofulvin from *Penicillium*), amino acids and peptides (e.g., echinocandins from *Aspergillus* and β-lactams from *Penicillium*), benzofuranoids (e.g., mycophenolic acid from *Penicillium*), carbohydrate NPs (e.g., ß-glucans from basidiomycetes), polycyclic aromatic NPs (e.g., parietin from lichens), and polyketides (e.g., aflatoxins and lovastatin from *Aspergillus*) (Fig. 4) (Goyal et al. 2016). This plethora of diverse fungal metabolites is associated with various biological activities ranging from pharmacological agents to mycotoxins (Peláez 2005; Goyal et al. 2016).

With more than 13,000 compounds, the Ascomycota division is the largest producer of NPs in the Fungi kingdom (Suppl. Mat. S1C). This group is also the largest phylum of Fungi (64,000 known species) and one of the most ubiquitous phyla of eukaryotes (Schoch et al. 2009). It comprises filamentous fungi (e.g., *Aspergillus* and *Penicillium*) but also yeasts (e.g., *Saccharomyces cerevisiae*), endophytic species (e.g., *Alternaria*, *Phoma*, *Stachybotrys* and *Trichoderma*), and lichen-forming fungi representing 40% of all ascomycetes (e.g., *Cladonia*, *Lecanora* and *Parmelia*) (Bérdy 2005; Blackwell 2011).

One of the milestones in the history of NPs from ascomycetes is the discovery of penicillin from the fungus *Penicillium chrysogenum* in 1928 by Alexander Fleming. This discovery led to the production and commercialization of synthetic penicillins in the early 1940s which saved a huge number of lives during WWII and ushered in the "Golden Age of Antibiotics" from the 1940s to the 1970s (Cragg et al. 2014; Bernardini et al. 2017). Besides the important discovery of β-lactams from *Penicillium* species, other NPs with pharmaceutical activities have also been isolated from this genus including mycophenolic acid, an immunosuppressant agent produced by *P. brevicompactum*; griseofulvin, an antifungal agent isolated from *P. griseofulvum*, and compactins known for their cholesterol-lowering activities produced by *P. brevicompactum* and *P. citrinum* (Frisvad et al. 2004; Chakravarti and Sahai 2004).

While the *Penicillium* genus is the second-largest producer of NPs in the Ascomycota division, *Aspergillus* is the most represented genus with *A. terreus* being the most prolific species from this genus in the DNP. Indeed, *Aspergillus* is renowned for its medical, pathogenic, and industrial importance (Sanchez et al. 2012). As an example, *A. terreus* is a significant cause of aspergillosis, but it is also the main source of lovastatin, the first commercially marketed statin which was approved by the FDA in 1987 for the treatment of hypercholesterolemia (Demain and Sanchez 2009). Nowadays, synthetic derivatives of lovastatin are among the world's most sold drugs (Katz and Baltz 2016). Another example is that of *A. nidulans*, a model organism widely used to study genetics and cell biology, but it is also the main producer of echinocandins, a family of lipopeptides used in the treatment of candidiasis (Denning 2003). The first licensed semisynthetic echinocandin derivatives were caspofungin, micafungin and anidulafungin since 2001 (Butler 2008). Of all *Aspergillus* species, *A. fumigatus* is the primary causative agent of human infection, and gliotoxin is the main mycotoxin involved in mycosis (Dagenais and Keller 2009). *A. fumigatus* also produces fumagillin, first isolated in 1949 and used in the treatment of microsporidiosis (Mishra and Tiwari 2011). Finally, *Aspergillus* species (especially *A. flavus* and *A. parasiticus*) are also a unique source of aflatoxins, a major class of

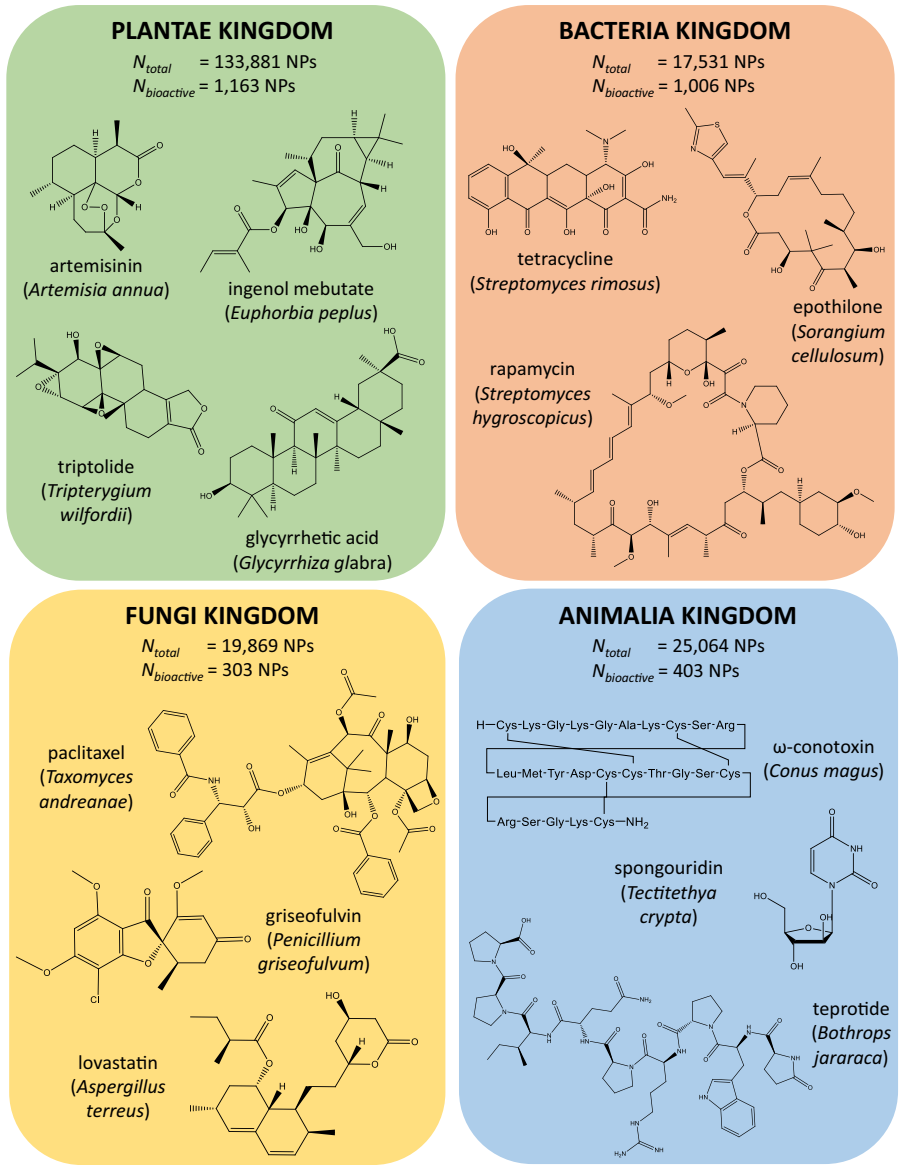


Fig. 4 Structures of biologically active compounds of plant, bacterial, fungal and animal origin

mycotoxins that have been described as human carcinogens and implicated in hepatocellular carcinoma (Henry et al. 2002; Mishra and Das 2003).

Endophytic fungi are also well represented in the DNP with *Fusarium* (i.e., *F. oxysporum* and *F. solani*), *Trichoderma* (i.e., *T. harzianum* and *T. viride*), *Alternaria* (i.e., *A. alternata*), and *Chaetomium* genera (i.e., *C. globosum*) ranking 3rd, 4th, 5th and 6th respectively in terms of NPs per genus. Endophytic fungi are defined by their occurrence within internal tissues of plants without causing any immediate overtly negative effects (Stone et al. 2000). They represent a polyphyletic group of ascomycetous fungi and are found in various plants including liverworts, mosses, ferns, and seed plants (Aly et al. 2010). Interest in endophytic strains began with the detection of paclitaxel (an anti-tumour agent originally isolated from the plant species *Taxus brevifolia*) from the endophytic fungus *Taxomyces andreanae* (Stierle et al. 1993). This discovery highlighted the potential of endophytic fungi as alternative sources of plant secondary metabolites, and led to a shift in the discovery of new compounds from fungal sources in the early 1990s (Bérdy 2005). In the last decade, about

half of the newly discovered fungal metabolites (approximately 5000 compounds) were from endophytic fungi (Aly et al. 2010; Bérdy 2012). This includes various compounds belonging to different chemical classes such as aldehydes (e.g., chaetopyranin from *Chaetomium globosum*), alkaloids (e.g., camptothecin from *Fusarium solani*), lignans (e.g., podophyllotoxin from *Fusarium oxysporum*), peptides (e.g., beauvericin from *F. oxysporum*), polyketides (e.g., alternariol from *Alternaria* species), steroids (e.g., wortmannins from *Talaromyces wortmannii*), and terpenoids (e.g., pestalotiopsins from *Pestalotiopsis* spp.) (Tan and Zou 2001; Kharwar et al. 2011; Gao et al. 2018). Moreover, in a literature review of 135 metabolites isolated from endophytic fungi, it was shown that the proportion of novel chemical structures produced by endophytes is significantly higher (51%) than that produced by fungi isolated from soil (38%), thus confirming that endophytes are a good source of bioactive metabolites (Schulz et al. 2002).

In addition to ascomycetes, basidiomycetes are also of great importance since amongst the fungi kingdom they are the second-largest producer of NPs in the DNP (4820 NP). This group is also the second-largest phylum of fungi (22,000 known species), and it includes most of the macroscopic fungi (Bills et al. 2005). One of the main features of basidiomycetes is the presence of compounds with high molecular weight such as polysaccharides, proteins, and lipids, as well as low molecular weight metabolites including alkaloids, peptides, polyketides, steroids, and terpenoids (De Silva et al. 2013). Polysaccharides have been widely studied, and in particular the study of ß-glucans has led to the development of anticancer drugs used in Asia such as lentinan from *Lentinula edodes*, schizophyllan from *Schizophyllum commune* and two specific proteoglycans [polysaccharide krestin (PSK) and polysaccharopeptide (PSP)] from *Trametes versicolor* (Lindequist et al. 2005). Secondary metabolites produced by basidiomycetes have also been investigated and have proved to be a source of bioactive compounds with, e.g., strobilurins isolated from *Strobilurus tenacellus* used as fungicides; pleuromutilins produced by *Pleurotus mutilus* and its semisynthetic derivative (retapamulin) used in the topical treatment of impetigo; and illudins isolated from *Omphalotus* spp. and their analogue (irofulven) currently under investigation as anticancer drugs (Stadler and Hoffmeister 2015).

Ganoderma species account for almost 10% of the total compounds isolated from basidiomycetes, and in the DNP half of these compounds are from *G. lucidum*. Also known as "Lingzhi", this species has been widely used in traditional Chinese medicine to promote good health and to treat many diseases including allergy, arthritis, cancer, hypertension, and inflammation (Paterson 2006). Polysaccharides (e.g., β-glucans) and triterpenoids (e.g., ganoderic acids) are the two major groups of compounds that exhibit pharmacological effects, especially anticancer and immunomodulatory activities (Yuen and Gohel 2005). Extracts of *G. lucidum*, and the medicinal peptide, Ling Zhi-8, are currently under investigation for use as chemopreventive and adjuvant agents in the treatment of cancer (Cheng and Sliva 2015; Chen et al. 2017a).

NPs from the Animalia kingdom

Of all the kingdoms Animalia ranks as the second-largest producer of NPs in the DNP. More than 25,000 NPs have been identified from animal sources, and approximately half of these compounds belong to the terpenoid and alkaloid groups.

In the DNP, one quarter of the bioactive NPs from animal sources were isolated from their venoms and toxins, among which amino acid and peptide NPs are highly represented (97%) (Fig. 5). The best-known example of a successful venom-based drug is that of the angiotensin-converting enzyme (ACE) inhibitors which act as antihypertensive agents. A venom peptide (teprotide) was isolated from the snake *Bothrops jararaca* in the 1960s and led to the development of ACE inhibitors (i.e., captopril, enalapril and lisinopril) which are currently among the top best-selling drugs in the world (Lewis and Garcia 2003; Antunes et al. 2016). Other venomous animals have also been investigated for their pharmaceutical potential such as the Gila monster *Heloderma suspectum* which afforded exendin-4, an antidiabetic agent approved by the FDA in 2005; the medicinal leech *Hirudo medicinalis* which led to the development of bivalirudin, a synthetic analog of hirudin with anticoagulant activity approved by the FDA in 2000; and the cone snail *Conus magus* which produces ω-conotoxins, the synthetic form ziconotide was approved in 2004 for the treatment of refractory neuropathic pain (King 2011; Zambelli et al. 2016). Interestingly, among all of

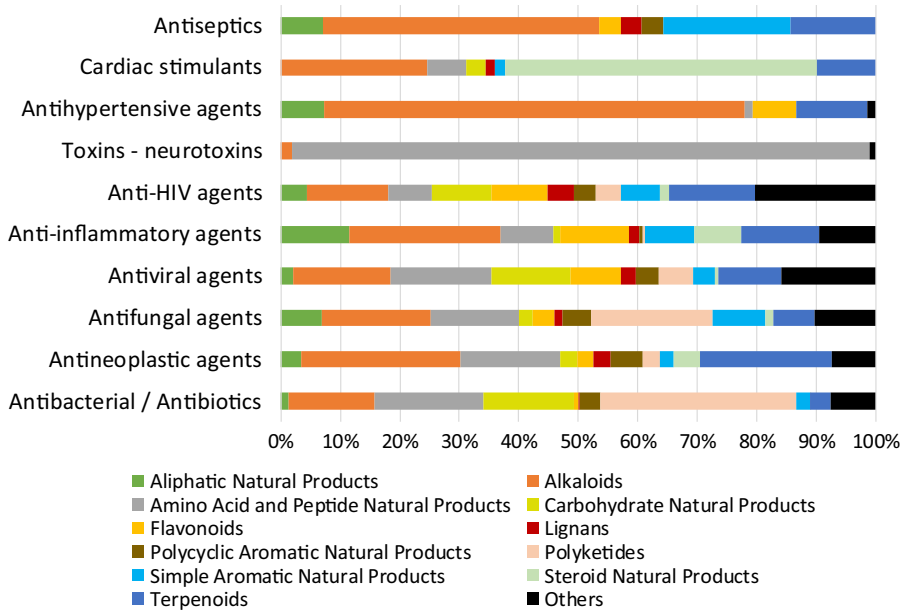


Fig. 5 Distribution of the main chemical classes of natural products in the most important biological activities

the kingdoms *Conus* is one of the genera that produces the largest number of bioactive compounds, since more than 100 bioactive NPs (accounting for more than 70% of the total NPs from this genus) have been recorded in the DNP. This high number of diverse pharmacologically active compounds results from the high rate of hypermutations as well as the remarkable number of post-translational modifications in this genus which allow little overlap in conopeptides between *Conus* species (Buczek et al. 2005). To date, it is estimated that the number of conopeptides per species might exceed 1000, and that a total of 35,000 compounds could be available from the 700 *Conus* species identified (Davis et al. 2009).

An interesting feature of animal-derived products is the high proportion of compounds originating from marine sources, with more than half of the animal-derived NPs in the DNP belonging to sponges (i.e., Ceractinomorpha, Tetractinomorpha, and Homoscleromorpha) and cnidarians (i.e., Octocorallia and Hexacorallia) (Suppl. Mat. S1D). It is noteworthy that species from the Animalia kingdom are highly represented in aquatic environments, with 32 of the 33 animal phyla being marine (Cragg and Newman 2013). Historically, the first discovery of biologically active compounds from a marine environment was that of the nucleosides spongouridine and spongothymidine, isolated from the Caribbean sponge, *Tectitethya crypta* in the early 1950s (Dias et al. 2012). Two decades later, the systematic investigation of bioactive products from marine sources began thanks to the development of scuba diving techniques as well as the use of submersibles (Cragg and Newman 2013).

With 8656 compounds in the DNP, the Porifera phylum (sponges) is by far the most represented amongst the Animalia kingdom. This phylum, comprising about 8500 species, has also been reported to be the most prolific marine producer of natural compounds (approximately 30% of all NPs from marine sources) with an average of 200 compounds described each year (Van Soest et al. 2012; Blunt et al. 2015, 2017). In the DNP, *Dysidea* is the genus with the highest number of NPs from the Porifera phylum, while *Theonella swinhoei* (also from the Porifera phylum) is the largest producer of NPs from the Animalia kingdom. In a review of marine sponge-derived NPs, Mehbub et al. (2014) suggested that the high number of NPs found in key orders of the sponges including Dictyoceratida (including *Dysidea* spp.), Haplosclerida, Halichondrida, Poecilosclerida, and Astrophorida was the result of high species diversity in these orders. Another reason could be the diversity and host specificity of microbial symbionts found in these species, which are responsible for the synthesis of some of the metabolites isolated from these sponges (Thomas et al. 2010). This is the case for *Dysidea* spp. which host a distinct cyanobacterial clade that may be responsible for the variable patterns in secondary

metabolites found in this genus (Thacker and Starnes 2003). More generally, terpenoids, alkaloids, as well as peptides represent the main chemical classes of compounds found in the Porifera phylum, and more than half of the biological activities investigated were anticancer properties (Mehbub et al. 2014). Three compounds derived from marine sponges have been approved by the FDA including the anti-tumour compound eribulin mesylate, which is an analogue of halichondrin B first isolated from *Halichondria okadai* (Agrawal et al. 2016).

According to the DNP, Cnidarians are the second-largest producer of NPs from animal sources (5249 NPs). This group of relatively simple animals comprises over 11,000 species, and includes reef-forming corals, sea anemones, soft corals, jellyfishes, and marine hydroids (Daly et al. 2007; Appeltans et al. 2012). The Cnidarian phylum contains the *Sinularia* genus which, according to the DNP, is by far the biggest producer of NPs from the Animalia kingdom, while *Clavularia viridis* is the most prolific Cnidarian species. Both taxa belong to the Alcyonacea order (soft corals) which have been described as the main source of NPs from Cnidarians (Leal et al. 2012). With more than 160 known species, *Sinularia* are one of the richest genera from the Alcyonacea order (Roskov et al. 2018). They produce a wide range of secondary metabolites including sesquiterpenes, polyhydroxylated steroids, polyamine compounds, as well as cembranoid diterpenes which are the most frequently isolated NPs from *Sinularia* (Chen et al. 2012). Some of these compounds display various bioactivities such as antiulcer properties (e.g., sinulide), anti-inflammatory activities (e.g., gibberoketosterol) and anti-tumour activities (e.g., flexilarin D) (Rocha et al. 2011). Altogether, most of the bioactive compounds found in Cnidarians belong to the terpenoids (i.e., monoterpenoids, diterpenoids, and sesquiterpenoids). Most of the interest in NPs isolated from Cnidarians surrounds their potential anti-tumour and anti-inflammatory activities (Rocha et al. 2015). To date, none of the compounds isolated from Cnidarians have been developed as drugs, however some compounds have been exhaustively investigated in preclinical studies, such as the diterpene glycoside pseudopterosin extracted from the sea whip *Pseudopterogorgia elisabethae*; eleutherobin derived from the soft coral *Eleutherobia* sp.; and the diterpene sarcodictyn, found in some corals (Mariottini and Grice 2016). In addition, more than 70 compounds extracted from Cnidarians have been reported to possess promising bioactivities and thus might be of interest for future clinical studies (Rocha et al. 2011).

Targeting the most prolific biological organisms: popular taxa versus unexplored phyla

Targeting "popular" taxonomic groups seems to be a common way to uncover new natural compounds. In the DNP, a large number of NPs originate from species-rich biological groups. As an example, in the Plantae kingdom the highest number of natural compounds has been isolated from the Compositae and Leguminosae families, and they are also amongst the three largest botanical families in the world. In the Animalia kingdom, a high proportion of compounds originate from marine sources, but 32 of the 33 phyla in the Animalia kingdom include species from aquatic environments. Thus, the accessibility of biological sources (i.e., wide abundance and distribution) probably plays an important role in the selection of candidate species for further investigation. This is especially true for some marine taxa, for which the abundance, size, and colour have been reported to influence the selection process (Blunt et al. 2008). Indeed, "popular" taxa provide a wide range of NPs with high chemical diversity, thus suggesting the potential of these taxa as sources for drug discovery (Leal et al. 2012). The popularity of some organisms might also be related to their high-yield of bioactive compounds. For instance, *Streptomyces* spp., the most investigated genus in our analysis, exhibited the highest yield of bioactive compounds. Although it has been widely investigated, *Streptomyces* spp. continue to deliver novel scaffolds, such as the antibiotic platensimycin launched in 2006, as well as many other compounds currently under investigation in clinical trials (de Lima Procópio et al. 2012; Genilloud 2017). Furthermore, an increasing number of compounds thought to originate from plants or animals can now be reattributed to microbial organisms (e.g., endophytes, symbionts). Therefore, the high number of compounds found in one particular taxonomic group could in fact be produced by microbial sources. As an example, 31 of the 32 polyketide metabolites isolated from the sponge *T. swinhoei*, the most represented species from the

Animalia kingdom in our study, have been reattributed to an uncultured bacterium (Wilson et al. 2014).

While some researchers will prefer to focus on "popular" taxonomic groups exhibiting high chemical diversity, others will select unexplored phyla or geographical regions (Leal et al. 2012). Estimates for the total number of non-microbial species living on Earth range from 2 to 100 million (May 2010). However, some authors suggested that this number might be narrower and range between 5 and 10 million (Mora et al. 2011; Costello et al. 2013). If we compare the number of species catalogued with those predicted, the most interesting groups with a large number of species still to be discovered are likely to be fungi ($\approx$ 100,000 species catalogued vs. 0.8–5 million predicted) and animals (1–2 million vs. 5–10 million) (Scheffers et al. 2012; Pimm et al. 2014). In contrast, more than two thirds of the predicted number of plant species ($\approx$ 450,000) are already known (Pimm and Joppa 2015). Regarding bacteria, the total number of species is quite difficult to evaluate and estimates range between 10^3 and 10^{12} species (Pedrós-Alió 2006; Locey and Lennon 2016). In addition, some authors argued that marine environments and biodiversity hotspots could be a great source of new taxa (Mora et al. 2011; Scheffers et al. 2012). Species from specific taxonomic groups and geographical regions could be a source of NPs with potential bioactivity. In a recent paper, Pye et al. (2017) demonstrated that identifying new biological organisms leads to a burst in the discovery of structurally novel compounds, while investigating already studied classes of organisms is likely to yield existing classes of compounds. Therefore, further studies should be performed to identify biological organisms with great potential. This could be done by mapping the distribution area of un-investigated (or less investigated) phyla, based on a datamining-guided search of an exhaustive as possible NP database.

NP research in the pharmaceutical industry

Historically, NPs have formed the basis of the therapeutic arsenal developed by pharmaceutical companies, and today they still represent a huge reservoir of inspirational bioactive chemodiversity. Out of the 293,798 NPs compiled in the DNP database, only 3882 ($\approx$ 1%) were reported to have biological activities, viz. compounds with established activity and being used as drugs or under investigation for drug use. This raises concerns about the future potential of NPs as drug candidates.

Historical background

Plants and NPs have historically been valuable sources of effective medicine. In the beginning of the nineteenth century, pure NPs started to be isolated from medicinal plants and administered as pure and unique active principles at precise dosages. This was the beginning of the single and "magic bullet" paradigm (Strebhardt and Ullrich 2008). In the meantime, herbalism with its mixture of multi-active NPs lost credit. Natural products then reached their moment of glory when hundreds of pharmaceutical companies were trying to associate clinical activity with the presence of a particular NP (Fig. 6, phase 1800–1980). Inspired by the seminal huge screening campaigns initiated in 1938 by Jonathan Hartwell and the Cancer program at the National Cancer Institute in 1955, many companies ran large-scale drug discovery screening programs. The pharmaceutical industry turned towards the thousands of NPs found in genetic resources for their high throughput screening (HTS) drug discovery programs, made possible with the advent of miniaturization and robotics. During this phase (1980–2005), drug discovery strategies made a 180° turn, starting with target molecules (e.g., enzymes, receptors, and ion channels) and moving up to cells, groups of cells, tissues, animals, and finally human patients. High throughput screening and molecular assays were performed in more and more dense microtiter plates (96, 384 or 1536 wells) at the farthest distance from the patient to conduct more and more experiments (up to 100,000 wells tested per day). Despite huge throughputs, and the advantage of screening very large numbers of compounds at a time, high attrition rates were observed when moving from hits active on a biological molecule towards the complexity of a human being. A HTS hit can easily access the target molecule in solution, but this is more difficult with cells, tissues, and complex biological organisms like laboratory animals (David and Ausseil 2014). Besides, by the end of the 1990s new techniques such as combinatorial chemistry, virtual screening, and even the deciphering of the human genome monopolized financial investments diverting

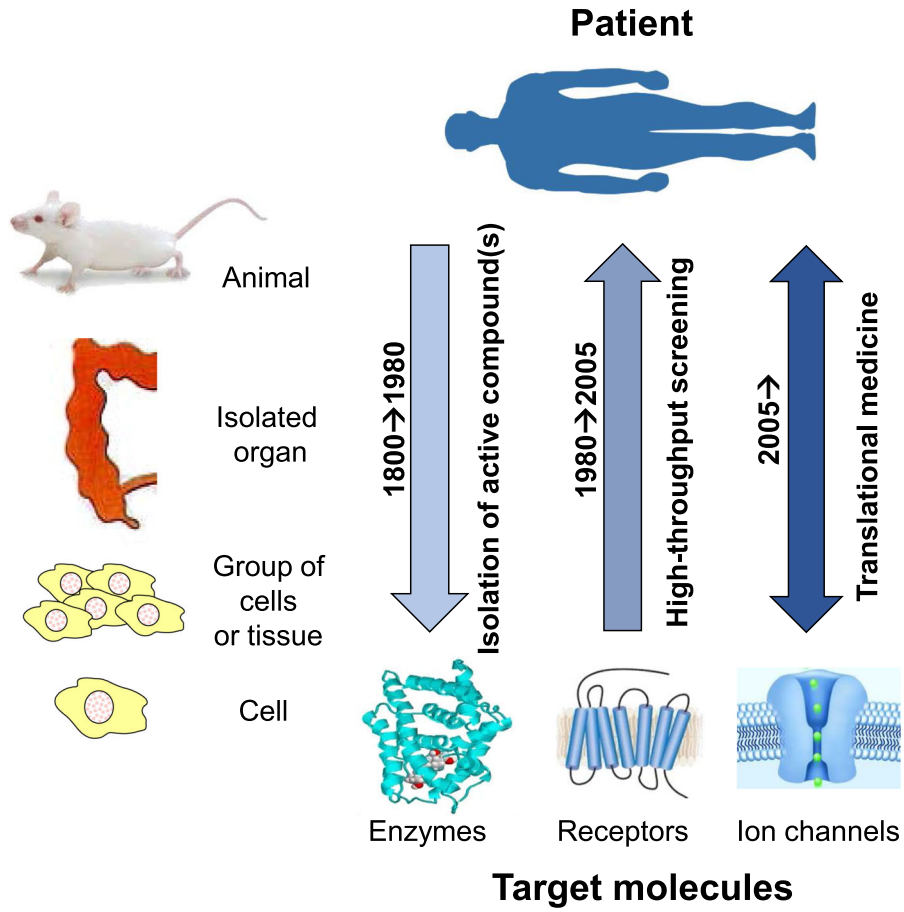


Fig. 6 Evolution of the research strategies utilized by pharma companies

funds away from NP screening programs. The Human Genome Project was supposed to generate thousands of new druggable proteins as pharmacological targets. But in fact, the number of targets rose from 300 to around 600 (Rask-Andersen et al. 2014).

Current state of NP research in Pharma companies

Since the beginning of the twenty-first century, research into NPs has largely been downsized. The probability of a hit from a HTS screening program reaching the market is around one in a million and the cost from hit to approval about $ 3 billion (Oprea 2000; DiMasi et al. 2016).

Paradigms need to change as a single approach is no longer possible. In drug discovery hits at a molecular level need to be validated up to the patient level. Drug discovery needs to connect the patients' scale to the scale of molecules by translational medicine. Moreover, computational and combinatorial chemistry techniques are now being combined and applied to drug discovery.

Nowadays studies on traditional knowledge are restricted to academic research mostly on tropical diseases. But in the meantime, the economic and social demand for "green" food supplements, botanical drugs, and NP-based medicine is expanding.

Main reasons for the decline in NP research

It is clear that for decades (or even centuries) traditional medicines have been raided by pharmaceutical companies and turned into active molecules. Therefore, most of the easiest "low-hanging fruits"

have been picked (David and Ausseil 2014). An analysis of new structures described in the commercial database *AntiMarin* (measured by the Tanimoto index) reveals that newly reported NPs are becoming more and more similar to structures already described (Pye et al. 2017).

High throughput screening and even phenotypic high content screening on small animals have not been able to generate a return on investment with the introduction of new chemical entities on the market. In this strategy, researchers wait for an improbable "alignment of planets" between the druggability of NPs and the specific biological target screened. The number of potential targets is not boundless, even after the deciphering of the human genome (around 600 druggable targets). Even huge phenotypic screening campaigns on small animals were not able to optimize this passive "lottery" approach. This strategy was based on the premise that a large library of NPs should contain the perfect gem dedicated to the pharmacological target being studied.

The complexity and inherent slowness of working with NPs was one of the reasons for the downsizing or shutdown of NP drug discovery programs. The supply, re-supply and authentication of genetic resources, as well as the isolation and identification of active molecules through bio-guided fractionation is not an easy game (Atanasov et al. 2015). The implementation of tedious access laws for genetic resources also had a negative effect on bioprospection and NP research, along with paradoxical effects on biodiversity conservation (David 2018).

Trends for the future

For the last two decades, research efforts into NPs have gone phases of decline and renaissance. Nevertheless, NPs still represent a huge source of potential drug leads, but research strategies need to evolve with the times and new paradigms need to be invented.

Since it was discontinued in the beginning of the 2000s, bioprospection will remain marginal in Big Pharma. Only Novartis and Pierre Fabre are still working on higher plants for drug discovery. The few companies involved in NP research are rather focusing on microorganisms or marine organisms which offer more "low-hanging fruits" in underexplored biotopes. More than 90% of microorganisms are not cultivable, but heterologous models make it possible to produce NPs from these organisms should allow the production of novel or "cryptic" NPs (Palazzolo et al. 2017).

For scientific and administrative reasons, for two centuries pharmaceutical companies have focused their research efforts on the discovery of pure active compounds. This is the "magic bullet" paradigm. But plants act in more subtle ways. However, it should be admitted that the "single bullet" approach is not always the best approach, and indeed multitherapy is favoured in the treatment of AIDS and other infective diseases. Thus, in some chronic or metabolic diseases the single high affinity bullet should be replaced by a swarm of less active bullets as in phytotherapy and traditional African, Indian or Chinese medicines. The synergetic combination approach is gaining more and more interest. It is likely that paradigms will evolve away from single molecules to well-defined extracts simultaneously addressing multiple pharmacological targets, as is the proposed mechanism of action for botanical drugs.

Novel scientific methods for the discovery, validation, characterization, and standardization of these multicomponent botanical drugs will surely become more and more recognised by Health agencies. Products from vegetal origin such as Acheflan®, Angipars®, Epogam®, Fulyzaq®, Iberogast®, Picato®, Rosaderm® and Veregen® are paving the way towards the new paradigm.

Meanwhile, the economic and social need for food supplements, phytotherapy, and botanical drugs is growing. The cultivation of medicinal plants for the production of standardized enhanced traditional medicines is recommended by the World Health Organization and was advocated in 2008 with the Beijing declaration (World Health Organization 2008). This makes sense in terms of cost, availability, and environmental impact. Pharmacological screening is generally conducted with a huge number of compounds on one specific biological target. But the opposite is possible with selectivity profiling, *id est* the screening of one NP on a lot of targets (Pouny et al. 2014). Re-purposing old NPs which benefit from decades of pharmacovigilance is also a fruitful approach (Cragg et al. 2014).

Fragment-based drug discovery looks very promising. This ground-breaking technique uses a restricted library of approximately 500 small natural molecules (150–250 Da called "fragments"). Optimized libraries of natural fragments cover a larger chemical

space than synthetic fragments and offer straightforward growing possibilities to cover a huge chemodiversity space (Pascolutti et al. 2015).

Translational medicine, neural network analyses, and the generation of large data sets have started to allow molecular interactome studies, and links between biological systems to disease mechanisms are very encouraging (Capriotti et al. 2018). New trends in NP research include microbial genomics, synthetic biology, control of NP biosynthesis and bioinformatics. Computational approaches that can analyse large amounts of data are rapidly developing in every field of NP research (Sarker and Nahar 2018).

Limitations

This study has several limitations which are mainly linked to the composition of the DNP. Indeed, numerous biases related to the collection, isolation, and identification of biota and their chemical constituents can be found in this database. First of all, taxonomic data stems usually from the first report, hence sources can sometimes be wrongly assigned. As an example, the genus *Lyngbya* has been completely revised during the last five years. However, most of the species cited in this work are well known and readily identifiable (e.g., *Aspergillus terreus*, *Azadirachta indica*, *Helianthus annuus*, *Nicotiana tabacum*, and *S. cerevisiae*) and so we expect to have very few incidences of misinformation. Moreover, for most of the analyses we have intentionally limited our study to the family level or higher taxonomic rank. Another limitation is that some biological sources cited in this work (e.g., *Bacillus subtilis, E. coli*, and *Streptomyces* spp.) are used as hosts for the production of drugs and small molecules. Therefore, compounds reported in our study as being produced by these species may not originate from these species. Given the restricted number of species used as hosts and the recent development of this technique for some species (i.e., *Streptomyces* spp.), this limitation does not significantly affect our results. In addition, not all compounds reported in the DNP to have biological activities (3882 NPs in total) have been used as drugs. Therefore, for some compounds, preclinical and clinical data may be limited, and their biological activities need to be confirmed. Finally, this dictionary is not exhaustive, and our study cannot be taken as a descriptive study of all known NP.

Despite these limitations, our results correlate with the findings of Bérdy (2005) who aimed to provide a summary of bioactive compounds from microbial sources, as well as those from Blunt et al. (2015), Mehbub et al. (2014) and Hu et al. (2011) who reviewed the NPs derived from marine sources.

Conclusion

This review aimed to characterize the global diversity of compounds isolated from natural sources. We based our analysis on the DNP which is one of the most comprehensive libraries of NPs, and one of the few to link chemical entities to their natural sources.

Our study of the diversity of natural compounds shows that a high proportion of NPs are from plant origin, with terpenoids being the most represented chemical class (except in the Bacteria kingdom), and antibacterial as well as antineoplastic activities are the two main biological activities reported. Besides the importance of plant-derived compounds, we also highlight the large number of NP isolated from *Streptomyces* spp. as well as from the Ascomycota phylum and marine sources (i.e., sponges and soft corals).

Although a large number of compounds have been isolated from natural sources, few are used as medicines today. More integrative and comprehensive approaches should unravel the limitations in drug discovery from NPs outlined in this review. Drug discovery from NPs is still in its infancy, and NPs remain relevant in modern drug discovery.

Acknowledgements We would like to thank the French National Research Institute for Sustainable Development (IRD) documentation service for subscribing to the DNP. We would also like to thank Fiona Macdonald, Taylor and Francis group, for the clarifications she provided concerning the information given in the DNP.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Adams JD, Wang R, Yang J, Lien EJ (2006) Preclinical and clinical examinations of *Salvia miltiorrhiza* and its tanshinones in ischemic conditions. Chin Med 1:3. https://doi.org/10.1186/1749-8546-1-3

Agrawal S, Adholeya A, Deshmukh SK (2016) The pharmacological potential of non-ribosomal peptides from marine sponge and tunicates. Front Pharmacol. https://doi.org/10.3389/fphar.2016.00333

Aly AH, Debbab A, Kjer J, Proksch P (2010) Fungal endophytes from higher plants: a prolific source of phytochemicals and other bioactive natural products. Fungal Divers 41:1–16. https://doi.org/10.1007/s13225-010-0034-4

Antunes AMS, Guerrante RDS, Ávila JPC et al (2016) Case study of patents related to captopril, Squibb's first blockbuster. Expert Opin Ther Pat 26:1449–1457. https://doi.org/10.1080/13543776.2016.1227321

Appeltans W, Ahyong ST, Anderson G et al (2012) The magnitude of global marine species diversity. Curr Biol 22:2189–2202. https://doi.org/10.1016/j.cub.2012.09.036

Atanasov AG, Waltenberger B, Pferschy-Wenzig E-M et al (2015) Discovery and resupply of pharmacologically active plant-derived natural products: a review. Biotechnol Adv 33:1582–1614. https://doi.org/10.1016/j.biotechadv.2015.08.001

Baker D, Chu M, Oza U, Rajgarhia V (2007) The value of natural products to future pharmaceutical discovery. Nat Prod Rep 24:1225–1244. https://doi.org/10.1039/B602241N

Banerjee P, Erehman J, Gohlke B-O et al (2015) Super Natural II—a database of natural products. Nucleic Acids Res 43:D935–D939. https://doi.org/10.1093/nar/gku886

Bérdy J (2005) Bioactive microbial metabolites. J Antibiot 58:1–26. https://doi.org/10.1038/ja.2005.1

Bérdy J (2012) Thoughts and facts about antibiotics: where we are now and where we are heading. J Antibiot 65:385–395. https://doi.org/10.1038/ja.2012.27

Bernardini S, Tiezzi A, Masci VL, Ovidi E (2017) Natural products for human health: an historical overview of the drug discovery approaches. Nat Prod Res. https://doi.org/10.1080/14786419.2017.1356838

Bills G, Spatafora JW, Blackwell M (2005) Phylogeny of the fungal kingdom and fungal-like eukaryotes. In: An Z (ed) Handbook of industrial mycology. Marcel Dekker, New York, pp 27–47

Blackwell M (2011) The Fungi: 1, 2, 3 … 5.1 million species? Am J Bot 98:426–438. https://doi.org/10.3732/ajb.1000298

Blunt JW, Copp BR, Hu WP et al (2008) Marine natural products. Nat Prod Rep 25:35–94. https://doi.org/10.1039/b701534h

Blunt J, Munro M, Upjohn M (2012) The role of databases in marine natural products research. In: Fattorusso E, Gerwick WH, Taglialatela-Scafati O (eds) Handbook of marine natural products. Springer, Dordrecht, pp 389–421

Blunt JW, Copp BR, Keyzers RA et al (2015) Marine natural products. Nat Prod Rep 32:116–211. https://doi.org/10.1039/c4np00144c

Blunt JW, Copp BR, Keyzers RA et al (2017) Marine natural products. Nat Prod Rep 34:235–294. https://doi.org/10.1039/c6np00124f

Brinker AM, Ma J, Lipsky PE, Raskin I (2007) Medicinal chemistry and pharmacology of genus *Tripterygium* (Celastraceae). Phytochemistry 68:732–766. https://doi.org/10.1016/j.phytochem.2006.11.029

Buczek O, Bulaj G, Olivera BM (2005) Conotoxins and the posttranslational modification of secreted gene products. Cell Mol Life Sci 62:3067–3079. https://doi.org/10.1007/s00018-005-5283-0

Butler MS (2008) Natural products to drugs: natural product-derived compounds in clinical trials. Nat Prod Rep 25:475–516. https://doi.org/10.1039/B514294F

Cantisani C, Gado FD, Ulrich M et al (2013) Actinic keratosis: review of the literature and new patents. Recent Pat Inflamm Allergy Drug Discov 7:168–175

Capriotti E, Ozturk K, Carter H (2018) Integrating molecular networks with genetic variant interpretation for precision medicine. Wiley Interdiscip Rev Syst Biol Med. https://doi.org/10.1002/wsbm.1443

Chakravarti R, Sahai V (2004) Compactin—a review. Appl Microbiol Biotechnol 64:618–624. https://doi.org/10.1007/s00253-003-1553-7

Chang K-W, Tsai T-Y, Chen K-C et al (2011) iSMART: an integrated cloud computing web server for traditional Chinese medicine for online virtual screening, de novo evolution and drug design. J Biomol Struct Dyn 29:243–250. https://doi.org/10.1080/073911011010524988

Chao J, Dai Y, Verpoorte R et al (2017) Major achievements of evidence-based traditional Chinese medicine in treating major diseases. Biochem Pharmacol 139:94–104. https://doi.org/10.1016/j.bcp.2017.06.123

Chen BJ (2001) Triptolide, a novel immunosuppressive and anti-inflammatory agent purified from a Chinese Herb *Tripterygium wilfordii* Hook F. Leuk Lymphoma 42:253–265. https://doi.org/10.3109/10428190109064582

Chen W, Li Y, Guo Y (2012) Terpenoids of *Sinularia* soft corals: chemistry and bioactivity. Acta Pharm Sin B 2:227–237. https://doi.org/10.1016/j.apsb.2012.04.004

Chen S-J, Lin H-H, Huang W-C et al (2017a) Ling-Zhi-8 protein (LZ-8) suppresses the production of pro-inflammatory mediators in murine microglial BV-2 cells. Food Agric Immunol 28:1393–1407. https://doi.org/10.1080/09540105.2017.1346062

Chen Y, de Bruyn Kops C, Kirchmair J (2017b) Data resources for the computer-guided discovery of bioactive natural products. J Chem Inf Model 57:2099–2111. https://doi.org/10.1021/acs.jcim.7b00341

Chen S-R, Dai Y, Zhao J et al (2018) A mechanistic overview of triptolide and celastrol, natural products from *Tripterygium wilfordii* Hook F. Front Pharmacol. https://doi.org/10.3389/fphar.2018.00104

Cheng S, Sliva D (2015) *Ganoderma lucidum* for cancer treatment: we are close but still not there. Integr Cancer Ther 14:249–257. https://doi.org/10.1177/1534735414568721

Clardy J, Walsh C (2004) Lessons from natural molecules. Nature 432:829–837. https://doi.org/10.1038/nature03194

Clardy J, Fischbach MA, Walsh CT (2006) New antibiotics from bacterial natural products. Nat Biotechnol 24:1541–1550. https://doi.org/10.1038/nbt1266

Costello MJ, May RM, Stork NE (2013) Can we name Earth's species before they go extinct? Science 339:413–416. https://doi.org/10.1126/science.1230318

Cragg GM, Newman DJ (2013) Natural products: a continuing source of novel drug leads. Biochim Biophys Acta Gen Subj 1830:3670–3695. https://doi.org/10.1016/j.bbagen.2013.02.008

Cragg GM, Grothaus PG, Newman DJ (2014) New horizons for old drugs and drug leads. J Nat Prod 77:703–723. https://doi.org/10.1021/np5000796

Dagenais TRT, Keller NP (2009) Pathogenesis of *Aspergillus fumigatus* in invasive aspergillosis. Clin Microbiol Rev 22:447–465. https://doi.org/10.1128/CMR.00055-08

Daly M, Brugler MR, Cartwright P et al (2007) The phylum Cnidaria: a review of phylogenetic patterns and diversity 300 years after Linnaeus. Zootaxa 1668:127–182

David B (2018) New regulations for accessing plant biodiversity samples, what is ABS? Phytochem Rev 17:1211–1223. https://doi.org/10.1007/s11101-018-9573-1

David B, Ausseil F (2014) High-throughput screening of plant chemodiversity. In: Meyers RA (ed) Encyclopedia of analytical chemistry. American Cancer Society, pp 1–24

David B, Wolfender J-L, Dias DA (2015) The pharmaceutical industry and natural products: historical status and new trends. Phytochem Rev 14:299–315. https://doi.org/10.1007/s11101-014-9367-z

Davis J, Jones A, Lewis RJ (2009) Remarkable inter- and intra-species complexity of conotoxins revealed by LC/MS. Peptides 30:1222–1227. https://doi.org/10.1016/j.peptides.2009.03.019

de Lima Procópio RE, da Silva IR, Martins MK et al (2012) Antibiotics produced by *Streptomyces*. Braz J Infect Dis 16:466–471. https://doi.org/10.1016/j.bjid.2012.08.014

De Silva DD, Rapior S, Sudarman E et al (2013) Bioactive metabolites from macrofungi: ethnopharmacology, biological activities and chemistry. Fungal Divers 62:1–40. https://doi.org/10.1007/s13225-013-0265-2

Demain AL, Sanchez S (2009) Microbial drug discovery: 80 years of progress. J Antibiotics 62:5–16. https://doi.org/10.1038/ja.2008.16

Denning DW (2003) Echinocandin antifungal drugs. Lancet 362:1142–1151. https://doi.org/10.1016/S0140-6736(03)14472-8

Dhakal D, Pokhrel AR, Shrestha B, Sohng JK (2017) Marine rare actinobacteria: isolation, characterization, and strategies for harnessing bioactive compounds. Front Microbiol. https://doi.org/10.3389/fmicb.2017.01106

Dias DA, Urban S, Roessner U (2012) A historical overview of natural products in drug discovery. Metabolites 2:303–336. https://doi.org/10.3390/metabo2020303

DiMasi JA, Grabowski HG, Hansen RW (2016) Innovation in the pharmaceutical industry: new estimates of R&D costs. J Health Econ 47:20–33. https://doi.org/10.1016/j.jhealeco.2016.01.012

Dischinger J, Josten M, Szekat C et al (2009) Production of the novel two-peptide lantibiotic lichenicidin by *Bacillus licheniformis* DSM 13. PLoS ONE 4:e6788. https://doi.org/10.1371/journal.pone.0006788

Ernst M, Grace OM, Saslis-Lagoudakis CH et al (2015) Global medicinal uses of *Euphorbia* L. (Euphorbiaceae). J Ethnopharmacol 176:90–101. https://doi.org/10.1016/j.jep.2015.10.025

Falk H, Wolkenstein K (2017) Natural product molecular fossils. In: Kinghorn D, Falk H, Gibbons S, Kobayashi J (eds) Progress in the chemistry of organic natural products, vol 104. Springer, Cham, pp 1–126. https://doi.org/10.1007/978-3-319-45618-8

Frisvad JC, Smedsgaard J, Larsen TO, Samson RA (2004) Mycotoxins, drugs and other extrolites produced by species in *Penicillium* subgenus *Penicillium*. Stud Mycol 49:201–241

Gao H, Li G, Lou H-X (2018) Structural diversity and biological activities of novel secondary metabolites from endophytes. Molecules 23:646. https://doi.org/10.3390/molecules23030646

Gaudêncio S, Pereira F (2015) Dereplication: racing to speed up the natural products discovery process. Nat Prod Rep 32:779–810. https://doi.org/10.1039/C4NP00134F

Genilloud O (2017) Actinomycetes: still a source of novel antibiotics. Nat Prod Rep 34:1203–1232. https://doi.org/10.1039/C7NP00026J

Gershenzon J, Dudareva N (2007) The function of terpene natural products in the natural world. Nat Chem Biol 3:408–414. https://doi.org/10.1038/nchembio.2007.5

Goldbach-Mansky R (2009) Comparison of *Tripterygium wilfordii* Hook F versus sulfasalazine in the treatment of rheumatoid arthritis: a randomized trial. Ann Intern Med 151:229. https://doi.org/10.7326/0003-4819-151-4-200908180-00005

Goyal S, Ramawat KG, Mérillon JM (2016) Different shades of fungal metabolites: an overview. In: Mérillon JM, Ramawat KG (eds) Fungal metabolites. Springer, Cham, pp 1–29

Gross H, Loper JE (2009) Genomics of secondary metabolite production by *Pseudomonas* spp. Nat Prod Rep 26:1408–1446. https://doi.org/10.1039/B817075B

Haas D, Défago G (2005) Biological control of soil-borne pathogens by fluorescent pseudomonads. Nat Rev Microbiol 3:307–319. https://doi.org/10.1038/nrmicro1129

Harvey AL, Edrada-Ebel R, Quinn RJ (2015) The re-emergence of natural products for drug discovery in the genomics era. Nat Rev Drug Discov 14:111–129. https://doi.org/10.1038/nrd4510

Henkel T, Brunne RM, Müller H, Reichel F (1999) Statistical investigation into the structural complementarity of natural products and synthetic compounds. Angew Chem Int Ed 38:643–647. https://doi.org/10.1002/(SICI)1521-3773(19990301)38:5%3c643:AID-ANIE643%3e3.0.CO;2-G

Henry SH, Bosch FX, Bowers JC (2002) Aflatoxin, hepatitis and worldwide liver cancer risks. In: DeVries JW, Trucksess MW, Jackson LS (eds) Mycotoxins and food safety. Springer, Boston, pp 229–233

Hu G-P, Yuan J, Sun L et al (2011) Statistical research on marine natural products based on data obtained between 1985 and 2008. Mar Drugs 9:514–525. https://doi.org/10.3390/md9040514

Huang M, Lu J-J, Huang M-Q et al (2012) Terpenoids: natural products for cancer therapy. Expert Opin Investig Drugs

21:1801–1818. https://doi.org/10.1517/13543784.2012.727395

Ivanescu B, Miron A, Corciova A (2015) Sesquiterpene lactones from *Artemisia* genus: biological activities and methods of analysis. J Anal Methods Chem. https://doi.org/10.1155/2015/247685

Jassbi AR (2006) Chemistry and biological activity of secondary metabolites in *Euphorbia* from Iran. Phytochemistry 67:1977–1984. https://doi.org/10.1016/j.phytochem.2006.06.030

Jose PA, Jebakumar SRD (2013) Non-streptomycete actinomycetes nourish the current microbial antibiotic drug discovery. Front Microbiol. https://doi.org/10.3389/fmicb.2013.00240

Katz L, Baltz RH (2016) Natural product discovery: past, present, and future. J Ind Microbiol Biotechnol 43:155–176. https://doi.org/10.1007/s10295-015-1723-5

Kayser O, Kiderlen AF, Croft SL (2003) Natural products as antiparasitic drugs. Parasitol Res 90:S55–S62. https://doi.org/10.1007/s00436-002-0768-3

Kharwar RN, Mishra A, Gond SK et al (2011) Anticancer compounds derived from fungal endophytes: their importance and future challenges. Nat Prod Rep 28:1208–1228. https://doi.org/10.1039/C1NP00008J

King GF (2011) Venoms as a platform for human drugs: translating toxins into therapeutics. Expert Opin Biol Ther 11:1469–1484. https://doi.org/10.1517/14712598.2011.621940

Koch MA, Schuffenhauer A, Scheck M et al (2005) Charting biologically relevant chemical space: a structural classification of natural products (SCONP). Proc Natl Acad Sci 102:17272–17277. https://doi.org/10.1073/pnas.0503647102

Kong D-X, Guo M-Y, Xiao Z-H et al (2011) Historical variation of structural novelty in a natural product library. Chem Biodivers 8:1968–1977. https://doi.org/10.1002/cbdv.201100156

Kurtböke Dİ (2012) Biodiscovery from rare actinomycetes: an eco-taxonomical perspective. Appl Microbiol Biotechnol 93:1843–1852. https://doi.org/10.1007/s00253-012-3898-2

Langenheim JH (1994) Higher plant terpenoids: a phytocentric overview of their ecological roles. J Chem Ecol 20:1223–1280. https://doi.org/10.1007/BF02059809

Lawton EM, Cotter PD, Hill C, Ross RP (2007) Identification of a novel two-peptide lantibiotic, Haloduracin, produced by the alkaliphile *Bacillus halodurans* C-125. FEMS Microbiol Lett 267:64–71. https://doi.org/10.1111/j.1574-6968.2006.00539.x

Lazzarini A, Cavaletti L, Toppo G, Marinelli F (2000) Rare genera of actinomycetes as potential producers of new antibiotics. Antonie Van Leeuwenhoek 78:399–405. https://doi.org/10.1023/A:1010287600557

Leal MC, Puga J, Serôdio J et al (2012) Trends in the discovery of new marine natural products from invertebrates over the last two decades—where and what are we bioprospecting? PLoS ONE 7:e30580. https://doi.org/10.1371/journal.pone.0030580

Lewis RJ, Garcia ML (2003) Therapeutic potential of venom peptides. Nat Rev Drug Discov 2:790–802. https://doi.org/10.1038/nrd1197

Lindequist U, Niedermeyer THJ, Jülich W-D (2005) The pharmacological potential of mushrooms. Evid Based Complement Alternat Med 2:285–299. https://doi.org/10.1093/ecam/neh107

Locey KJ, Lennon JT (2016) Scaling laws predict global microbial diversity. Proc Natl Acad Sci 113:5970–5975. https://doi.org/10.1073/pnas.1521291113

Mangal M, Sagar P, Singh H et al (2013) NPACT: naturally occurring plant-based anti-cancer compound–activity–target database. Nucleic Acids Res 41:D1124–D1129. https://doi.org/10.1093/nar/gks1047

Mariottini GL, Grice ID (2016) Antimicrobials from cnidarians. A new perspective for anti-infective therapy? Mar Drugs 14:48. https://doi.org/10.3390/md14030048

May RM (2010) Tropical arthropod species, more or less? Science 329:41–42. https://doi.org/10.1126/science.1191058

Mehbub MF, Lei J, Franco C, Zhang W (2014) Marine sponge derived natural products between 2001 and 2010: trends and opportunities for discovery of bioactives. Mar Drugs 12:4539–4577. https://doi.org/10.3390/md12084539

Miller CM, Miller RV, Garton-Kenny D et al (1998) Ecomycins, unique antimycotics from *Pseudomonas viridiflava*. J Appl Microbiol 84:937–944. https://doi.org/10.1046/j.1365-2672.1998.00415.x

Mishra HN, Das C (2003) A review on biological control and metabolism of aflatoxin. Crit Rev Food Sci Nutr 43:245–264. https://doi.org/10.1080/10408690390826518

Mishra BB, Tiwari VK (2011) Natural products: an evolving role in future drug discovery. Eur J Med Chem 46:4769–4807. https://doi.org/10.1016/j.ejmech.2011.07.057

Mora C, Tittensor DP, Adl S et al (2011) How many species are there on earth and in the ocean? PLoS Biol 9:e1001127. https://doi.org/10.1371/journal.pbio.1001127

Newman DJ, Cragg GM (2016) Natural products as sources of new drugs from 1981 to 2014. J Nat Prod 79:629–661. https://doi.org/10.1021/acs.jnatprod.5b01055

Ntie-Kang F, Zofou D, Babiaka SB et al (2013) AfroDb: a select highly potent and diverse natural product library from African medicinal plants. PLoS ONE 8:e78085. https://doi.org/10.1371/journal.pone.0078085

Oprea TI (2000) Current trends in lead discovery: are we looking for the appropriate properties? Mol Divers 5:199–208. https://doi.org/10.1023/A:1021368007777

Palazzolo AME, Simons CLW, Burke MD (2017) The natural productome. Proc Natl Acad Sci 114:5564–5566. https://doi.org/10.1073/pnas.1706266114

Pascolutti M, Campitelli M, Nguyen B et al (2015) Capturing nature's diversity. PLoS ONE 10:e0120942. https://doi.org/10.1371/journal.pone.0120942

Paterson RRM (2006) Ganoderma—a therapeutic fungal biofactory. Phytochemistry 67:1985–2001. https://doi.org/10.1016/j.phytochem.2006.07.004

Patridge E, Gareiss P, Kinch MS, Hoyer D (2016) An analysis of FDA-approved drugs: natural products and their derivatives. Drug Discov Today 21:204–207. https://doi.org/10.1016/j.drudis.2015.01.009

Pedrós-Alió C (2006) Marine microbial diversity: can it be determined? Trends Microbiol 14:257–263. https://doi.org/10.1016/j.tim.2006.04.007

Peláez F (2005) Biological activities of fungal metabolites. In: An Z (ed) Handbook of industrial mycology. Marcel Dekker, New York, pp 49–92

Pimm SL, Joppa LN (2015) How many plant species are there, where are they, and at what rate are they going extinct? Ann Mo Bot Gard 100:170–176. https://doi.org/10.3417/2012018

Pimm SL, Jenkins CN, Abell R et al (2014) The biodiversity of species and their rates of extinction, distribution, and protection. Science 344:1246752. https://doi.org/10.1126/science.1246752

Pouny I, Batut M, Vendier L et al (2014) Cytisine-like alkaloids from *Ormosia hosiei* Hemsl. & E.H. Wilson. Phytochemistry 107:97–101. https://doi.org/10.1016/j.phytochem.2014.07.022

Pye CR, Bertin MJ, Lokey RS et al (2017) Retrospective analysis of natural products provides insights for future discovery trends. Proc Natl Acad Sci 114:5601–5606. https://doi.org/10.1073/pnas.1614680114

Quinn RJ, Carroll AR, Pham NB et al (2008) Developing a drug-like natural product library. J Nat Prod 71:464–468. https://doi.org/10.1021/np070526y

Rask-Andersen M, Masuram S, Schiöth HB (2014) The druggable genome: evaluation of drug targets in clinical trials suggests major shifts in molecular class and indication. Annu Rev Pharmacol Toxicol 54:9–26. https://doi.org/10.1146/annurev-pharmtox-011613-135943

Reed JL, Palsson BØ (2003) Thirteen years of building constraint-based in silico models of *Escherichia coli*. J Bacteriol 185:2692–2699. https://doi.org/10.1128/JB.185.9.2692-2699.2003

Rocha J, Peixe L, Gomes NCM, Calado R (2011) Cnidarians as a source of new marine bioactive compounds—an overview of the last decade and future steps for bioprospecting. Mar Drugs 9:1860–1886. https://doi.org/10.3390/md9101860

Rocha J, Calado R, Leal M (2015) Marine bioactive compounds from cnidarians. In: Kim SK (ed) Springer handbook of marine biotechnology. Springer, Berlin, pp 823–849

Rosén J, Gottfries J, Muresan S et al (2009) Novel chemical space exploration via natural products. J Med Chem 52:1953–1962. https://doi.org/10.1021/jm801514w

Roskov Y, Abucay L, Orrell T et al (2018) Species 2000 & ITIS catalogue of life, 2018 annual checklist. www.catalogueoflife.org/col. Accessed 4 Jun 2018

Russo M, Russo GL, Daglia M et al (2016) Understanding genistein in cancer: the "good" and the "bad" effects: a review. Food Chem 196:589–600. https://doi.org/10.1016/j.foodchem.2015.09.085

Sanchez JF, Somoza AD, Keller NP, Wang CCC (2012) Advances in *Aspergillus* secondary metabolite research in the post-genomic era. Nat Prod Rep 29:351–371. https://doi.org/10.1039/C2NP00084A

Sarker SD, Nahar L (2018) Chapter 1—an introduction to computational phytochemistry. In: Sarker SD, Nahar L (eds) Computational phytochemistry. Elsevier, Amsterdam, pp 1–41

Scheffers BR, Joppa LN, Pimm SL, Laurance WF (2012) What we know and don't know about Earth's missing biodiversity. Trends Ecol Evol 27:501–510. https://doi.org/10.1016/j.tree.2012.05.008

Schmidt U, Struck S, Gruening B et al (2009) SuperToxic: a comprehensive database of toxic compounds. Nucleic Acids Res 37:D295–D299. https://doi.org/10.1093/nar/gkn850

Schoch CL, Sung G-H, López-Giráldez F et al (2009) The Ascomycota tree of life: a phylum-wide phylogeny clarifies the origin and evolution of fundamental reproductive and ecological traits. Syst Biol 58:224–239. https://doi.org/10.1093/sysbio/syp020

Schulz B, Boyle C, Draeger S et al (2002) Endophytic fungi: a source of novel biologically active secondary metabolites. Mycol Res 106:996–1004. https://doi.org/10.1017/S0953756202006342

Sertuerner F (1817) Ueber das Morphium, eine neue salzfähige Grundlage, und die Mekonsäure, als Hauptbestandtheile des Opiums. Ann Phys 55:56–89. https://doi.org/10.1002/andp.18170550104

Shi Q-W, Su X-H, Kiyota H (2008) Chemical and pharmacological research of the plants in genus *Euphorbia*. Chem Rev 108:4295–4327. https://doi.org/10.1021/cr078350s

Stadler M, Hoffmeister D (2015) Fungal natural products—the mushroom perspective. Front Microbiol. https://doi.org/10.3389/fmicb.2015.00127

Stein T (2005) *Bacillus subtilis* antibiotics: structures, syntheses and specific functions. Microbiol Mol 56:845–857. https://doi.org/10.1111/j.1365-2958.2005.04587.x

Stierle A, Strobel G, Stierle D (1993) Taxol and taxane production by *Taxomyces andreanae*, an endophytic fungus of Pacific yew. Science 260:214–216. https://doi.org/10.1126/science.8097061

Stone JK, Bacon CW, White JF (2000) An overview of endophytic microbes: endophytism defined. In: Bacon CW, White JF (eds) Microbial endophytes. Marcel Dekker, New York, pp 3–29

Strebhardt K, Ullrich A (2008) Paul Ehrlich's magic bullet concept: 100 years of progress. Nat Rev Cancer 8:473–480. https://doi.org/10.1038/nrc2394

Strobel G, Daisy B, Castillo U, Harper J (2004) Natural products from endophytic microorganisms. J Nat Prod 67:257–268. https://doi.org/10.1021/np030397v

Su C-Y, Ming Q-L, Rahman K et al (2015) *Salvia miltiorrhiza*: traditional medicinal uses, chemistry, and pharmacology. Chin J Nat Med 13:163–182. https://doi.org/10.1016/S1875-5364(15)30002-9

Tan RX, Zou WX (2001) Endophytes: a rich source of functional metabolites. Nat Prod Rep 18:448–459. https://doi.org/10.1039/B100918O

Tan D, Wu J, Zhang X et al (2018) Sodium tanshinone II a sulfonate injection as adjuvant treatment for unstable angina pectoris: a meta-analysis of 17 randomized controlled trials. Chin J Integr Med 24:156–160. https://doi.org/10.1007/s11655-017-2424-x

Tao X, Lipsky PE (2000) The Chinese anti-inflammatory and immunosuppressive herbal remedy *Tripterygium wilfordii* Hook. f. Rheum Dis Clin 26:29–50. https://doi.org/10.1016/S0889-857X(05)70118-6

Thacker RW, Starnes S (2003) Host specificity of the symbiotic cyanobacterium *Oscillatoria spongeliae* in marine sponges, *Dysidea* spp. Mar Biol 142:643–648. https://doi.org/10.1007/s00227-002-0971-x

Thomas TRA, Kavlekar DP, LokaBharathi PA (2010) Marine drugs from sponge-microbe association—a review. Mar Drugs 8:1417–1468. https://doi.org/10.3390/md8041417

Tiwari K, Gupta RK (2012) Rare actinomycetes: a potential storehouse for novel antibiotics. Crit Rev Biotechnol 32:108–132. https://doi.org/10.3109/07388551.2011.562482

Valli M, dos Santos RN, Figueira LD et al (2013) Development of a natural products database from the biodiversity of Brazil. J Nat Prod 76:439–444. https://doi.org/10.1021/np3006875

Van Soest RWM, Boury-Esnault N, Vacelet J et al (2012) Global diversity of sponges (Porifera). PLoS ONE 7:e35105. https://doi.org/10.1371/journal.pone.0035105

Vasas A, Hohmann J (2014) *Euphorbia* diterpenes: isolation, structure, biological activity, and synthesis (2008–2012). Chem Rev 114:8579–8612. https://doi.org/10.1021/cr400541j

Veitch NC (2010) Flavonoid chemistry of the Leguminosae. In: Santos-Buelga C, Escribano-Bailon MT, Lattanzio V (eds) Recent advances in polyphenol research. Wiley-Blackwell, Hoboken, pp 23–58

Wang J, Zhang L, Teng K et al (2014) Cerecidins, novel lantibiotics from *Bacillus cereus* with potent antimicrobial activity. Appl Environ Microbiol 80:2633–2643. https://doi.org/10.1128/AEM.03751-13

Watve MG, Tickoo R, Jog MM, Bhole BD (2001) How many antibiotics are produced by the genus *Streptomyces*? Arch Microbiol 176:386–390. https://doi.org/10.1007/s002030100345

Whittle M, Willett P, Klaffke W, van Noort P (2003) Evaluation of similarity measures for searching the dictionary of natural products database. J Chem Inf Comput Sci 43:449–457. https://doi.org/10.1021/ci025591m

Wilson MC, Mori T, Rückert C et al (2014) An environmental bacterial taxon with a large and distinct metabolic repertoire. Nature 506:58–62. https://doi.org/10.1038/nature12959

Wink M (2013) Evolution of secondary metabolites in legumes (Fabaceae). S Afr J Bot 89:164–175. https://doi.org/10.1016/j.sajb.2013.06.006

World Health Organization (2008) WHO|"Beijing declaration." In: WHO. http://www.who.int/medicines/areas/traditional/congress/beijing_declaration/en/. Accessed 16 Dec 2018

Wu Y-B, Ni Z-Y, Shi Q-W et al (2012) Constituents from *Salvia* species and their biological activities. Chem Rev 112:5967–6026. https://doi.org/10.1021/cr200058f

Xue R, Fang Z, Zhang M et al (2013) TCMID: traditional Chinese medicine integrative database for herb molecular mechanism analysis. Nucleic Acids Res 41:D1089–D1095. https://doi.org/10.1093/nar/gks1100

Ye H, Ye L, Kang H et al (2011) HIT: linking herbal active ingredients to targets. Nucleic Acids Res 39:D1055–D1059. https://doi.org/10.1093/nar/gkq1165

Yuen JWM, Gohel MDI (2005) Anticancer effects of *Ganoderma lucidum*: a review of scientific evidence. Nutr Cancer 53:11–17. https://doi.org/10.1207/s15327914nc5301_2

Zambelli VO, Pasqualoto KFM, Picolo G et al (2016) Harnessing the knowledge of animal toxins to generate drugs. Pharmacol Res 112:30–36. https://doi.org/10.1016/j.phrs.2016.01.009

Zhu F, Qin C, Tao L et al (2011) Clustered patterns of species origins of nature-derived drugs and clues for future bioprospecting. Proc Natl Acad Sci 108:12943–12948. https://doi.org/10.1073/pnas.1107336108

Phytochemistry 70 (2009) 75–85

Contents lists available at ScienceDirect

Phytochemistry

journal homepage: www.elsevier.com/locate/phytochem

Antiplasmodial benzophenones from the trunk latex of *Moronobea coccinea* (Clusiaceae)

Guillaume Marti [a], Véronique Eparvier [b], Christian Moretti [c], Sophie Susplugas [d], Soizic Prado [e], Philippe Grellier [d], Pascal Retailleau [a], Françoise Guéritte [a], Marc Litaudon [a,*]

[a] Institut de Chimie des Substances Naturelles, 1, avenue de la Terrasse, CNRS, 91198 Gif-sur-Yvette, France
[b] UPS 2561, CNRS, 16, avenue André Aron, 97300 Cayenne, Guyane Française, France
[c] Unité N° 84 de l'IRD: Biodival, IRD, Technoparc, 5, rue du Carbone, 45072 Orleans Cedex 2, France
[d] Museum National d'Histoire Naturelle, USM504-EA3335, CP52, 61, rue Buffon, 75231 Paris Cedex 05, France
[e] Museum National d'Histoire Naturelle, USM502-UMR 5152 CNRS, 63, rue Buffon, 75231 Paris Cedex 05, France

ARTICLE INFO

Article history:
Received 1 August 2008
Received in revised form 25 September 2008
Available online 4 December 2008

Keywords:
Moronobea coccinea
Clusiaceae
Benzophenone
Coccinone
Antiplasmodial

ABSTRACT

In an effort to find antimalarial drugs, a systematic *in vitro* evaluation on a chloroquine-resistant strain of *Plasmodium falciparum* (FcB1) was undertaken on sixty plant extracts collected in French Guiana. The methanol extract obtained from the latex of *Moronobea coccinea* exhibited a strong antiplasmodial activity (95% at 10 μg/ml). The phytochemical investigation of this extract led to the isolation of eleven polycyclic polyprenylated acylphloroglucinols (PPAPs), from which eight showed potent antiplasmodial activity with IC50 ranged from 3.3 μM to 37.2 μM.

1. Introduction

With over 300 millions of new cases each year, resulting in more than one million deaths annually, malaria remains one of the most important infectious disease of the developing world. Fatal cases are generally caused by the most virulent human malaria parasite, *Plasmodium falciparum*. Symptoms occur during the asexual parasite development in human red blood cells. Current clinical treatment involves the use of inexpensive antimalarial drugs such as chloroquine and their derivatives. However, the emergence of resistant populations of *Plasmodium* sp. to this class of drugs has rendered most of them ineffective (Laufer and Plowe, 2004). Since few years, World Health Organization recommends the use of a combinatorial therapy to limit resistance phenomenon. The "Artemisinin Combination Therapy" is currently the most effective treatment. Although no clear clinical resistance was yet reported, there are accumulating evidences for increasing artemisinin resistance *in vitro* of *P. faciparum* isolates (Gay et al., 1994; Jambou et al., 2005). There is therefore an urgent need for the discovery of new efficient antimalarial drugs to cure this disease and to prevent the emergence of resistance.

In an effort to find new naturally antimalarial drugs, we screened a series of sixty plant extracts prepared from various species belonging to the French Guiana biodiversity. The results of this screening led us to investigate the latex of *Moronobea coccinea* Aubl. (Clusiaceae). Indeed, ethyl acetate extracts of trunk bark and fruits, having a rich latex secretion system, exhibited a strong antiplasmodial activity (75% and 83% of inhibitory growth at 10 μg/ml, respectively) whereas the leave and trunk root extracts did not show any significant activity (4% and 5% of inhibitory growth at 10 μg/ml, respectively). The latex was therefore collected separately by scraping the trunk bark. The methanol extract of the latex showed 95% of inhibitory growth at 10 μg/ml. In this report we described the isolation, structure elucidation and biological activities of eleven new polycyclic polyprenylated acylphloroglucinols (PPAPs) (**3–8, 10–14**) along with the known isogarcinol, cycloxanthochymol and garcinol (**1, 2, 9,** respectively).

The Clusiaceae family consists of 40 genera and 1200 species essentially mostly tropical. In French Guyana several species are well-known in the manufacture of hulls, for the quality of their wood *(Calophyllum brasiliense)* and for the healing properties of their latex (*Mahurea, Vismia, Symphonia* and *Moronobea* spp.) used traditionally for their effectiveness against dermatoses (Grenand et al., 2004). From a phytochemical point of view, this family is known to be a rich source of PPAPs with a large spectrum

* Corresponding author. Tel.: +33 1 69 82 30 85; fax: +33 1 69 07 72 47.
E-mail address: marc.litaudon@icsn.cnrs-gif.fr (M. Litaudon).

0031-9422/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.phytochem.2008.10.005

of biological activities (Cuesta-Rubio et al., 2005). However, to our knowledge only guttiferone A has been screened for its antiplasmodial activity (Ngouela et al., 2006). The genus *Moronobea* consists of some seven species strictly distributed in Amazonia. Only *Moronobea pulchra* has been investigated and the geranyl benzophenone marupone was isolated from the trunk wood (Dias et al., 1974).

2. Results and discussion

One hundred milligram of the latex methanolic extract was filtered on polyamide to remove tannins and 15 mg of the filtered extract was fractionated on a semipreparative C-18 column to give nine fractions according to a standardized method (see Section 4) (Bousserouel et al., 2005). Biological assay on *P. falciparum* allowed us to identify the active fraction (fraction 6, t_R from 42 to 50 min). A large amount of the extract (2.5 g) was then submitted to silica gel chromatography to give 18 fractions (F1–F18). Comparative study on analytical HPLC of these fractions with fraction 6 allowed us to target those containing the active compounds. Subsequent preparative HPLC purification resulted in the isolation of compounds **1**–**14** (Fig. 1).

Compounds **1**–**14** shared several common spectral characteristics. The UV spectra showed absorption bands at 230 and 280 nm consistent with aromatic rings and conjugated carbonyl groups. The IR spectra exhibited bands for hydroxyl (3350 cm^{-1}), α,β-unsaturated carbonyl groups (1650, 1668 cm^{-1}), ketone (1720 cm^{-1}) and aromatic rings (1600 cm^{-1}). The ^1H and ^{13}C NMR spectra recorded in CD_3OD appeared quite complex due to a keto-enol tautomeric balance within the molecules (Fuller et al., 1999; Gustafson et al., 1992; Williams et al., 2003). In contrast, only one set of signals was observed when pyridine-d_5 was used (Table 1). All the structures reported herein follow the IUPAC convention for numbering bicyclo[3.3.1]nonanes in which C(1)–C(9) and C(5)–C(9) bonds as bold or hashed wedges according to

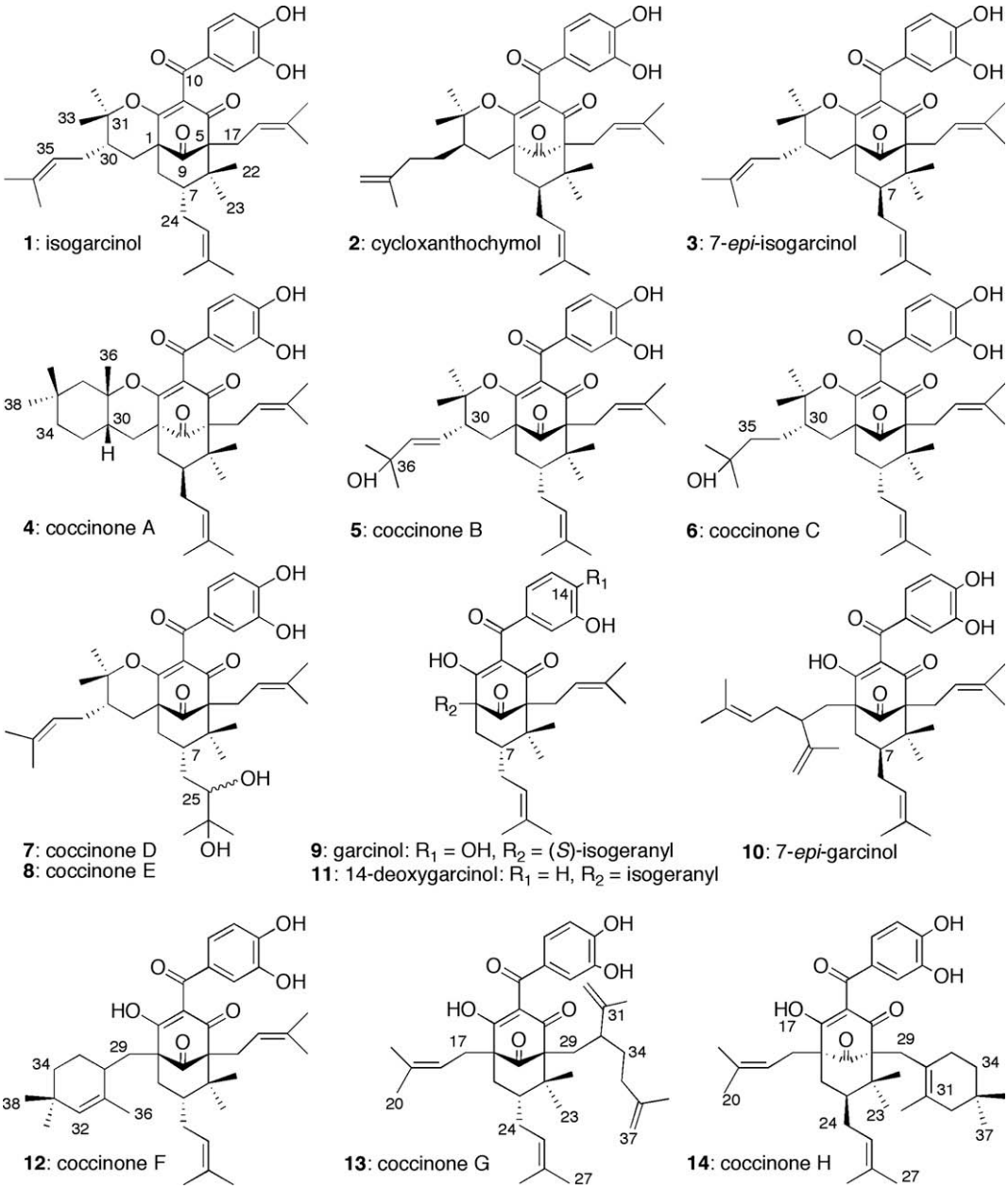


Fig. 1. Structures of compounds **1**–**14**.

Table 1
NMR spectroscopic data for compound **1**, **3**–**8** and **10**–**14** (500 MHz for ^{1}H NMR and 125 MHz for ^{13}C NMR in pyridine-d_5.

Position	Compound **1**			Compound **3**			Compound **4**			Compound **5**		
	δ_C	δ_H	*J* (Hz)	δ_C	δ_H	*J* (Hz)	δ_C	δ_H	*J* (Hz)	δ_C	δ_H	*J* (Hz)
1	52.2			52.2			48.4			51.7		
2	171.4			170.8			171.3			171.1		
3	127.2			129.1			128.5			127.0		
4	195.0			194.9			195.6			195.0		
5	69.2			71.4			69.4			69.3		
6	46.8			46.8			46.8			46.9		
7	47.0	1.59	*dt* (6.2, 6.1)	42.2	2.44	*m*	46.7	1.62	*m*	47.1	1.58	*m*
8	39.8	2.11	*dd* (14.1, 7.3)	43.1	1.77	*m*	42.1	2.05	*brd* (13.7)	40.0	2.10	*m*
–		2.43	*brd* (14.1)		2.43	*m*		2.78	*dd* (13.7, 5.8)		2.48	*m*
9	207.9			207.3			208.6			207.9		
10	193.0			193.2			192.5			193.0		
11	130.9			131.0			130.4			130.9		
12	116.6	8.05	*d* (2.0)	117.1	8.13	*brs*	117.1	8.04	*d* (2.2)	116.7	8.06	*d* (1.8)
13	147.8			147.9			147.5			147.8		
14	153.7			153.7			153.4			153.8		
15	116.5	7.28	*d* (8.1)	116.5	7.31	*brd* (7.6)	116.1	7.28	*d* (8.4)	116.6	7.29	*d* (8.0)
16	124.3	7.68	*dd* (8.1, 2)	124.4	7.72	*brd* (7.7)	124.5	7.60	*dd* (8.4, 2.2)	124.4	7.72	*dd* (8.0, 1.8)
17	26.7	2.76	*dd* (13.7, 5.6)	25.9	2.77	*dd* (13.6, 4.8)	26.4	2.81	*dd* (13.6, 6.8)	26.8	2.79	*dd* (13.3, 7.3)
–		2.95	*dd* (13.5, 7.6)		2.97	*dd* (13.6, 6.4)		2.93	*m*		2.85	*dd* (13.3, 7.3)
18	121.7	5.42	*brt* (6.5)	122.1	5.38	*brt* (6.2)	121.4	5.48	*brt* (7.0)	121.7	5.46	*brt* (5.7)
19	134.5			134.2			134.4			134.7		
20	26.6	1.57	*s*	26.2	1.55	*s*	26.5	1.76	*brs*	26.7	1.67	*s*
–												
21	18.8	1.71	*s*	18.8	1.73	*s*	18.7	1.76	*brs*	18.8	1.74	*s*
22	27.1	1.05	*s*	16.6	0.83	*s*	26.7	1.08	*s*	27.1	1.08	*s*
23	23.1	1.30	*s*	22.8	1.27	*s*	22.5	1.28	*s*	23.1	1.31	*s*
24	30.4	1.82	*ddd* (14.2, 9.5, 9.5)	28.5	1.82	*dd* (14.1, 8.5)	30.2	2.45	*m*	29.8	2.43	*m*[a]
–		3.21	*ddd* (14.4, 10.7, 9.5)		2.26	*dd* (14.1, 2.5)		3.03	*m*		3.19	*m*[a]
25	126.4	5.09	*brt* (6.5)	123.8	5.18	*brt* (6.2)	125.9	5.15	*brt* (7.0)	126.4	5.08	*brt* (5.7)
26	132.9			132.9			133.1			133.4		
27	26.5	1.74	*s*	26.2	1.62	*s*	26.1	1.71	*s*	26.5	1.73	*s*
28	19.0	1.91	*s*	18.4	1.62	*s*	18.5	1.86	*s*	19.1	1.86	*s*
29	29.1	1.14	*dd* (13.9, 13.7)	28.5	1.21	*dd* (13.9, 13.7)	29.8	1.40	*m*	30.4	1.41	*m*[a]
–		3.27	*dd* (13.9, 3.1)		3.28	*dd* (13.6, 3.2)		3.08	*dd* (14.5, 5.4)		3.18	*m*[a]
30	43.8	1.66	*dt* (9.9, 5.0)	43.8	1.65	*m*	37.1	1.49	*m*[a]	46.0	2.30	*m*
31	87.2			87.6			83.9			86.0		
32	21.7	1.23	*s*	21.7	1.14	*s*	48.8	1.04	*m*[a]	22.1	1.19	*s*
–								1.46	*m*[a]			
33	29.4	1.07	*s*	29.2	1.08	*s*	30.6			29.4	0.99	*s*
–												
34	30.4	1.96	*brd* (14.1)	30.4	1.79	*m*	37.9	1.03	*m*[a]	124.8	5.70	*d* (15.6)
–		2.42	*brd* (14.1)		1.96	*m*		1.35	*m*			
35	122.8	5.09	*brt* (6.5)	122.9	5.13	*brt* (6.6)	26.4	1.37	*m*	144.4	5.71	*d* (15.6)
–								1.89	*m*			
36	133.7			133.7			29.0	0.91	*s*	69.9		
37	26.2	1.68	*s*	26.2	1.71	*s*	32.1	0.72	*s*	31.0[b]	1.43	*s*
38	18.3	1.56	*s*	18.3	1.56	*s*	27.4	0.97	*s*	30.9[b]	1.45	*s*

(*continued on next page*)

Table 1 (*continued*)

Position	Compound **6**			Compound **7**			Compound **8**			Compound **10**		
	δ_C	δ_H	*J* (Hz)	δ_C	δ_H	*J* (Hz)	δ_C	δ_H	*J* (Hz)	δ_C	δ_H	*J* (Hz)
1	52.3			52.3			52.7			61.5		
2	171.5			171.2			171.6			192.4		
3	127.4			127.2			127.5			120.6		
4	195.1			195.1			195.3			196.2		
5	69.4			69.5			69.6			70.2		
6	47.4			46.9			48.1			48.2		
7	46.9	1.54	*m*	43.2	2.29	*m*[a]	44.5	2.24	*m*	43.2	2.17	*m*
8	40.8	2.20	*brd* (14.6)	40.3	2.28	*m*[a]	45.5	2.47	*dd* (13.9-5.9)	45.4	1.63	*m*
–		2.51	*dd* (14.6, 7.1)		2.58	*brd* (13.3)		2.87	*brd* (13.9)		2.40	*m*
9	207.9			207.9			208.2			210.4		
10	193.4			193.7			193.7			195.8		
11	130.9			130.9			130.9			131.1		
12	116.6	8.08	*d* (1.7)	116.5	8.08	*d* (1.8)	116.5	8.1	*d* (1.8)	117.8	7.86	*brs*
13	147.8			147.8			147.8			147.4		
14	153.8			153.9			153.8			153.3		
15	117.0	7.27	*d* (8.2)	116.7	7.28	*d* (8.0)	116.7	7.30	*d* (8.0)	115.8	7.20	*d* (8.1)
16	124.2	7.70	*dd* (8.2, 1.7)	125.5	7.73	*dd* (8.0, 1.8)	121.8	7.71	*dd* (8.0, 1.8)	125.2	7.61	*brs*
17	26.6	2.75	*dd* (13.3, 5.0)	26.7	2.78	*m*[a]	26.4	2.74	*dd* (13.4, 5.0)	26.2	2.92	*m*
–		2.94	*dd* (13.3, 7.0)		2.95	*dd* (13.5, 7.4)		2.96	*dd* (13.4, 7.4)		3.04	*m*
18	121.9	5.42	*brt* (5.9)	121.8	5.43	*brt* (6.1)	121.8	5.45	*brt* (6.0)	122.8	5.55	*brt* (6.0)
19	134.4			134.3			134.3			133.3		
20	26.6	1.57	*s*	26.6	1.57	*s*	27.0	1.58	*s*	26.2	1.79	*s*
21	18.9	1.71	*s*	18.8	1.71	*s*	18.8	1.73	*s*	18.9	1.85	*s*
22	27.1	1.04	*s*	27.3	1.08	*s*	26.9	1.08	*s*	16.7	0.88	*s*
23	22.8	1.31	*s*	23.2	1.43	*s*	22.9	1.36	*s*	24.2	1.31	*s*
24	30.6	1.80	*m*	33.7	2.05	*dd* (13.9, 11.4)	36.9	2.25	*m*	29.2	1.78	*m*
–		1.97	*m*		2.80	*m*		2.63	*m*		2.15	*m*
25	122.9	5.08	*brt* (5.9)	78.5	3.80	*brd* (10.4)	83.1	3.69	*brd* (10.4)	124.5	5.08	*brt* (6.0)
26	133.7			73.4			73.5			133.9		
27	26.2	1.67	*s*	27.2[b]	1.74	*s*	26.6[b]	1.59	*s*	26.2	1.59	*s*
28	18.3	1.54	*s*	25.0[b]	1.66	*s*	25.2[b]	1.73	*s*	18.4	1.57	*s*
29	29.5	1.14	*dd* (14.3, 9.3)	29.2	1.14	*m*[a]	28.9	1.08	*m*[a]	36.8	2.23	*m*
–		3.29	*dd* (14.3, 3.5)		3.32	*dd* (14.3, 3.3)		3.37	*dd* (14.3, 3.3)		2.33	*m*
30	44.0	1.64	*m*	43.9	1.67	*m*	43.9	1.64	*m*[a]	44.9	3.16	*m*
31	87.2			87.2			87.3			149.5		
32	21.6	1.22	*s*	21.5	1.22	*s*	21.6	1.13	*s*	113.9	4.82	*brs*
–											4.93	*brs*
33	29.4	1.06	*s*	29.4	1.06	*s*	29.4	1.06	*s*	18.9	1.77	*s*
34	26.9	2.15	*m*	30.4	1.78	*m*	30.3	1.76	*m*[a]	33.6	2.28	*m*
–		2.43	*m*		1.98	*m*		1.94	*m*		2.37	*m*
35	46.6	1.89	*m*	122.8	5.09	*brd* (6.1)	122.8	5.07	*brd* (6.0)	124.9	5.31	*brt* (6.0)
–		1.64	*m*									
36	70.5			133.7			133.6			131.9		
37	30.5 [b]	1.56	*s*	26.2	1.67	*s*	26.2	1.66	*s*	26.3	1.64	*s*
38	30.2 [b]	1.56	*s*	18.3	1.54	*s*	18.3	1.52	*s*	18.6	1.68	*s*

Position	Compound **11**			Compound **12**			Compound **13**			Compound **14**		
1	60.4			62.2			62.6			62.8		
2	193.0			192.2			192.2			191.0		
3	119.5			119.3			119.6			118.0		
4	198.2			190.3			196.2			195.7		
5	68.6			66.1			66.2			66.3		
6	49.0			49.4			50.2			50.5		
7	47.8	1.56	*m*	47.9	1.58	*m*	47.0	1.51	*dt* (12.5, 5.1)	47.5	1.57	*m*
8	43.1	2.17	*dd* (13.9, 6.7)	40.4	2.22	*dd* (14.0, 7.0)	42.0	2.10	*o*	42.2	2.14	*dd* (14.0, 6.9)
–		2.48	*m*		2.40	*m*[a]		2.41	*m*		2.43	*dd* (14.0, 1.2)
9	206.4			211.3			210.9			210.1		
10	195.7			196.7			195.9			195.0		
11	142.7			131.4			130.1			129.9		
12	116.6	7.75	*brs*	117.9	7.95	*d* (2.0)	118.0	7.90	*s*	117.9	7.84	*d* (2.1)
13	159.3			147.3			147.4			147.4		
14	120.1	7.29	*brd*	153.2			153.4			153.7		
15	129.7	7.35	*m*	116.1	7.21	*d* (8.4)	116.0	7.19	*m*	115.9	7.20	*d* (8.2)
16	121.4	7.53	*brd*	124.4	7.70	*dd* (8.4, 2.0)	125.9	7.59	*m*	126.0	7.53	*dd* (8.2, 2.2)
17	27.0	2.91	*m*	32.2	2.87	*brs*	32.2	2.71	*dd* (13.6, 9.1)	32.2	2.78	*dd* (13.8, 5.6)
–		2.98	*m*		2.89	*brs*		2.91	*dd* (13.6, 5.7)		2.91	*dd* (13.8, 8.0)
18	122.8	5.60	*brt* (5.6)	122.2	5.81	*brt* (6.0)	122.5	5.67	*m*	121.7	5.69	*brt* (7.1)
19	133.6			133.8			133.8			134.7		
20	26.7	1.84	*s*	26.6	1.78	*m*[a]	26.7	1.88	*s*	26.7	1.86	*s*
21	18.9	1.83	*s*	18.7	1.78	*m*[a]	18.6	1.79	*s*	18.4	1.81	*s*
22	27.7	1.11	*s*	27.8	1.15	*s*	27.4	1.09	*s*	27.2	1.13	*s*
23	23.6	1.36	*s*	23.6	1.39	*s*	23.8	1.41	*s*	23.7	1.44	*s*
24	33.51	2.42	*m*	30.2	2.45	*m*	33.3	1.59	*m*	30.15	2.43	*dd* (15.0, 1.2)
–		2.67	*m*		2.78	*m*		1.65	*m*		2.50	*dd* (15.0, 9.5)
25	126.4	5.05	*brt* (5.6)	126.4	5.06	*brt* (6.0)	125.9	5.00	*m*	125.8	5.01	*brt* (7.0)
26	132.5			132.6			132.7			132.8		
27	26.3	1.62	*s*	26.4	1.62	*s*	26.2	1.61	*s*	26.3	1.63	*s*
28	18.5	1.60	*s*	18.6	1.63	*s*	18.4	1.56	*s*	18.4	1.58	*s*
29	37.6	2.28	*m*	29.5	2.30	*m*	32.2	2.09	*o*	30.2	3.10	*d* (15.0)
–					2.40	*m*		2.56	*m*		3.14	*d* (15.0)
30	44.8	3.14	*m*	35.9	2.66	*m*	45.4	2.52	*m*	128.8		
31	149.8			137.4			147.1			125.9		
32	112.7	4.81	*brs*	134.0	5.21	*brs*	113.5	4.81	*brs*	47.4	1.67	*d* (19.3)
–		4.92	*brs*					4.87	*brs*		1.74	*d* (19.3)
33	19.1	1.78	*s*	32.3			18.4	1.94	*s*	29.2		
34	27.7	2.32	*m*	34.5	1.62	*m*[a]	33.3	1.59	*m*	37.0	1.29	*m*
–		2.42	*m*		1.13	*m*[a]		1.65	*m*			
35	124.7	5.31	*brt* (5.6)	25.5	1.59	*m*[a]	36.6	1.87	*m*	29.3	1.96	*m*
–					1.78	*m*[a]		1.94	*m*		2.20	*m*
36	131.9			23.4	2.11	*s*	146.7			21.1	1.86	*s*
37	26.3	1.67	*s*	31.5	0.94	*s*	110.5	4.72	*brs*	28.3	0.69	*s*
–								4.76	*brs*			
38	18.7	1.71	*s*	29.9	0.84	*s*	22.5	1.65	*s*	28.9	0.84	*s*

[a] Overlap.
[b] Assignments are interchangeable, *brt*: broad triplet, *brs*: broad singulet.

Grossman's recommendations (Ciochina and Grossman, 2006). While several PPAPs have been isolated in both enantiomeric forms (Baggett et al., 2005; Gustafson et al., 1992; Iinuma et al., 1996; Krishnamurthy et al., 1981; Rao and Venkatswamy, 1980; Roux et al., 2000; Sahu et al., 1989) only few of them have their absolute configuration confirmed experimentally (Blount and Williams, 1976; Krishnamurthy et al., 1982). Therefore, it has been postulated that a positive or negative $[\alpha]_D$ value is due to the bicyclo[3.3.1]nonane system orientation. PPAPs presented here follow this rule.

Compound **1** has been isolated as a brown crystal, $[\alpha]_D$ −168° (*c* 1.0, $CHCl_3$). The HREIMS indicated a $[M+Na]^+$ ion peak at 625.3499, which suggested a molecular formula of $C_{38}H_{50}O_6Na$ (calc. 625.3505). Examination of the 1D and 2D NMR spectra associated with the value of optical rotation confirmed that **1** was isogarcinol (Iinuma et al., 1996; Krishnamurthy et al., 1981). However, no detailed NMR data were available in the literature. We reported in Table 1 the complete assignment of ^1H and ^{13}C NMR data of **1**. The X-ray diffraction analysis from crystals of the dibrosylate derivative of **1,** (**1′**) allowed us to confirm the absolute configuration of isogarcinol previously established by Krishnamurthy and coworkers (Krishnamurthy et al., 1982; Rogers et al., 1981). X-ray data are available on the Cambridge Crystallographic Data Centre (see supplementary material section).

The spectroscopic data of compound **2** were similar to those of cycloxanthochymol isolated from *Garcinia subelliptica* and *Garcinia pyrifera* (Iinuma et al., 1996; Roux et al., 2000).

Compound **3**, a brown powder, exhibited $[\alpha]_D$ −158° (*c* 1.0, $CHCl_3$). The HREIMS indicated an ion peak at *m*/*z* 625.3519, which suggested the same molecular formula as **1** and **2** ($C_{38}H_{50}O_6Na$). The ^1H and ^{13}C NMR spectral data were almost identical to those of **1**. In the ^{13}C NMR, only the signals of carbons C-7, C-8, C-22, C-23 and C-24 shifted slightly suggesting a β-orientation for the prenyl side chain located at position 7. According to Ciochina and Grossman, the orientation of the substituent on C-7 (axial or equatorial) can be deduced from the ^1H and ^{13}C NMR spectral data (Ciochina and Grossman, 2006; Hamed et al., 2006; Piccinelli et al., 2005). The C-22 and C-23 methyl resonate at δ 27.1 and 23.1 respectively for isogarcinol (**1**) having an axial prenyl side chain on position C-7, whereas the *gem*-dimethyl group showed chemicals shifts at δ 16.6 and δ 22.8, respectively in compound **3**. The upfield shift of the C-22_{ax} signal resulted from a γ-gauche interaction between this carbon and the CH_2-24 of the prenyl group. In the NOESY spectrum, nOe interactions between H-8_{eq}, H-7 and Me-23_{eq} (δ_H 1.77, 2.44 and 1.27, respectively) confirmed the absolute configuration (*S*) for C-7. Compound **3** was thus identified as 7-*epi*-isogarcinol.

Compound **4** was isolated as a brown oil. The HREIMS showed the same ion peak $[M+Na]^+$ *m*/*z* 625.3521 as compound **1** indicating the same formula $C_{38}H_{50}O_6Na$. The ^1H and ^{13}C NMR spectra confirmed the presence of a 1,2,4-trisubstituted benzene ring (Table 1) and displayed resonances for nine methyl groups [δ_H 1.08 (3 H, *s*, Me-22), 1.28 (3 H, *s*, Me-23), 0.91 (3 H, *s*, Me-36), 0.72 (3 H, *s*, Me-37), 0.97 (3 H, *s*, Me-38), 1.76 (6 H, *brs*, Me-20 and Me-21), 1.71 (3 H, *s*, Me-27), and 1.86 (3 H, *s*, Me-28)], and two vinylic protons [δ_H 5.48 (1H, *brt*, *J* = 7.0 Hz, H_1-18), and 5.15 (1H, *brt*, *J* = 7.0 Hz, H_1-25)]. The COSY and HMBC spectra revealed the presence of an additional 1,1-dimethylcyclohexane fused to the tetrahydropyrane ring on positions C-30 and C-31 instead of a prenyl side chain observed in **1** (Fig. 2). The relative configuration of the asymmetric carbons delineated in structure **4** has been determined by nOe experiment. Resonances for Me-36 with H-29_{ax} (δ_H 3.08), H-30 (δ_H 1.49) and H-32_{ax} (δ_H 1.46) allowed us to place these groups in the same face of the trimethylperhydrochromene moiety. On the other hand, resonances of Me-38 with H-35_{ax} (δ_H 1.89) allowed us to locate these groups on the opposite face. These correlations suggested a β-position of H-30 and the Me-36 group indicating a *cis*-junction between the tetrahydropyrane and the dimethylcyclohexane rings. From a biogenetic point of view, compound **4** could be derived from a double intramolecular cyclisation of the ω-isolavandulyl side chain at C-1 position of the known xanthochymol, although this last has not been isolated (Dreyer, 1974). The trimethylperhydrochromene system has already been described in a type A benzophenone derivative isolated from *Clusia obdeltifolia* (Teixeira and Cruz, 2005). Compound **4** was named coccinone A.

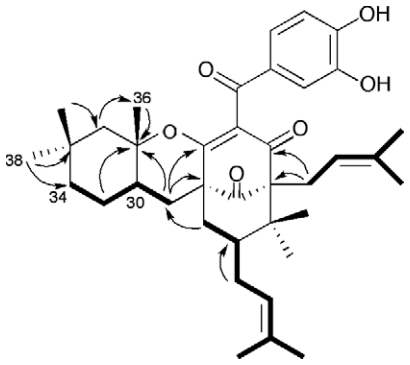


Fig. 2. Key HMBC (arrows) and COSY (bold) correlations for **4**.

Compound **5** was isolated as a brown oil. The HREIMS indicated a molecular formula $C_{38}H_{50}O_7$ deduced from the ion peak at *m*/*z* 641.3446 $M+Na^+$ (calc. 641.3454), which suggested the presence of an additional hydroxy group when compared with **1**. Examination of the 1D and 2D NMR spectral data of compound **5** confirmed that a benzoylphloglucinol skeleton is associated to a tetrahydropyrane ring. The presence of a 3-hydroxyl-3-methylbutenyl side chain in position C-30 was suggested by the following evidence. In the HMBC spectrum, correlations were observed from the quaternary carbon C-36 and the methine C-30 (δ_C 69.9 and 46.0, respectively) to the vinylic protons at δ_H 5.70 and 5.71 (each 1H, *d*, *J* = 15.6 Hz) H-34 and H-35, respectively, and from C-35 at δ_C 144.4 and C-36 to the methyl protons H_3-37 and H_3-38 (δ_H 1.43 and 1.45, respectively). The correlation between H-34 and H-30 (δ_H 2.30) in the COSY spectrum confirmed the location of the side chain at position C-30. These data, together with other results of 2D NMR analysis confirmed the structure of compound **5**, which was named coccinone B.

Compound **6** was isolated as a brown oil. The molecular formula $C_{38}H_{52}O_7$ has been deduced by HREIMS from the ion peak $[M+Na]^+$ *m*/*z* 643.3619 (calc. 643.3611) and suggested the presence of two additional protons when compared with **5**. The presence of a saturated side chain was confirmed in the ^1H and ^{13}C NMR spectra by signals for two additional methylene groups instead of two vinylic carbons. In the HMBC spectrum, correlations were observed from the quaternary carbon C-36 (δ_C 70.5) and the methylene C-35 (δ_C 46.6) to the methyl protons H_3-37 and H_3-38 (δ_H 1.56, 6 H, *s*). In the COSY spectrum, the correlations between H-35, H-34 and H-30 confirmed the location on position C-30 of the 3-hydroxy-3-methylbutyl side chain. This was further confirmed by NOESY interactions between Me-32 with H_α-34 (δ_H 2.43). Compound **6** was named coccinone C.

Compound **7,** which was isolated as a brown oil, has an $[\alpha]_D$ value of −76° (*c* 0.4, $CHCl_3$). The HREIMS revealed an ion peak at $[M+Na]^+$ *m*/*z* 659.3563 (calc. 659.3560), giving the molecular formula $C_{38}H_{52}O_8$ suggesting the presence of an additional hydroxy group when compared with compound **6**. Examination of the 1D and 2D NMR spectra showed that compound **7** shared several characteristics with the previous compounds. The

^{1}H NMR spectrum showed only two vinylic protons at δ_H 5.09 and 5.43 (1H each, *brt*, J = 6.1 Hz, H_1-35 and H_1-18, respectively), which suggested the presence of two prenyl side chains. The location of the first prenyl side chain at C-5 was suggested by correlations from C-4, C-5, C-6 and C-18 to the methylene H_2-17 (δ_H 2.78) in the HMBC spectrum. In the COSY spectrum, correlations between protons at δ_H 1.14/3.32, 1.67, 1.78/1.98 and 5.09 (H_2-29, H-30, H_2-34 and H-35, respectively) confirmed the location of the second prenyl side chain at C-30. The presence of a 2,3-dihydroxy-3-methylbutyl side chain on position C-7 was suggested by the following evidence. In the HMBC spectrum, correlations were observed from the quaternary carbon C-26 (δ_C 73.4) and the methine carbon C-25 (δ_C 78.5) to the methyl protons H_3-27 and H_3-28 (3H, *s* each, δ_H 1.74 and 1.66, respectively). In addition, correlations between H-24, H-7 and H-25 in the COSY spectrum allowed to locate the side chain on position C-7. The chemical shifts of the carbons C-25 and C-26 observed in the ^{13}C NMR spectrum suggested the presence of two hydroxy groups. The chemical shifts of the *gem*-dimethyl group C-22 and C-23 (δ_C 27.3 and 23.2, respectively) suggested an axial position of the side chain. This was confirmed by the correlations between H-7$_{eq}$ (δ_H 2.29), H-8$_{eq}$ (δ_H 2.58) and Me-22 (δ_H 1.08) observed in the NOESY spectrum. On the other hand, despite the presence of correlations between H-25 (δ_H 3.80) with Me-27, Me-28, H-24$_\alpha$ (δ_H 2.80), H-7$_{eq}$ (δ_H 2.29) and H-8$_{eq}$ (δ_H 2.58), we cannot resolve the relative configuration at C-25. Compound **7** was named coccinone D.

Compound **8,** which was isolated as a brown oil, has an $[\alpha]_D$ value of −70° (*c* 0.3, $CHCl_3$). The HREIMS presented the same ion peak as compound **7** at *m*/*z* 659.3560 $[M+Na]^+$ (calc. 659.3560) indicating the same formula $C_{38}H_{52}O_8$. The ^{1}H and ^{13}C spectra of both compounds **7** and **8** were very similar except for the "south region" implying carbons C-6, C-7, C-8, C-24 and C-25. A close examination of the 1D and 2D NMR data allowed us to deduce that compound **8** possessed the same skeleton as compound **7** with an axial 2,3-dihydroxy-3-methylbutyl side chain at position C-7. In contrast with compound **7**, other correlations were observed between the protons H-25 and Me-27, Me28, H-24$_\beta$ (δ_H 2.25), H-7$_{eq}$ (δ_H 2.24) and Me-23$_{eq}$ (δ_H 1.36) in the NOESY spectrum. Compound **8** was named coccinone E.

The preparation of Mosher esters should confirm the absolute configuration at C-25. Unfortunately the amount of both compounds was not sufficient to obtain the desired products. To our knowledge, it is the first report of a 2,3-dihydroxy-3-methylbutyl side chain located at C-7 in the type B PPAPs.

Compound **9** was isolated as yellow crystals. The HREIMS $[M+Na]^+$ *m*/*z* 625.3494 indicated a molecular formula of $C_{38}H_{50}O_6$ (calc. 625.3505). The ^{1}H and ^{13}C NMR (Table 1) data and the value of $[\alpha]_D$ −135° (*c* 1.0, $CHCl_3$) were identical with those of garcinol isolated from *Garcinia xanthochymus* (Krishnamurthy et al., 1981; Rao and Venkatswamy, 1980). A complete set of data obtained from the X-ray crystallographic analysis is available on Cambridge Crystallographic Data Centre (see supplementary material section).

Compound **10**, a yellow oil, exhibited $[\alpha]_d$ −86° (c 0.8, $CHCl_3$). The HREIMS indicated an ion peak at *m*/*z* 625.3517, which suggested the same molecular formula as **9** ($C_{38}H_{50}O_6$). The ^{1}H and ^{13}C NMR spectral data were almost identical to those of **9**. In the ^{13}C NMR, only the signals of carbons C-7, C-8, C-22, C-23 and C-24 shifted slightly suggesting a β-orientation for the prenyl side chain located at position **7**. This was confirmed by nOe data (see discussion for compound **3**). Compound **10** was deduced to be the 7-*epi*-garcinol.

Compound **11** was purified as a yellow oil, the HREIMS indicated a molecular formula of $C_{38}H_{50}O_5$ $[M+Na]^+$ *m*/*z* = 609.3538 (calc. 609,3556). Examination of the 1D and 2D NMR data suggested the presence of a disubstituted benzene ring instead of the catechol previously observed for compounds **1–9**. All other data are fully similar to those of garcinol (**9**) indicating that compound **11** was 14-deoxygarcinol.

Compound **12**, a yellow oil, exhibited an $[\alpha]_D$ −32° (*c* 0.8, $CHCl_3$). The HREIMS indicated the same molecular formula as garcinol ($C_{38}H_{50}O_6$). The ^{1}H and ^{13}C NMR spectral data (Table 1) were very close to those of garcinol. The COSY and HMBC spectrum suggested that the ω-isolavandulyle side chain was replaced by a 1,3,3-trimethylcyclohex-1-enyl (Fig. 3). In the HMBC spectrum, a *gem*-dimethyl group is deduced from the correlations observed from the quaternary carbon C-33 (δ_C 32.3), the methylene carbon C-34 (δ_C 34.5) and the methine C-32 (δ_C 134.0) to the methyl protons H_3-37 and H_3-38. Correlations were also observed from the two methines C-32 and C-30 (δ_C 35.9) to the methyl protons H_3-36 (δ_H 2.11) and from the methylene carbon C-35 (δ_C 25.5) and the quaternary carbons C-1, C-2 and C-31 (δ_C 62.2, 192.2 and 137.4, respectively) to the methylene protons H_2-29, which confirmed the presence of the C10 moiety located at position C-1. In the NOESY spectrum, the correlation observed between H-30 (δ_H 2.66) and H-34$_{ax}$ (δ_H 1.62) suggested a β-orientation for H-30. This relative configuration was confirmed by correlations between H-34$_{ax}$ and H_3-38 from one side and H_3-37 with H-35$_{ax}$ (δ_H 1.59) on the other side. These data, together with other results of 2D NMR analysis confirmed the structure of compound **12**, which was named coccinone F.

Compound **13** was isolated as a yellow oil. The molecular formula of $C_{38}H_{50}O_6$ has been deduced from the ion peak at $[M+Na]^+$ *m*/*z* = 625.3524 (calc. 625.3505) obtained by HREIMS. The ^{1}H and ^{13}C NMR spectra suggested the presence of a 1,2,4-trisubstituted benzene ring (Table 1). In addition, resonances for eight methyl at δ_H 1.09 (3 H, *s*, Me-22), 1.41 (3 H, *s*, Me-23), 1.94 (3 H, *s*, Me-33), 1.65 (3 H, *s*, Me-38), 1.61 (3 H, *s*, Me-27), 1.56 (3 H, *s*, Me-28), 1.79 (3 H, *s*, Me-21) and 1.88 (3H, *s*, Me-20), two trisubstituted double bonds at δ_H 5.67 (1 H, *m*, H-18) and 5.00 (1 H, *m*, H-25), two olefinic exomethylenes at δ_H 4.81 and 4.87 (2 H,

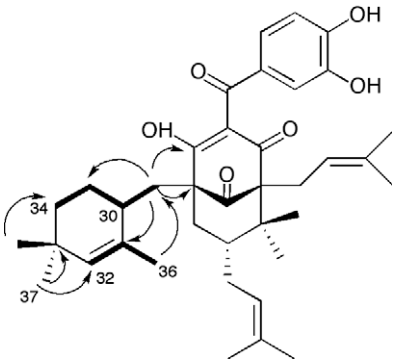


Fig. 3. Key HMBC (arrows) and COSY (bold) correlations for **12**.

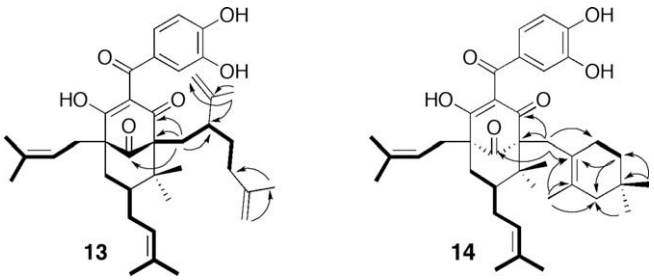


Fig. 4. Key HMBC (arrows) and COSY (bold) correlations for **13** and **14**.

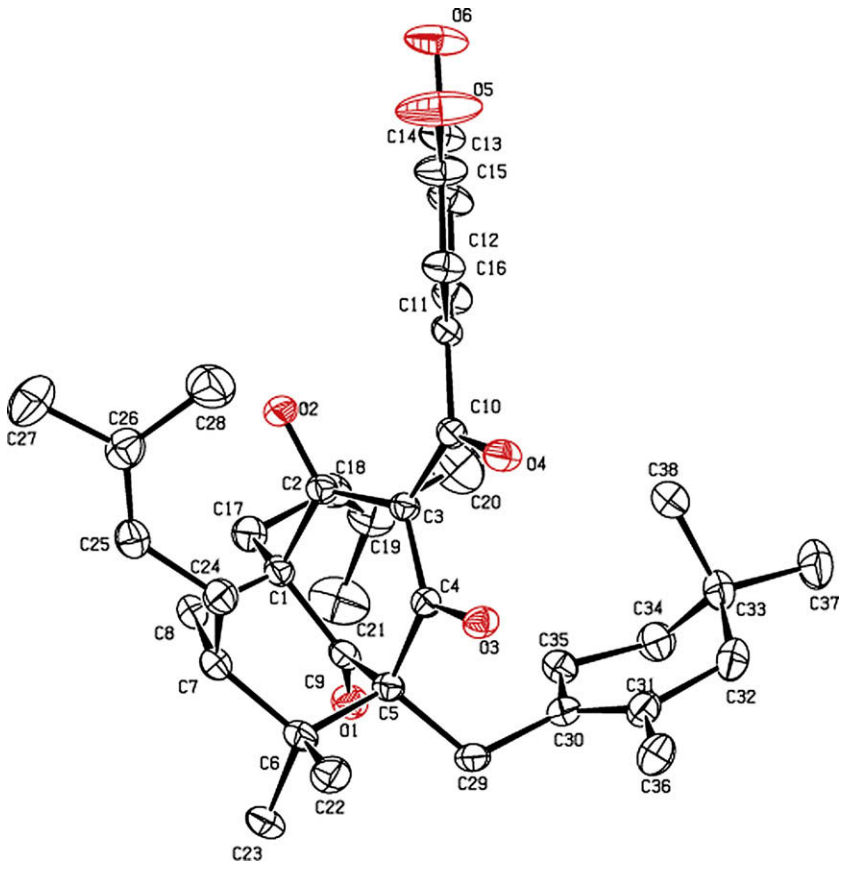


Fig. 5. X-ray crystal structure for compound **14**.

brs, H-32) and δ_H 4.72 (1 H, *m*, H-37), 4.76 (1 H, *m*, H-37), and seven allylic protons at δ_H 2.1-2.9 (10 H, *m*) were observed (Table 1). These data associated with the correlations observed in the COSY and the HMBC spectra, suggested the presence of a ω-isolavandulyl and two prenyl side chains (Fig. 4). The location of the prenyl side chain, α-oriented, at position C-7 was deduced from the chemical shifts of the *gem*-dimethyl at δ_C 27.4 and 23.8 (H_3-22 and H_3-23, respectively) observed in the ^{13}C NMR spectra. The location of the ω-isolavandulyl side chain at position C-5 was deduced from the correlations from the carbons C-4, C-5, C-9 and C-30 to the methylene protons H_2-29 observed in the HMBC spectrum (Fig. 4). The relative configuration of the methine C-30 remains unknown. Compound **13** was named coccinone G.

Compound **14** was purified as a yellow crystal. The ion peak in HREIMS $[M+Na]^+$ *m*/*z* 625.3499 indicated the same molecular formula as compound **13**, $C_{38}H_{50}O_6Na$ (calc. 625.3505). The ^1H and ^{13}C NMR spectra showed the same benzoylphloroglucinol skeleton as the previous one (Table 1). The 2D NMR data were very similar with those of compound **11**, except for the ω-isolavandulyl side chain, which is replaced by a 2,4,4-trimethylcyclohex-1-enyl moiety in the case of compound **14**. In the HMBC spectrum, an additional *gem*-dimethyl group is deduced from the correlations observed from the quaternary carbon C-33 (δ_C 29.2) and the methylenes C-32 and C-34 (δ_C 47.4 and 37.0, respectively) to the methyl protons H_3-37 and H_3-38. In addition, correlations from the carbons C-30 (δ_C 128.8), C-31 (δ_C 125.9) and C-32 (δ_C 47.4) to the methyl protons H_3-36, and from the carbons C-32 and C-35 (δ_C 29.3) to the methylene protons H_2-34 confirmed the presence of a 2,4,4-trimethylcyclohex-1-enyl moiety. Finally, correlations from the carbons C-4, C-5, C-9, C-30 and C-35 to the methylene protons H_2-29 indicated that the C10 moiety is attached at position C-5 (Fig. 4). The relative configuration of compound **14** was confirmed by the X-ray crystallographic analysis (Fig. 5, see supplementary material section). Compound **14** was named coccinone H. To our knowledge, coccinones G and H are the first example of PPAPs having a C10 side chain attached at C-5.

2.1. Biological activity

The antiplasmodial activity of compounds **1**–**14** was evaluated by their ability to inhibit the growth of *P. falciparum* following

Table 2
IC50 value of compounds **1**–**14** tested against FcB1 strain of *P. falciparum* and MRC5 cells.

Compound	IC50 *P.f.* FcB1 (μM) ± SD (*n* = 3)	IC50 MRC 5 (μM) (*n* = 3)
Isogarcinol (**1**)	3.5 ± 1.1	3.5 ± 0.4
Cycloxanthochymol (**2**)	2.1 ± 0.5	3.6 ± 0.6
7-*epi*-isogarcinol (**3**)	5.1 ± 1.3	2.3 ± 0.5
Coccinone A (**4**)	4.3 ± 0.5	3.3 ± 0.8
Coccinone B (**5**)	5.5 ± 0.4	11.7 ± 0.6
Coccinone C (**6**)	9.0 ± 1.2	9.3 ± 1.1
Coccinone D (**7**)	7.0 ± 0.9	10.9 ± 0.3
Coccinone E (**8**)	4.9 ± 0.7	9.1 ± 1.3
Garcinol (**9**)	12.6 ± 4.8	11.0 ± 0.2
7-*epi*-garcinol (**10**)	10.1 ± 4.6	11.3 ± 0.6
14-deoxygarcinol (**11**)	37.2 ± 9.7	19.1 ± 0.2
Coccinone F (**12**)	17.0 ± 9.4	21.3 ± 2.8
Coccinone G (**13**)	19.2 ± 5.9	11.5 ± 0.4
Coccinone H (**14**)	16.6 ± 6.1	10.5 ± 0.6
Chloroquine	0.078 ± 0.006	25 ± 3.8
Taxotere	0.010 ± 0.005	31.5 ± 4.5

the method of Desjardins et al. (1979). The anti-plasmodial activity against the chloroquine-resistant strain of *P. falciparum* FcB1 and the cytotoxicity on the human MRC-5 cell line are summarized in Table 2. The results demonstrated that compounds **1**–**14** exhibited potent antiplasmodial activity. Compounds **1**–**8** have shown the best IC50s between 3.3 and 9.0 μM, indicating that the benzophenones having a tetrahydropyrane ring are the most potent compounds. In contrast, compounds **9**–**14** showed moderate antiplasmodial activities with IC50s over 10 μM. Moreover, hydroxyl position on the benzoyl groups seems to play an important role for the antiplasmodial activity regarding the IC50 values for garcinol and dehydroxygarcinol. The analysis of the IC50s obtained on the MRC-5 cell line indicated that compounds having a hydroxylated side chain (**5**–**8**) exhibited a lowest cytotoxicity. Further mechanistic and structure-activity relationship studies have to be done in order to better understand the structural features needed for developing new anti-malarial drugs based on PPAPs.

3. Conclusions

Phytochemical investigation of *M. coccinea* led to the isolation of eleven new PPAPs derivatives. This is the first report on the isolation of PPAPs with a C_{10} side chain attached at C5. Furthermore, *in vitro* antiplasmodial assays reveal that PPAPs having a furan ring fused to the phloroglucinol moiety exhibited potent biological activity. Biological and phytochemical investigations of other tropical Clusiaceae species are underway in order to confirm this hypothesis.

4. Experimental

4.1. General

The NMR spectra were recorded with a Bruker 500 MHz (advance 500) spectrometer with pyridine-d_5 as solvent. ESIMS were obtained on a Navigator mass Thermoquest. HRESIMS were run on a MALDI-TOF spectrometer (Voyager-De STR; Perspective Biosystems). Kromasil analytic, semi-preparative and preparative C-18 columns (250 × 4.5 mm; 250 × 10 mm and 250 × 21.2 mm I.D, 5 μm Thermo®) were used for preparative HPLC separations using a Waters autopurification system equipped with a binary pump (Waters 2525), a UV–vis diode array detector (190–600 nm, Waters 2996) and a PL-ELS 1000 ELSD detector Polymer Laboratory. IR spectra were obtained on a Nicolet FTIR 205 spectrophotometer. The UV spectra were recorded on a Perkin–Elmer Lambda 5 spectrophotometer. Specific rotation was obtained in $CHCL_3$ with a JASCO P-1010 polarimeter. Melting point for isolated crystals was done on a Büchi B-540. Silica gel 60 (35–70 μm) and analytical TLC plates (Si gel 60 F 254) were purchased from SDS (France). Polyamide DC 6 and polyamide cartridge was purchased from Macherey–Nagel (Chromabond PA®, 1 g). All other chemicals and solvents were of analytical grade and purchased from SDS (France).

4.2. Plant material

Trunk barks, roots barks, fruits, leaves and latex of *M. coccinea* were collected in the dense rain forest of French Guyana by one of us (G.M). Latex was collected by scoring trunk tissue and then scraping the latex as it oozed from the wound. This yellow resin was transferred to Teflon-stoppered glass vials. A herbarium sample (GM-008) was deposited in the Guyane Herbarium of Cayenne, and identified by M-F. Prevost (Institut de Recherche pour le Développement).

4.3. Extraction and isolation procedures

Trunk bark, roots bark, fruits, and leaves (20 g each) were extracted three times using the automatic extractor Dionex® ASE 300 with EtOAc (3 × 100 ml) followed by MeOH (3 × 100 ml) at 40 °C and 100 bar. Each extract were concentrated *in vacuo* at 40 °C to yield 2.69 g (trunk bark, EtOAc), 1.48 g (trunk bark, MeOH), 5.14 g (roots bark, EtOAc) 0.96 g (trunk bark, MeOH), 2.57 g (fruits, EtOAc), 5.23 g (fruits, MeOH), 1.71 g (leaves, EtOAc) and 1.13 g (leaves, MeOH). Aliquot of 100 mg of each extract and yellow resin were dissolved in MeOH and then were filtered on polyamide cartridge before being tested at 10 μg/ml on *P. falciparum* FcB1 strain. The resin-filtered extract was then fractionated on a semi-preparative C-18 column according to a standardized method previously detailed (Bousserouel et al., 2005). Fraction 6 was shown to possess antiplasmodial activity.

A larger amount of latex (3.51 g) was dissolved in MeOH and filtered on polyamide DC 6 to remove bark residue and tannins. The filtered extract was dried under vacuum at 40 °C to give a yellow gummy residue (2.35 g), which was subjected to silica gel chromatography using cyclohexane, mixtures of cyclohexane–ethyl acetate (9:1–1:9) and ethyl acetate as eluents. According to their TLC profiles, 18 fractions (F1–F18) were obtained. Analytical HPLC was used to compare these fractions with the active fraction 6 previously identified and allowed us to identify those containing the active compounds. Fraction 12 (740 mg) was submitted to a preparative C-18 column using an isocratic mobile phase consisting of MeCN–H_2O 90:10 + 0.1% formic acid, leading to the isolation of 7-*epi*-isogarcinol **3** (109 mg; w/w 4.6%), isogarcinol **1** (332 mg; w/w 14.1%), cycloxanthochymol **2** (11 mg; w/w 0.5%), coccinone A **4** (18 mg; w/w 0.76%), 14-dehydroxygarcinol **11** (4 mg; w/w 0.17%) garcinol **9** (134 mg; w/w 5.7%), 7-*epi*-garcinol **10** (20 mg; w/w 0.85%) coccinone F **12** (16 mg; w/w 0.68%), coccinone G **13** (61 mg; w/w 2.6%) and coccinone H **14** (30 mg; w/w 1.4%), with retention times of 4.7, 5.2, 5.5, 6.0, 9.3, 9.4, 16.5, 13.5 and 12.6 min, respectively. Fraction 13 (100 mg) using an isocratic system consisting of MeCN–H_2O 70:30 + 0.1% formic acid on the same column, afforded coccinone B **5** (4 mg; 0.17%), coccinone C **6** (3 mg; 0.12%), coccinone D **7** (3 mg; 0.12%) and coccinone E **8** (4 mg, 0.17%) with retention times of 6.5, 7.2, 10.6 and 11.4 min, respectively.

4.3.1. Isogarcinol ***1***

Brown crystal, HR-EIMS $[M+Na]^+$ *m/z* 625.3499, $C_{38}H_{50}O_6Na$ requires 625.3505; $[\alpha]_D$ −158° (*c* 1, $CHCl_3$); mp 251 °C; UV λ_{max} ($\log\varepsilon$): 317 (3.82), 277 (4.14), 233 (4.07); IR ν_{max} (ns) 3290, 2920, 2850, 1730, 1670, 1590, 1520, 1440, 1370, 1290, 1170 cm^{-1}; ^1H NMR and ^{13}C NMR (Table 1). CCDC (Deposit No: 689070) contains the supplementary crystallographic data (Krishnamurthy et al., 1982).

4.3.2. 13-*O*,14-*O*-Bis(4-bromobenzoylsulfonyl)-isogarcinol ***1′***

At a pyridine solution of isogarcinol (32 mg in 350 μl, 0.053 mmol) were added DMAP (3 mg, 0.024 mmol) and 4-bromobenzenesulfonyl chloride (44 mg, 0.17 mmol). After stirring for 4 h at 0 °C, the solution was extracted with diisopropyl ether and washed with HCl 1 N. Four products were separated using preparative TLC plates (cyclohexane/EtOAc 7:3): starting materials (17 mg); 13-*O*-(4-bromobenzoylsulfonyl)-isogarcinol (5 mg); 14-*O*-(4-bromobenzoylsulfonyl)-isogarcinol (9 mg); and the desired product 13-*O*,14-*O*-bis(4-bromobenzoylsulfonyl)-isogarcinol (7 mg). The last was crystallized under MeOH/$H_2$0 95:5. mp 138 °C; CCDC (Deposit No: 688119) contains the supplementary crystallographic data.

4.3.3. Cycloxanthochymol ***2***

Yellow oil, $[\alpha]_D$ +112° (*c* 1, $CHCl_3$), HR-EIMS $[M+Na]^+$ *m/z* 625.3512, $C_{38}H_{50}O_6Na$ requires 625.3505; UV λ_{max} ($\log\varepsilon$): 315 (3.87), 277 (4.18), 229 (4.16); IR ν_{max} 3330, 2920, 1730, 1650, 1590, 1520, 1440, 1370, 1290, 1190, 1120 cm^{-1}.

4.3.4. 7-Epi-isogarcinol ***3***

Brown powder, $[\alpha]_D$ −158° (*c* 1.0, $CHCl_3$); HR-EIMS $[M+Na]^+$ *m/z* 625.3519, $C_{38}H_{50}O_6Na$ requires 625.3505; UV λ_{max} ($\log\varepsilon$): 319 (3.85), 276 (4.15), 233 (4.12); IR ν_{max} 3320, 2970, 2930, 1730, 1650, 1590, 1520, 1440, 1370, 1290, 1170 cm^{-1}; ^1H NMR and ^{13}C NMR (Table 1).

4.3.5. Coccinone A ***4***

Brown oil, $[\alpha]_D$ +28° (*c* 1.0, $CHCl_3$); HR-EIMS $[M+Na]^+$ *m/z* 625.3521, $C_{38}H_{50}O_6Na$ requires 625.3505; UV λ_{max} ($\log\varepsilon$): 276 (3.23), 232 (3.19); IR ν_{max} 3280, 2920, 2850, 1730, 1650, 1550, 1440, 1370, 1290, 1170 cm^{-1}; ^1H NMR and ^{13}C NMR (Table 1).

4.3.6. Coccinone B ***5***

Brown oil, $[\alpha]_D$ −55° (*c* 0.3, $CHCl_3$); HR-EIMS $[M+Na]^+$ *m/z* 641.3446, $C_{38}H_{50}O_7Na$ requires calc. 641.3454; UV λ_{max} ($\log\varepsilon$): 316 (3.71), 278 (4.03), 233 (3.99); IR ν_{max} 3330, 2930, 1730, 1650, 1590, 1440, 1350, 1290, 1120 cm^{-1}; ^1H NMR and ^{13}C NMR (Table 1).

4.3.7. Coccinone C ***6***

Brown oil, $[\alpha]_D$ −60° (*c* 0.2, $CHCl_3$); HR-EIMS $[M+Na]^+$ *m/z* 643.3619, $C_{38}H_{52}O_7$ (calc. 643.3611); UV λ_{max} ($\log\varepsilon$): 316 (3.83), 277 (4.15), 233 (4.11); IR ν_{max} 3350, 2930, 1730, 1640, 1590, 1440, 1350, 1290, 1120 cm^{-1}; ^1H NMR and ^{13}C NMR (Table 1).

4.3.8. Coccinone D ***7***

Brown oil, $[\alpha]_d$ −76° (*c* 0.4, $CHCl_3$); HR-EIMS $[M+Na]^+$ *m/z* 659.3563, $C_{38}H_{52}O_8$ (calc. 659.3560); UV λ_{max} ($\log\varepsilon$): 313 (3.83), 278 (4.14), 233 (4.08); IR ν_{max} 3410, 2930, 1730, 1590, 1440, 1350, 1290, 1120 cm^{-1}; ^1H NMR and ^{13}C NMR (Table 1).

4.3.9. Coccinone E ***8***

Brown oil, $[\alpha]_D$ −70° (*c* 0.3, $CHCl_3$); HR-EIMS $[M+Na]^+$ *m/z* 659.3560, $C_{38}H_{52}O_8$ (calc. 659.3560); UV λ_{max} ($\log\varepsilon$): 313 (3.82), 278 (4.14), 233 (4.08); IR ν_{max} 3430, 2930, 1730, 1590, 1440, 1350, 1290, 1120 cm^{-1}; ^1H NMR and ^{13}C NMR (Table 1).

4.3.10. Garcinol ***9***

Yellow crystal, $[\alpha]_D$ −135° (*c* 1.0, $CHCl_3$); HR-EIMS $[M+Na]^+$ *m/z* 625.3494, $C_{38}H_{50}O_6Na$ requires 625.3505; mp 132 °C; UV λ_{max} ($\log\varepsilon$): 279 (4.18), 232 (4.04); IR ν_{max} 3300, 2920, 1720, 1640, 1590, 1440, 1370, 1290, 1190 cm^{-1}; ^1H NMR and ^{13}C NMR are consistent with published data (Krishnamurthy et al., 1981; Rao and Venkatswamy, 1980; Sahu et al., 1989). CCDC (Deposit No: 688120) contains the supplementary crystallographic data (see supplementary material section).

4.3.11. 7-epi-garcinol ***10***

Yellow oil, $[\alpha]_D$ −86° (*c* 0.8, $CHCl_3$); HR-EIMS $[M+Na]^+$ *m/z* 625.3517, $C_{38}H_{50}O_6Na$ requires 625.3505; UV λ_{max} ($\log\varepsilon$): 280 (4.32), 230 (4.17); IR ν_{max} 3300, 2920, 1720, 1640, 1520, 1440, 1375, 1290, 1200 cm^{-1}; ^1H NMR and ^{13}C NMR (Table 1).

4.3.12. 14-deoxygarcinol ***11***

Yellow oil, $[\alpha]_D$ −42° (*c* 0.3, $CHCl_3$); HR-EIMS $[M+Na]^+$ *m/z* 609.3538, $C_{38}H_{50}O_5$ (calc. 609.3556); UV λ_{max} ($\log\varepsilon$): 283 (4.05), 258 (4.03); IR ν_{max} 3290, 2920, 1720, 1640, 1560, 1440, 1370, 1290, 1190 cm^{-1}; ^1H NMR and ^{13}C NMR (Table 1).

4.3.13. Coccinone F ***12***

Yellow oil, $[\alpha]_D$ −32° (*c* 0.7, $CHCl_3$), HR-EIMS $[M+Na]^+$ *m/z* 625.3490, $C_{38}H_{50}O_6Na$ requires 625.3505; UV λ_{max} ($\log\varepsilon$): 280 (4.09), 229 (4.03); IR ν_{max} 3300, 2920, 1720, 1660, 1590, 1440, 1375, 1290, 1200 cm^{-1}; ^1H NMR and ^{13}C NMR (Table 1).

4.3.14. Coccinone G ***13***

Yellow oil, $[\alpha]_D$ −16° (*c* 1.0, $CHCl_3$); HR-EIMS $(M+Na^+)$ *m/z* 625.3524, $C_{38}H_{50}O_6Na$ requires 625.3505; UV λ_{max} ($\log\varepsilon$): 279 (4.18), 233 (4.10); IR ν_{max} 3280, 2920, 2850, 1730, 1640, 1600, 1440, 1380, 1290, 1180 cm^{-1}; ^1H NMR and ^{13}C NMR (Table 1).

4.3.15. Coccinone H ***14***

Yellow crystal, $[\alpha]_D$ +2° (*c* 1.0, $CHCl_3$); HR-EIMS $(M+Na^+)$ *m/z* 625.3499, $C_{38}H_{50}O_6Na$ requires 625.3505; mp 169 °C; UV λ_{max} ($\log\varepsilon$): 282 (3.96), 232 (3.88); IR ν_{max} 3310, 2940, 2870, 1720, 1590, 1440, 1380, 1290, 1190 cm^{-1}; ^1H NMR and ^{13}C NMR (Table 1).

4.3.16. The crystal data of ***14***

$C_{38}H_{50}O_6$, $M = 602.78$, monoclinic, space group C2, $a = 31.987(5)$, $b = 12.928(2)$, $c = 8.847(2)$ Å, $\beta = 98.54(2)°$, $V = 3617.9(11)$ $Å^3$, $Z = 4$, $D_c = 1.107$ g cm^{-3}, μ(Mo–Ka) = 0.073 mm^{-1}, $F(000) = 1304$, prismatic crystal, colorless, size = 0.20 × 0.15 × 0.08 mm, 21043 reflections measured [$R_{int} = 0.022$], 3376 unique, $wR_2 = 0.115$ for all data, conventional $R = 0.044$ [$(D/s)_{max} = 0.001$] on *F*-values of 2587 reflections with $I > 2\ s(I)$, GooF = 1.039 for all data and 407 parameters. Unit cell determination and intensity data collection ($q = 25.1°$) were performed on an Enraf–Nonius KAPPA CCD diffractometer at 293(2) K. Structure solution by direct methods and refinement by full-matrix least-squares on F^2. Programs: COLLECT, (Nonius, 1997–2000) HKL2000 (Otwinowski and Minor, 1997), SHELX97 (Sheldrick, 2008). CCDC [Deposit No: 688118] contains the supplementary crystallographic data for this paper (see supplementary material section).

4.4. Biological activities

Extracts were tested against the chloroquine-resistant FcB1/Colombia strain of *P. falciparum* (Frappier et al., 1996). *P. falciparum* was maintained continuously *in vitro* in human erythrocytes according to Trager and Jensen (2005). The antiplasmodial activity was determined according to Desjardins et al. (1979). The extracts were dissolved in dimethylsulfoxide (DMSO) and tested at a concentration of 10 μg/ml. Compounds showing significant inhibition rates were submitted to serial dilutions with culture medium before being added to asynchronous parasite cultures (1% parasitemia and 1% final hematocrite) in 96-well microplates for 24 h at 37 °C. A concentration of 0.5 μCi of [^{3}H] hypoxanthine was then added to each well, and parasites were maintained for an additional 24 h. The growth inhibition for each compound concentration was determined by comparing the radioactivity incorporated in the treated culture with that in the control culture maintained on the same plate. The concentrations causing 50% inhibition of parasite growth (IC50) were calculated from the drug concentration-response curves.

The human diploid embryonic lung cells MRC-5 were seeded into 96-well microplates at 2000 cells per well. The cytotoxicity assays were performed according to a published procedure (Tempete et al., 1995). Taxotere® was used as a control compound.

Supplementary material

CCDC 689070 (**1**), CCDC 688119 (**1′**), CCDC 688120 (**9**) and CCDC 688118 (**14**) contain the supplementary crystallographic data for this paper. These data can be obtained free of charge via http://

www.ccdc.cam.ac.uk/conts/retrieving.html (or from the CCDC, 12 Union Road, Cambridge CB2 1EZ, UK; fax: +44 1223 336033; e-mail: deposit@ccdc.cam.ac.uk).

Acknowledgements

This works was supported by an ICSN–CNRS grant to one of us (G.M.). We are very grateful to Bernard Delpech and Philippe Nuhant (ICSN) for their helpful advises, Marie-Françoise Prevost (IRD) for her assistance in the identification and the plant collection, Marie-Thérèse Martin for her assistance in the structural determination and Geneviève Aubert (ICSN) who performed the cytotoxic assays on MRC5 cells.

References

Baggett, S., Protiva, P., Mazzola, E.P., Yang, H., Ressler, E.T., Basile, M.J., Weinstein, I.B., Kennelly, E.J., 2005. Bioactive benzophenones from *Garcinia xanthochymus* Fruits. J. Nat. Prod. 68, 354–360.

Blount, J.F., Williams, T.H., 1976. Revised structure of xanthochymol. Tetrahedron Lett. 34, 2921–2924.

Bousseroue, H., Litaudon, M., Morleo, B., Martin, M.T., Thoison, O., Nosjean, O., Boutin, J.A., Renard, P., Sevenet, T., 2005. New biologically active linear triterpenes from the bark of three new-caledonian *Cupaniopsis* species. Tetrahedron 61, 845–851.

Ciochina, R., Grossman, R.B., 2006. Polycyclic polyprenylated acylphloroglucinols. Chem. Rev. 106, 3963–3986.

Cuesta-Rubio, O., Piccinelli, A.L., Rastrelli, L., 2005. Chemistry and biological activity of polyisoprenylated benzophenone derivatives. In: Atta-ur-Rahman (Ed.), Studies in Natural Products Chemistry, vol. 32. Elsevier, Amsterdam, pp. 671–720.

Dias, J.P., Gottlieb, O.R., Lins Mesquita, A.A., 1974. Marupone, a benzophenone from *Moronobea pulchra*. Phytochemistry 13, 1953–1955.

Desjardins, R.E., Canfield, C.J., Haynes, J.D., Chulay, J.D., 1979. Quantitative assessment of antimalarial activity in vitro by a semiautomated microdilution technique. Antimicrob. Agents Chemother. 16, 710–718.

Dreyer, D.L., 1974. Xanthochymol from *Clusia rosea* (guttiferae). Phytochemistry 13, 2883–2884.

Frappier, F., Jossang, A., Soudon, J., Calvo, F., Rasoanaivo, P., Ratsimamanga-Urverg, S., Saez, J., Schrevel, J., Grellier, P., 1996. Bisbenzylisoquinolines as modulators of chloroquine resistance in *Plasmodium falciparum* and multidrug resistance in tumor cells. Antimicrob. Agents Chemother. 40, 1476–1481.

Fuller, R.W., Blunt, J.W., Boswell, J.L., Cardellina, J.H., Boyd, M.R., Guttiferone, F., 1999. The first prenylated benzophenone from *Allanblackia stuhlmannii*. J. Nat. Prod. 62, 130–132.

Gay, F., Ciceron, L., Litaudon, M., Bustos, D., Astagneau, P., Diquet, B., Danis, B., Gentilini, M., 1994. In vitro resistance of *Plasmodium falciparum* to qinghaosu derivatives in West Africa. The Lancet 343, 350–351.

Grenand, P., Moretti, C., Jacquemin, H., Prévost, MF., 2004. Pharmacopées Traditionnelles en Guyane. Créoles, Palikur, Wayãpi.

Gustafson, K.R., Blunt, J.W., Munro, M.H.G., Fuller, R.W., McKee, T.C., Cardellina, I.I., John, H., McMahon, J.B., Cragg, G.M., Boyd, M.R., 1992. The guttiferones, HIV-inhibitory benzophenones from *Symphonia globulifera*, *Garcinia livingstonei*, *Garcinia ovalifolia* and *Clusia rosea*. Tetrahedron 48, 10093–10102.

Hamed, W., Brajeul, S., Mahuteau-Betzer, F., Thoison, O., Mons, S., Delpech, B., Hung, N.V., Sevenet, T., Marazano, C., Oblongifolins, A.-D., 2006. Polyprenylated benzoylphloroglucinol derivatives from *Garcinia oblongifolia*. J. Nat. Prod. 69, 774–777.

Iinuma, M., Tosa, H., Tanaka, T., Kanamaru, S., Asai, F., Kobayashi, Y., Miyauchi, K., Shimano, R., 1996. Antibacterial activity of some *Garcinia* benzophenone derivatives against methicillin-resistant *Staphylococcus aureus*. Biol. Pharm. Bull. 19, 311–314.

Jambou, R., Legrand, E., Niang, M., Khim, N., Lim, P., Volney, B., Ekala, M.T., Bouchier, C., Esterre, P., Fandeur, T., Mercereau-Puijalon, O., 2005. Resistance of *Plasmodium falciparum* field isolates to in vitro artemether and point mutations of the SERCA-type PfATPase6. The Lancet 366, 1960–1963.

Krishnamurthy, N., Lewis, Y.S., Ravindranath, B., 1981. On the structures of garcinol, isogarcinol and camboginol. Tetrahedron Lett. 22, 793–796.

Krishnamurthy, N., Ravindranath, B., Row, T.N.G., Venkatesan, K., 1982. Crystal and molecular structure of isogarcinol. Tetrahedron Lett. 23, 2233–2236.

Laufer, M.K., Plowe, C.V., 2004. Withdrawing antimalarial drugs: impact on parasite resistance and implications for malaria treatment policies. Drug Resist. Update 7, 279–288.

Ngouela, S., Lenta, B.N., Noungoue, D.T., Ngoupayo, J., Boyom, F.F., Tsamo, E., Gut, J., Rosenthal, P.J., Connolly, J.D., 2006. Anti-plasmodial and antioxidant activities of constituents of the seed shells of *Symphonia globulifera* Linn f. Phytochemistry 67, 302–306.

Nonius, COLLECT software, Nonius B.V., Delft, The Netherlands, 1997–2000.

Otwinowski, Z., Minor, W., 1997. Processing of X-ray diffraction data collected in oscillation mode. Meth. Enzymol. 276, 307–326.

Piccinelli, A.L., Cuesta-Rubio, O., Chica, M.B., Mahmood, N., Pagano, B., Pavone, M., Barone, V., Rastrelli, L., 2005. Structural revision of clusianone and 7-*epi*-clusianone and anti-HIV activity of polyisoprenylated benzophenones. Tetrahedron 61, 8206–8211.

Rao, A.V.R., Venkatswamy, G., 1980. Xanthochymol and isoxanthochymol, two novel polyisoprenylated benzophenones from *Garcinia xanthochymus*. Ind. J. Chem. B 19B, 627–633.

Rogers, D., McConway, J.C., Ramnachandia, B., Ramadas, R., 1981. Ind. J. Chem. B 20B, 915–916.

Roux, D., Hadi, H.A., Thoret, S., Guenard, D., Thoison, O., Pais, M., Sevenet, T., 2000. Structure–activity relationship of polyisoprenyl benzophenones from *Garcinia pyrifera* on the tubulin/microtubule system. J. Nat. Prod. 63, 1070–1076.

Sahu, A., Das, B., Chaterjee, A., 1989. Polyisoprenylated benzophenones from *Garcinia pedunculata*. Phytochemistry 28, 1233–1235.

Sheldrick, G.M., 2008. A short history of SHELX. Acta Cryst. A 64, 112–122.

Teixeira, J.S.A.R., Cruz, F.G., 2005. Polyisoprenylated benzophenone derivatives from *Clusia obdeltifolia*. Tetrahedron Lett. 46, 2813–2816.

Tempete, C., Werner, G.H., Favre, F., Rojas, A., Langlois, N., 1995. In vitro cytostatic activity of 9-demethoxyporothramycin B. Eur. J. Med. Chem. 30, 647–650.

Trager, W., Jensen, J.B., 2005. Human malaria parasites in continuous culture 1976. J. Parasitol. 91, 484–486.

Williams, R.B., Hoch, J., Glass, T.E., Evans, R., Miller, J.S., Wisse, J.H., Kingston, D.G., 2003. A novel cytotoxic guttiferone analogue from *Garcinia macrophylla* from the Suriname rainforest. Planta Med. 69, 861–864.

Phytochemistry 71 (2010) 964–974

Contents lists available at ScienceDirect

Phytochemistry

journal homepage: www.elsevier.com/locate/phytochem

Antiplasmodial benzophenone derivatives from the root barks of *Symphonia globulifera* (Clusiaceae)

Guillaume Marti [a], Véronique Eparvier [b], Christian Moretti [c], Soizic Prado [d], Philippe Grellier [d], Nathalie Hue [a], Odile Thoison [a], Bernard Delpech [a], Françoise Guéritte [a], Marc Litaudon [a,*]

[a] Centre de Recherche de Gif, Institut de Chimie des Substances Naturelles, CNRS, 1, Avenue de la Terrasse, 91198 Gif-sur-Yvette, France
[b] UPS 2561, CNRS, 16, Avenue André Aron, 97300 Cayenne, Guyane Française, France
[c] Unité N°. 84 de l'IRD: Biodival, IRD, Technoparc, 5, Rue du Carbone 45072 ORLEANS Cedex 2, France
[d] Muséum National d'Histoire Naturelle, FRE 3206 CNRS/MNHN, 61-63, Rue Buffon, 75231 Paris Cedex 05, France

ARTICLE INFO

Article history:
Received 21 January 2010
Received in revised form 4 March 2010
Available online 29 March 2010

Keywords:
Symphonia globulifera
Clusiaceae
Benzophenone
Symphonone
Antiplasmodial
LC/ESI-MS

ABSTRACT

In an effort to find antimalarial drugs, a systematic *in vitro* evaluation on a chloroquine-resistant strain of *Plasmodium falciparum* (FcB1) was undertaken on sixty plant extracts collected in French Guiana. The ethyl acetate extract obtained from the root barks of *Symphonia globulifera* exhibited a strong antiplasmodial activity (97% at 10 μg/ml). The phytochemical investigation of this extract led to the isolation of nine polycyclic polyprenylated acylphloroglucinol (PPAPs) compounds and two oxidized derivatives. All compounds showed antiplasmodial activity with IC_{50}s ranged from 2.1 to 10.1 μM. A LC/ESI-MS^n study performed on polyprenylated benzophenones previously isolated from *Moronobea coccinea* provided a reliable method for their detection in the extract and structural elucidation.

1. Introduction

Malaria remains one of the most important infectious disease of the developing world causing 2–3 million deaths every year. Current clinical treatment involves the use of inexpensive antimalarial drugs such as chloroquine and their derivatives. Due to the rising prevalence of *Plasmodium falciparum* resistance to these drugs, the treatment of malaria is becoming increasingly difficult (Laufer and Plowe, 2004). Artemisinin in combination therapy recommended by the World Health Organization is currently the most effective treatment. However, there are accumulating evidences for increasing artemisinin resistance *in vitro* of *P. faciparum* isolates (Gay et al., 1994). Clinical resistance has been recently detected in Cambodia (Noedl et al., 2008). There is therefore an urgent need to diversify the antimalarial therapeutic arsenal. Most of the antimalarial drugs are natural product and natural product analogues (artemisinin derivatives) or possess a pharmacophore from natural products (quinoline ring). Their discovery has encouraged the possibility of finding new antimalarial agents from higher plants (Bourdy et al., 2008). In the continuation of our screening program for the discovery of new antimalarial compounds in French Guiana flora, the *Symphonia globulifera* L.f. (Clusiaceae) ethyl acetate root barks extract exhibited a strong antiplasmodial activity (97% of inhibitory growth at 10 μg/ml) and led us to investigate this extract.

The genus *Symphonia* consists of some 17 species distributed from South America to Africa and Madagascar. The species *S. globulifera* is a medium to tall tree with characteristic yellow latex and presents stout, stilt-like roots at its lower part. This species is also characterized by important morphological variations which seem to be dependant in its ecological distribution, mainly in the neotropic (Dick et al., 2003). Paleobotanical studies have established the common origin of *Symphonia* spp. to Madagascar Island (Germeraad et al., 1968). Phylogenetical analyses have demonstrated that colonization of the new world by this species has occurred in three distinguish events between −15 and −1.5 mya. This led to ecological diversification of South America species and increased their genetic and morphological diversity (Dick et al., 2003). *S. globulifera* grows mainly in flooded forests, but in our collection area, near the town of Sinnamary, our sample (Herbarium collection: Marie-Françoise Prévost CAY-3527) belongs to a population growing in terra firma forests and can thus be regarded as an ecotype of *S. globulifera* (M.-F. Prévost pers. comm.) or as a new species of *Symphonia* as thought by some botanists. The healing properties of the latex of *S. globulifera* are used

* Corresponding author. Tel.: +33 1 69 82 30 85; fax: +33 1 69 07 72 47.
E-mail address: marc.litaudon@icsn.cnrs-gif.fr (M. Litaudon).

0031-9422/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.phytochem.2010.03.008

traditionally for their effectiveness against dermatoses. The boiled and dried sap produces a wax used in tool manufacture and as general-purpose glue (Grenand et al., 2004).

Antimicrobial prenylated xanthones, glaboxanthones A–E (Nkengfack et al., 2002a,b), and benzophenones, guttiferones A–D, which have exhibited *in vitro* HIV-inhibitory activities (Gustafson et al., 1992) were isolated from the African *S. globulifera* roots extract. To our knowledge only a seed shell extract has been screened for its antiplasmodial activity and led to the isolation of guttiferone A and three prenylated xanthones, glaboxanthone, symphonin and globuliferin with IC_{50} in the ranges of 1.5–4.0 μM (Ngouela et al., 2006).

In a previous study, we have reported the isolation and characterization of several polycyclic polyprenylated acylphloroglucinol (PPAPs) from the latex of *Moronobea coccinea* Aubl. (Clusiaceae) (Marti et al., 2009). *In vitro* antiplasmodial assays on the *P. falciparum* chloroquine-resistant strain (FcB1) revealed IC_{50}s from 2.1 to 9.0 μM for PPAPs having a tetrahydropyran ring fused to the benzophenone moiety, and above 10 μM for PPAPs with an enolic β-diketo functionality at C2–C4. In this report, we describe the isolation, structure elucidation and biological activities of eleven new PPAPs (**2**–**8**, **10**–**11**) including two new oxidized derivatives (**12**–**13**) from *S. globulifera*.

2. Results and discussion

The ethyl acetate extract from the root barks (100 mg) was filtered on polyamide to remove tannins and the filtered extract (15 mg) was fractionated on a semi-preparative C-18 column to give nine fractions according to a standardized method (see Section 4) (Bousserouel et al., 2005). Biological assay on the *P. falciparum* resistant strain FcB1 allowed us to identify one active fraction (fraction 5, t_R from 34 to 42 min). In order to determine the main structural features of compounds contained in this fraction, we applied a rapid and sensitive LC/ESI-MS^2 method using a negative ionization mode, which was developed with the benzophenone derivatives previously isolated from *M. coccinea* latex (Marti et al., 2008). Briefly, hypotheses of fragmentation patterns are as follow (Fig. 1a–d). A major fragment at *m/z* 465 $[M-H-C_7H_4O_3]^-$ was shown for most of the polyprenylated benzophenones including 7-*epi*-isogarcinol (**1**), 7-*epi*-garcinol (**9**), isogarcinol (**14**), cycloxanthochymol (**15**), coccinone A (**16**) and garcinol (**21**) (compounds **1**, **9** and **14**–**25** were isolated from *M. coccinea*, Fig. 1S Supplementary data). Only 14-deoxygarcinol (**22**) failed to produce this ion because of the lack of hydroxyl group in *para* position on the benzoyl moiety. The departure of a neutral fragment of 136 amu ($C_7H_4O_3$) was explained by the α-cleavage of the bond adjacent to the C-10 carbonyl group (Fig. 1a). The corresponding ion was observed for all derivatives with an *ortho*-dihydroxybenzoyl group (coccinones A–H, **16**–**20**, **23**–**25**) (Marti et al., 2009). Interestingly, some of the benzophenones having the enolic β-dicarbonyl system showed an additional intense fragment at *m/z* 409 corresponding to the loss of aliphatic sides chains (Fig. 1b). The departure of the neutral fragments of 192 amu ($C_{14}H_{24}$) was shown for garcinol (**21**), 7-*epi*-garcinol (**9**), 14-deoxygarcinol (**22**) and coccinone F (**23**) but not for compounds having the tetrahydropyran ring fused to the benzophenone core (**14**–**20**). This is important insofar as cyclized derivatives are slightly more potent on *P. falciparum* growth inhibition (Marti et al., 2009). In addition, MS/MS spectra of compounds having a C_{10} side chain at C-5, e.g. coccinones G, H (**24**, **25**) showed two intense fragments at *m/z* 423 and *m/z* 396 instead of one intense ion at *m/z* 409 for compound having the C_{10} side chain at C-1 (e.g. garcinol series, **21**–**23**). The fragmentation hypotheses depicted in Fig. 1c involved the departure of a prenyl side chain by homolytic cleavage for fragment *m/z* 396. The α-cleavage of the bond adjacent to the C-10 carbonyl group, followed by the departure of a prenyl side chain, could explain the fragment *m/z* 423 (Fig. 1d). Thus, these fragmentation patterns allowed the distinction between benzophenones having a dimethyltetrahydropyran ring fused to the phloroglucinol core, such as compounds **1**–**8** and **14**–**20**, and for compounds having a keto–enol moiety at C2–C4, those possessing a C_{10} side chain at C1 (**9**–**11**, **21**–**23**) or C5 (**24**, **25**). LC–MS^2 analysis of HPLC-fraction 5 allowed the detection of one benzophenone having a "free" keto–enol function at C2–C4 and nine cyclized derivatives structurally close to those of the "coccinone series", one of which with a monohydroxylated benzoyl. Two other derivatives showed different fragmentation patterns.

The root barks (1.3 kg) were extracted using ethyl acetate on a static high-pressure high-temperature extractor and the solution was filtrated on polyamide. The filtered extract (8.4 g) was then subjected to a C-18 flash chromatography to give 10 fractions (FrA–FrJ). Comparative studies with LC/ESI-MS/MS of these fractions with fraction 5 (*vide infra*) allowed us to target those containing the supposed active compounds. Subsequent preparative HPLC purification resulted in the isolation of compounds **1**–**13** (Fig. 2), of which compounds **2**–**8** and **10**–**13** were new.

Compounds **1**–**13** shared several common spectral characteristics. The UV spectra showed absorption bands at 230, 280 and 320 nm consistent with aromatic rings and conjugated carbonyl groups. The IR spectra exhibited bands for hydroxyl (3350 cm^{-1}), α,β-unsaturated carbonyl groups (1650, 1668 cm^{-1}), ketone (1720 cm^{-1}) and aromatic rings (1600 cm^{-1}).

According to Ciochina and Grossman (2006), the geometry of the bridged bicyclic system is responsible for the bulk of the chiroptical properties. Therefore, the absolute configuration of the bicyclo[3.3.1]nonanes with β C(1)–C(9) and C(5)–C(9) bonds, were deduced from their negative $[\alpha]_D$ values (see Section 4). In addition, in the ^{13}C NMR spectra, all isolated compounds exhibited an upfield signal at ca. 17 ppm for C-22_{ax} resulting from a γ-gauche interaction between this methyl group and the CH_2-24. In the NOESY spectrum, NOE interactions between H-8_{eq}, H-7 and Me-23_{eq} confirmed the configuration with an equatorial substituent at C-7 (Ciochina and Grossman, 2006; Hamed et al., 2006; Piccinelli et al., 2005).

The spectroscopic data of compound **1** were identical with those of 7-*epi*-isogarcinol isolated from the latex of *M. coccinea* (Marti et al., 2009).

The molecular formula $C_{38}H_{50}O_5$ of compound **2** was deduced by HREIMS from the ion peak $[M+Na]^+$ *m/z* = 609.3589 (calc. 609.3556). The ^1H NMR spectra revealed important similarities with those of 7-*epi*-isogarcinol (**1**), except for the aromatic region which showed the presence of four aromatic protons instead of 3 for **1**. The missing 14-hydroxyl group was indirectly confirmed by the hypsochrome effect observed in the UV spectra (276 nm for compound **1** and 260 nm for compound **2**). The analysis of coupling constants and ^2D NMR data suggested the presence of a 1,3-disubstituted aromatic ring with four aromatic protons appearing as two doublets at δ_H 7.38 (1H, *d*, *J* = 7.9 Hz, H-14) and δ_H 7.70 (1H, *brd*, *J* = 7.2 Hz, H-16), one broad triplet at δ_H 7.45 (1H, *brt*, *J* = 7.7 Hz, H-15), and a broad singlet at δ_H 7.98 (1H, *brs*, H-12). To our knowledge, 14-deoxy-7-*epi*-isogarcinol (**2**) and 14-deoxygarcinol (**22**) isolated from *M. coccinea* (Marti et al., 2009) are the only type B PPAPs substituted by a *meta*-phenol. Another example of this substitution pattern is a type A benzophenone, hydroxynemorosone, isolated from *Clusia nemorosa* (de Oliveira et al., 1996).

Compound **3**, which was isolated as a brown oil, exhibited an $[\alpha]_D$ value of −37° (*c* 0.7, $CHCl_3$). The UV spectra revealed five absorption bands at 229, 255, 261, 276 and 320 nm, which suggested an additional conjugated system compared to 7-*epi*-isogarcinol (**1**). HREIMS indicated a $[M+Na]^+$ ion peak at *m/z* 623.3342,

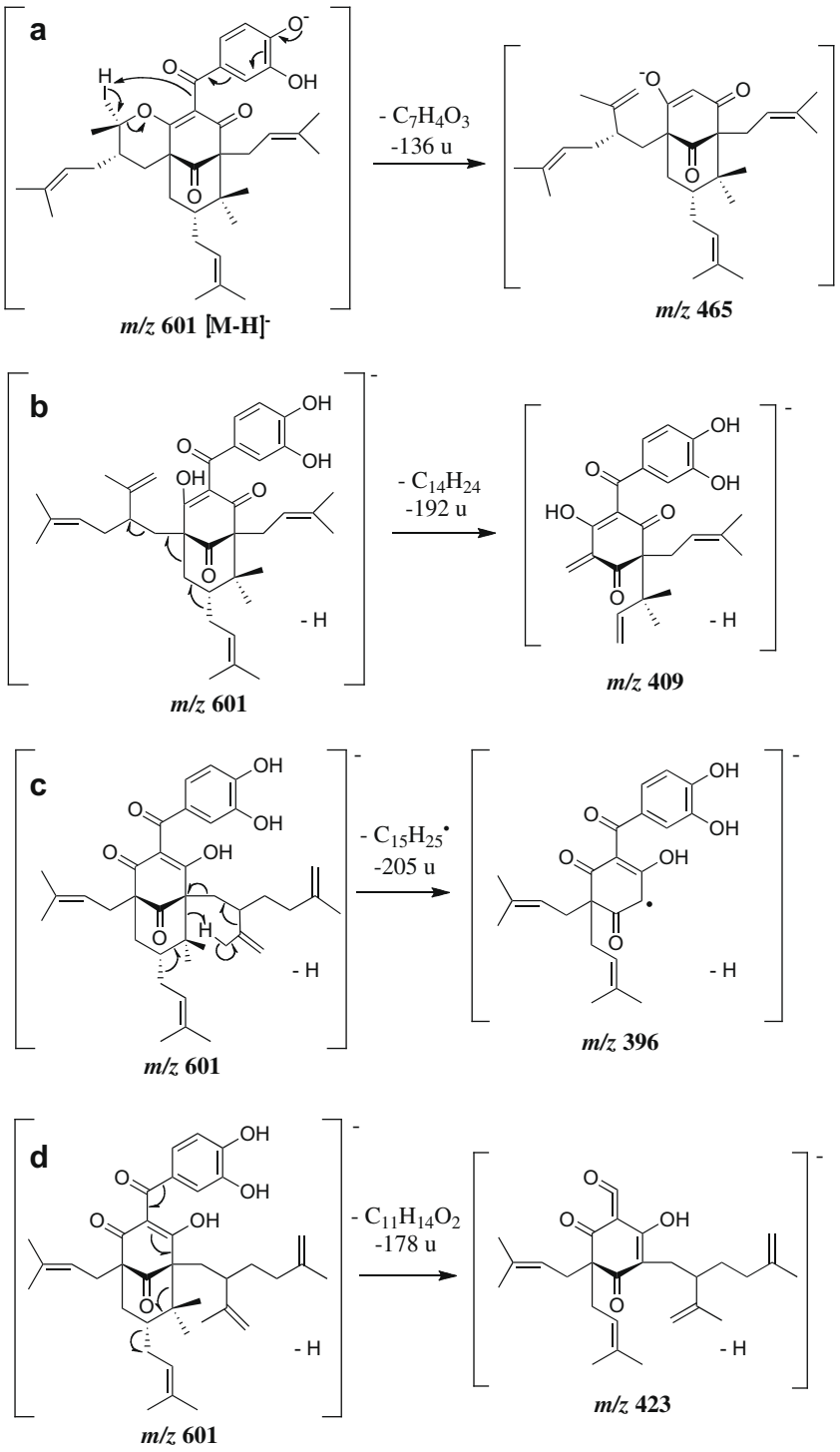


Fig. 1. Fragmentation hypotheses using MS/MS (ESI, negative mode) for 7-*epi*-garcinol (**9**, Fig. 1a), garcinol (**21**, Fig. 1b), and coccinone G (**24**, Fig. 1c,d).

giving the molecular formula of $C_{38}H_{48}NaO_6$ (calc. 623.3349), and confirmed the presence of an additional double bond. The ^1H and ^{13}C NMR data were very close to those of compound **1**. The 3-methylbuta-1,3-dien-1-yl side chain was deduced from correlations of Me-28 (δ_H 1.86, 3H, *s*) with the C-27 methylene group (δ_C 117.3), and of the quaternary C-26 carbon (δ_C 142.8) with the C-25 methine carbon (δ_C 137.6) observed in the HMBC spectrum. In the COSY spectrum correlations between H-25 (δ_H 6.26, 1H, *brd*, J = 15.5 Hz) and H-24 (δ_H 5.55, 1H, *dd*, J = 15.5; 8.4 Hz), and between H-24 and H-7 (δ_H 3.10, 1H, *m*) confirmed the position of the side chain at C-7. The coupling constant between H-24 and H-25 indicated the *E* geometry for the C-24–C-25 double bond. NOESY correlations between H-8$_{eq}$ (δ_H 2.21, 1H, *dd*, J = 14.0; 4.5 Hz), H-7 (δ_H 3.10, 1H, *m*) and Me-23$_{eq}$ (δ_H 1.18, 3H, *s*) established the equatorial position of the C-7 side chain, and those between H-30 (δ_H 1.68, 1H, *m*) and Me-33 (δ_H 1.12, 3H, *s*) indicated an equatorial position at C-30 for the prenyl group. Thus, compound **3**, named symphonone A possesses the structure depicted in Fig. 2.

The HREIMS of compound **4** showed a molecular peak $[M+Na]^+$ at *m/z* 693.4116 which fits the molecular formula $C_{43}H_{58}NaO_6$ (calc. 693.4131), and suggested the presence of an additional

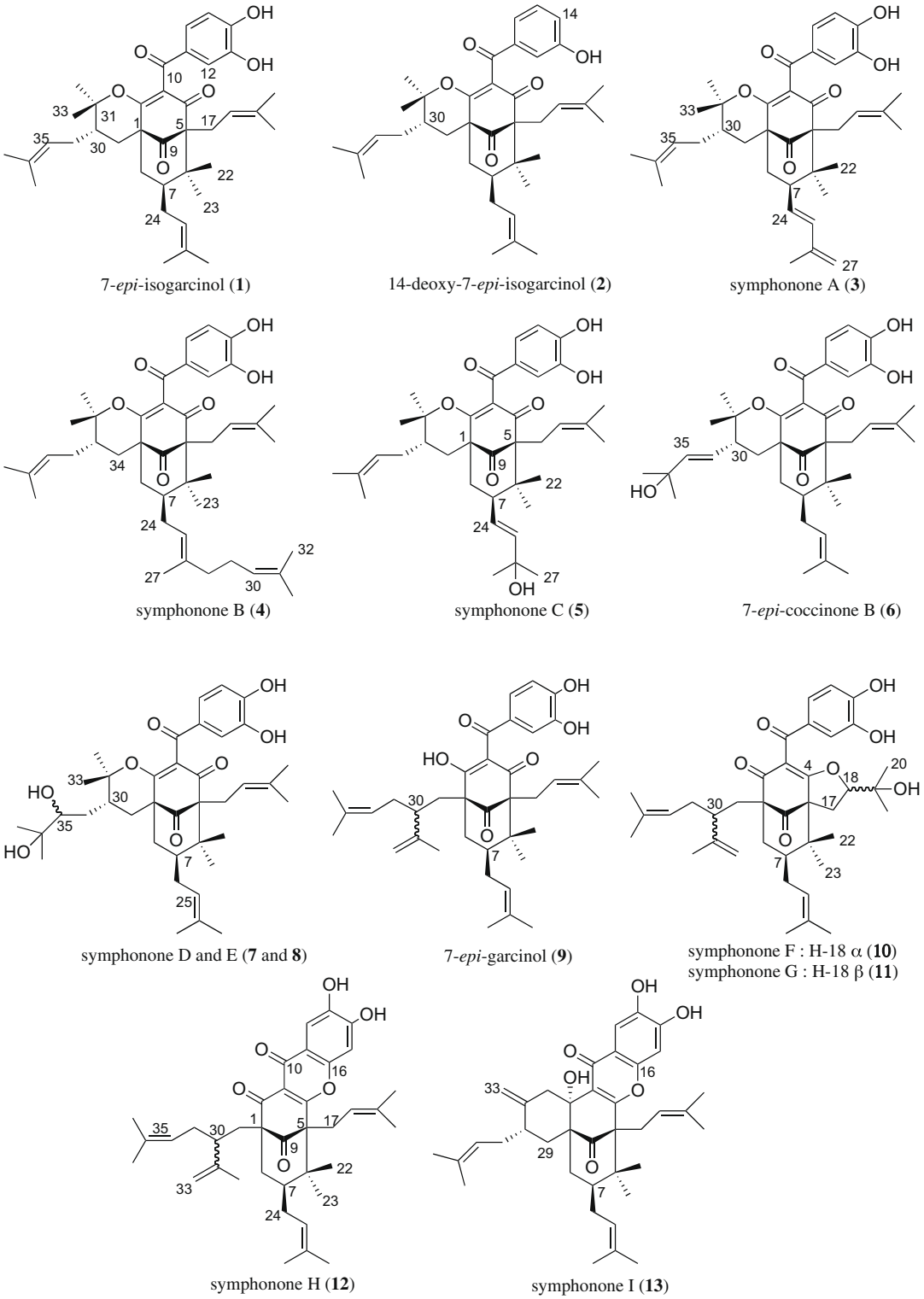


Fig. 2. Structures of compounds **1**–**13**.

prenyl side chain (C_5H_8) when compared with compound **1**. Analysis of NMR spectra confirmed that the structure of **4** was very close to that of 7-*epi*-isogarcinol (**1**). A geranyl side chain located at C-7 was deduced from correlations observed in COSY and HMBC spectra: ^{1}H–1H correlations between the vinyl proton H-30 (δ_H 5.22, 1H, *m*), and Me-32 and Me-33 (δ_H 1.73 and 1.62, 3H each, *s*, H-32 and H-33, respectively) and CH_2-29 (δ_H 2.15, 2H, *m*), and ^{1}H–13C correlations of C-29 at δ_C 27.7 with Me-27 (δ_H 1.69, 3H, *s*), and of C-7 with H-25 at δ_H 5.28 (1H, *m*) and one CH-24 at δ_H 2.27 (1H, *m*). Other COSY, NOESY and HMBC correlations were identical with those of compound **1** and allowed to propose the structure **4** depicted in Fig. 2. Compound **4** was named symphonone B.

The molecular formula of compound **5** was determined as $C_{38}H_{50}O_7$ on the basis (HREIMS) of the $[M+Na]^+$ ion peak at *m/z* 641.3444 (calc. 641.3454), and suggested the presence of an

additional hydroxyl group when compared with **1**. The presence of a 3-hydroxy-3-methylbut-1-en-1-yl side chain at position C-7 was suggested by the following evidence. In the HMBC spectrum, the correlations observed between the methyl protons Me-27 and Me-28 at δ_H 1.47 (6H, *s*) and the quaternary carbon C-26 (δ_C 70.1) and the vinyl carbon C-25 (δ_C 144.4) confirmed the presence of a hydroxylated side chain. In the COSY spectrum, cross peaks were observed between the vinyl protons H-25 at δ_H 5.83 (1H, *brd*, *J* = 15.0 Hz), H-24 at δ_H 5.81 (1H, *dd*, *J* = 15.0; 8.7 Hz) and H-7 (δ_H 3.11, 1H, *m*), and confirmed the position of the 3-hydroxy-3-methylbut-1-en-1-yl side chain at C-7. The geometry of the C-24–C-25 double bond was determined as *E* due to the coupling constant between the olefinic protons. These data, together with other results of ^{2}D NMR analysis, confirmed the structure of compound **5**, which was named symphonone C.

The HREIMS of compound **6** showed a $[M+Na]^+$ ion peak at *m/z* 641.3463, corresponding to the molecular formula $C_{38}H_{50}O_7$ for compound **5**. However, the spectroscopic data were closely comparable to those of coccinone B previously isolated from the latex of *M. coccinea* (Marti et al., 2009), but slight differences suggested that the prenyl side chain attached at C-7 was equatorial. In the ^{13}C NMR spectrum the upshield shift for C-22 for compound **6** at δ_C 16.6 (27.1 for coccinone B) is in favour of an α-configuration of H-7. This was confirmed by correlations observed in NOESY spectrum between Me-23$_{eq}$ (δ_H 1.28) and H-7 (δ_H 2.43). The *E* configuration of the C-34–C-35 double bond was confirmed by the couplings constants observed for the olefinic protons H-34 and H-35 (δ_H 5.72, *dd*, *J* = 14.8, 7.2 Hz and 5.74, 1H, *brd*, *J* = 14.8 Hz). Compound **6** is therefore 7-*epi*-coccinone B.

The molecular formula $C_{38}H_{52}O_8$ of compound **7** was deduced by HREIMS from the $[M+Na]^+$ ion peak at *m/z* 659.3582 (calc. 659.3560), and suggested the presence of two additional hydroxyl groups when compared with 7-*epi*-isogarcinol (**1**). It was confirmed by the presence of a large absorption band at 3400 cm^{-1} in the IR spectrum of **7**. From this formula and the NMR data, we deduced that compound **7** possesses a 2,3-dihydroxy-3-methylbut-1-yl side chain attached at C-30. Coccinones D and E (**19**, **20**) previously isolated from *M. coccinea* latex (Marti et al., 2009) possess the same type of side chain. This was confirmed in the COSY spectrum of **7**, in which cross peaks between H-30 at δ_H 2.01 (1H, *m*) and H$_2$-34 (δ_H 1.50, 1H, *m* and 2.03, 1H, *m*) and H$_2$-29 (δ_H 1.46, 1H, *m*, and 3.92, 1H, *m*) on one hand, and between H$_2$-34 and H-35 (δ_H 3.81; 1H, *brd*, *J* = 8.8 Hz) on the other hand, were observed. In addition, the chemical shifts observed for carbons C-22, C-23, C-7, C-8 and C-24 were in agreement with an equatorial position of the prenyl side chain at C-7. Compound **7** was named symphonone D.

Compound **8** was isolated as a brown oil. Its HREIMS indicated a $[M+Na]^+$ ion peak at *m/z* 659.3563, which suggested a molecular formula $C_{38}H_{52}NaO_8$ (calc. 659.3560), identical with that observed for compound **7**. Although their optical activities were almost identical −50 (**8**, *c* 0.4, $CHCl_3$) and −41 (**7**, *c* 0.4, $CHCl_3$), they showed different retention times at 9.1 and 7.8 min, respectively (see Section 4 for HPLC conditions). The spectroscopic data of these two compounds were almost identical but slight differences were noticed for the dihydroxylated prenyl side chain (see Table 1). NOESY spectrum of compound **8** showed cross peaks between H-35 at δ_H 4.12 (1H, *dd*, *J* = 9.1; 2.3 Hz), Me-37 and Me-38 (δ_H 1.56, 3H, *s*, H-37; 1.53, 3H, *s*, H-38), H$_2$-34 (δ_H 1.67, 1H, *m*; 1.75, 1H, *m*) and H-29$_{eq}$ at δ_H 3.49 (1H, *dd*, *J* = 14.3; 3.0 Hz). In contrast, no correlation was observed between H-35 and H-29$_{eq}$ for compound **7**. Compound **8**, which was named symphonone E is thus the 35-epimer of symphonone D (**7**). Preparation of Mosher esters could perhaps allow the determination of the absolute configuration at C-35 for both compounds, but the small amounts isolated were not sufficient to obtain the desired products.

Compound **9**, a yellow oil, exhibited an $[\alpha]_D$ value of −86° (*c* 1.0, $CHCl_3$). Its HREIMS indicated an ion peak at *m/z* 625.3517 $[M+Na]^+$, from which we have deduced the molecular formula $C_{38}H_{50}O_6$ identical to that of compound **1**. The ^{1}H and ^{13}C NMR spectral data were identical to those of 7-*epi*-garcinol isolated from the latex of *M. coccinea* (Marti et al., 2009).

Compound **10** was isolated as a brown oil. Its molecular formula $C_{38}H_{50}O_7$ was deduced by HREIMS of the $[M+Na]^+$ ion peak at *m/z* 641.3459 (calc. 641.3454), which suggested the presence of an additional hydroxyl group when compared with **9**. The presence of a benzoylphloroglucinol skeleton substituted by a lavandulyl side chain at C-1 was deduced by comparison of ^{1}D and ^{2}D NMR data with those of compound **9**. The COSY spectrum showed correlations between H-35 at δ_H 5.24 (1H, *m*) and Me-37 and Me-38 at δ_H 1.62 and 1.61, respectively (3H each, *s*), and CH$_2$-34 (δ_H 2.18, 1H, *m* and 2.23, 1H, *m*) on one hand, and between the methine H-30 at δ_H 3.09 (*m*) and CH$_2$-34 and CH$_2$-29 (δ_H 2.09, 1H, *dd*, *J* = 14.0; 4.6 Hz; 2.35, 1H, *m*) on the other hand (Fig. 2S Supplementary data). In the HMBC spectrum, cross peaks between H$_3$-33 (δ_H 1.71, 3H, *s*) and carbons C-30 (δ_C 44,5), C-31 (δ_C 148.7) and C-32 (δ_C 114.3) confirmed the presence of an isopropenyl side chain attached at C-30. A 2-(2′-hydroxyprop-2′-yl)-tetrahydrofuran ring fused to the benzophenone moiety at C-4–C-5 is confirmed by correlations of C-19 (δ_C 71.6) with the *gem*-dimethyl group (δ_H 1.34 and 1.36, 3H each, *s* Me-20 and Me-21) and CH$_2$-17 (δ_H 2.48, 1H, *dd*, *J* = 14.3; 8.1 Hz; 3.22, 1H, *dd*, *J* = 14.0; 8.0 Hz), and of C-5, C-6, C-9, C-18, C-19 at δ_C 70.3, 47.5, 207.6, 93.7 and 71.6, respectively, with CH$_2$-17 observed in the HMBC spectrum (Fig. 2S Supplementary data). The configuration at C-18 was deduced from the correlations observed in the NOESY spectrum (Fig. 3) between H-18 at δ_H 4.81 (1H, *brt*, *J* = 7.8 Hz), and H-17α (δ_H 2.48, 1H, *dd*, *J* = 14.3, 8.1 Hz), Me-23$_{eq}$ (δ_H 1.18, 3H, *s*) and the *gem*-dimethyl Me-20 and Me-21, which allowed to locate H-18 on the α-face. The configuration at C-30 remained undetermined. This compound was named symphonone F.

Compound **11**, which was isolated as a brown oil, presented in HREIMS the same ion peak as compound **10** at *m/z* 641.3450 (calc. 641.3454 for $C_{38}H_{50}NaO_7$). The spectroscopic data of these two compounds were almost identical but slight differences were noticed for the ^{1}H and ^{13}C chemical shifts of the tetrahydrofuran ring moiety. In contrast with symphonone F (**10**), cross peaks between H-18 at δ_H 4.29 (1H, *brt*, *J* = 8.7 Hz) and H-17β (3.17, 1H, *dd*, *J* = 12.8, 8.1) and the *gem*-dimethyl at C-19, observed in the NOESY spectrum of **11** allowed to place H-18 on the β face. Compound **11**, which was named symphonone G is thus the 18-*epi*-symphonone F if it is assumed that the configuration of the lavandulyl chain is the same for **10** and **11**.

Compound **12** was obtained as a yellow oil. The UV spectrum showed two intense absorptions bands at 349 and 250 nm, which suggested the presence of an extended conjugated system. From an examination of NMR data we have deduced that compound **12** was structurally close to 7-*epi*-garcinol (**9**). However, the molecular formula $C_{38}H_{48}O_6$ of compound **12** deduced from the HREIMS of the ion peak $[M+Na]^+$ *m/z* 623.3355 (calc. 623.3349), indicated an additional degree of unsaturation when compared with compound **9**. The ^{1}H NMR spectrum of compound **12** showed only two aromatic protons at δ_H 8.06 (1H, *brs*, H-12), and 7.36 (1H, *brs*, H-15). The analysis of the aromatic region of the ^{13}C NMR spectrum revealed three oxygenated carbons at δ_C 147.9 (C-13); 155.5 (C-14) and 151.3 (C-16), and two sp2 methine carbons at δ_C 109.8 (C-12) and 104.2 (C-15). These data confirmed the presence of a 1,2,4,5-tetrasubstituted aromatic ring. Moreover, the upfield shifts of C-4, C-10 and the downfield shift of C-16 (δ_C 172.1, 173.1 and 151.3, respectively) compared to the 7-*epi*-garcinol (δ_C 196.2, 195.8 and 125.2, respectively) indicated that one oxygen of the keto–enol system was linked to C-16 forming an additional

Table 1
NMR spectroscopic data for compounds **1–8** and **10–13** (500 MHz for ^{1}H NMR and 125 MHz for ^{13}C NMR in pyridine-d_5).

Position	**1**			**2**			**3**			**4**		
	δ_C	δ_H	*J* (Hz)	δ_C	δ_H	*J* (Hz)	δ_C	δ_H	*J* (Hz)	δ_C	δ_H	*J* (Hz)
1	52.2			53.8			53.7			53.8		
2	170.8			171.6			170.8			170.8		
3	129.1			130.0			128.4			131.0		
4	194.9			195.6			194.8			195.4		
5	71.4			71.4			71.9			71.4		
6	46.8			46.8			47.5			46.8		
7	42.2	2.44	*m*	42.3	2.37	*m*	45.5	3.10	*m*	42.1	2.44	*m*
8	43.1	1.77	*m*	43.1	1.79	*m*	43.1	2.03	*m*	43.1	1.79	*m*
–		2.43	*m*		2.45	*m*		2.21	*dd* (14.0, 4.5)		2.44	*m*
9	207.3			207.3			206.7			207.4		
10	193.2			194.8			193.5			193.2		
11	131.0			140.2			129.1			130.1		
12	117.1	8.13	*brs*	116.4	7.98	*brs*	117.2	8.16	*brs*	116.8	8.13	*brs*
13	147.9			159.9			148.0			147.9		
14	153.7			121.8	7.38	*d* (7.8)	153.9			153.7		
15	116.5	7.31	*brd* (7.7)	130.8	7.45	*brt* (7.8)	117.2	7.31	*brd* (7.8)	116.4	7.31	*brd* (8.0)
16	124.4	7.72	*brd* (7.7)	121.2	7.70	*d* (7.8)	124.7	7.76	*brd* (7.8)	125.2	7.73	*brd* (8.0)
17	25.9	2.77	*dd* (13.6, 4.8)	25.9	2.76	*dd* (13.7, 5.4)	25.9	2.78	*m*	25.9	2.78	*dd* (13.2, 5.9)
–		2.97	*dd* (13.6, 6.4)		2.97	*dd* (13.7, 7.4)		2.96	*dd* (13.5, 5.9)		2.97	*dd* (13.2, 7.1)
18	122.1	5.38	*brt* (6.4)	121.9	5.38	*brt* (5.4)	122.7	5.52	*brt* (5.9)	122.0	5.39	*brt* (5.9)
19	134.2			134.3			134.9			133.7		
20	26.2	1.55	*s*	26.6	1.58	*s*	26.6	1.58	*s*	26.5	1.56	*s*
21	18.8	1.73	*s*	18.8	1.74	*s*	18.5	1.73	*s*	18.8	1.74	*s*
22	16.6	0.83	*s*	16.6	0.83	*s*	17.1	0.86	*s*	16.7	0.85	*s*
23	22.8	1.27	*s*	22.8	1.27	*s*	23.8	1.18	*s*	22.8	1.27	*s*
24	28.5	1.82	*dd* (14.1, 8.5)	28.4	1.84	*m*	128.4	5.55	*dd* (15.5, 8.4)	28.4	2.27	*m*
–		2.26	*dd* (14.1, 2.5)		2.27	*m*					1.88	*m*
25	123.8	5.18	*brt* (6.2)	123.8	5.20	*brt* (5.4)	137.6	6.26	*brd* (15.5)	125.2	5.28	*m*[a]
26	132.9			133.7			142.8			137.4		
27	26.2	1.62	*s*	26.2	1.65	*s*	117.3	5.03	*brs*	16.8	1.69	*s*
28	18.4	1.62	*s*	18.4	1.63	*s*	19.0	1.86	*s*	40.5	2.06	*m*[a]
–											2.16	*m*[a]
29	28.5	1.21	*dd* (13.9, 13.7)	28.5	1.21	*m*	28.4	1.14	*m*[a]	27.7	2.05	*m*[a]
–		3.28	*dd* (13.6, 3.2)		3.29	*dd* (14.3, 3.5)		3.29	*dd* (14.2, 2.6)		2.15	*m*[a]
30	43.8	1.65	*m*	43.8	1.62	*m*	43.9	1.68	*m*	125.9	5.22	*m*[a]
31	87.6			87.8			88.2			131.8		
32	21.7	1.14	*s*	21.6	1.14	*s*	21.4	1.21	*s*	26.2	1.73	*s*
33	29.2	1.08	*s*	29.1	1.04	*s*	29.8	1.12	*s*	18.3	1.62	s
34	30.4	1.79	*m*	30.3	1.79	*m*	30.5	1.82	*m*	28.5	1.21	*m*[a]
–		1.96	*m*		1.96	*m*		1.98	*m*		3.29	*dd* (14.1, 2.7)
35	122.9	5.13	*brt* (6.6)	122.7	5.12	*brt* (5.4)	123.6	5.12	*brt* (5.9)	43.8	1.67	*m*
36	133.7			133.8			133.3			87.4		
37	26.2	1.71	*s*	26.2	1.69	*s*	26.5	1.69	*s*	21.6	1.20	*s*
38	18.3	1.56	*s*	18.3	1.56	*s*	18.3	1.57	*s*	29.2	1.10	*s*
39										30.4	1.80	*m*
–											1.96	*m*
40										122.8	5.13	*brt* (5.9)
41										133.6		
42										26.3	1.69	*s*
43										18.2	1.56	*s*

[a] Overlap.
[b] Assignments are interchangeable, *brt*: broad triplet, *brs*: broad singlet.

heterocyclic ring. In the HMBC spectrum, correlations between H_2-29 (δ_H 2.10, 1H, *m*; 2.46, 1H, *m*) and C-2, C-8 and C-9 (δ_C 193.3, 46.9 and 208.4, respectively), and between H_2-17 (δ_H 3.04, 1H, *dd*, *J* = 13.6, 8.7 Hz; 3.15, 1H, *m*) and C-4 at δ_C 172,1 suggested that the oxygenated bridge was located between C-4 and C-16 (Fig. 3S Supplementary data). The configuration at C-30 remains undetermined. Compound **12** showed an optical activity, with $[\alpha]_D$ −37° (*c* 0.2, $CHCl_3$), suggesting a β-position for the C-9 carbonyl bridge, according to Ciochina and Grossman (2006). The prenyl side chain at C-7 has been placed in equatorial position taking into account the chemical shift at δ_C 16.4 for C-22 and the observation of cross peaks between H-8$_{eq}$, H-7 and Me-23$_{eq}$ in the NOESY spectrum. Compound **12** was named symphonone H. Compound **12** is the C-7 epimer of GDPPH-1 (Sang et al., 2002). GDPPH-1 is a product obtained by reaction of garcinol with DPPH (2,2-diphenyl-1-picrylhydrazyl) and is the result of a single electron transfer followed by a deprotonation of the hydroxy group of the keto–enol system. From a biogenetical point of view, symphonone H (**12**) is the natural isomer of GDDPH-1 and could be formed by an oxidative cyclization of 7-*epi*-garcinol (**9**) catalyzed by a xanthone synthase (Peters et al., 1997).

Compound **13** was isolated as a yellow oil. Its molecular formula $C_{38}H_{48}O_6$, which was deduced from the HREIMS of the $[M-H]^-$ ion peak at *m/z* 599.3470 (calc. 599.3373), is identical to that of compound **12**. Analysis of the UV, IR, ^{1}H and ^{13}C NMR spectral data suggested the presence of the same carbon skeleton as that of the tetracyclic compound **12**. The COSY and HMBC spectra of **13** revealed the presence of three prenyl side chains attached to C-5, C-7 and C-30 (Fig. 4S Supplementary data). In addition, in the HMBC spectrum, cross peaks between H_2-29 (δ_H 1.52, 1H, *m*, and 2.35, 1H, *dd*, *J* = 13.8, 4.9 Hz) and C-1 (δ_C 56.4), C-2 (δ_C 76.5), C-8 (δ_C 42.3) and C-30 (δ_C 39.2), and between

Table 1 (*continued*)

Position	**5**			**6**			**7**			**8**		
	δ_C	δ_H	*J* (Hz)	δ_C	δ_H	*J* (Hz)	δ_C	δ_H	*J* (Hz)	δ_C	δ_H	*J* (Hz)
1	53.8			53.2			53.6			53.8		
2	170.8			170.4			171.2			171.2		
3	129.2			129.9			127.3			127.3		
4	194.8			194.6			195.1			194.9		
5	71.2			71.5			71.4			71.4		
6	46.5			46.8			46.7			46.7		
7	45.4	3.11	*m*	42.2	2.41	*m*	42.1	2.42	*m*	42.3	2.43	*m*
8	43.1	2.00	*m*	43.2	1.73	*m*	43.1	1.72	*m*	43.4	1.72	*m*
–		2.23	*m*		2.50	*m*		2.49	*m*		2.48	*m*
9	206.8			207.3			207.4			207.5		
10	193.2			193.1			193.3			193.2		
11	130.9			130.8			130.8			131.1		
12	116.6	8.15	*brs*	116.9	8.12	*brs*	116.9	8.15	*brs*	116.9	8.16	*d* (1.9)
13	147.9			147.3			147.8			147.9		
14	153.8			153.9			153.9			153.6		
15	117.1	7.29	*d* (7.6)	116.5	7.31	*brd* (7.8)	116.7	7.24	*brd* (7.8)	116.7	7.21	*brd* (7.7)
16	121.5	7.75	*brd* (7.6)	124.4	7.75	*brd* (7.8)	125.5	7.70	*brd* (7.8)	124.5	7.66	*brd* (7.7)
17	25.9	2.76	*dd* (13.9, 5.0)	25.9	2.81	*dd* (13.6, 6.8)	26.7	2.80	*m*[a]	26.0	2.76	*dd* (13.7, 4.3)
–		2.96	*dd* (13.9, 7.3)		2.99	*dd* (13.6, 6.8)		2.98	*dd* (13.2, 7.6)		2.99	*dd* (13.7, 7.7)
18	121.9	5.37	*m*	122.0	5.43	*m*	122.2	5.39	*brt* (5.7)	121.9	5.33	*m*
19	134.2			134.4			135.4			134.3		
20	26.2	1.57	*s*	26.7	1.64	*s*	26.6	1.54	*s*	26.3	1.72	*s*[a]
21	18.8	1.72	*s*	18.8	1.77	*s*	18.8	1.75	*s*	18.8	1.72	*s*[a]
22	16.9	0.83	*s*	16.6	0.86	*s*	16.6	0.83	*s*	16.7	0.85	*s*
23	23.5	1.22	*s*	22.8	1.28	*s*	22.9	1.27	*s*	22.9	1.28	*s*
24	124.5	5.81	*dd* (15.0, 8.7)	28.5	1.82	*m*	28.4	1.83	*m*	28.6	1.84	*m*
–					2.25	*m*		2.26	*m*		2.27	*m*
25	144.4	5.83	*brd* (15.0)	124.1	5.20	*m*	123.8	5.20	*m*	124.1	5.23	*m*
26	70.1			133.7			133.5			133.6		
27	31.2[b]	1.47	*s*[a]	26.2	1.62	*s*	26.1	1.62	*s*[a]	26.2	1.62	*s*[a]
28	31.3[b]	1.47	*s*[a]	18.4	1.63	*s*	18.4	1.62	*s*[a]	18.5	1.62	*s*[a]
29	28.4	1.14	*m*	29.2	1.49	*m*[a]	30.3	1.46	*m*	28.7	1.32	*m*
–		3.26	*m*		3.22	*dd* (13.9, 2.3)		3.92	*m*		3.49	*dd* (14.3, 3.0)
30	43.9	1.67	*m*	45.9	2.32	*m*	43.1	2.01	*m*	40.2	2.31	*m*
31	87.7			86.2			87.7			88.1		
32	21.7	1.15	*s*	22.2	1.09	*s*	22.0	1.18	*s*	22.0	1.16	*s*
33	29.3	1.09	*s*	29.2	1.01	*s*	29.3	1.14	*s*	28.9	1.17	*s*
34	30.3	1.81	*m*	124.6	5.72	*dd* (14.8, 5.2)	34.4	1.50	*m*	34.2	1.67	*m*
–		1.97	*m*					2.03	*m*		1.75	*m*
35	122.8	5.11	*m*	144.7	5.74	*brd* (14.8)	79.6	3.81	*brd* (8.8)	76.1	4.12	*dd* (9.1, 2.3)
36	133.8			69.9			73.2			72.9		
37	26.6	1.69	*s*	30.9[b]	1.46	*s*	26.6[b]	1.48	*s*	27.3[b]	1.56	*s*
38	18.3	1.56	*s*	31.1[b]	1.44	*s*	25.9[b]	1.44	*s*	26.0[b]	1.53	*s*

[a] Overlap.
[b] Assignments are interchangeable, *brt*: broad triplet, *brs*: broad singlet.

H_2-32 (δ_H 2.20, 1H, *m*, and 3.36, 1H, *brd*, *J* = 13.3 Hz) and C-2, C-1, C-3 (δ_C 124.2), C-29 (δ_C 38.8), C-31 (δ_C 147.5) and C-33 (δ_C 109.1) confirmed the presence of a cyclohexane, bearing an *exo*-methylene at C-31 and a prenyl side chain at C-30 fused to the tetracyclic moiety at positions C-1 and C-2. (Fig. 4S Supplementary data). Moreover, the chemical shift of C-2 at δ_C 76.5 and the absence of any signal in the DEPT experiment confirmed the presence of an oxygenated quaternary carbon, instead of a carbonyl in compound **12**. The relative configuration of the cyclohexane moiety was deduced from NOESY interactions. Correlations observed between H-32_{ax} at δ_H 2.20 (1H, *d*, 13.3 Hz) and H-30, and between CH_2-33 and CH_2-34 are in favour of an equatorial prenyl side chain at C-30. Finally, in the NOESY spectrum run in DMF-d_6 at 253 K, a correlation between the proton of the 2-hydroxy group at δ_H 6.55 and H-8_{eq} at δ_H 2.54 confirmed a *cis* junction between the cyclohexane moiety and the phloroglucinol part of the bridged bicycle (Fig. 4).

Compound **13**, which was named symphonone I, is a new example of oxidized benzophenone along with symphonone H, oxy-gutttiferone K isolated from *Garcinia cambodgia* fruits (Masullo et al., 2008), and garcinialone isolated from *Garcinia multiflora* roots (Chien et al., 2008). Symphonone I (**13**) could be seen as a cyclization product of symphonone H (**12**).

2.1. Biological activity

The antiplasmodial activity of compounds **1**–**13** against the chloroquine-resistant strain of *P. falciparum* FcB1 and the cytotoxicity on the human fibroblast cell line MRC-5 are summarized in Table 2. All the $PPAP_S$ exhibited moderate cytotoxicity on the MRC-5 cell line with IC_{50}s ranging between 3.7 and 19.5 μM, and exhibited significative antiplasmodial activities on FcB1 strain with IC_{50}s ranging from 2.1 to 10.1 μM. These results confirmed those obtained from the study of *M. coccinea* (Marti et al., 2009) indicating that compounds having an additional heterocycle fused to the bicyclo[3.3.1]nonane core are slightly more active; only 7-*epi*-garcinol (**9**) showed a weak antiplasmodial activity. However, since the IC_{50}s are very close to each other, it is not possible to draw any structure–activity relationships from this chemical series.

3. Conclusion

Phytochemical investigation of *S. globulifera* led to the isolation of eleven new PPAPs derivatives. Biological assays revealed that PPAPs having a heterocycle fused to the phloroglucinol moiety exhibited potent antimalarial activity. It is interesting to note that,

Table 1 (*continued*)

Position	**10**			**11**			**12**			**13**			
	δ_C	δ_H	*J* (Hz)	δ_C	δ_H	*J* (Hz)	δ_C	δ_H	*J* (Hz)	δ_C	δ_H	*J* (Hz)	δ_H[c]
1	62.8			62.6			63.9			56.4			
2	195.3			196.0			193.3			76.5			
3	120.1			122.0			122.1			124.1			
4	174.5			174.2			172.1			161.4			
5	70.3			70.1			66.1			64.6			
6	47.5			49.3			49.5			50.0			
7	42.1	2.42	*m*	43.0	2.35	*m*	43.5	1.82	*m*	43.2	2.14	*m*	1.72
8	44.3	1.64	*m*[a]	44.9	1.66	*m*[a]	46.9	1.52	*m*	42.3	1.22	*m*[a]	0.97
–		2.35	*m*[a]		2.37	*m*		2.34	*m*		3.02	*m*[a]	2.54
9	207.6			207.2			208.4			209.7			
10	192.4			190.8			173.1			177.7			
11	131.2			131.1			116.9			117.2			
12	117.3	8.19	*brs*	117.2	8.17	*brs*	109.8	8.06	*brs*	108.9	7.99	*s*	8.21
13	147.9			147.9			147.9			147.9			
14	153.9			153.9			155.5			156.1			
15	116.3	7.25	*brd* (7.8)	116.3	7.28	*brd* (8.0)	104.2	7.36	*brs*	104.2	7.39	*s*	7.35
16	125.3	7.87	*brd* (7.8)	124.3	7.80	*brd* (8.0)	151.3			152.0			
17	26.6	2.48	*dd* (14.2, 7.9)	26.3	2.62	*dd* (12.7, 8.5)	26.1	3.04	*dd* (13.6, 8.7)	25.3	3.02	*m*[a]	2.70
–		3.22	*dd* (14.2, 7.9)		3.17	*dd* (12.7, 8.5)		3.15	*m*		3.04	*m*[a]	2.91
18	93.7	4.81	*brt* (7.9)	93.2	4.29	*brt* (8.5)	121.1	5.18	*m*[a]	121.1	5.19	*brt* (6.5)	4.78
19	71.6			70.3			134.7			134.8			
20	26.1[b]	1.34	*s*	26.3[b]	1.28	*s*	26.3	1.48	*s*	26.3	1.48	*s*	1.40
21	26.4[b]	1.36	*s*	27.8[b]	1.33	*s*	18.8	1.85	*s*	18.9	1.82	*s*	1.68
22	16.5	0.95	*s*	16.6	0.86	*s*	16.4	0.90	*s*	17.4	0.85	*s*	0.77
23	23.4	1.18	*s*	24.4	1.38	*s*	24.4	1.19	*s*	24.4	1.23	*s*	1.15
24	29.5	1.75	*m*	29.9	1.70	*m*	29.2	1.71	*m*	30.2	1.67	*m*	1.62
–		2.21	*m*		2.10	*m*[a]		2.03	*m*		2.12	*m*	2.04
25	124.8	5.21	*m*[a]	124.4	5.08	*brt* (6.3)	123.6	4.94	*brt* (6.0)	124.2	5.12	*brt* (6.5)	4.96
26	133.9			134.7			133.6			132.8			
27	26.3	1.59	*s*	26.3	1.54	*s*	26.1	1.57	*s*	26.2	1.56	*s*	1.59
28	18.6	1.64	*s*	18.7	1.57	*s*	18.3	1.52	*s*	18.4	1.52	*s*	1.51
29	36.9	2.09	*dd* (14.0, 4.6)	37.1	2.08	*m*[a]	37.8	2.10	*m*	38.8	1.52	*m*	1.04
–		2.35	*m*[a]		2.38	*m*		2.46	*m*		2.35	*dd* (13.8, 4.9)	1.97
30	44.5	3.09	*m*	45.0	3.02	*m*	44.5	3.12	*m*	39.2	2.91	*m*	2.30
31	148.7			149.4			150.8			147.5			
32	114.3	4.89	*brs*	115.1	4.90	*brs*	18.5	1.84	*s*	50.3	2.20	*brd* (13.3)	1.72
–		4.93	*brs*		4.92	*brs*					3.36	*brd* (13.3)	2.88
33	18.7	1.71	*s*	18.7	1.72	*s*	114.2	4.74	*brs*	109.1	4.93	*brs*	4.71
–								4.76	*brs*		4.96	*brs*	4.74
34	33.4	2.18	*m*	34.0	2.22	*m*	33.0	2.13	*m*	31.9	2.18	*m*[a]	1.98
–		2.23	*m*		2.29	*m*		2.16	*m*		2.44	*m*	2.29
35	124.2	5.24	*m*[a]	125.5	5.29	*brt* (6.3)	123.9	5.20	*m*[a]	124.4	5.43	*brt* (6.5)	5.24
36	131.9			135.5			132.2			132.5			
37	26.5	1.62	*s*	26.6	1.63	*s*	26.2	1.61	*s*	26.3	1.67	*s*	1.71
38	18.4	1.61	*s*	18.5	1.64	*s*	18.4	1.60	*s*	18.4	1.62	*s*	1.65
2-OH													6.55

[a] Overlap.
[b] Assignments are interchangeable, *brt*: broad triplet, *brs*: broad singlet.
[c] ^{1}H NMR (600 MHz in DMF-d_6 at 253 K).

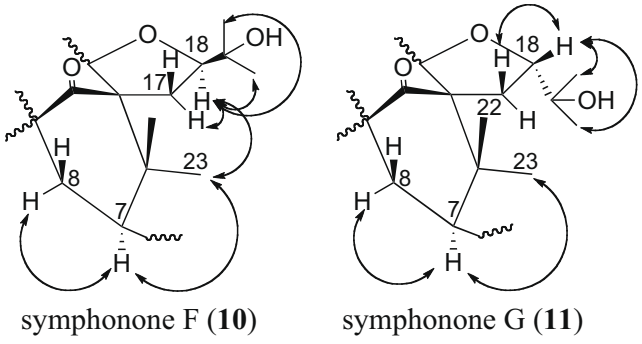


Fig. 3. Key NOESY (arrows) correlations for **10** and **11**.

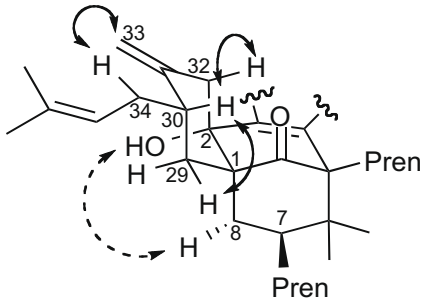


Fig. 4. Key NOESY (arrows, hashed for the spectrum in DMF-d_6, 253 K) correlations observed for the cyclohexane moiety of compound **13**.

in contrast with guttiferones A–E (Gustafson et al., 1992), all the compounds of this chemical series possessed a unique orientation of the bridged bicycle core and a C-7 equatorial substituent. These results could reflect the survival of different ecospecies through the neotropic leading to specific chemotypes. Hyphenated techniques such as LC/MS/MS could be applied to crude Clusiaceae extracts in order to resolve these chemotaxonomical interrogations or to find potent new biological actives PPAPs.

Table 2
IC_{50} values of compounds **1**–**13** assayed against the FcB1 strain of *P. falciparum* and MRC-5 cell line.

Compound	IC_{50} *P.f.* FcB1 (μM) ± SD (n = 3)	IC_{50} MRC-5[a] (μM) (n = 2)
7-*epi*-Isogarcinol (**1**)	3.2 ± 1.3	4.1
14-Deoxy-7-*epi*-isogarcinol (**2**)	2.5 ± 0.9	19.5
Symphonone A (**3**)	2.8 ± 0.8	5.8
Symphonone B (**4**)	3.3 ± 1.1	11.8
Symphonone C (**5**)	2.6 ± 0.6	10.8
7-*epi*-Coccinone B (**6**)	3.3 ± 0.3	11.5
Symphonone D (**7**)	2.1 ± 0.7	11.9
Symphonone E (**8**)	2.7 ± 1.0	10.4
7-*epi*-Garcinol (**9**)	10.1 ± 4.6	11.3
Symphonone F (**10**)	3.2 ± 1.2	6.9
Symphonone G (**11**)	2.1 ± 0.9	3.7
Symphonone H (**12**)	3 ± 0.4	4.9
Symphonone I (**13**)	6.7 ± 0.5	15.9
Chloroquine	0.078 ± 0.006	25
Taxotere	0.010 ± 0.005	31.5

P.f. FcB1: chloroquine-resistant strain of *Plasmodium falciparum*.
[a] IC_{50}s (μM) are mean values from two replicates (the variation is a maximum of 20%) MRC-5: human fibroblast cell line.

4. Experimental

4.1. General

IR spectra were obtained on a Nicolet FT-IR 205 spectrophotometer and UV spectra were recorded on a Perkin–Elmer Lamba 5 spectrophotometer. Optical rotations were measured at 25 °C on a JASCO™ P-1010 polarimeter. The NMR spectra were recorded on a Bruker 500 MHz (Avance 500) for ^{1}H and 125 MHz for ^{13}C using pyridine-d_5 as solvent. Chemical shifts (relative to TMS) are in ppm, and coupling constants in Hz. ESI-MS was obtained on a LCT mass spectrometer. HR-ESI-MS were run on a ESI-TOF spectrometer (LCT; Waters). LC/ESI-MS analyses were performed using Finnigan Surveyor HPLC system and ion trap MS^n (Finnigan LCQdeca) detector. Kromasil C-18 column (250 × 4.6 mm I.D., 5 μm) was used for analytical analysis. A gradient mobile phase consisting of acetonitrile/water 75:25–95:5 at 1.0 ml/min in 35 min was used to analyse fractions. The Finnigan detector was equipped with an electrospray ionization (ESI) source operating under the following conditions: sheath gas flow rate: 80 (arbitrary units), auxiliary gas flow rate: 20 (arbitrary units), N_2, capillary temperature: 300 °C, capillary voltage: 30 V. Full scan mass spectra (50–1500 amu) were recorded in negative mode. For MS^2, the data dependent program was used so that the most abundant ions in each scan were selected and subject to MS^2 analyses. The collision-induced dissociation (CID) energy was adjusted to 40%. The isolation width of precursor ion was 1.5 mass units.

Kromasil analytic, semi-preparative and preparative C-18 columns (250 × 4.5 mm; 250 × 10 mm and 250 × 21.2 mm I.D., 5 μm Thermo®) were used for preparative HPLC separations using a "Waters autopurification system" equipped with a sample manager (Waters 2767), a column fluidics organizer, a binary pump (Waters 2525), a UV–vis diode array detector (190–600 nm, Waters 2996) and a PL-ELS 1000 ELSD detector Polymer Laboratory. Companion® was used for flash chromatography with prepacked column of silica-gel 15–40 μm, 40 × 150, Versapack™. Silica-gel 60 (35–70 μm) and analytical TLC plates (Si gel 60 F 254) were purchased from SDS (France). Polyamide DC 6 and polyamide cartridge was purchased from Macherey–Nagel (Chromabond PA®, 1 g). All other chemicals and solvents were of analytical grade and purchased from SDS (France).

4.2. Plant material

Trunk barks, root barks, leaves and latex of *S. globulifera* were collected by one of us (G.M.) in the dense rain forest of French Guiana, in May 2006, under the reference CAY-3527. This plant was identified by Dr. M.-F. Prévost of the "Institut de Recherche pour le Développement (IRD)", Cayenne, French Guiana. Herbarium specimen has been deposited at the IRD Centre (Cayenne, French Guyana).

4.3. Extraction and isolation procedures

Root barks (36 g) were extracted using the automatic extractor Dionex® ASE 300 with EtOAc (3 × 100 ml) followed by MeOH (3 × 100 ml) at 40 °C and 100 bar. The extracts were concentrated *in vacuo* at 40 °C to yield 4.53 g (EtOAc) and 2.65 g (MeOH). Aliquot of 100 mg of each extract were dissolved in EtOAc or MeOH and then filtered on polyamide cartridge before being tested on *P. falciparum* FcB1 strain at 10 μg/ml. The root barks ethyl acetate extract was then fractionated on a semi-preparative C-18 column according to a standardized method previously detailed (Bousserouel et al., 2005). Fraction 5 (t_R from 34 to 42 min) exhibited 95% of inhibitory growth on *P. falciparum*.

A larger amount of root barks (1.3 kg) was extracted with EtOAc (3 × 1 l) at 40 °C and 1450 psi on a static high-pressure high-temperature extractor, Zippertex®, developed in the ICSN Pilot Unit. The extract was concentrated *in vacuo* at 40 °C to yield 110 g of a brown powder. A large amount of the extract (10.0 g) was then filtered on polyamide DC 6 with MeOH/CH_2CL_2 1/1 to remove bark residue and tannins. The filtered extract was dried under *vacuum* at 40 °C to give a brown gummy residue (8.4 g), which was subjected to flash chromatography using *n*-heptane, mixtures of *n*-heptane-EtOAc (9:1–1:9) and EtOAc as eluents and a 20 ml/min flux. According to their TLC profiles, 10 fractions (FrA–FrJ) were obtained. A comparative study on analytical HPLC and LC/ESI-MS/MS of FrA–J with the reference fraction F5 allowed us to target those containing the active compounds. FrD (330 mg) was subjected to a preparative C-18 column using an isocratic mobile phase consisting of MeCN–H_2O 85:15 + 0.1% formic acid, leading to the isolation of symphonone A **3** (6 mg; w/w 0.07%), 7-*epi*-isogarcinol **1** (180 mg; w/w 2.1%), 14-deoxy-7-*epi*-isogarcinol **2** (8 mg; w/w 0.09%), symphonone H **12** (4 mg; w/w 0.04%), symphonone B **4** (5 mg; w/w 0.06%), 7-*epi*-garcinol **9** (30 mg; w/w 0.35%), with t_R of 11.5, 12.8, 18.5, 23.9, 33.1 and 34.1 min, respectively. Using the same solvent, FrE (180 mg) led to symphonone A **3** (16 mg; w/w 0.19%), 7-*epi*-isogarcinol **1** (90 mg; w/w 1.07%) and 8 mg of a mixture (t_R 18.1 min), which was subjected to a semi-preparative C-18 column with MeCN–H_2O 88:12 + 0.1% formic acid to afford symphonone I **13** (2 mg; w/w 0.02%, t_R 17.4 min). FrF (93 mg) was subjected on the same C-18 preparative column as for FrD and FrE with MeCN–H_2O 70:30 + 0.1% formic to afford symphonone C **5** (8 mg; w/w 0.09%), 7-*epi*-coccinone B **6** (6 mg; w/w 0.07%), symphonone F **10** (4 mg; w/w 0.04%), symphonone G **11** (3 mg; w/w 0.03%), with t_R of 12.5, 14.0, 26.1 and 26.8 min, respectively. FrG was subjected to the same isocratic system as for FrF and led to symphonone D **7** (4 mg; w/w 0.04%, t_R 7.8 min) and symphonone E **8** (3 mg; w/w 0.03%, t_R 9.1 min).

4.3.1. 7-epi-Isogarcinol **1**

Brown powder, $[\alpha]_D$ −158° (*c* 1.0, $CHCl_3$); HREIMS $[M+Na]^+$ *m/z* 625.3519, $C_{38}H_{50}NaO_6$ requires 625.3505; UV λ_{max} (log ε): 319 (3.85), 276 (4.15), 233 (4.12); IR ν_{max} 3320, 2970, 2930, 1730, 1650, 1590, 1520, 1440, 1370, 1290, 1170 cm^{-1}; ^{1}H NMR and ^{13}C NMR (Table 1).

4.3.2. 14-Deoxy-7-epi-isogarcinol ***2***

Brown oil, HREIMS $[M+Na]^+$ *m/z* 609.3589, $C_{38}H_{50}NaO_5$ requires 609.3556; $[\alpha]_D$ −77° (*c* 0.9, $CHCl_3$); UV λ_{max} (log ε): 322 (3.69), 260 (4.22), 218 (4.31); IR ν_{max} 3310, 2970, 2920, 1730, 1670, 1590, 1450, 1370, 1290, 1120 cm^{-1}; ^1H NMR and ^{13}C NMR (Table 1).

4.3.3. Symphonone A ***3***

Brown oil, HREIMS $[M+Na]^+$ *m/z* 623.3342, $C_{38}H_{48}NaO_6$ requires 623.3349; $[\alpha]_D$ −37° (*c* 0.7, $CHCl_3$); UV λ_{max} (log ε): 320 (3.52), 276 (3.84), 261 (3.91), 255 (3.86), 229 (4.15); IR ν_{max} 3290, 2970, 2930, 1730, 1660, 1590, 1440, 1370, 1290, 1120 cm^{-1}; ^1H NMR and ^{13}C NMR (Table 1).

4.3.4. Symphonone B ***4***

Brown oil, HREIMS $[M+Na]^+$ *m/z* 693.4116, $C_{43}H_{58}NaO_6$ requires 693.4131; $[\alpha]_D$ −81° (*c* 1.0, $CHCl_3$); UV λ_{max} (log ε): 318 (3.70), 274 (4.03), 234 (4.05); IR ν_{max} 3300, 2970, 2930, 1730, 1670, 1590, 1440, 1370, 1290, 1120 cm^{-1}; ^1H NMR and ^{13}C NMR (Table 1).

4.3.5. Symphonone C ***5***

Brown oil, $[\alpha]_D$ −67° (*c* 1.0, $CHCl_3$); HREIMS $[M+Na]^+$ *m/z* 641.3444, $C_{38}H_{50}NaO_7$ requires calc. 641.3454; UV λ_{max} (log ε): 320 (3.51), 274 (3.85), 258 (3.91), 234 (3.91); IR ν_{max} 3350, 2970, 2920, 1730, 1650, 1590, 1440, 1370, 1290 cm^{-1}; ^1H NMR and ^{13}C NMR (Table 1).

4.3.6. 7-epi-Coccinone B ***6***

Brown oil, $[\alpha]_D$ −50° (*c* 0.9, $CHCl_3$); HREIMS $[M+Na]^+$ *m/z* 641.3463, $C_{38}H_{50}NaO_7$ (calc. 641.3454); UV λ_{max} (log ε): 318 (3.65), 276 (3.91), 261 (3.99), 258 (4.02), 237 (4.00); IR ν_{max} 3340, 2970, 2950, 1730, 1660, 1590, 1440, 1370, 1290, 1220 cm^{-1}; ^1H NMR and ^{13}C NMR (Table 1).

4.3.7. Symphonone D ***7***

Brown oil, $[\alpha]_D$ −41° (*c* 0.4, $CHCl_3$); HREIMS $[M+Na]^+$ *m/z* 659.3582, $C_{38}H_{52}NaO_8$ (calc. 659.3560); UV λ_{max} (log ε): 319 (3.85), 276 (4.14), 260 (4.25), 255 (4.28), 250 (4.26), 238 (4.27); IR ν_{max} 3400, 2970, 2930, 1720, 1590, 1440, 1370, 1290, 1190 cm^{-1}; ^1H NMR and ^{13}C NMR (Table 1).

4.3.8. Symphonone E ***8***

Brown oil, $[\alpha]_D$ −50° (*c* 0.4,$CHCl_3$); HREIMS $[M+Na]^+$ *m/z* 659.3563, $C_{38}H_{52}NaO_8$ (calc. 659.3560); UV λ_{max} (log ε): 320 (3.82), 274 (4.12), 260 (4.19), 255 (4.22), 250 (4.21), 236 (4.23); IR ν_{max} 3370, 2970, 2920, 1730, 1590, 1440, 1370, 1290, 1220 cm^{-1}; ^1H NMR and ^{13}C NMR (Table 1).

4.3.9. 7-epi-Garcinol ***9***

Yellow oil, $[\alpha]_D$ −86° (*c* 0.8, $CHCl_3$); HREIMS $[M+Na]^+$ *m/z* 625.3517, $C_{38}H_{50}NaO_6$ requires 625.3505; UV λ_{max} (log ε): 280 (4.32), 230 (4.17); IR ν_{max} 3300, 2920, 1720, 1640, 1520, 1440, 1375, 1290, 1200 cm^{-1}; ^1H NMR and ^{13}C NMR (Table 1).

4.3.10. Symphonone F ***10***

Brown oil, $[\alpha]_D$ −9° (*c* 1.0, $CHCl_3$); HREIMS $[M+Na]^+$ *m/z* 641.3459, $C_{38}H_{50}NaO_7$ (calc. 641.3454); UV λ_{max} (log ε): 322 (3.77), 261 (4.05), 236 (4.12); IR ν_{max} 3370, 2980, 2920, 1730, 1590, 1440, 1370, 1290, 1190 cm^{-1}; ^1H NMR and ^{13}C NMR (Table 1).

4.3.11. Symphonone G ***11***

Brown oil, $[\alpha]_D$ −4° (*c* 0.4, $CHCl_3$); HREIMS $[M+Na]^+$ *m/z* 641.3450, $C_{38}H_{50}NaO_7$ (calc. 641.3454); UV λ_{max} (log ε): 322 (3.77), 261 (4.05), 236 (4.12); IR ν_{max} 3330, 2980, 2920, 1730, 1590, 1450, 1370, 1280, 1190 cm^{-1}; ^1H NMR and ^{13}C NMR (Table 1).

4.3.12. Symphonone H ***12***

Yellow oil, $[\alpha]_D$ −37° (*c* 0.2, $CHCl_3$); HREIMS $[M+Na]^+$ *m/z* 623.3355, $C_{38}H_{48}NaO_6$ requires 623.3349; UV λ_{max} (log ε): 349 (3.60), 250 (3.94); IR ν_{max} 3280, 2960, 2920, 1730, 1680, 1620, 1470, 1390, 1290, 1140 cm^{-1}; ^1H NMR and ^{13}C NMR (Table 1).

4.3.13. Symphonone I ***13***

Yellow oil, $[\alpha]_D$ −22° (*c* 0.4, $CHCl_3$); HREIMS $[M-H]^-$ *m/z* 599.3470, $C_{38}H_{47}O_6$ requires 599.3373; UV λ_{max} (log ε): 344 (3.54), 254 (3.84), 230 (3.98); IR ν_{max} 3320, 2960, 2920, 1720, 1620, 1590, 1470, 1380, 1290, 1120 cm^{-1}; ^1H NMR and ^{13}C NMR (Table 1).

4.4. Biological activities

The chloroquine-resistant strain FcB1/Colombia of *P. falciparum* was maintained *in vitro* on human erythrocytes in RPMI 1640 medium supplemented by 8% (v/v) heat-inactivated human serum, at 37 °C, under an atmosphere of 3% CO_2, 6% O_2, 91% N_2. *In vitro* drug susceptibility was measured by [^{3}H]-hypoxanthine incorporation as described (Guillon et al., 2004). Stock solutions of drugs were prepared in DMSO. Compounds were serially diluted twofold with 100 µl of culture medium in 96-well plates. Asynchronous parasite cultures (100 µl, 1% parasitaemia and 1% final hematocrite) were then added to each well and incubated for 24 h at 37 °C prior to the addition of 0.5 µCi of [^{3}H]-hypoxanthine (GE Healthcare, France, 1–5 Ci mmol/ml) per well. After a further incubation of 24 h, plates were frozen and thawed. Cell lysates were then collected onto glass-fiber filters and counted in a liquid scintillation spectrometer. The growth inhibition for each drug concentration was determined by comparison of the radioactivity incorporated in the treated culture with that in the control culture maintained on the same plate. The concentration causing 50% growth inhibition (IC_{50}) was obtained from the drug concentration–response curve and the results were expressed as the mean values ± standard deviations determined from three independent experiments.

The human diploid embryonic lung cells MRC-5 were seeded into 96-well microplates at 2000 cells per well. The cytotoxicity assays were performed according to a published procedure (Tempete et al., 1995). Taxotere® was used as a control compound.

Acknowledgements

This works was supported by an ICSN-CNRS grant to one of us (G.M.). We are very grateful to Marie-Françoise Prévost (IRD) for her assistance in the identification of the species, Marie-Thérèse Martin for her help in NMR studies, Geneviève Aubert (ICSN) who performed the cytotoxic assays on MRC-5 cells, Olivier Laprévote and Isabelle Schmitz-Afonso for their helpful advises and technical support.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.phytochem.2010.03.008.

References

Bourdy, G., Willcox, M.L., Ginsburg, H., Rasoanaivo, P., Graz, B., Deharo, E., 2008. Ethnopharmacology and malaria: new hypothetical leads or old efficient antimalarials? Int. J. Parasitol. 38, 33–41.

Bousserouel, H., Litaudon, M., Morleo, B., Martin, M.T., Thoison, O., Nosjean, O., Boutin, J.A., Renard, P., Sevenet, T., 2005. New biologically active linear triterpenes from the bark of three new-caledonian Cupaniopsis species. Tetrahedron 61, 845–851.

Chien, S.-C., Chyu, C.-F., Chang, I.S., Chiu, H.-L., Kuo, Y.-H., 2008. A novel polyprenylated phloroglucinol, garcinialone, from the roots of *Garcinia multiflora*. Tetrahedron Lett. 49, 5276–5278.

Ciochina, R., Grossman, R.B., 2006. Polycyclic polyprenylated acylphloroglucinols. Chem. Rev. 106, 3963–3986.

de Oliveira, C.M.A., Porto, A., Bittrich, V., Vencato, I., Marsaioli, A.J., 1996. Floral resins of *Clusia* spp.: chemical composition and biological function. Tetrahedron Lett. 37, 6427–6430.

Dick, C.W., Abdul-Salim, K., Bermingham, E., 2003. Molecular systematic analysis reveals cryptic tertiary diversification of a widespread tropical rain forest tree. Am. Nat. 162, 691–703.

Gay, F., Ciceron, L., Litaudon, M., Bustos, D., Astagneau, P., Diquet, B., Danis, B., Gentilini, M., 1994. *In vitro* resistance of *Plasmodium falciparum* to qinghaosu derivatives in West Africa. Lancet 343, 350–351.

Germeraad, J.H., Hopping, C.A., Muller, J., 1968. Palynology of tertiary sediments from tropical areas. Rev. Palaeobot. Palynol. 6, 189–348.

Grenand, P., Moretti, C., Jacquemin, H., Prévost, M.F., 2004. Pharmacopées Traditionnelles en Guyane, IRD Éditions, Paris, 317.

Guillon, J., Grellier, P., Labaied, M., Sonnet, P., Leger, J.-M., Deprez-Poulain, R., Forfar-Bares, I., Dallemagne, P., Lemaitre, N., Pehourcq, F., Rochette, J., Sergheraert, C., Jarry, C., 2004. Synthesis, antimalarial activity, and molecular modeling of new pyrrolo[1,2-a]quinoxalines, bispyrrolo[1,2-a]quinoxalines, bispyrido[3,2-e]pyrrolo[1,2-a]pyrazines, and bispyrrolo[1,2-a]thieno[3,2-e]pyrazines. J. Med. Chem. 47, 1997–2009.

Gustafson, K.R., Blunt, J.W., Munro, M.H.G., Fuller, R.W., McKee, T.C., Cardellina, I.I., John, H., McMahon, J.B., Cragg, G.M., Boyd, M.R., 1992. The guttiferones, HIV-inhibitory benzophenones from *Symphonia globulifera*, *Garcinia livingstonei*, *Garcinia ovalifolia* and *Clusia rosea*. Tetrahedron 48, 10093–10102.

Hamed, W., Brajeul, S., Mahuteau-Betzer, F., Thoison, O., Mons, S., Delpech, B., Hung, N.V., Sevenet, T., Marazano, C., 2006. Oblongifolins A–D, polyprenylated benzoylphloroglucinol derivatives from *Garcinia oblongifolia*. J. Nat. Prod. 69, 774–777.

Laufer, M.K., Plowe, C.V., 2004. Withdrawing antimalarial drugs: impact on parasite resistance and implications for malaria treatment policies. Drug Resist. Update 7, 279–288.

Marti, G., Litaudon, M., Moretti, C., Susplugas, S., Grellier, P., Thoison, O., Hue, N., Guéritte, F., 2008. Role of LC/UV/MS2 in the search for antiplasmodial polycyclic polyprenylated acylphloroglucinols. In: Seventh Joint Meeting of GA, AFERP, ASP, PSE and SIF, vol. 74. Planta Med. Athen.

Marti, G., Eparvier, V., Moretti, C., Susplugas, S., Prado, S., Grellier, P., Retailleau, P., Guéritte, F., Litaudon, M., 2009. Antiplasmodial benzophenones from the trunk latex of *Moronobea coccinea* (Clusiaceae). Phytochemistry 70, 75–85.

Masullo, M., Bassarello, C., Suzuki, H., Pizza, C., Piacente, S., 2008. Polyisoprenylated benzophenones and an unusual polyisoprenylated tetracyclic xantone from the fruits of *Garcinia cambodgia*. J. Agric. Food Chem. 56, 5205–5210.

Ngouela, S., Lenta, B.N., Noungoue, D.T., Ngoupayo, J., Boyom, F.F., Tsamo, E., Gut, J., Rosenthal, P.J., Connolly, J.D., 2006. Anti-plasmodial and antioxidant activities of constituents of the seed shells of *Symphonia globulifera* Linn f. Phytochemistry 67, 302–306.

Nkengfack, A.E., Mkounga, P., Fomum, Z.T., Meyer, M., Bodo, B., 2002a. Globulixanthones A and B, two new cytotoxic xanthones with isoprenoid groups from the root bark of *Symphonia globulifera*. J. Nat. Prod. 65, 734–736.

Nkengfack, A.E., Mkounga, P., Meyer, M., Fomum, Z.T., Bodo, B., 2002b. Globulixanthones C, D and E: three prenylated xanthones with antimicrobial properties from the root bark of *Symphonia globulifera*. Phytochemistry 61, 181–187.

Noedl, H., Se, Y., Schaecher, K., Smith, B.L., Socheat, D., Fukuda, M.M., 2008. The artemisinin resistance in Cambodia 1 study, C.: evidence of artemisinin-resistant malaria in Western Cambodia. N. Engl. J. Med. 359, 2619–2620.

Peters, S., Schmidt, W., Beerhues, L., 1997. Regioselective oxidative phenol couplings of 2,3′,4,6-tetrahydroxybenzophenone in cell cultures of *Centaurium erythraea* RAFN and *Hypericum androsaemum* L. Planta 204, 64–69.

Piccinelli, A.L., Cuesta-Rubio, O., Chica, M.B., Mahmood, N., Pagano, B., Pavone, M., Barone, V., Rastrelli, L., 2005. Structural revision of clusianone and 7-*epi*-clusianone and anti-HIV activity of polyisoprenylated benzophenones. Tetrahedron 61, 8206–8211.

Sang, S., Liao, C.-H., Pan, M.-H., Rosen, R.T., Lin-Shiau, S.-Y., Lin, J.-K., Ho, C.-T., 2002. Chemical studies on antioxidant mechanism of garcinol: analysis of radical reaction products of garcinol with peroxyl radicals and their antitumor activities. Tetrahedron 58, 10095–10102.

Tempete, C., Werner, G.H., Favre, F., Rojas, A., Langlois, N., 1995. *In vitro* cytostatic activity of 9-demethoxyporothramycin B. Eur. J. Med. Chem. 30, 647–650.

Molecules **2010**, *15*, 7106-7114; doi:10.3390/molecules15107106

OPEN ACCESS

molecules

ISSN 1420-3049

www.mdpi.com/journal/molecules

Article

A New Xanthone from the Bark Extract of *Rheedia acuminata* and Antiplasmodial Activity of Its Major Compounds

Guillaume Marti [1], Véronique Eparvier [2,*], Marc Litaudon [1], Philippe Grellier [3] and Françoise Guéritte [1]

[1] Centre de Recherche de Gif, Institut de Chimie des Substances Naturelles, CNRS, 1 avenue de la Terrasse, 91198 Gif-sur-Yvette Cedex, France; E-Mails: guillaume.marti@icsn.cnrs-gif.fr (G.M.); marc.litaudon@icsn.cnrs-gif.fr (M.L.); francoise.gueritte@icsn.cnrs-gif.fr (F.G.)

[2] UPS2561, CNRS, 2, avenue Gustave Charlery, 97300 Cayenne, France

[3] Muséum National d'Histoire Naturelle, FRE 3206 CNRS, CP52, 61 rue Buffon, 75231 Paris cedex 05, France; E-Mail: grellier@mnhn.fr (P.G.)

* Author to whom correspondence should be addressed; E-Mail: veronique.eparvier@guyane.cnrs.fr; Tel.: +594-594-29-75-16; Fax: +594-594-28-47-86.

Received: 21 August 2010; in revised form: 23 September 2010 / Accepted: 3 October 2010 / Published: 14 October 2010

Abstract: Bioassay-guided fractionation of the ethyl acetate bark extract of *Rheedia acuminata* led to the isolation of the new compound 1,5,6-trihydroxy-3-methoxy-7-geranyl-xanthone (**1**), together with four known compounds **2-5**. These compounds were tested *in vitro* for their antiplasmodial activity on a chloroquine-resistant strain of *Plasmodium falciparum* (FcB1) and for their cytotoxicity against the human diploid embryonic lung cell line MRC-5.

Keywords: *Rheedia acuminata*, Clusiaceae; xanthones; antiplasmodial activity; cytotoxicity

1. Introduction

In South America several species of Clusiaceae are widely used in the manufacture of hulls, and well known for quality of their wood and for the healing properties of their latex used traditionally used for their effectiveness against dermatoses [1]. One of these, *Rheedia acuminata* (Ruiz & Pavon) Planchon and Triana, a tree growing in Amazonian rainforest, possesses an abundant latex used for

various medicinal purposes by the Guianese Amerindians (Palikur) in the form of patches or breakdowns applied to the wrinkling muscle. In addition, *Rheedia acuminata* fruits are commonly consumed in South America [1].

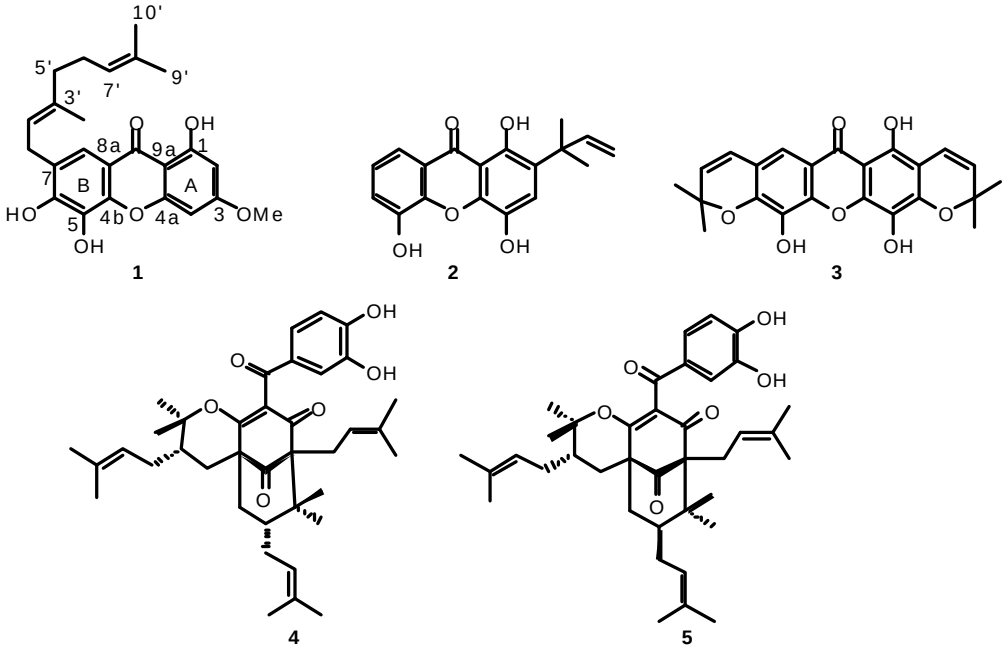
Prenylated xanthones and **p**oly**p**renylated **a**cyl**p**hloroglucinols (PPAPs) are widely distributed in the Clusiaceae, the genus *Rheedia* being a rich source of them [2-5]. Some biflavonoids were also previously isolated from this species [6]. Their biological activities include antibacterial, analgesic and cytotoxic properties [5,7,8].

In an effort to find new natural antimalarial drugs and as part of investigation of French Guiana plants, we found that the ethyl acetate extract of the bark of *Rheedia acuminata* showed antiplasmodial activity (92% of inhibitory growth of *Plasmodium falciparum* (FcB1) at 10 µg/mL), whereas leaves and fruits extracts showed no significant and weak activity respectively (28% and 50% of inhibitory growth at 10 µg/mL, respectively). We report here the bioassay-guided fractionation of this extract on the basis of this antiplasmodial activity.

2. Results and Discussion

The bioguided fractionation of the ethyl acetate extract of trunk bark of *Rheedia acuminata* led to the isolation of the new 1,5,6-trihydroxy-3-methoxy-7-geranyl-xanthone (**1**) along with two xanthone analogues (**2** and **3**) and two PPAPs (**4** and **5**) (Figure 1).

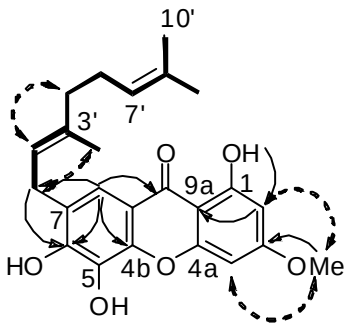
Figure 1. Structures of compounds **1**-**5**.



Compound **1** was obtained as a yellow oil. The HREIMS indicated a $[M+H]^-$ ion peak at *m/z* 409.1652, giving the molecular formula $C_{24}H_{25}O_6$ (calc. 409.1651). The IR spectrum displayed free hydroxyl (3,410 cm^{-1}), chelated hydroxyl (3,170 cm^{-1}), conjugated carbonyl (1,640 cm^{-1}) and aromatic ring (1,570 cm^{-1}) peaks. These data, together with absorption bands observed at λ_{max} 251, 283 and 327 nm in the UV spectrum and those obtained from 1D and 2D NMR experiments of compound **1**

suggested the presence of a 1,3,5,6-tetraoxygenated xanthone system [10,11]. Examination of the ^{1}H- NMR spectrum showed the presence of three hydroxyl groups with two broad signals at δ_H 7.88, 12.68 and one singlet at δ_H 14.02 ppm, corresponding to a chelated proton, one singlet at δ_H 8.08, two *meta*-coupled protons at δ_H 6.56 and 6.04 (1H, d, J = 2.3 Hz), and one methoxyl (δ_H 3.63). In addition two vinyl protons at δ_H 5.77 and 5.26 (1H, *brt*, J = 6.8 Hz), three methylene groups at δ_H 3.83 (2H, *d*, J = 7.3 Hz), 2.21 (2H, *m*) and 2.15 (2H, *m*), and three methyl groups at δ_H 1.84, 1.69, 1.59 (3H, *s*) suggested the presence of a geranyl side chain. HSQC experiments allowed the assignment of all protonated carbons. A combination of HMBC, COSY and NOESY experiments were used to establish the position of the substituents. In the HMBC spectrum (Figure 2), the chelated proton OH-1 showed correlations with C-2 at δ_c 97.6 on one hand, and the proton H-2 with carbons C-9a (δ_C 104.2), C-4 (δ_C 92.8), C-1 (δ_C 164.7) and C-3 (δ_C 166.7) on the other hand. A methoxy group was deduced from the correlation between the methyl protons at δ_H 3.63 with C-3. NOESY correlations between the methoxy at C-3 with protons H-2 (δ_H 6.56) and H-4 (δ_H 6.04) confirmed the position of the substituents on ring A (Figure 2). Furthermore, HMBC correlations observed between H-8 and C-9 (δ_C 181.5) were indicative of a *peri* location of the carbonyl. In addition, proton H-8 showed correlations with C-6 (δ_C 153.4), C-4b (δ_C 146.6), and CH_2-1' (δ_C 29.5) and a long-range correlation with C9a (δ_C 104.2). The geranyl side chain was deduced from the COSY spectrum with, on one hand, correlations observed between Me-9' and Me-10' [δ_H 1.59 (3H, *s*, H-9') and 1.69 (3H, *s*, H-10')] and the vinylic proton H-7' at δ_H 5.28 (1H, *brt*, J = 6.8 Hz), which in turn showed a cross peak with H_2-6'(δ_H 2.21, 2H, *m*). On the other hand, correlations observed between H_2-5' (δ_H 2.16, 2H, *m*) and H_2-6'and H-2' at δ_H 5.77 (1H, *brt*, J = 6.8 Hz), which in turn showed correlations with Me-4'(δ_H 1.84, 3H, *s*) and H_2-1'(δ_H 3.83, 2H, *d*, J = 7.3 Hz) confirmed the presence of a geranyl side chain. The *E* configuration of the C-2'-C-3' double bond was established by NOESY correlations observed between H_2-5' with H-2', and Me-4'with H_2-1' (Figure 2). Finally, the position of the geranyl side chain at C-7 (δ_C 127.8) was confirmed by HMBC correlation from H_2-1' to C-6 (δ_c 153.4) (Figure 2). Compound **1**, which was named 1,5,6-trihydroxy-3-methoxy-7-geranylxanthone, is a positional isomer of cowaxanthone and rubraxanthone isolated from *Garcinia cowa* [12,13].

Figure 2. Key COSY (bold), HMBC (plain arrows) and NOESY (dashed arrows) correlations for **1**.



Compound **2** was identified as 2-(1'-1'-dimethylprop-2'-enyl)-1,4,5-trihydroxyxanthone and compound **3** as pyrojacareubin by comparison of the UV, HRESIMS, 1D and 2D NMR with the

literature data [13-15]. The 1D and 2D NMR spectra associated with their optical rotation values confirmed that **4** was isogarcinol [16-18] and **5**, 7-*epi*-isogarcinol [16].

The antiplasmodial activity against the chloroquine-resistant strain of *P. falciparum* FcB1 and the cytotoxicity upon the human diploid embryonic cell line MRC-5 of compounds **1-5** are summarized in Table 1. The three xanthones **1**, **2** and **3** isolated from *Rheedia acuminata* bark have IC_{50}s against *Plasmodium falciparum* exceeding 10 µM. Compounds **4** and **5** have shown the best activity with IC_{50}s around 3 µM. The IC_{50} values obtained on MRC-5 cell line indicated that PPAPs **4** and **5** exhibited significant cytotoxicity whereas xanthones (**1** to **3**) showed weak cytotoxicity. Several xanthones, which have been isolated from various species of the family Clusiaceae, showed cytotoxic and antiplasmodial activities [9,10,19,20] and structure-activity relationships have been proposed in this series. For example, Winter *et al.* have demonstrated that xanthones with a hydroxyl group in *peri* positions with regards to the carbonyl (such as **1**, **2** and **3)**, possess weak antiplasmodial activity due to their low affinity for the heme [21,22]. In conclusion, the antiplasmodial activity detected in the crude extract is probably due to the presence of high quantity of the two PPAPs **4** and **5**.

Table 1. IC_{50} values of compounds **1-5** tested against FcB1 strain of *P. falciparum* and MRC-5 cell.

Compound	IC_{50} *P.f.* FcB1 (µM) ± SD (n=3)	IC_{50} MRC-5 (µM) ± SD (n=3)
1	10.5[a]	36[a]
2	15.1[a]	b
3	11.4[a]	29[a]
4	3.5 ± 1.1	3.5 ± 0.4
5	3.2 ± 1.3	2.3 ± 0.5
Chloroquine	0.078 ± 0.006	
Taxotere		31.5 ± 4.5

[a] biological assays in duplicates; [b] not enough material.

3. Experimental

3.1. General

The NMR spectra were recorded on a Bruker 500 MHz (Avance 500) or 300 MHz (Aspect DPX 300 MHz) spectrometer. ESIMS were obtained on a Thermoquest Navigator mass spectrometer. HRESIMS were obtained on a ESI-TOF spectrometer (LCT; Waters). Kromasil analytical, semi-preparative and preparative C-18 columns (250 × 4.5 mm; 250 × 10 mm and 250 × 21.2 mm I.D, 5 µM Thermo®) were used for preparative HPLC separations using a "Waters autopurification system" equipped with a sample manager (Waters 2767), a column fluidics organizer, a binary pump (Waters 2525), a UV-Vis diode array detector (190-600nm), Waters 2996) and PL-ELS 1000 ELSD detector Polymer laboratory. IR spectra were obtained on a Nicolet FTIR 205 spectrophotometer. The UV spectra were recorded on a Perkin-Elmer Lambda 5 spectrophotometer. Specific rotation was obtained in $CHCl_3$ with a JASCO P-1010 polarimeter. Silica gel 60 (35–70 µm) and analytical TLC plates (Si gel 60 F 254) were purchased from SDS (France). Polyamide DC 6 and polyamide cartridge were purchased from Macherey-Nagel (Chromabond PA, 1 g).

3.2. Plant Material

Trunk bark of *R. acuminata* were collected in French Guiana. A voucher specimen (PMF-258) was deposited in the Herbarium of Cayenne (French Guiana), and identified by M-F. Prévost (IRD).

3.3. Extraction and Isolation

Trunk bark (370 g) were extracted three times with EtOAc (3 × 1L) at 40 °C and 1,450 psi on a Zippertex® static high-pressure high-temperature extractor developed in the ICSN Pilot Unit. This extract (10 g) was filtered on polyamide. Filtered extract (9.3 g) was subjected to silica gel chromatography using heptane/EtOAc mixtures (100/0-0/100). According to their TLC profile, 10 fractions were obtained (F1-F10). Fraction F4 (1.1 g), which showed antiplasmodial activity, was subjected to a preparative C-18 column using an isocratic mobile phase consisting of ACN/water + 0.1 % formic acid (flow rate 21 mL/min). This resulted in the isolation of **4** (72 mg, 0.019%), **5** (65 mg, 0.017%), **3** (2 mg, 0.0008%), **1** (8 mg, 0.0021%) and **2** (18 mg, 0.004%) with retention times of 15.5, 13.5, 10.4, 9.4, and 4.6 mins, respectively.

*1,5,6-Trihydroxy-3-methoxy-7-geranylxanthone (**1**):* Yellow oil; UV (MeOH) λ_{max} (log ε): 327 (3.99), 283 (3.96), 251 (4.35); IR ν_{max} (ns) 3410, 1640, 1570, 1450, 1370, 1290, 1220 cm^{-1}; ^{1}H-NMR (pyridine-d_5, 500 MHz): δ 14.02 (1H, *s*, OH-1), 12.68 (1H, *brs*, OH), 8.08 (1H, *s*, H-8), 7.88 (1H, *brs*, OH) 6.56 (1H, *d*, *J* = 2.3 Hz, H-2), 6.04 (1H, *d*, *J* = 2.3 Hz, H-4), 5.77 (1H, *brt*, *J* = 6.8 Hz, H-2'), 5.28 (1H, *brt*, *J* = 6.8 Hz, H-7'), 3.83 (2H, *brd*, *J* = 7.3 Hz, H-1'), 3.63 (3H, *s*, OMe-3), 2.21 (2H, *m*, H-6'), 2.15 (2H, *m*, H-5'), 1.84 (3H, *s*, H-4'), 1.69 (3H, *s*, H-10'), 1.59 (3H, *s*, H-9'); ^{13}C-NMR (pyridine-d_5, 125 MHz): δ 181.5 (C-9), 166.7 (C-3), 164.7 (C-1), 158.5 (C-4a), 153.4 (C-6), 146.6 (C-4b), 137.2 (C-3'), 133.9 (C-5), 131.8 (C-8'), 127.8 (C-7), 125.3 (C-7'), 123.5 (C-2'), 116.7 (C-8), 113.8 (C-8a), 104.2 (C-9a), 97.6 (C-2), 92.8 (C-4), 56.2 (O-CH_3), 40.5 (C-5'), 29.5 (C-1'), 27.5 (C-6'), 26.2 (C-9'), 17.7 (C-4'), 18.2 (C-10'); HREIMS $[M-H]^-$ *m/z* 409.1697, $C_{24}H_{26}O_6$ requires 409.1651.

2-(1',1'-Dimethylprop-2'-enyl)-1,4,5-Trihydroxyxanthone (**2**): Yellow powder; UV (MeOH) λ_{max} (log ε): 320 (3.61), 263 (4.12), 249 (4.12), 236 (4.07); IR ν_{max} (ns) 3570, 1720, 1680, 1540, 1270, 1120 cm^{-1}; ^{1}H-NMR (pyridine-d_5, 500 MHz): δ 13.55 (1H, *s*, -OH), 7.99 (1H, *dd*, *J* = 8.0 Hz, H-8), 7.72 (1H, *s*, H-3), 7.53 (1H, *dd*, *J* = 8.3, 1.5 Hz, H-6), 7.27 (1H, *t*, *J* = 8.0 Hz, H-7), 6.47 (1H, *dd*, *J* = 17.1, 10.4 Hz, H-2'), 5.15 (2H, *m*, H-3'), 1.66 (6H, H-4' , H-5'); ^{13}C- NMR (pyridine-d_5, 125 MHz): δ 184.4 (C-9), 153.2 (C-1), 149.7 (C-5), 148.7 (C-2'), 148.6 (C-4b), 144.2 (C-4a), 138.1 (C-4), 128.7 (C-2), 124.7 (C-3), 124.6 (C-7), 123.4 (C-8a), 122.2 (C-6), 116.1 (C-8), 112.3 (C-3'), 110.1 (C-9a), 41.1 (C-1'), 27.2 (C-4', C-5'); HREIMS $[M-H]^-$ *m/z* 311,0934, $C_{18}H_{16}O_5$ requires 311.0919.

Pyrojacareubine (**3**): Orange-yellow powder; UV (MeOH) λ_{max} (log ε): 329 (3.42), 276 (4,18), 220 (4.08); IR ν_{max} (ns) 3330, 1650, 1580, 1490, 1450, 1290, 1220 cm^{-1}; ^{1}H-NMR (pyridine-d_5, 500 MHz): δ 13.05 (1H, *s*, 1-OH), 7.46 (1H, *s*, H-8), 6.90 (1H, *d*, *J* = 10.0 Hz, H-2'), 6.45 (1H, *d*, *J* = 10.1 Hz, H-7'), 6.27 (1H, *s*, H-4), 5.74 (1H, *d*, *J* = 10.1 Hz, H-8'), 5.62 (1H, *d*, *J* = 10.0 Hz, H-3'), 1.55 (6H, *s*, H-10', H-11'), 1.50 (6H, *s*, H-5', H-6'); ^{13}C-NMR (pyridine-d_5, 125 MHz): δ 180.8 (C-9), 164.2 (C-1), 161.5 (C-3), 146.4 (C-4a), 131.0 (C-8'), 127.2 (C-3'), 121.4 (C-7'), 118.2 (C-7), 115.2 (C-2'), 113.5

(C-8), 110.4 (C-8a), 103.3 (C-2), 101.6 (C-9a), 99.4 (C-4), 78.9 (C-9'), 78.2 (C-4'), 28.4 (C-10', C-11'), 28.3 (C-5', C-6'); HREIMS $[M-H]^-$ *m/z* 391.1177, $C_{23}H_{19}O_6$ requires 391.1182.

Isogarcinol (**4**): Brown powder; $[\alpha]^{25}_D = -160$ (*c* = 1.0, $CHCl_3$); UV (MeOH) λ_{max} (log ε): 317 (3.82), 277 (4,14), 233 (4.07); IR ν_{max} (ns) 3290, 2920, 2850, 1730, 1670, 1590, 1520, 1440, 1370, 1290, 1170 cm^{-1}; ^{1}H- NMR (pyridine-d_5, 500 MHz): δ 8.05 (1H, *d*, *J* = 2.0 Hz, H-12), 7.68 (1H, *dd*, *J* = 8.1, 2.0 Hz, H-16), 7.28 (1H, *d*, *J* = 8.1 Hz, H-15), 5.42 (1H, *tl*, *J* = 6.5 Hz, H-18), 5.09 (2H, *tl*, *J* = 6.5 Hz, H-35, H-25), 3.27 (1H, *dd*, *J* = 13.9, 3.1 Hz, H-29), 3.21 (1H, *ddd*, *J* = 14.4, 10.7, 9.5 Hz, H-24), 2.76 (1H, *dd*, *J* = 13.7, 5.6 Hz, H-17), 2.43 (1H, *dl*, *J* = 14.1 Hz, H-8), 2.42 (1H, *dl*, *J* = 14.1 Hz, H-24), 2.11 (1H, *dd*, *J* = 14.1, 7.3 Hz, H-8), 1.96 (1H, *dl*, *J* = 14.1 Hz, H-34), 1.91 (3H, *s*, H-28), 1.82 (1H, *ddd*, *J* = 14.2, 9.5, 9.5 Hz, H-34), 1.74 (3H, *s*, H-27), 1.71 (3H, *s*, H-21), 1.68 (3H, *s*, H-37), 1.66 (1H, *dt*, *J* = 9.9, 5.0 Hz, H-30), 1.57 (3H, *s*, H-20), 1.56 (3H, *s*, H-38), 1.54 (1H, *m*, H-7), 1.30 (3H, *s*, H-23), 1.23 (3H, *s*, H-32), 1.14 (1H, *dd*, *J* = 13.9, 13.7 Hz, H-29), 1.07 (3H, *s*, H-33), 1.05 (3H, *s*, H-22); ^{13}C-NMR (pyridine-d_5, 125 MHz): δ 207.9 (C-9), 195.0 (C-4), 193.0 (C-10), 172.2 (C-3), 171.4 (C-2), 153.7 (C-14), 147.8 (C-13), 134.5 (C-19), 133.7 (C-36), 132.9 (C-26), 130.9 (C-11), 126.4 (C-25), 124.3 (C-16), 122.8 (C-35), 121.7 (C-18), 116.6 (C-12), 116.5 (C-15), 87.2 (C-31), 69.2 (C-5), 52.2 (C-1), 47.0 (C-7), 46.8 (C-6), 43.8 (C-30), 39.8 (C-8), 30.4 (C-34, C-24), 29.4 (C-33), 29.1 (C-29), 27.1 (C-22), 26.7 (C-17), 26.6 (C-20), 26.5 (C-27), 26.2 (C-37), 23.1 (C-23), 21.7 (C-32), 19.0 (C-28), 18.8 (C-21), 18.3 (C-38); HREIMS $[M-Na]^+$ *m/z* 625.3499, $C_{38}H_{50}O_6$ requires 625.3505.

7-epi-Isogarcinol (**5**): Brown powder; $[\alpha]^{25}_D = -158$ (*c* = 1.0, $CHCl_3$); UV (MeOH) λ_{max} (log ε): 319 (3.85), 276 (4,15), 233 (4.12); IR ν_{max} (ns) 3320, 2970, 2930, 1730, 1650, 1590, 1520, 1440, 1370, 1290, 1170 cm^{-1}; ^{1}H- NMR (pyridine-d_5, 500 MHz): δ 8.13 (1H, *sl*, H-12), 7.72 (1H, *dl*, *J* = 7.7 Hz, H-16), 7.31 (1H, *dl*, *J* = 7.6 Hz, H-15), 5.38 (1H, *tl*, *J* = 6.2 Hz, H-18), 5.18 (1H, *tl*, *J* = 6.2 Hz, H-25), 5.13 (1H, *tl*, *J* = 6.6 Hz, H-35), 3.28 (1H, *dd*, *J* = 13.9, 3.2 Hz, H-29), 2.97 (1H, *dd*, *J* = 13.6, 6.4 Hz, H-17), 2.77 (1H, *dd*, *J* = 13.6, 4.8 Hz, H-17), 2.44 (1H, *m*, H-7), 2.43 (1H, *m*, H-8), 2.26 (1H, *dd*, *J* = 14.1, 2.5 Hz, H-24), 1.96 (1H, *m*, H-34), 1.82 (1H, *dd*, *J* = 14.1, 8.5 Hz, H-24), 1.79 (1H, *m*, H-34), 1.77 (1H, *m*, H-8), 1.73 (3H, *s*, H-21), 1.71 (3H, *s*, H-37), 1.65 (1H, *m*, H-30), 1.62 (6H, *s*, H-27, H-28), 1.56 (3H, *s*, H-38)1.55 (3H, *s*, H-20), 1.27 (3H, *s*, H-23), 1.21 (1H, *dd*, *J* = 13.9, 13.7 Hz, H-29), 1.14 (3H, *s*, H-32), 1.08 (3H, *s*, H-33), 0.83 (3H, *s*, H-22); ^{13}C-NMR (pyridine-d_5, 125 MHz): δ 207.3 (C-9), 194.9 (C-4), 193.2 (C-10), 170.8 (C-2), 153.7 (C-14), 147.9 (C-13), 134.2 (C-19), 133.7 (C-36), 132.9 (C-26), 131.0 (C-11), 129.1 (C-3), 124.4 (C-16), 123.8 (C-25), 122.9 (C-35), 122.1 (C-18), 117.1 (C-12), 116.5 (C-15), 87.6 (C-31), 71.4 (C-5), 52.2 (C-1), 46.8 (C-6), 43.8 (C-30), 43.1 (C-8), 42.2 (C-7), 30.4 (C-34), 29.2 (C-33), 28.5 (C-24, C-29), 26.2 (C-27), 25.9 (C-17), 26.2 (C-37, C-20), 22.8 (C-23), 21.7 (C-32), 18.8 (C-21), 18.4 (C-28), 18.3 (C-38), 16.6 (C-22); HREIMS $[M-Na]^+$ *m/z* 625.3519, $C_{38}H_{50}O_6$ requires 625.3505.

3.4. Biological Activities

The extracts and compounds were tested against the chloroquine-resistant FcB1/ Colombia strain of *Plasmodium falciparum* in 96-well plates by measuring [3H]-hypoxanthine incorporation by parasite as previously described [23]. The growth inhibition for each compound concentration was determined by comparing the radioactivity incorporated in the treated culture with that in the control culture

maintained on the same plate. The concentrations causing 50% inhibition of parasite growth (IC50) were calculated from the drug concentration-response curves. Chloroquine® was used as a control compound.

The human diploid embryonic lung cells MRC-5 were seeded into 96-well microplates at 2000 cells per well. The cytotoxicity assays were performed according to a published procedure [24]. Taxotere® was used as a control compound.

4. Conclusions

A chemical investigation of *Rheedia acuminata* bark was carried out in the framework of a global investigation on French Guiana flora. This study showed that the bark contained a new xanthone, 1,5,6-trihydroxy-3-methoxy-7-geranylxanthone, together with 2-(1'-1'-dimethylprop-2'-enyl)-1,4,5-trihydroxyxanthone. Pyrojacareubin, isogarcinol and 7-*epi*-isogarcinol were isolated from the *Rheedia* genus, for the first time. The two PPAPs isolated from *Rheedia acuminata,* which exhibited cytotoxic and antiplasmodial properties, were likely responsible of the biological activity found in the crude extract.

Acknowledgements

This work was supported by an ICSN-CNRS grant. We are grateful to Marie-Françoise Prévost (IRD) for her help in the identification of plant collection, Marie-Thérese Martin for her assistance in the structural determination and Geneviève Aubert (ICSN) who performed the cytotoxic assays on MRC-5 cells. All work carried out and reported in this manuscript is based on the program "Valorization of French Guiana flora" (CNRS, IRD, UMR Ecofog) with the financial support of the "Institut de Chimie des Substances Naturelles" (CNRS, France).

References

1. Grenand, P.; Moretti, C.; Jacquemin, H.; Prévost, M-F. *Pharmacopées Traditionnelles en Guyane. Créoles*, *Palikur*, *Wayãpi*; 2nd ed.; IRD: Marseille, France, 2004; pp. 309-319.
2. Peres, V.; Nagem, T.J.; Oliveira, F.F. Tetraoxygenated naturally occurring xanthones. *Phytochemistry* **2000**, *55*, 683-710.
3. Peres, V.; Nagem, T.J. Trioxygenated naturally occurring xanthones. *Phytochemistry* **1997**, *44*, 191-214.
4. Torrico, F.; Velasco, P.; Gimenez, A.; Almanza, G.R. An antibacterial xanthone and triterpenes of *Rheedia acuminata*. *Rev. Boliv. Quim.* **2001**, *18*, 38-42.
5. Cao, S.; Brodies, P.J.; Miller, J.S.; Ratovoson, F.; Birkinshaw, C.; Randrianasolo, S.; Rakotobe, E.; Rasamison, V.E.; Kingston, D.G.I. Guttiferones K and L, Antiproliferative compounds of *Rheedia calicola* from the Madagascar rain forest. *J. Nat. Prod.* **2007**, *70*, 686-688.
6. Li, X-C.; Joshi, A.S.; Tan, B.; ElSohly, H.N.; Walker, L.A.; Zjawiony, J.K.; Ferreira, D. Absolute configuration, conformation, and chiral properties of flavanone-(3→8")-flavone biflavonoids from *Rheedia acuminata*. *Tetrahedron* **2002**, *58*, 8709-8717.

7. Almeida, L.S.B.; Murata, R.M.; Yatsuda, R.; Dos Santos, M.H.; Nagem, T.J.; Alencar, S.M.; Koo, H.; Rosalen, P.L. Antimicrobial activity of *Rheedia brasiliensis* and 7-epiclusianone against *Streptococcus mutans*. *Phytomedicine* **2008**, *15*, 886-891.
8. Verdi, L.G.; Pizzolatti, M.G.; Montanher, A.B.P.; Brighente, I.M.; Smânia Júnior, A.; Smânia, E.D.F.A.; Simionatto, E.L.; Delle Monache, F. Antibacterial and brine shrimp lethality tests of biflavonoids and derivatives of *Rheedia gardneriana. Fitoterapia* **2004**, *75*, 360-363.
9. Zelefack ,F.; Guilet, D.; Fabre, N.; Bayet, C.; Chevalley, S.; Ngouela, S.; Ndjakou Lenta, B.; Valentin, A.; Tsamo, E.; Dijoux-Franca, M-J. Cytotoxic and antiplasmodial xanthones from *Pentadesma butyracea*. *J. Nat. Prod.* **2009**, *72*, 954-957.
10. Han, A.R.; Kim, J.A.; Lantvit, D.D.; Kardono, L.B.S.; Riswan, S.; Chai, H.; De Blanco, E.J.C.; Farnsworth, N.R.; Swanson, S.M.; Kinghorn, A.D. Cytotoxic xanthone constituents of the stem bark of *Garcinia mangostana* (mangosteen). *J. Nat. Prod.* **2009**, *72*, 2029-2031.
11. Pattalung, P.; Thongtheeraparp, W.; Wiriyachitra, P.; Taylor, W.C. Xanthones of *Garcinia cowa*. *Planta Medica* **1994**, *60*, 365-368.
12. Lee, H-H.; Chan, H-K. 1,3,6-Trihydroxy-7-methoxy-8-(3,7-dimethyl-2,6-octadienyl)xanthone from *Garcinia cowa*. *Phytochemistry* **1977**, *16*, 2038-2040.
13. Sordat-Diserens, I.; Marston, A.; Hamburger, M.; Hostettman, K. Novel prenylated xanthones from *Garcinia gerrardii* Harvey. *Helv. Chim. Acta* **1989**, *72*, 1001-1007.
14. Monach, G.D.; Monache, F.D.; Waterman, P.G.; Crichton, E.G.; De Limas, R.A. Minor xanthones from *Rheedia gardenaria*. *Phytochemistry* **1984**, *23*, 1757-1759.
15. Harrison, L.J.; Leong, L.S.; Sia, G.L.; Sim, K.Y.; Tan, H.T.W. Xanthones from *Garcinia forbesii. Phytochemistry* **1993**, *33*, 727-728.
16. Marti, G.; Eparvier, V.; Moretti, C.; Susplugas, S.; Prado, S. ; Grellier, P.; Retailleau, P., Guéritte, F.; Litaudon M. Antiplasmodial benzophenones from the trunk latex of *Moronobea coccinea* (Clusiaceae). *Phytochemistry* **2009**, *70*, 75-85.
17. Ciochina, R.; Grossman, R.B. Polycyclic polyprenylated acylphloroglucinols. *Chem. Rev.* **2006**, *106*, 3963-3986.
18. Hamed, W.; Brajeul, S.; Mahuteau-Betzer, F.; Toison, O.; Mons, S.; Delpech, B.; Hung, N.V.; Sevenet, T.; Marazano, C. Polyprenylated benzoylphloroglucinol derivatives from *Garcinia oblongifolia*. *J. Nat. Prod.* **2006**, *69*, 774-777.
19. Azebaze, A.G.B.; Meyer, M.; Valentin, A.; Nguemfo, E.L.; Fomum, Z.T.; Nkengfack, A.E. Prenylated xanthone derivatives with antiplasmodial activity from *Allanblackia monticola* Staner L.C. *Chem. Pharm. Bull.* **2006**, *54*, 111-113.
20. Tao, S.J; Guan, S.H.; Wang, W.; Lu, Z.Q.; Chen, G.T.; Sha, N.; Yue, Q.X.; Liu, X.; Guo, D.A. Cytotoxic polyprenylated xanthones from the resin of *Garcinia hanburyi*. *J. Nat. Prod.* **2009**, *72*, 117-124.
21. Ignatushchenko, M.V.; Winter, R.W.; Riscoe, M. Xanthones as antimalarial agents: Stage specificity. *Am. J. Trop. Med. Hyg.* **2000**, *62*, 77-81.
22. Ignatushchenko, M.V.; Winter, R.W.; Bächinger, H.P.; Hinrichs, D.J. Riscoe, M.K. Xanthones as antimalarial agents; studies of a possible mode of action. *FEBS Letters* **1997**, 409, 67-73.

23. Frappier, F.; Jossang, A.; Soudon, J.; Calvo, F.; Rasoinevo, P.; Ratsimamanga-Urverg, S.; Saez, J.; Schrevel, J.; Grellier, P. Bisbenzylisoquinolines as modulators of chloroquine resistance in *Plasmodium falciparum* and multidrug resistance in tumor cells. *Antimicrob. Agents Chemother.* **1996**, *40*, 1476-1481.
24. Tempete, C.; Werner, G.H.; Favre, F.; Rojas, A.; Langlois, N. *In vitro* cytostatic activity of 9-demethoxyporothramycin B. *Eur. J. Med. Chem.* **1995**, *30*, 647-650.

Sample Availability: Samples of the compounds are available from the authors.

Metabolomics (2017) 13:96
DOI 10.1007/s11306-017-1227-6

ORIGINAL ARTICLE

Dereplication of natural products from complex extracts by regression analysis and molecular networking: case study of redox-active compounds from *Viola alba* subsp. *dehnhardtii*

Justine Chervin[1,2,4] · Pierre Perio[1] · Nathalie Martins-Froment[3] · Chiobouaphong Pharkeovilay[1] · Karine Reybier[1] · Françoise Nepveu[1] · Nicolas Fabre[1] · Thierry Talou[2] · Valérie Bonzon-Ponnet[4] · Guillaume Marti[1]

Received: 5 May 2017 / Accepted: 19 June 2017

Abstract

Introduction In natural product research, bioassay-guided fractionation was previously widely employed but is now judged to be inadequate in terms of time and cost, particularly if only known compounds are ultimately isolated. The development of metabolomics, along with improvements in analytical tools, allows comprehensive metabolite profiling. This enables dereplication to target unknown active compounds early in the purification workflow.

Objectives Starting from an ethanolic extract of violet leaves, this study aims to predict redox active compounds within a complex matrix through an untargeted metabolomics approach and correlation analysis.

Methods Rapid fractionation of crude extracts was carried out followed by multivariate data analysis (MVA) of liquid chromatography–high resolution mass spectrometry (LC–HRMS) profiles. In parallel, redox active properties were evaluated by the capacity of the molecules to reduce 2,2-diphenyl-1-picrylhydrazyl ($DPPH^{\cdot}$) and superoxide ($O_2^{\cdot -}$) radicals using UV–Vis and electron spin resonance spectroscopies (ESR), respectively. A spectral similarity network (molecular networking) was used to highlight clusters involved in the observed redox activities.

Results Dereplication on *Viola alba* subsp. *dehnhardtii* highlighted a reproducible pool of redox active molecules. Polyphenols, particularly *O*-glycosylated coumarins and *C*-glycosylated flavonoids, were identified and de novo dereplicated through molecular networking. Confirmatory analyses were undertaken by thin layer chromatography (TLC)–DPPH–MS assays and nuclear magnetic resonance (NMR) spectra of the most active compounds.

Conclusion Our dereplication strategy allowed the screening of leaf extracts to highlight new biologically active metabolites in few steps with a limited amount of crude material and reduced time-consuming manipulations.

Electronic supplementary material The online version of this article (doi:10.1007/s11306-017-1227-6) contains supplementary material, which is available to authorized users.

✉ Guillaume Marti
guillaume.marti@univ-tlse3.fr

Justine Chervin
justine.chervin@ensiacet.fr

Pierre Perio
pierre.perio@univ-tlse3.fr

Nathalie Martins-Froment
martins@chimie.ups-tlse.fr

Chiobouaphong Pharkeovilay
chiobouaphong.phakeovilay@univ-tlse3.fr

Karine Reybier
karine.reybier@univ-tlse3.fr

Françoise Nepveu
francoise.nepveu@univ-tlse3.fr

Nicolas Fabre
nicolas.fabre@univ-tlse3.fr

Thierry Talou
thierry.talou@ensiacet.fr

Valérie Bonzon-Ponnet
valerie.bonzonponnet@parfumsberdoues.com

[1] UMR 152 Pharmadev, Université de Toulouse, IRD, UPS, Toulouse, France

[2] Laboratoire de Chimie Agro-industrielle, Université de Toulouse, INRA, INPT, Toulouse, France

[3] Service commun de spectrométrie de masse, Université de Toulouse, ICT, UPS, Toulouse, France

[4] Berdoues SAS, Cugnaux, France

Published online: 03 July 2017

This approach could be applied to any kind of natural extract for the study of various biological activities.

Keywords Dereplication · Metabolomics · Molecular networking · Natural products · Redox active properties · UHPLC–HRMS

1 Introduction

In natural product research, identification and characterisation of bioactive compounds from crude extracts remains challenging. To identify active molecules, conventional treatments include extraction, bio-guided fractionation, isolation, and characterisation steps. However, these procedures are time consuming as the composition of natural extracts can be complex. There is also a risk of redundant results, such as isolating known compounds or loss of bioactivity during manipulations (Ayouni et al. 2016). Holistic methods such as metabolomics can enhance the classical reductionist workflow based on iterative bioassays and paralleled fractionation to reach active compounds. This approach, referred to as untargeted metabolite profiling, and subsequent statistical analysis was revealed as a valuable tool in natural product research (Cox et al. 2014).

Ultra-high performance liquid chromatography–high resolution mass spectrometry (UHPLC–HRMS) has been used to study complex mixtures, taking advantage of sensitive detection and high chromatographic and mass resolutions. This approach consists of the pre-identification of known compounds in a single analytical run, which is of particular interest in the dereplication process (Wolfender et al. 2015). As a preliminary step, chemical compound databases are interrogated based on accurate mass of all detected features (*m*/*z*-retention time (RT) pairs), which produces molecular formula hits. However, several possibilities may be obtained, mainly due to isobaric matches. Therefore, additional filters are needed to rank database hits, such as cross-research with chemotaxonomic data or tandem mass spectrometry (MS/MS) fragmentation patterns. Indeed, development of *in silico* tools to mirror predicted fragmentation behaviour with experimental data limits the number of matches for a given feature (Xiao et al. 2012; Kind and Fiehn 2016). Additionally, a mass spectral similarity network (e.g., molecular networking) was developed to help interpret LC–MS/MS spectra (Yang et al. 2013; Grapov et al. 2015), complementing this dereplication strategy with the propagation of assigned peaks (Allard et al. 2016).

MS/MS-based dereplication or de novo identification strategies are necessary to target active compounds early in the purification process as untargeted metabolite profiling methods generate tens to thousands of variables, depending of the analytical workflow (Wolfender et al. 2010). Multivariate data analysis (MVA) is commonly used to exploit the variable space in order to rank the features most involved in sample separation. Orthogonal partial least square (OPLS) regression models enable the prediction of features involved in a given activity for each sample (Worley and Powers 2013; Bylesjö et al. 2006).

Violets contain various natural products with diverse biological activities (Muhammad et al. 2012). These include anti-inflammatory, antimicrobial (Witkowska-Banaszczak et al. 2005) and, predominantly, antioxidant activity (Vukics et al. 2008). Previous work describes over 200 compounds in this genus, mainly flavonoids, terpenoids and phenylpropanoids (Zhu et al. 2015). However, only 30 of the 500 species were studied. Thus, a potential supply of new biologically active compounds remains to be discovered, particularly non-volatile metabolites. Many interesting activities are related to the redox active functions of the molecules. Monitoring the ability to reduce free radicals by a single electron reaction is straightforward and automatable in vitro (López-Alarcón and Denicola 2013). This makes it possible to identify reducing agents (also known as antioxidants), which are useful in liquid or soft medium as preservatives. In vivo, the antioxidant character of a compound is not predominantly linked to its ability to scavenge free radicals due to kinetic constraints. The redox property of the molecule, in its reversibility, is involved. It is the oxidised form that paradoxically activates the Nrf2 (NF-E2-related factor 2) signalling pathway and allows the expression of antioxidant enzymes and proteins (Forman et al. 2014). Thus evaluating in vitro reductants is the first step in identifying new structures capable of having active oxidised forms in vivo. The aim of this work was to identify metabolite clusters most probably involved in reduction reactions starting from a complex extract of leaves of violet of Toulouse (*Viola alba* subsp. *Dehnhardtii*, Violaceae).

2 Materials and methods

2.1 Plant material

In spring 2016, seven distinct flowerpots of violet of Toulouse (*Viola alba* subsp. *dehnhardtii*) were collected from the Toulouse Municipal Greenhouses, France. Samples were washed before separation of flowers, aerial parts and roots. To stabilise the vegetable matter, each plant part was lyophilised and ground into powder.

2.2 Leaf extraction

Metabolites were extracted by adding 10 volumes of 80% ethanol (EtOH) to the powdered leaf (10.0 ± 0.2 g). The

solutions were then sonicated in a bath (Fisher Scientific) at room temperature for 30 min and filtered. This procedure was repeated once with fresh solvent. Each extract was evaporated under reduced pressure (Buchi rotavapor R-114) and 100 mg of the crude extracts were finally suspended in 1 mL of methanol prior to solid phase extraction (SPE) (1 g Sep-Pak® C_{18} cartridge, Waters, Milford, MA, USA). Rapid fractionation was achieved using five aqueous methanolic solutions of increasing organic solvent concentrations (F1: 95/5 H_2O/MeOH; F2: 75/25; F3: 50/50; F4: 25/75; F5: 0/100). The five collected fractions of each extract were dried under reduced pressure. One part was put aside for redox properties analysis and the other part was dissolved to 1 mg/mL for UHPLC–HRMS analysis.

2.3 Redox properties assay

The reductive properties of the crude extracts and their respective fractions were firstly evaluated by the DPPH radical scavenging assay (Nguyen et al. 2013). Each sample was analysed in triplicate and five concentrations were tested in order to graphically determine the half maximal inhibitory concentration (IC_{50}). Trolox (Fluka, purity >98%) and rutoside (Sigma, purity >94%) were used as positive controls. Secondly, the capacity of the molecules to reduce the superoxide radical ($O_2^{\cdot -}$) was analysed by electron spin resonance (ESR) (Hubert et al. 2008). Rutoside was employed as a positive control. The results were analysed with WinEPR software (v. 2.11b, Bruker): a baseline correction was firstly done before second order integration of the signal.

2.4 UHPLC–HRMS profiling

All extracts were profiled using a UHPLC–DAD–LTQ Orbitrap XL instrument (Ultimate 3000, Thermo Fisher Scientific, Hemel Hempstead, UK). The UV detection was performed by a diode array detector (DAD) from 210 to 400 nm. Mass detection was performed using an electrospray source in positive ionisation (PI) and negative ionisation (NI) modes at 15,000 resolving power (full width at half maximum (FWHM) at 400 *m/z*). The mass scanning range was *m/z* 100–1500 Da. The capillary temperature was 300 °C and ISpray voltage was fixed at 4.2 kV (positive mode) and 3.0 kV (negative mode). Mass measurement was externally calibrated before starting the experiment. Each full MS scan was followed by data dependant MS/MS on the three most intense peaks using stepped collision-induced dissociation (CID) (35% normalised collision energy, isolation width 2 Da, activation Q 0.250). The LC–MS system was run in binary gradient mode using a BEH C_{18} Acquity column (100×2.1 mm i.d., 1.7 µm, Waters, MA, USA) equipped with a guard column. Mobile phase A (MPA) was 0.1% formic acid (FA) in water and mobile phase B (MPB) was 0.1% FA in acetonitrile. Gradient conditions were: 0 min, 95% MPA; 0.5 min 95% MPA; 12 min, 5% MPA; 15 min, 5% MPA, 15.5 min, 95% MPA; 19 min, 95% MPA. The flow rate was 0.3 mL/min, column temperature 40 °C and injection volume was 2 µL.

2.5 Data processing

The UHPLC–HRMS raw data were converted to abf files (Reifycs Abf Converter) and processed with MS-DIAL version 2.54 (Tsugawa et al. 2015) for mass signal extraction between 100 and 1500 Da from 0 to 12.5 min. Respective MS1 and MS2 tolerance were set to 0.01 and 0.4 Da in centroid mode. The optimised detection threshold was set to 2×10^4 for MS1 and 10 for MS2. Adducts and complexes were identified to exclude them from the final peak list. Finally, the peaks were aligned on a quality control (QC) reference file with a retention time tolerance of 0.1 min and a mass tolerance of 0.025 Da. The resulting peak list was then exported to comma-separated value (CSV) format prior to MVA using SIMCA-P+ (version 14.0, Umerics, Umea, Sweden).

2.6 Statistical analysis

Comma-separated value files were directly imported into SIMCA-P+ (version 14.0, Umerics, Umea, Sweden). For MVA, all data were log transformed and pareto scaled. The OPLS regression analysis was done with DPPH IC_{50} values as Y input. Coefficient scores were used to rank variables according to their reductive potential. For each model, a leave-one-subject-out cross-validation was performed to assess the model fit. The validity of the discriminant model was verified using permutation tests (Y-scrambling).

2.7 Identification of significant features

Molecular formulae of significant features were calculated with MS-FINDER 2.10 (Tsugawa et al. 2016). Various parameters were used in order to reduce the number of potential candidates, such as the element selection exclusively including C, H, O; mass tolerance fixed to 10 ppm and the isotopic ratio tolerance set to 20%. Only natural product databases focused on plants were selected from Universal Natural Products Database (UNPD), KNApSAc, PlantCyc, Dictionary of Natural Products (DNP, CRC press, v25:2) and CheBI. The results were presented as a list of compounds sorted according to the score value of the match. This value encompassed uncertainty on accurate mass, the isotopic pattern score and the experimental MS/MS fragmentation mirrored to *in silico* matches. Chemical classes were retained for identified features with a score

above 5 and only structures with a score above 7 were retained for thorough analysis.

2.8 Mass spectral similarity network

The text file format exported from MS-DIAL was cleaned-up by eliminating the identified adducts before importing into MetamapR (version 1.4.0) (Grapov et al. 2015). A mass spectral similarity network was created with a maximum of ten connections between nodes, a cut off fixed at 0.3 and a retention time filter at 1.5 min. The calculated edge list was then downloaded and processed with Cytoscape 2.8.3 (Shannon et al. 2003). An attribute file containing all processed information, in particular *m*/*z* values, OPLS coefficients and chemical classes of identified features was imported to improve network visualisation.

2.9 Validation of the model by TLC–DPPH–MS

A TLC method was developed to confirm the statistical results. TLC separation was undertaken for fractions F2 and F3 with respective migration solvent composition: ethyl acetate, formic acid, acetic acid, water (50:5.5:5.5:13.5) and ethyl acetate, formic acid, water (60:6.5:6.5). 675 µL of a 10 mg/mL solution was deposed in band of 180 mm in a 20 cm TLC plate. The plate was then placed in an oven at 70 °C for 15 min and then observed under UV at 254 nm for F2 and 366 nm for F3. A small part of the plate was revealed with a purple DPPH solution at 600 mol/L. Then, yellow active spots were collected on the non-treated surface with the use of a TLC–MS interface (Camag, Muttenz, Switzerland) and directly injected into the LTQ-Orbitrap instrument.

2.10 Purification of compounds of interest

Purification of active compounds was undertaken starting from 200 mg of F2. LC–UV separation was performed using a XBridge™ C_{18} prep column (4.6 × 150 mm i.d., 5 µm, Waters, MA, USA) on an HPLC-DAD–QTOF-MS instrument (HPLC Alliance 2695- QTof Premier, Waters, MA, USA). Mobile phase A (MPA) was 0.1% formic acid (FA) in water and mobile phase B (MPB) was 0.1% FA in acetonitrile. The linear gradient program was as follows: 0 min, 95% MPA; 15 min 75% MPA; 16 min, 50% MPA; 18 min, 50% MPA, 18.5 min, 95% MPA; 23.5 min, 95% MPA. The flow rate was 20 mL/min and injection volume was 250 µL. For all analyses, detection was performed by UV at 325 and 254 nm and collection was automatically done by filling 30% of the tube. Samples were kept at ambient temperature during the whole analysis.

All collected fractions were dried using a Speed-vac (SpeedVac plus, Thermo Savant™) and dissolved in DMSO-d_6 for NMR analysis (Bruker cryo 500 MHz, Germany) (see Supplementary Table S1).

3 Results

3.1 Radical scavenging properties in vitro

Redox properties of extracts were evaluated by studying their ability to scavenge 2,2-diphenyl-1-picrylhydrazyl ($DPPH^{\cdot}$) radical. The analyses were first carried out on ethanolic extracts of flowers, leaves, and roots. The results showed that roots presented weak radical scavenging activity, with an IC_{50} of 1525 ± 122 mg/L. Flowers and leaves had greater activity, with IC_{50} values of 475 ± 15 and 467 ± 14 mg/L, respectively (see Supplementary Data S1). Leaves were chosen for further investigations as flowers are more fragile and less abundant.

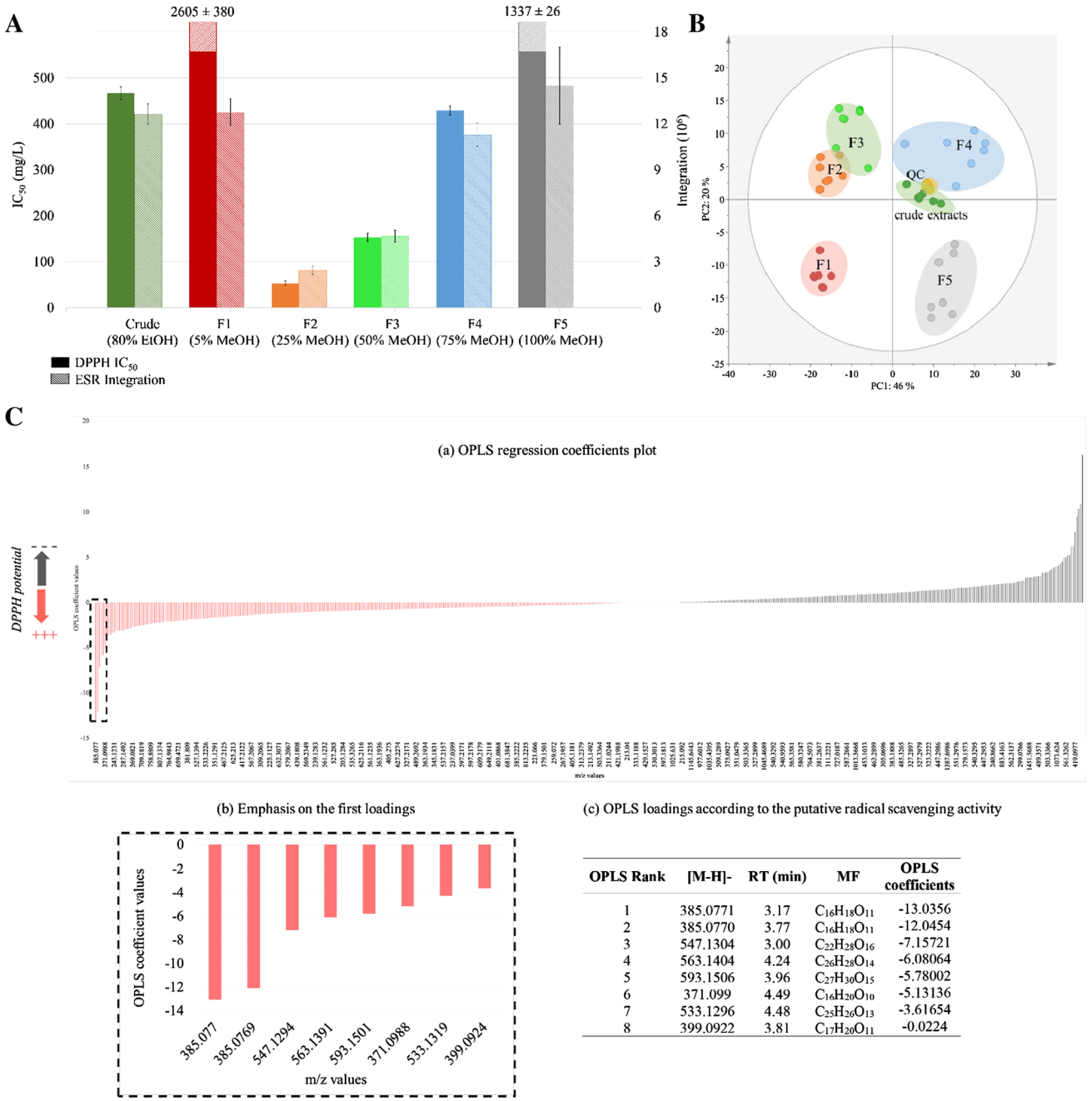
Radical scavenging activity was measured for the crude ethanolic leaf extracts of seven individual plants of violet of Toulouse and respective fractions (N = 42) of decreasing polarity (F1–F5) (Fig. 1A; plain bar plots). While the seven crude extracts displayed an IC_{50} value of 467 ± 14 mg/L, high IC_{50} values were obtained for SPE fractions F1 and F5 (2605 ± 380 and 1337 ± 26 mg/L, respectively) demonstrating weak radical scavenging activity in vitro in these cases. Significant activities were measured for fractions F2, F3 and F4, with IC_{50} values of 53 ± 5, 153 ± 9 and 429 ± 10 mg/L, respectively. For comparison, the positive controls Trolox and rutoside possess IC_{50} values of approximately 7 and 14 mg/L, respectively. Therefore, fractions F2 and F3 seemed promising for the discovery of antioxidant compounds in vitro as preservatives.

To confirm these results, ESR experiments were conducted to measure the capacity of extracts (final concentration of 50 mg/L) to trap superoxide radical ($O_2^{\cdot -}$). The corresponding values of the ESR signal double integrations are presented in Fig. 1A (striped bar plots). As for DPPH assay, F2 and F3 appeared to have the most redox potential as demonstrated by the low intensity of the remaining signal.

3.2 UHPLC–HRMS-based metabolomics approach

UHPLC–HRMS profiles of all the 46 extracts (7 crude extracts, 35 SPE fractions, and 6 QC samples prepared by pooling an aliquot of all fractions) afforded 434 and 527 features (*m*/*z*-RT pairs) in NI and PI modes, respectively.

The C_{18}-SPE procedure is clearly effective at resolving the extracts according to polarity, with polar fractions 1–3 eluting early (<6 min in our UHPLC conditions) and fractions 4 and 5, containing apolar compounds, eluting at later retention times (see Supplementary Data S2).



OPLS Rank	[M-H]-	RT (min)	MF	OPLS coefficients
1	385.0771	3.17	$C_{16}H_{18}O_{11}$	-13.0356
2	385.0770	3.77	$C_{16}H_{18}O_{11}$	-12.0454
3	547.1304	3.00	$C_{22}H_{28}O_{16}$	-7.15721
4	563.1404	4.24	$C_{26}H_{28}O_{14}$	-6.08064
5	593.1506	3.96	$C_{27}H_{30}O_{15}$	-5.78002
6	371.099	4.49	$C_{16}H_{26}O_{10}$	-5.13136
7	533.1296	4.48	$C_{25}H_{26}O_{13}$	-3.61654
8	399.0922	3.81	$C_{17}H_{20}O_{11}$	-0.0224

Fig. 1 MVA workflow: **A** comparison of radical scavenging properties of leaf crude extracts and respective fractions of violet of Toulouse obtained by C_{18}-SPE using DPPH assays (*plain bar plots*) and ESR (*striped bar plots*). **B** PCA score plot of the ESI–NI dataset (QC: quality controls, F1–F5 denote respective C_{18}-SPE fractions). **C** (a) Coefficient plot obtained by OPLS regression; (b) emphasis of the first loadings; (c) ranking of the first seven loadings estimated as radical scavenger compounds according to the OPLS regression

As a preliminary step, principal component analysis (PCA) was applied as an exploratory data analysis to provide an unsupervised overview of the LC–MS fingerprints. PCA clustered all independent biological replicates from the same fraction. Thus, one fraction was related to one position on the plot with slight variability highlighting a reproducible response. As expected, PCA grouped crude extracts and QC near the plot centre. By contrast, single fractions were well distributed (Fig. 1B). These results indicated that SPE fractionation and LC–MS workflows highlighted variability in the data set, and were highly reproducible. Moreover, it demonstrated a stable chemical composition within aerial parts of the seven independent plants under study.

After PCA, we applied OPLS regression analysis to obtain a classification of the loadings (i.e. *m/z*-RT pairs) regarding the DPPH IC_{50} value (input Y). The quality of model prediction was good (R2Y = 0.977, Q2Y = 0.952) and a permutation test assessed its validity (see Supplementary Data S3). This supervised method allowed the ranking of potential redox-active compounds according to their regression coefficient values (Fig. 1C(a, b)): negative coefficients were correlated to potential redox-active compounds and positive coefficients to less-active redox compounds. From the list containing the first hits (Fig. 1C(c)), we carried out the identification procedure based on their MS/MS spectra.

3.3 Identification of potential radical scavenger compounds based on in silico fragmentation

Using the OPLS regression analysis results, the top eight compounds were tentatively identified by interrogating local natural product databases integrated in MS-FINDER: PlantCyc, UNPD, KNApSAcK and ChEBI (Table 1, Hit NP databases). For each compound, the results afforded several candidates and ranked them according to their similarity score, which was based on comparison between experimental MS/MS fragments and *in silico* spectra of candidates. Interestingly, for compounds **1, 2, 3, 6** and **8**, the top three candidates ranked by MS-FINDER each contained a coumarin core, and for compounds **4, 5** and **7**, each contained a flavonoid derivative. UV spectra of each peak (Table 1; UV λ_{max}) supported our assumptions. The coumarin nucleus displays two absorption bands near 270 and 310 nm. Substitutions on this nucleus tend to produce a bathochromic shift (Masrani et al. 1974). For the flavonoids, two absorption bands of interest are around 250 to 295 and 310 to 370 nm, depending on the flavonoid class and the substitution pattern (Olsen et al. 2009). Therefore, the higher wavelength band gives differentiation between coumarins and flavonoids. Still, the candidates were numerous and not always relevant, e.g., compound **5** had hundreds of candidates and feruloyl derivatives were proposed for compound **1**. To refine the results, we carried out a second identification step by creating a user-defined database restricting results of UNPD to compounds found in *Viola* genus (Table 1, Hit UNPD-Viola). Thus, fewer and more pertinent candidates could be proposed. In our case, this led to the identification of compounds **4, 5** and **7** as flavonoids, indicating that all the other significant features have never been described in *Viola* genus. Moreover, compounds **3** and **8** remained totally unknown as no hit matched with any database.

3.4 Extending the dereplication process via a mass spectral similarity network

The mass spectral similarity network displayed two clusters of radical scavenger compounds, i.e., antioxidants (Fig. 2). Each cluster was closely related to one SPE fraction: according to our previous identification results, flavonoids were mainly found in F3 while coumarins were mainly found in F2.

The flavonoid cluster displayed three compounds with related structures active in reducing DPPH$^{\cdot}$ (Fig. 3). Compounds **4** and **5** presented a statistically significant radical scavenging potential according to their coefficient score (Table 1; OPLS coefficient). Interestingly, the molecular formulae of the compounds from this cluster matched structures already identified in Violaceae. Experimental MS/MS patterns mirrored to *in silico* fragmentation allowed the preliminary identification of schaftoside (**4**), isovitexin 6″-*O*-β-D-glucopyranoside (**5**) and 6,8-di-*C*-α-L-arabinopyranosylapigenin (**7**) (see Supplementary Data S4). These structures were based on a user-defined *Viola* database, as hits were more pertinent and the final candidate had the highest total score, generally above 7.

The other compounds of interest belonged to the coumarin cluster (Fig. 3). Compounds **3** and **8** were connected with the identified glycosylated coumarins $C_{16}H_{18}O_{11}$ (compound **2**), confirmed by 1D and 2D NMR (see Supplementary Data S5). Thus, we applied a de novo dereplication based on the interpretation of mass loss to propose potential chemical structures. We interpreted a difference of 162 Da between **3** and **2** as a glucose substituent. We suggest a new 8-*O*-diglycosylated coumarin structure for **3**, based on the MS/MS spectrum and the loss of a 324.12 Da, correlated to two linked glucose units (see Supplementary Data S4). Regarding **8**, a difference of 14 Da was indicative of a di-methoxy-8-*O*-glycosylated coumarin.

Compound **6** was tentatively identified as an 8-*C*-glycolsylated coumarin from comparison with *in silico* fragmentation data using MS-FINDER local databases.

3.5 Validation of the model

To validate our OPLS model, we used a TLC–DPPH–MS assay for fractions F2 (Fig. 4a) and F3 (Fig. 4b). Overall, the four coumarins (**1, 2, 3** and **8)** and the three flavonoids (**4, 5** and **7**) were identified and matched with the OPLS ranking.

We then tested purified compounds **1** and **2** for their respective scavenging activities against DPPH (plain bar plots) and superoxide radicals (striped bar plots) (Fig. 5). Trolox and rutoside were used as references for both assays. The difference in radical scavenging capacity of Trolox against DPPH and superoxide radicals can be easily

Table 1 Summary of all dereplicated compounds

Rank*	RT (min)	$[M–H]^-$	Molecular formula	Error (mDa)	Hit NP databases**	Hit UNPD *Viola****	Main MS/MS fragments	UV λ_{max} (nm)	Potential id[a,b,c]	MSI level	Chemical class	Biological source	OPLS coefficient	TLC-DPPH assay	Ref.
1	3.17	385.0771	$C_{16}H_{18}O_{11}$	0.4349	2	0	206.9150 222.0869 370.1008	260 326	7-Methoxy-5,6,8-dihydroxycoumarin-5-β-glucopyranoside[a,b]	1	Coumarin	*Tetraphis pellucida*	−13.03	Detected	Jung et al. (1995)
2	3.77	385.0770	$C_{16}H_{18}O_{11}$	0.7349	2	0	207.9790 223.1457 370.1048	258 325	7-Methoxy-5,6,8-trihydroxycoumarin-8-β-glucopyranoside[a,b]	1	Coumarin	*Tetraphis pellucida*	−12.04	Detected	Jung et al. (1995)
3	3.00	547.1304	$C_{22}H_{28}O_{16}$	1.1584	0	0	208.0135 223.0087 385.1274	268 320	7,8-dimethoxy-5,6,8-trihydroxycoumarin-8-β glucopyranosyl-glucopyranoside[c]	2	Coumarin	–	−7.15	Detected	–
4	4.24	563.1404	$C_{26}H_{28}O_{14}$	0.5291	88	3	353.0987 383.2114 443.1452 473.0864	271 333	Schaftoside[b]	2	Flavonoid	Numerous plants	−6.08	Detected	Chen Xie et al. (2003)
5	3.96	593.1501	$C_{27}H_{30}O_{15}$	1.2938	154	2	353.0222 383.19O3 473.0770 503.1120	272 332	Isovitexin 6″-*O*-β-D glucopyranoside[b]	2	Flavonoid	*Viola tricolor*	−5.78	Detected	Vukics et al. (2008)
6	4.49	371.099	$C_{16}H_{20}O_{10}$	0.5704	11	8	120.976 231.124 248.972	265 320	8-*C*-Glucosyl-3,4-dihydro-5,7-dihydroxy-6-methoxy-coumarin[b]	2	Coumarin	*Diceratella elliptica*	−5.13	Not detected	Marzouk et al. (2012)
7	4.48	533.1296	$C_{25}H_{26}O_{13}$	0.8644	25	1	353.1044 383.2066 443.1201 473.1409	271 331	6,8-Di-*C*-α-L-arabinopyranosylapigenin[b]	2	Flavonoid	*Viola yedoensis*	−3.61	Detected	Xie et al. (2003)
8	3.81	399.0922	$C_{17}H_{20}O_{11}$	0.2850	3	0	207.0253 221.9839 237.0366	270 320	7,8-Dimethoxy-5,6,8-trihydroxycoumarin-8-β-glucopyranoside[a,c]	1	Coumarin	–	−0.02	Detected	–

*Ranking based on OPLS regression coefficients, served as compound number

**Hits found when interrogating natural product databases (UNPD, KnapSack, CheBI, and PlantCyc) in MS-FINDER

***Hits found when interrogating the user-defined database (UNPD and DNP) composed of known compounds in *Viola* genus

[a]Confirmed by NMR

[b]Top ranked hit determined by *in silico* MS/MS fragmentation with MS-FINDER

[c]Dereplicated with molecular networking

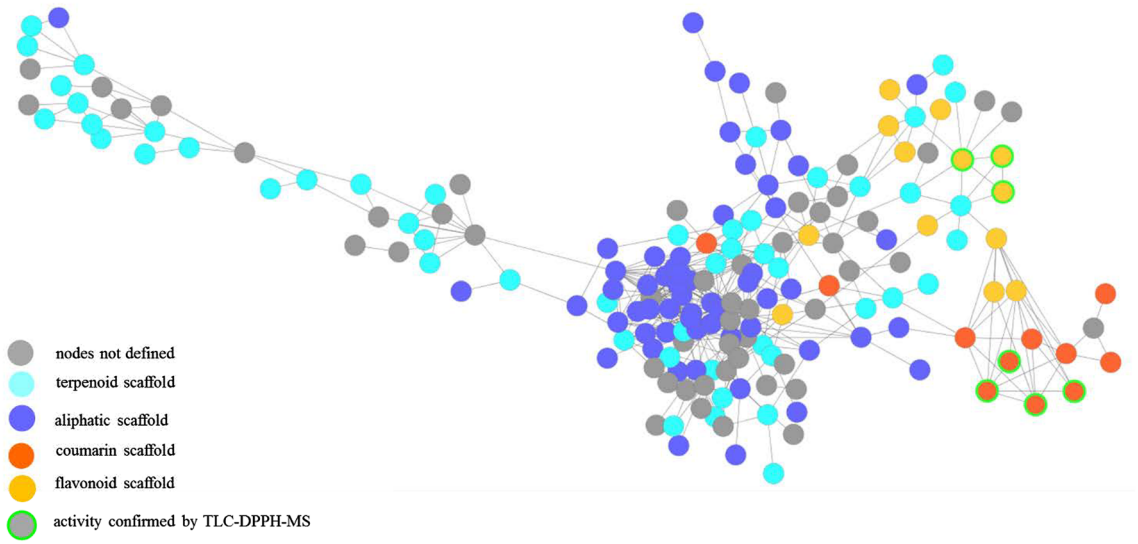


Fig. 2 Mass spectral similarity network of *Viola* ethanolic extracts with coloration based on the chemical class

explained by considering the rate constants of the scavenging reactions: $k_{Trolox/DPPH} = 1.1 \times 10^4$ M^{-1} s^{-1} (Friaa and Brault 2006) and $k_{Trolox/O_2^{\cdot-}} = 0.1$ M^{-1} s^{-1} (Cabelli and Bielski 1986).

The radical scavenging properties of compound **1** (97% pure) were confirmed by comparison with the active controls, Trolox (DPPH IC_{50}: **1** 9 ± 2 mg/L vs. Trolox 7 mg/L), and rutoside (ESR double integration: **1** 0.88 ± 0.01 vs. rutoside 1.78 ± 0.10).

For compound **2** (90% pure), the DPPH IC_{50} of 49 ± 1 mg/L and ESR double integration of 15 ± 1, suggests this compound is a less active radical scavenger than compound **1**, meeting the statistical ranking.

4 Discussion

The objective of this work was to characterise redox active metabolites from the leaves of *Viola alba* subsp. *dehnhardtii* together with de novo dereplication of active compounds and the establishment of a non-volatile metabolite fingerprint.

In order to decipher the redox potential in few steps, we correlated DPPH results to UHPLC–HRMS fingerprints by OPLS regression. This method allowed the ranking of detected features according to their potential radical scavenging properties. In addition, a mass spectral similarity network allowed de novo dereplication of top ranked features, based on the acquisition of UHPLC–HRMS profiles in data dependent analysis mode. This provided accurate mass-to-charge ratio for molecular formula determination along with MS/MS fragments used for peak assignments. NI mode MS and MS/MS spectra were mainly used as they were of better quality than PI spectra, thus improving *in silico* identification using MS-FINDER. Processing with MS-DIAL allowed the acquisition of a clean peak list with deconvoluted MS/MS data for each feature. This meant we could easily remove adducts and other ionisation artefacts from files uploaded to MetamapR. This resulted in a cleaner molecular network compared with uploading raw data files. The mass spectral similarity network allowed the organisation of features in clusters and dereplication results highlighted mainly coumarins in SPE fraction F2 and mainly flavonoids in fraction F3. These scaffolds are known for having several biological activities, including redox activity (Procházková et al. 2011; Gacche and Jadhav 2012). We identified a mixture of *C*-glycosylated flavonoids along with *O*-glycosylated coumarins. The enrichment of these molecules in the two fractions explains the improved antioxidant properties of the fractions over the crude extracts. Although some of the compounds, e.g., compounds **5** and **7**, are well-known in *Viola* genus (Vukics et al. 2008; Xie et al. 2003), the other compounds we identified have not previously been described in this genus.

The protective effect of phenolic coumarins against oxidative damage depends on the hydrogen-donating capacity of hydroxyl groups (Kostova 2006; Borges et al. 2005) and then on their oxidisability. Previous structure–activity relationship studies have identified the importance of the number and location of the phenolic hydroxyl groups (Kancheva et al. 2010). As compounds **1** and **2** have the same substituents, energy minimisation calculations were carried out using the MM2 forcefield (Chemdraw 3D, Cambridge Soft, USA) to understand the difference between their radical reducing capacity. These calculations demonstrate that

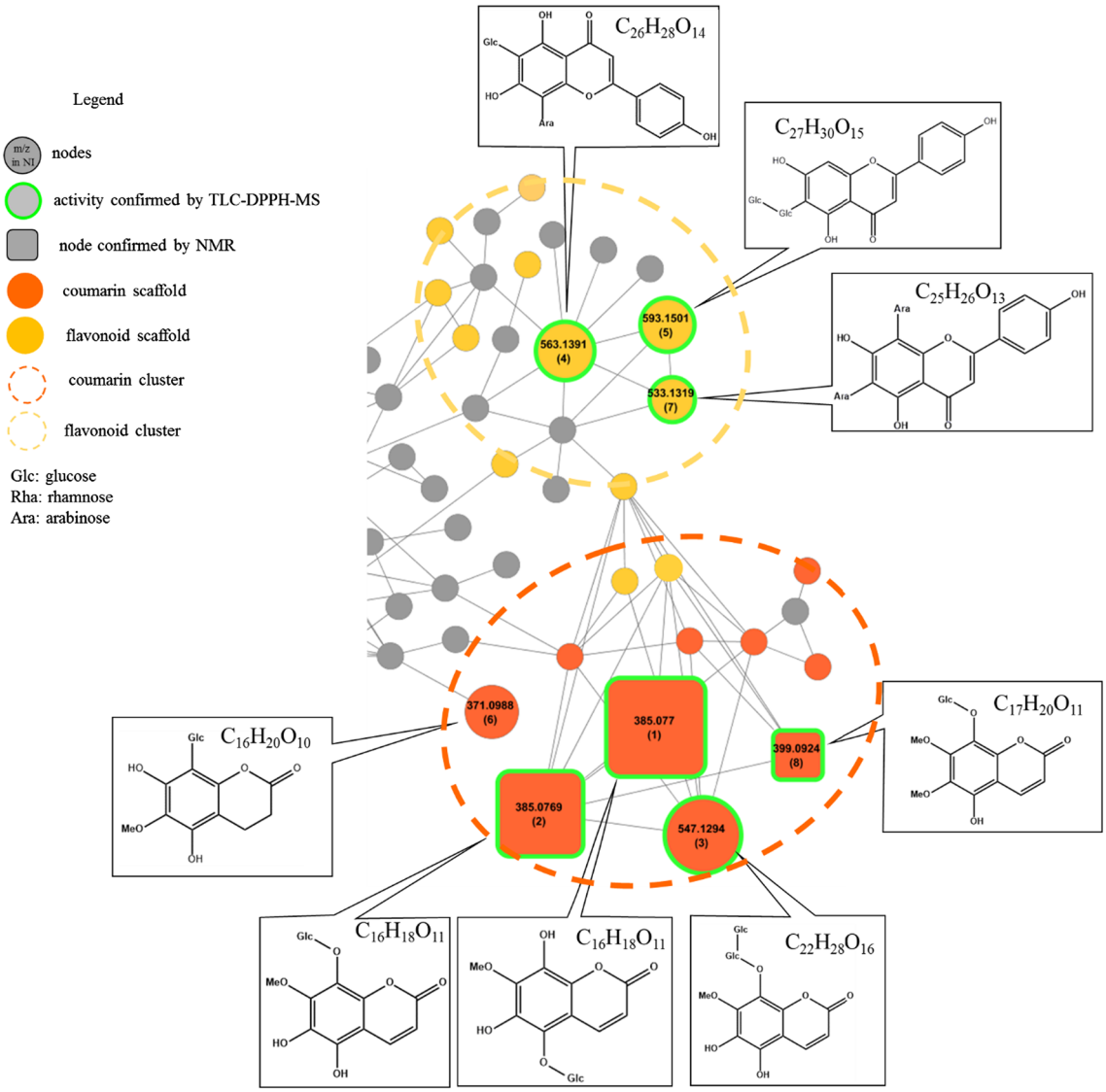


Fig. 3 Expansion on the redox-active compounds clusters. Putative coumarin and flavonoids scaffold correspond to top ranked hits from MS-FINDER. Node size was emphased based on OPLS coefficient value. Compound number is indicated in brackets in each node in correspondence to Table 1

the hydroxyl group in position 6 in compound **1** was the most readily oxidisable into OH^{-+} and thus responsible for the strong antioxidant capacity measured for **1** compared with **2**. Purification of coumarins $C_{16}H_{18}O_{11}$**1** and **2** tested in DPPH and ESR redox assays confirmed the workflow used in this study: as predicted by the model, compound **1** appeared as a reducer that is a good antioxidant compound in vitro.

Our statistical analysis was validated by TLC–DPPH–MS assays, with detection of seven DPPH-active compounds ranked in top positions of the analysis. A big advantage of our approach is the limited manipulation steps (crude extraction followed by a single SPE fractionation) and low sample demand. For instance, 1 g of dried plant material should be sufficient to conduct the entire workflow. The workflow could easily be implemented to any kind of extract with other biological activities. However, ranking results of variables may not accurately reflect the biological activity. Ranked positions may vary depending on scaling and normalisation processes. The first 20 ranked variables should be considered for thorough analysis. Another drawback of this method concerns the correlation between biological assay results (Y) and the detection method used to generate variables (X). In our case, we only took into account ionisable compounds in electrospray ionisation (ESI) mode for OPLS modelling. The ionisation process is compound dependent and does not reflect a quantitative measure of each detected feature. We could reach

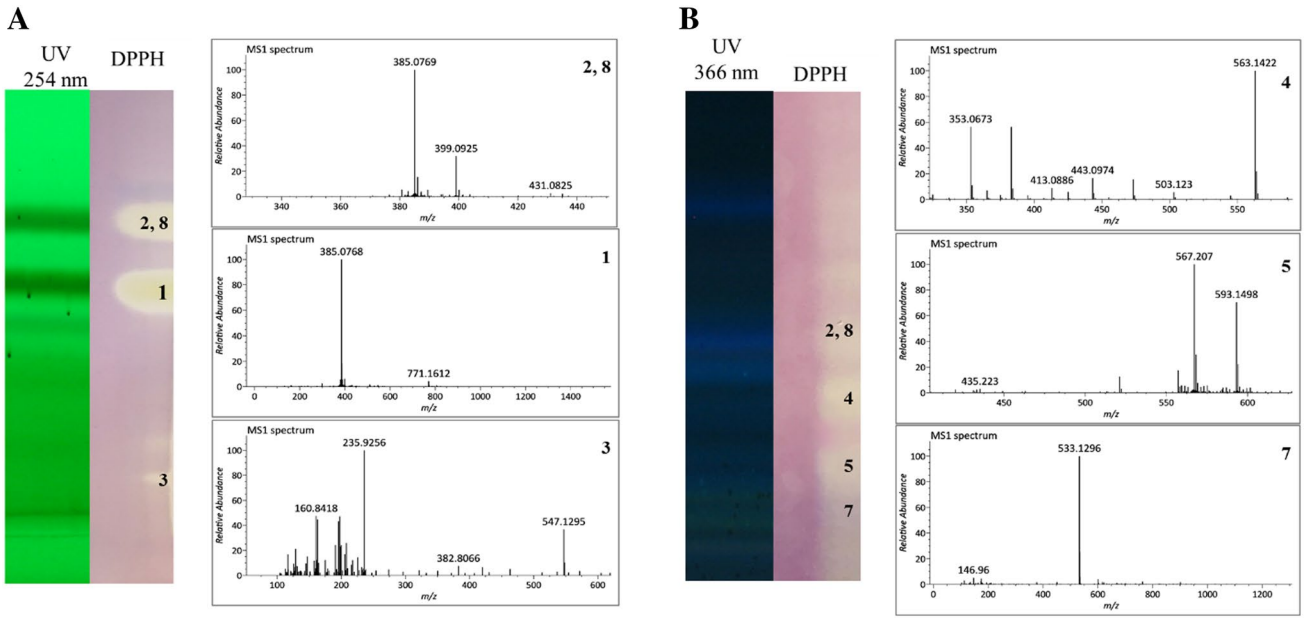


Fig. 4 DPPH TLC–MS dereplication of redox active compounds of F2 (**a**) and F3 (**b**)

more realistic models by using universal-type detectors such as Evaporating Light Scattering Detector (ELSD) or Charged Aerosol Detector (CAD). While ^{1}H- or ^{13}C-NMR could be an interesting alternative, the lack of sensitivity of this method is not compatible with complex samples like crude natural extracts.

Finally, the framework of our dereplication strategy capitalised on UHPLC–HRMS–MS/MS profiles, using agreement between modelled and experimental MS/MS fragmentation patterns and mass spectral similarity network to propagate assignments of clustered features. As illustrated with the coumarin cluster, using the molecular network and propagating from a known compound allows identification of others belonging to the same class by interpreting the neutral mass loss.

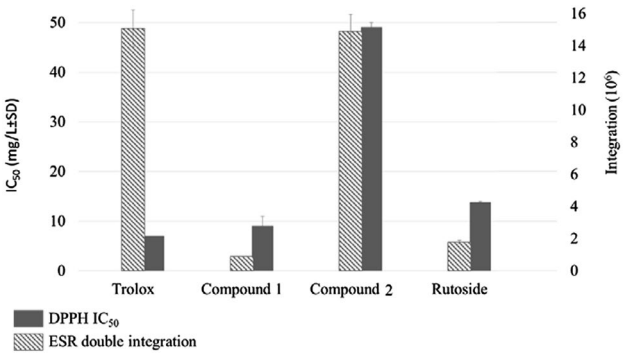


Fig. 5 Capacity of purified compounds **1** and **2** to scavenge free radicals

5 Conclusion

A UHPLC–HRMS based metabolomic study combined with molecular networking tools allowed the dereplication of redox-active metabolites present in the violet of Toulouse. These compounds were mainly flavonoid and coumarin derivatives. Our approach highlighted two main metabolite clusters displaying functional groups with redox-active properties. We identified seven compounds of interest, of which five were found through MS/MS fragmentation and comparison with references found in databases. Two unknown coumarins were de novo dereplicated through a molecular networking approach. Overall, the workflow proposed here allowed early identification of redox active compounds within a complex mixture with limited effort and crude materials.

In this work, it was chosen to correlate redox-active properties with LC–MS profiles to identify compounds of interest but more generally, provided variability is brought to the data set, this method can be implemented in any study of other chemical and biological properties employing different types of assays.

Acknowledgements The authors would like to thank the Regional Council Midi-Pyrénées for sponsoring this project (Project CLE 13053062) as well as the Toulouse Municipal Greenhouses for making the violet collection available to its collaborators.

Funding This study was funded by the Regional Council Midi-Pyrénées (Project CLE 13053062).

Author contributions GM proposed the study and designed the experiments. JC performed the extraction of crude material, the radical scavenging assays under the supervision of PP, processed the LC-HRMS profiles acquired by NMF and performed MVA and molecular networking analysis under the supervision of GM. JC analysed the results and wrote the manuscript which was reviewed by all the collaborators of the work.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Research involving human participants and/or animals No human participants or animals were involved in this study.

Informed consent No human participants were involved in this study.

References

Allard, P.-M., Péresse, T., Bisson, J., Gindro, K., Marcourt, L., Pham, V. C., et al. (2016). Integration of molecular networking and *in-silico* MS/MS fragmentation for natural products dereplication. *Analytical Chemistry, 88*(6), 3317–3323. doi:10.1021/acs.analchem.5b04804.

Ayouni, K., Berboucha-Rahmani, M., Kim, H. K., Atmani, D., Verpoorte, R., & Choi, Y. H. (2016). Metabolomic tool to identify antioxidant compounds of *Fraxinus angustifolia* leaf and stem bark extracts. *Industrial Crops and Products, 88*, 65–77. doi:10.1016/j.indcrop.2016.01.001.

Borges, F., Roleira, F., Milhazes, N., Santana, L., & Uriarte, E. (2005). Simple coumarins and analogues in medicinal chemistry: Occurrence, synthesis and biological activity. *Current Medicinal Chemistry, 12*(8), 887–916.

Bylesjö, M., Rantalainen, M., Cloarec, O., Nicholson, J. K., Holmes, E., & Trygg, J. (2006). OPLS discriminant analysis: Combining the strengths of PLS-DA and SIMCA classification. *Journal of Chemometrics, 20*(8-10), 341–351.

Cabelli, D. E., & Bielski, B. H. (1986). Studies of the reactivity of trolox with Mn^{3+}/Fe^{3+} complexes by pulse radiolysis. *Journal of Free Radicals in Biology & Medicine, 2*(1), 71–75.

Cox, D. G., Oh, J., Keasling, A., Colson, K. L., & Hamann, M. T. (2014). The utility of metabolomics in natural product and biomarker characterization. *Biochimica et Biophysica Acta (BBA)-General Subjects, 1840*(12), 3460–3474.

Forman, H. J., Davies, K. J., & Ursini, F. (2014). How do nutritional antioxidants really work: Nucleophilic tone and para-hormesis versus free radical scavenging in vivo. *Free Radical Biology and Medicine, 66*, 24–35.

Friaa, O., & Brault, D. (2006). Kinetics of the reaction between the antioxidant Trolox® and the free radical DPPH in semi-aqueous solution. *Organic & Biomolecular Chemistry, 4*(12), 2417–2423.

Gacche, R. N., & Jadhav, S. G. (2012). Antioxidant activities and cytotoxicity of selected coumarin derivatives: Preliminary results of a structure–activity relationship study using computational tools. *Journal of Experimental & Clinical Medicine, 4*(3), 165–169.

Grapov, D., Wanichthanarak, K., & Fiehn, O. (2015). MetaMapR: Pathway independent metabolomic network analysis incorporating unknowns. *Bioinformatics*. doi:10.1093/bioinformatics/btv194.

Hubert, J., Berger, M., Nepveu, F., Paul, F., & Daydé, J. (2008). Effects of fermentation on the phytochemical composition and antioxidant properties of soy germ. *Food Chemistry, 109*(4), 709–721.

Jung, M., Geiger, H., & Zinsmeister, H. D. (1995). Tri-and tetrahydroxycoumarin derivatives from *Tetraphis pellucida*. *Phytochemistry, 39*(2), 379–381.

Kancheva, V. D., Saso, L., Boranova, P. V., Khan, A., Saroj, M. K., Pandey, M. K., et al. (2010). Structure–activity relationship of dihydroxy-4-methylcoumarins as powerful antioxidants: Correlation between experimental & theoretical data and synergistic effect. *Biochimie, 92*(9), 1089–1100.

Kind, T., & Fiehn, O. (2016). Strategies for dereplication of natural compounds using high-resolution tandem mass spectrometry. *Phytochemistry Letters*. doi:10.1016/j.phytol.2016.11.006.

Kostova, I. (2006). Synthetic and natural coumarins as antioxidants. *Mini Reviews in Medicinal Chemistry, 6*(4), 365–374.

López-Alarcón, C., & Denicola, A. (2013). Evaluating the antioxidant capacity of natural products: A review on chemical and cellular-based assays. *Analytica Chimica Acta, 763*, 1–10.

Marzouk, M. M., Elkhateeb, A., Ibrahim, L. F., Hussein, S. R., & Kawashty, S. A. (2012). Two Cytotoxic Coumarin Glycosides from the aerial parts of Diceratella elliptica (DC.) Jonsell growing in Egypt. *Records of Natural Products, 6*(3), 237.

Masrani, K., Rama, H., & Bafna, S. (1974). Ultraviolet absorption spectra: Some substituted coumarins. *Journal of Chemical Technology and Biotechnology, 24*(6), 331–341.

Muhammad, N., Saeed, M., Aleem, A., & Khan, H. (2012). Ethnomedicinal, phytochemical and pharmacological profile of genus Viola. *Phytopharmacol, 3*(1), 214–226.

Nguyen, H. Y. T., Vo, B. H. T., Nguyen, L. T. H., Bernad, J., Alaeddine, M., Coste, A., et al. (2013). Extracts of *Crinum latifolium* inhibit the cell viability of mouse lymph oma cell line EL4 and induce activation of anti-tumour activity of macrophages in vitro. *Journal of Ethnopharmacology, 149*(1), 75–83.

Olsen, H., Aaby, K., & Borge, G. I. A. (2009). Characterization and quantification of flavonoids and hydroxycinnamic acids in curly kale (Brassica oleracea L. convar. acephala var. sabellica) by HPLC DADESI-MS. *Journal of Agricultural and Food Chemistry, 57*(7), 2816–2825.

Procházková, D., Boušová, I., & Wilhelmová, N. (2011). Antioxidant and prooxidant properties of flavonoids. *Fitoterapia, 82*(4), 513–523.

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research, 13*(11), 2498–2504.

Tsugawa, H., Cajka, T., Kind, T., Ma, Y., Higgins, B., Ikeda, K., et al. (2015). MS-DIAL: Data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nature Methods*. doi:10.1038/nmeth.3393.

Tsugawa, H., Kind, T., Nakabayashi, R., Yukihira, D., Tanaka, W., Cajka, T., et al. (2016). Hydrogen rearrangement rules: Computational MS/MS fragmentation and structure elucidation using MS-FINDER software. *Analytical Chemistry*. doi:10.1021/acs.analchem.6b00770.

Vukics, V., Kery, A., Bonn, G. K., & Guttman, A. (2008). Major flavonoid components of heartsease (*Viola tricolor* L.) and their antioxidant activities. *Analytical and Bioanalytical Chemistry, 390*(7), 1917–1925.

Vukics, V., Ringer, T., Kery, A., Bonn, G. K., & Guttman, A. (2008). Analysis of heartsease (*Viola tricolor* L.) flavonoid glycosides by micro-liquid chromatography coupled to multistage mass spectrometry. *Journal of Chromatography A, 1206*(1), 11–20.

Witkowska-Banaszczak, E., Bylka, W., Matławska, I., Goślińska, O., & Muszyński, Z. (2005). Antimicrobial activity of *Viola tricolor* herb. *Fitoterapia, 76*(5), 458–461.

Wolfender, J. L., Marti, G., & Ferreira Queiroz, E. (2010). Advances in techniques for profiling crude extracts and for the rapid identificationof natural products: Dereplication, quality control and metabolomics. *Current Organic Chemistry, 14*, 1808–1832.

Wolfender, J.-L., Marti, G., Thomas, A., & Bertrand, S. (2015). Current approaches and challenges for the metabolite profiling of complex natural extracts. *Journal of Chromatography A, 1382*, 136–164. doi:10.1016/j.chroma.2014.10.091.

Worley, B., & Powers, R. (2013). Multivariate analysis in metabolomics. *Current Metabolomics, 1*(1), 92–107.

Xiao, J. F., Zhou, B., & Ressom, H. W. (2012). Metabolite identification and quantitation in LC-MS/MS-based metabolomics. *TrAC Trends in Analytical Chemistry, 32*, 1–14.

Xie, C., Veitch, N. C., Houghton, P. J., & Simmonds, M. S. (2003). Flavone C-glycosides from *Viola yedoensis* Makino. *Chemical and Pharmaceutical Bulletin, 51*(10), 1204–1207.

Yang, J. Y., Sanchez, L. M., Rath, C. M., Liu, X., Boudreau, P. D., Bruns, N., et al. (2013). Molecular networking as a dereplication strategy. *Journal of Natural Products, 76*(9), 1686–1699.

Zhu, H., Qin, S., Zhang, N., Yang, D., Han, H., Wei, K., & Li, M. (2015). Chemical constituents and biological activities of plants from the genus Viola. *Chemistry & Biodiversity, 12*(12), 1777–1808.

Molecules **2013**, *18*, 3018-3027; doi:10.3390/molecules18033018

OPEN ACCESS

molecules

ISSN 1420-3049

www.mdpi.com/journal/molecules

Article

Natural Aristolactams and Aporphine Alkaloids as Inhibitors of CDK1/Cyclin B and DYRK1A

Guillaume Marti [1], Véronique Eparvier [1,*], Barbara Morleo [1], Jessica Le Ven [1], Cécile Apel [1], Bernard Bodo [2], Séverine Amand [2], Vincent Dumontet [1], Olivier Lozach [3], Laurent Meijer [3,4], Françoise Guéritte [1] and Marc Litaudon [1]

[1] Centre de Recherche de Gif, Institut de Chimie des Substances Naturelles, CNRS, 1 avenue de la Terrasse, 91198 Gif-sur-Yvette Cedex, France; E-Mails: Guillaume.Marti@icsn.cnrs-gif.fr (G.M.); barbara.morleo@icsn.cnrs-gif.fr (B.M.); jessica.leven@gmail.com (J.L.V.); cecile.apel@icsn.cnrs-gif.fr (C.A.); vincent.dumontet@icsn.cnrs-gif.fr (V.D.); francoise.gueritte@icsn.cnrs-gif.fr (F.G.); marc.litaudon@icsn.cnrs-gif.fr (M.L.)

[2] Muséum National d'Histoire Naturelle, UMR 7245 CNRS, 63 rue Buffon, Paris 75005, France; E-Mails: bodo@mnhn.fr (B.B.); amand@mnhn.fr (S.A.)

[3] Protein Phosphorylation & Human Disease' group, CNRS, Station Biologique, Place G. Teissier, Roscoff 29680, France; E-Mails: lozach@sb-roscoff.fr (O.L.); meijer@sb-roscoff.fr (L.M.)

[4] ManRos Therapeutics, Centre de Perharidy, Roscoff 29680, France

* Author to whom correspondence should be addressed; E-Mail: veronique.eparvier@icsn.cnrs-gif.fr; Tel.: +33-169-823-679; Fax: +33-169-077-247.

Received: 29 January 2013; in revised form: 28 February 2013 / Accepted: 1 March 2013 / Published: 6 March 2013

Abstract: In an effort to find potent inhibitors of the protein kinases DYRK1A and CDK1/Cyclin B, a systematic *in vitro* evaluation of 2,500 plant extracts from New Caledonia and French Guyana was performed. Some extracts were found to strongly inhibit the activity of these kinases. Four aristolactams and one lignan were purified from the ethyl acetate extracts of *Oxandra asbeckii* and *Goniothalamus dumontetii*, and eleven aporphine alkaloids were isolated from the alkaloid extracts of *Siparuna pachyantha*, *S. decipiens*, *S. guianensis* and *S. poeppigii*. Among these compounds, velutinam, aristolactam AIIIA and medioresinol showed submicromolar IC_{50} values on DYRK1A.

Keywords: *Oxandra asbeckii*; *Goniothalamus dumontetii*; *Siparuna* spp.; aporphinoids and aristolactams; kinases inhibitors

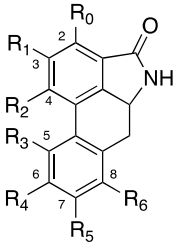
1. Introduction

Dual-specificity tyrosine phosphorylation-regulated kinase 1A (DYRK1A) and cyclin-dependent kinases (CDKs) play an important role in the regulation of various cellular processes by phosphorylating serine, threonine and/or tyrosine residues [1]. However, they also can be involved in several human diseases such as cancer and neurodegenerative disorders [1–5]. DYRK1A is a protein kinase with diverse functions, and it is implicated in neuronal development and adult brain physiology. High levels of DYRK1A are associated with neurodegenerative diseases and are also believed to be involved in the neurobiological abnormalities observed in Down-Syndrome, such as mental retardation [5]. The cyclin-dependent kinases (CDKs) are implicated in viral infections, cancer and neurodegenerative pathologies (e.g., Alzheimer's and Parkinson's diseases) [6–8]. Cyclin-dependent kinases, which are composed of a catalytic subunit (such as CDK1) and a regulatory subunit (such as Cyclin B), play an important role in the regulation of cell cycle progression. For example, the CDK1/Cyclin B complex is known to govern the entry into M-phase [9,10]. For the discussed reasons, these two families of kinases have been extensively used as targets to identify new pharmacological inhibitors of potential therapeutic interest [11]. In this context, and in continuation of our screening program [12] of plant extracts from French Guiana and New Caledonia for the discovery of bioactive natural products, 2,500 extracts (New Caledonian species) were screened against CDK1/Cyclin B, and 720 extracts (French Guiana species) were screened against DYRK1A. The EtOAc extract obtained from *Goniothalamus dumontetii* (R.M.K. Sauders and Munzinger) [13] was selected for its ability to significantly inhibit the activity of CDK1/Cyclin B, while the EtOAc and alkaloid extracts obtained from *Oxandra asbeckii* Pulle (R.E. Fries) and *Siparuna pachyantha* (A.C.Sm.), respectively, were selected for their ability to significantly inhibit the activity of DYRK1A. The selection was then extended to other species of the genus *Siparuna*: *S. decipiens* (A. DC.), *S guianensis* (Aubl.) and *S. poeppigii* (A. DC.). The present paper reports the isolation of 16 compounds, including four aristolactams **1**–**4**, one lignan **5**, and **11** aporphines **6**–**16**, as well as their ability to act as kinase inhibitors.

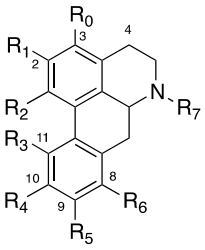
2. Results and Discussion

The chemical investigation of *O. asbeckii* afforded aristolactams AII (**1**) [14] and BII (**2**) [14,15] and velutinam (**3**) (Figure S1) [15,16]. Compounds **1** and **3**, aristolactam AIIIA (**4**) [17] and (−)-medioresinol (**5**) (Figure S2 and S3) [18,19] were isolated from *G. dumontetii*. (+)-Corydine (**6**) [20], (−)-roemerine (**7**) [21,22] and liriodenine (**8**) (Figure S4) [23] were isolated from *S. pachyantha*. (+)-bulbocapnine (**9**) [24,25], (+)-*N*-methyllindcarpine (**10**) [26,27], (+)-actinodaphnine (**11**) [28], liriodenine (8) and (+)-11-methoxynornoelistine (**12**) (Figure S5) [29] were obtained from *S. guianensis*, and the chemical investigation of *S. poeppigii* alkaloid extract yielded lysicamine (**13**) [30], (−)-*O*-methylisopiline (**14**) [31], (+)-*N*-nornuciferine (**15**) [21,32] and liriodenine (**8**). Finally, (+)-boldine (**16**) [33,34] and (+)-*N*-nornuciferine (**15**) were purified from *S. decipiens*. All compounds were identified by comprehensive analysis of spectroscopic and spectrometric data and were compared with data reported in the literature (Figures S1–S5; Tables 1 and S1) [14–34].

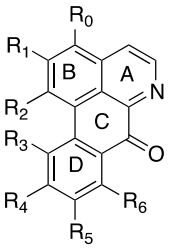
Table 1. Biological activities of the compounds isolated from the Annonaceae and Monimiaceae plant families.



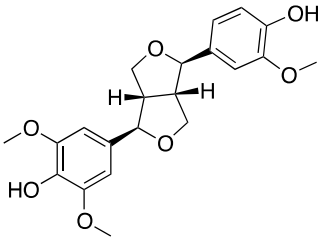
Aristolactam



Aporphines and *N*-methyl-aporphines



Oxo-aporphines



(-)-medioresinol

Compounds	R_0	R_1	R_2	R_3	R_4	R_5	R_6	R_7	DYRK1A (IC_{50} in μM) [a]	CDK1/Cyclin B (IC_{50} in μM) [a]
Aristolactams										
aristolactam AII (**1**)	H	OH	OCH_3	H	H	H	H	H	>30	>30
aristolactam BII (**2**)	H	OCH_3	OCH_3	H	H	H	H	H	>30	>30
velutinam (**3**)	H	OCH_3	OCH_3	H	H	H	OH	H	**0.6**	**1.5**
aristolactam AIIIA (**4**)	H	OH	OCH_3	H	OH	H	H	H	**0.08**	**0.2**
***N*-Methylaporphines**										
(+)-corydine (**6**)	H	OCH_3	OH	OCH_3	OCH_3	H	H	CH_3	>30	>30
(−)-roemerine (**7**)	H	O-CH_2-O		H	H	H	H	CH_3	15.0	>30
(+)-bulbocapnine (**9**)	H	O-CH_2-O		OH	OCH_3	H	H	CH_3	>30	>30
(+)-*N*-methyllindcarpine (**10**)	H	OH	OCH_3	OH	OCH_3	H	H	CH_3	>30	>30
(+)-boldine (**16**)	H	OH	OCH_3	H	OCH_3	OH	H	CH_3	>30	>30
Aporphines										
(+)-actinodaphnine (**11**)	H	O-CH_2-O		H	OCH_3	OH	H	H	>30	>30
(+)-11-methoxynorneolistine (**12**)	H	O-CH_2-O		OCH_3	O-CH_2-O		H	H	**2.5**	>30
(−)-*O*-methylisopiline (**14**)	OCH_3	OCH_3	OCH_3	H	H	H	H	H	>30	>30
(+)-*N*-nornuciferine (**15**)	H	OCH_3	OCH_3	H	H	H	H	H	**4.2**	>30
Oxo-aporphines										
liriodenine (**8**)	H	O-CH_2-O		H	H	H	H	-	**3.1**	>30
lysicamine (**13**)	H	OCH_3	OCH_3	H	H	H	H	-	**2.4**	>30
Other										
(−)-medioresinol (**5**)									**0.1**	**1.3**
6-bromoindirubin-3'-monoxime [b]									**0.52**	**0.32**

[a] IC_{50} are mean values from triplicates (the variation is a maximum of 20%); [b] positive control; IC_{50} = 3 μg/mL on CDK1-CyclinB for *G. dumontetii* EtOAc extract, and IC_{50} = 3.6 and 1.0 μg/mL on DYRK1A for *O. asbeckii* EtOAc extract and *S. pachyantha* total alkaloid extract, respectively.

Several aporphinoid alkaloids have been previously isolated from *Siparuna* spp. [35–38], but this is the first time that compounds **6**, **7** and **10**–**16** were described in this genus. In addition, this is only the second time that 11-methoxynornoelistine (**12**) is isolated from Nature [29]. Aristolactams are often found in the species of the genus *Aristolochia* and *Goniothalamus* [39,40], but this is the first time that this type of alkaloid is isolated from an *Oxandra* species [16].

Compounds **1**–**16** were subjected to the CDK1/Cyclin B and DYRK1A kinase inhibition assays (Table 1). Velutinam (**3**), aristolactam AIIIA (**4**) and (−)-medioresinol (**5**) showed the strongest

inhibition of CDK1/cyclin B activity, with IC_{50} values of 1.5, 0.2 and 1.3 μM, respectively. The IC_{50} values for inhibition of DYRK1A activity of **3**, **4** and **5** were 0.6, 0.08 and 0.1 μM, respectively. In the aporphine series, (−)-roemerine (**7**), (+)-11-methoxynorneolistine (**12**), (+)-*N*-nornuciferine (**15**), liriodenine (**8**) and lysicamine (**13**) were able to moderately inhibit DYRK1A activity, with IC_{50} values of 15.0, 2.5, 4.2, 3.1 and 2.4 μM, respectively, but these compounds were unable to inhibit CDK1/Cyclin B activity at 30 μM. All other compounds, the aristolactams **1** and **2**, the *N*-methylaporphines **6**, **7**, **9**, **10** and **16**; and the aporphines **11** and **14**, were found inactive at a concentration of 30 μM.

Some very interesting observations can be summarised from these results. In the first series (compounds **1**–**4**), it can be deduced that the presence of a hydroxy group at the C-6 (cf. **4** and **1**) or C-8 (cf. **3** and **2**) positions is critical for achieving inhibition of both CDK1/Cyclin B and DYRK1A activities. In contrast, the presence of a hydroxy group at the C-3 position (cf. **1** and **2**) is not required to inhibit either kinase. Aristolactams, which possess a phenanthrene chromophore, constitute an important alkaloid family due to their potent biological effects [41,42], including anti-inflammatory, anti-platelet, anti-mycobacterial and neuroprotective effects [39]. In particular, aristolactam BII (**2**) exhibited potent cytotoxic activity against human cancer cells [43,44]. In our study, aristolactam AIIIA (**4**) was found to be the most potent compound of the series. This result confirms previous studies in which compound **4** strongly inhibited various kinases such as CDK2, CDK4, and Aurora 2 kinase, with IC_{50} values of 0.14, 1.42 and 2.14 μM, respectively [17,45].

Furthermore, aristolactam AIIIA (**4**) was identified as a new ligand targeting the polo-box domain of Polo-like kinase 1. Bioassays indicated that this natural product could inhibit cancer cell proliferation and induce mitotic arrest at G2/M phase with spindle abnormalities and promote apoptosis [46]. Hedge *et al.* [45] also demonstrated that some lactam derivatives of aristolochic acid were inhibitors of CDK2 activity and that the presence of hydroxy groups at the C-6 and/or C-8 positions results in the enhanced ability to inhibit CDK.

In the second series of compounds (**6**–**15**), only alkaloids **8**, **12**, **13** and **15** were shown to inhibit DYRK1A, but not CDK1/Cyclin B activity, with IC_{50} values in the micromolar range. From these results, it can be deduced that the presence of an *N*-methyl group is detrimental for the interaction with DYRK1A; with the exception of (−)-roemerine, which showed a weak inhibitory activity, all *N*-methylated alkaloids were inactive at 30 μM. Finally, the presence of a methoxy group at the C-3 position (cf. **14** and **15**) abolished DYRK1A inhibition. Liriodenine (**8**), lysicamine (**13**) and *N*-nornuciferine (**15**) are known to possess many biological activities, including anti-microbial, anti-leishmanicidal, and cytotoxic effects on various cancer cells lines [32,47,48]. In addition, Chang *et al.* [49] have shown that liriodenine (**8**) at a concentration of 20 μM induced apoptosis by inhibiting the kinase activity of the CDK1/Cyclin B complex, resulting in G2/M cell cycle arrest. More recently, Chen *et al.* showed that this compound also inhibited the growth of human colon cancer cells and induced G1/S cell cycle arrest [50].

In addition to aristolactams and aporphinoids, the lignan, (−)-medioresinol (**5**), was a strong inhibitor of both kinases. Lignans are known to be cytotoxic and can induce G2/M cell cycle arrest and apoptosis [18,51,52].

3. Experimental

3.1. General

The NMR spectra were recorded with a Bruker 500 MHz (Avance 500) spectrometer with $CDCl_3$ or DMSO-d_6 as a solvent. ESIMS were obtained on a Navigator mass Thermoquest. HRESIMS were run on a MALDI-TOF spectrometer (Voyager-De STR; Perspective Biosystems). IR spectra were obtained on a Nicolet FTIR 205 spectrophotometer. The UV spectra were recorded on a Perkin-Elmer Lambda 5 spectrophotometer. Specific rotations were obtained in $CHCl_3$ with a JASCO P-1010 polarimeter.

The fractionation was performed in a Harrisson Research® Chromatotron using a rotating disk at 800 rpm. The Kromasil analytic and preparative C_{18} columns (250 × 4.6 mm and 250 × 21.2 mm, 5 μm Thermo®, with solvent elution at 1 and 15 mL/min) and the SymmetryShield RP18 preparative column (150 × 19.0 mm, 5 μm Waters®, with solvent elution at 17 mL/min) were used for HPLC separation using a Waters autopurification system® equipped with a binary pump (Waters 2525), a UV-vis diode array detector (190–600 nm, Waters 2996) and a Polymer Laboratory PL-ELS 1000 ELS detector. Silica gel 60 (35–70 μm) and analytical and preparative TLC plates (Si gel 60 F 254) were purchased from SDS (Peypin, France). All other chemicals and solvents were of analytical grade and purchased from SDS.

3.2. Plant Material

Bark of *O. asbeckii* was collected in 2007 by one of us (V.E.) in the dense forest of French Guiana (under the reference CAY-VE-128). The herbarium specimen was deposited at the IRD Center (Cayenne French Guiana) under the reference CAY-VE-110. Bark of *G. dumontetii* was collected in 2006 by one of us (V.D.) in New Caledonia, under the reference DUM-0558 [13]. Leaves of *S. pachyantha*, *S. decipiens*, *S. guianensis*, and *S. poeppigii*, were collected in 2005 in the dense forest of French Guiana. The herbarium specimens were deposited at the IRD Center (Cayenne French Guiana) under the references CAY: SD-4410, MP-1997, MFP-4517 and MP-1991.

3.3. Extraction and Isolation Procedures

Barks of *O. asbeckii* (2.0 Kg) and *G. dumontetii* (650 g) were dried in dry room for a week at 20% humidity. The samples were then crushed. Bark of *O. asbeckii* was extracted by maceration in EtOAc (3 × 2 L) at room temperature to yield 19.3 g. The bark of *G. dumontetii* was extracted using the Dionex® ASE 300 automatic extractor with EtOAc (3 × 100 mL) at 40 °C and 100 bar. The combined extracts were concentrated *in vacuo* at 35 °C to yield 11.7 g.

Leaves of *S. pachyantha* (3.1 Kg), *S. decipiens* (2.1 Kg), *S. guianensis* (6.5 Kg) and *S. poeppigii* (1.5 Kg), were dried in a dry room for a week at 20% humidity. The samples were then crushed. The dried and powdered plants materials were soaked separately in an alkaline solution of 25% NH_4OH and subjected to ethyl acetate extraction (1.5 L) for 12 h before filtration. The resulting alkaline extracts were then partitioned three times with 2% H_2SO_4 (250 mL). The aqueous layers were made alkaline (pH 10) with 25% NH_4OH and partitioned with $CHCl_3$ (1.5 L, at room temperature, 30 min) The chloroformic extracts were then washed three times with distilled water (250 mL) and dried with

sodium sulphate. The resulting solutions were evaporated to dryness to yield total alkaloids of *S. pachyantha* (1.6 g), *S. decipiens* (3.3 g), *S. guianensis* (12.5 g) and *S. poeppigii* (3.3 g).

The EtOAc extract of *G. dumontetii* (1.3 g) was subjected to flash chromatography using a gradient of CH_2Cl_2/MeOH (100:0 to 50:50) as the eluent and a flux rate of 20 mL/min. Twelve fractions were obtained. Fractions 2 and 3 were shown to inhibit CDK1 with IC_{50} values of 0.24 and 0.70 µg/mL. Fraction 2 (202.1 mg) was purified by preparative HPLC using a Kromasil column with an isocratic mobile phase (MeCN/H_2O 30:70) to yield compounds **4** (6.8 mg, w/w 0.0094%) and **5** (3.3 mg, w/w 0.0045%). Fraction 3 was purified twice. First, 120 mg was subjected to a Kromasil C-18 column with isocratic MeCN/H_2O (35:65 + 0.1% formic acid) for 30 min followed by 100% MeCN + 0.1% formic acid for 10 min to yield compound **3** (3.5 mg, w/w 0.0048%). Then, 185 mg was separated on a Symmetry Shield RP18 column with isocratic MeCN/H_2O (35:65 + 0.1% formic acid) for 25 min followed by 100% MeCN + 0.1% formic acid for 10 min to generate compound **2** (21.2 mg, w/w 0.0293%).

The EtOAc extract of the *O. asbeckii* bark (14.1 g) was subjected to flash chromatography using *n*-heptane/CH_2Cl_2 (100:0 to 0:100) followed by elution with CH_2Cl_2/MeOH (100:0–50:50) at a flux rate of 20 mL/min. Twenty fractions were obtained. Fractions 16 and 17 were shown to inhibit DYRK1A activity with IC_{50} values of 2.0 and 1.3 µg/mL. Fraction 16 was fractionated by flash chromatography using n-heptane/CH_2Cl_2 (100:0 to 0:100) followed by elution with CH_2Cl_2/MeOH (100:0–50:50) at a flux rate of 20 mL/min to give 25 fractions. The active fraction obtained was purified by TLC by using CH_2Cl_2/MeOH (97:3) as the eluent to give compound **2** (3.0 mg, w/w 0.0002%). Fraction 17 was purified twice. First, 1.2 g was fractionated by flash chromatography using petrol ether/CH_2Cl_2 (100:0 to 0:100) followed by CH_2Cl_2/MeOH (100:0–50:50). Then, 30.4 mg of the active fraction was separated on a Kromasil C-18 preparative column with isocratic MeCN/H_2O 35:65 + 0.1% formic acid for 30 min to give compounds **1** (1.7 mg, w/w 0.00012%) and **3** (1.0 mg, w/w 0.00007%).

The total alkaloid extracts of *Siparuna* spp. were fractionated on silica using centrifuge chromatography (Chromatotron™, Harrison Research). A solvent gradient from 100% $CHCl_3$ to $CHCl_3$/MeOH (90/10) was used as the first fractionation step. Alkaloids were isolated by preparative TLC using diverse mixtures (EtOAc/cyclohexane 50:50, $CHCl_3$/MeOH 30:70 and $CHCl_3$/MeOH 95:5 + NH_4). (+)-Corydine (**6**) (3.0 mg, w/w 0.00009%), (−)-roemerine (**7**) (2.5 mg, w/w 0.00008%) and liriodenine (**8**) (5.4 mg, 0.00017%) were isolated from *S. pachyantha*. From *S. guianensis*, (+)-bulbocapnine (**9**) (5.7 mg, w/w 0.00027%), (+)-*N*-methyllindcarpine (**10**) (3.5 mg, w/w 0.00016%), (+)-actinodaphnine (**11**) (15.2 mg, w/w 0.00072%), liriodenine (**8**) (20.8 mg, 0.00098%) and (+)-11-methoxynornoelistine (**13**) (3.2 mg, w/w 0.00015%) were isolated. Lysicamine (**13**) (8.0 mg, w/w 0.00012%), (−)-*O*-methylisopiline (**14**) (22.0 mg, w/w 0.00033%), (+)-*N*-nornuciferine (**15**) (11.3 mg, w/w 0.00017%) and liriodenine (**8**) (1.4 mg, w/w 0.00002%) were purified from *S. poeppigii*, and *N*-nornuciferine (**15**) (11.1 mg, w/w 0.00069%) and (+)-boldine (**16**) (5.6 mg, w/w 0.00035%) were purified from *S. decipiens*. The compounds were identified by spectroscopic analysis and compared to data reported in the literature (see supplementary material) [14–34].

3.4. Preparation and Assay of Protein Kinases

Kinase activities were assayed in buffer A or C (unless otherwise stated), at 30 °C, at a final ATP concentration of 15 µM. Values were background subtracted, and activities were calculated as pmoles

of phosphate incorporated during a 10 min incubation. The activities are usually expressed as % of the maximal activity, that is, in the absence of inhibitors. Controls were performed with appropriate dilutions of DMSO.

DYRK1A (rat, recombinant, expressed in *E. coli* as a GST fusion protein) was purified by affinity chromatography on glutathione-agarose and assayed as described [53].

CDK1/Cyclin B was extracted in homogenisation buffer from M phase starfish (*Marthasterias glacialis*) oocytes and purified by affinity chromatography on $p9^{CKShs1}$-Sepharose beads, from which it was eluted with free $p9^{CKShs1}$ as previously described [53]. The kinase activity was assayed in buffer C, with 1 mg of histone H1/mL, in the presence of 15 μM γ-^{33}P-ATP (3000 Ci/mmol; 1 mCi/mL) in a final volume of 30 μL. After a 10 min incubation at 30 °C, 25 μL aliquots of supernatant were spotted onto P81 phosphocellulose paper and treated as described above.

4. Conclusions

The search for inhibitors of DYRK1A and CDK1 led us to identify some active plants extracts. Afterward, the chemical investigations of *O. asbeckii* and *G. dumontetii* afforded a series of aristolactams of which the aristolactam AIIIA (**4**) was the strongest inhibitor of both CDK1/Cyclin B and DYRK1A activities. Eleven aporphinoid alkaloids were isolated from various *Siparuna* species, most of them for the first time in this genus. Liriodenine (**8**), 11-methoxynorneolistine (**12**), lysicamine (**13**) and *N*-nornuciferine (**15**) were moderate inhibitors of DYRK1A activity, but at a concentration of 30 μM, they did not inhibit CDK1/Cyclin B activity.

Supplementary Materials

Supplementary materials can be accessed at: http://www.mdpi.com/1420-3049/18/3/3018/s1.

Acknowledgments

The authors are very grateful to North and South provinces of New Caledonia who facilitated our field investigation. We express our thanks to J. Munzinger of the Botany Plant Ecology Department, IRD (*Institut de Recherche pour le Développement*) Nouméa and to the late M.-F. Prévot, French Guiana Herbarium, IRD Guyane, for their assistance in identifying the plants. We also thank C. Moretti of "*Centre Polynésien de Recherche et de Biodiversité Insulaire*", UMR7138 Systématique, adaptation, evolution, IRD Papeete, French Polynesia and V. Bultel MNHN, Paris, French, for their participation in this work. This work has also benefited from an "*Investissement d'Avenir*" grant managed by Agence Nationale de la Recherche (CEBA, ref. ANR-10-LABX-0025). This research was also supported by grants from the "*Fonds Unique Interministériel*" (FUI) PHARMASEA Project (L.M.).

References and Notes

1. Hunter, T. Protein kinases and phosphatases: The yin and yang of protein phosphorylation and signaling. *Cell* **1995**, *80*, 225–236.
2. Ferrer, I.; Barrachina, M.; Puig, B.; Martinez de Lagran, M.; Marti, E.; Avila, J.; Dierssen, M. Constitutive Dyrk1A is abnormally expressed in Alzheimer disease, Down syndrome, Pick disease, and related transgenic models. *Neurobiol. Dis.* **2005**, *20*, 392–400.

3. Becker, W.; Sippl, W. Activation, Regulation and Inhibition of DYRK1A. *FEBS J.* **2011**, *278*, 246–256.
4. Wegiel, J.; Gong, C.-X.; Hwang, Y-W. The role of DYRK1A in neurodegenerative diseases. *FEBS J.* **2011**, *278*, 236–245.
5. Harper, J.W.; Adams, P.D. Cyclin-dependent kinases. *Chem. Rev.* **2001**, *101*, 2511–2526.
6. Knockaert, M.; Greengard, P.; Meijer, L. Pharmacological inhibitors of cyclin-dependent kinases. *Trends Pharmacol. Sci.* **2002**, *23*, 417–425.
7. Shapiro, G.I. Cyclin-Dependent Kinase Pathways as Targets for Cancer Treatment. *J. Clin. Oncol.* **2006**, *24*, 1770–1783.
8. Malumbres, M.; Barbacid, M. Mammalian cyclin-dependent kinases. *Trends Biochem. Sci.* **2005**, *30*, 630–641.
9. Kamb, A. Cyclin-dependent kinase inhibitors and human cancer. *Curr. Top. Microbiol. Immunol.* **1998**, *227*, 139–148.
10. Pines, J. Cyclins and cyclin-dependent kinases: Take your partners. *Trends Biochem. Sci.* **1993**, *18*, 195–197.
11. Savage, M.J.; Gingrich, D.E. Advances in the development of kinase inhibitor therapeutics for Alzheimer's disease. *Drug Dev. Res.* **2009**, *70*, 125–144.
12. Apel, C.; Dumontet, V.; Lozach, O.; Meijer, L.; Guéritte, F.; Litaudon, M. Phenanthrene derivatives from *Appendicula reflexa* as new CDK1/cyclin B inhibitors *Phytochem. Lett.* **2012**, *5*, 814–818.
13. The species *G. dumontetii*, which was the first representative of the genus *Goniothalamus* in New Caledonia was discovered in October 1997 by one of us (M.L.) in the "Special Reserve of Nodela Flora" (South Province), and was named in 2007 by R.M.K. Sauders and Munzinger (*Bot. J. Linn. Soc.* **2007**, *155*, 497–503), following the chemical and biological studies carried out.
14. Crohare, R.; Priestap, H.A.; Farina, M.; Cedola, M.; Ruveda, E.A. Aristololactams of *Aristolochia argentina*. *Phytochemistry* **1974**, *13*, 1957–1962.
15. Wang, E.-C.; Shih, M.-H.; Liu, M.-C.; Chen, M.-T.; Lee, G.-H. Studies on constituents of *Saururus chinensis*. *Heterocycle* **1996**, *43*, 969–976.
16. Omar, S.; Chee, C.L.; Ahmad, F.; Ni, J.X.; Jaber, H.; Huang, J.; Nakatsu, T. Phenanthrene lactams from *Goniothalamus velutinus*. *Phytochemistry* **1992**, *31*, 4395–4397.
17. Hedge, V.R.; Borges, S.; Patel, M.; Das, P.R.; Wu, B.; Gullo, V.P.; Chan, T.-Z. New potential antitumor compounds from the plant *Aristolochia manshuriensis* as inhibitors of the CDK2 enzyme. *Bioorg. Med. Chem.* **2010**, *20*, 1344–1346.
18. Li, N.; Wu, J.-L.; Hasegawa, T.; Sakai, J.-I.; Bai, L.-M.; Wang, L.-Y.; Kakuta, S.; Furuya, Y.; Ogura, H.; Kataoka, T.; *et al*. Bioactive lignans from *Peperomia duclouxii. J. Nat. Prod.* **2007**, *70*, 544–548.
19. Xiong, L.; Zhu, C.; Li, Y.; Tian, Y.; Lin, S.; Yuan, S.; Hu, J.; Hou, Q.; Chen, N.; Yang, Y.; Shi, J. Lignans and Neolignans from *Sinocalamus affinis* and Their Absolute Configurations. *J. Nat. Prod*. **2011**, *74*, 1188–1200.
20. Ferreira, M.L.R.; de Pascoli, I.C.; Nascimento, I.R.; Lopes, L.M.X.; Zukerman-Schpector, J. Aporphine and bisaporphine alkaloids from *Aristolochia lagesiana* var. *intermedia*. *Phytochemistry* **2010**, *71*, 469–478.

21. Guinaudeau, H.; Leboeuf, M.; Debray, M.; Cave, A.; Paris, R.R. Alkaloids of *Colubrina faralaotra* ssp. *faralaotra Planta Med*. **1975**, *27*, 304–316.
22. Bhakuni, D.S.; Tewari, S.; Dhar, M.M. Aporphine alkaloids of *annona squamosa*. *Phytochemistry* **1972**, *11*, 1819–1822.
23. Simas, N.K.; Ferrari, S.F.; Pereira, S.N.; Leitao, G.G. Chemicalecological characteristics of herbivory of *Siparuna guianensis* seeds by buffy-headed marmosets (*Callithrix flaviceps*) in the Atlantic forest of southeastern Brazil. *J. Chem. Ecol.* **2001**, *27*, 93–108.
24. Chen, K.-S.; Wu, Y.-C.; Teng, C.-M.; Ko, F.-N.; Wu, T.-S. Bioactive alkaloids from *Illigera luzonensis*. *J. Nat. Prod*. **1997**, *60*, 645–647.
25. Denisenko, O.N.; Israilov, I.A.; Chelombitko, V.A.; Yunusov, M.S. Alkaloids of *Corydalis marschalliana*. *Chem. Nat. Compd.* **1993**, *29*, 690–691.
26. Kamenati, T.; Sugahara, T.; Fukumoto, K. Studies on total photolytic synthesis of alkaloids—III: The products of photo-pschorr reaction—Total synthesis of isocorydine. *Tetrahedron* **1971**, *27*, 5367–5374.
27. Johns, S.R.; Lamberton, J.A. Alkaloids of *Phoebe clemensii* Allen (family Lauraceae). *Aust. J. Chem.* **1967**, *20*, 1277–1281.
28. Stévigny, C.; Block, S.; Pauw-Gillet, M.C.; de Hoffmann, E.; de Llabres, G.; Adjakidje, V.; Quetin-Leclercq, J. Cytotoxic aporphine alkaloids from *Cassytha filiformis*. *Planta Med.* **2002**, *68*, 1042–1044.
29. Buchanan, M.S.; Carroll, A.R.; Pass, D.; Quinn, R.J. Aporphine Alkaloids from the Chinese Tree *Neolitsea aurata* var. *paraciculata*. *Nat. Prod. Com.* **2007**, *2*, 255–259.
30. Orito, K.; Uchiito, S.; Satoh, Y.; Tatsuzawa, T.; Harada, R.; Tokuda, M. Aryl Radical Cyclizations of 1-(2'-Bromobenzyl)isoquinolines with AIBN-Bu3SnH: Formation of Aporphines and Indolo[2,1-a]isoquinolines. *Org. Lett.* **2000**, *2*, 307–310.
31. Hocquemiller, R.; Rasamizafy, S.; Cavé, A.; Moretti, C. Alcaloïdes des Annonacees XXXVII: Alcaloïdes du *Guatteria scandens*. *J. Nat. Prod.* **1983**, *46*, 335–341.
32. Montenegro, H.; Gutierrrez, M.; Romero, L.I.; Ortega-Barria, E.; Capson, T.L.; Cubilla Rios, L. Aporphine alkaloids from *Guatteria* spp. with leishmanicidal activity. *Planta Med.* **2003**, *69*, 677–679.
33. Gunawardana, Y.A.; Geewanda, P.; Huck-Meng, L.; Bick, I.; Ralph, C. Alkaloids of *Hedycarya angustifolia*. *Heterocycles* **1987**, *26*, 447–456.
34. Sobarzo-Sanchez, E.; Cassels, B.K.; Saitz-Barria, C.; Jullian, C. Oxazine- and oxazole-fused derivatives of the alkaloid boldine and their complete structural and spectral assignments by HMQC and HMBC experiments. *Magn. Reson. Chem.* **2001**, *39*, 361–368.
35. Braz-F, R.; Gabriel, S.J.; Gomes, C.M.R.; Gottlieb, O.R.; Bichara, M.D.G.A.; Maia J.G.S. Oxoaporphine alkaloids from *Fusea longifolia* and *Siparuna guianensis*. *Phytochemistry* **1976**, *15*, 1187–1188.
36. Guinaudeau, H.; Leboeuf, M.; Cavé, A. Aporphine Alkaloids. II. *J. Nat. Prod.* **1979**, *42*, 325–360.
37. Guinaudeau, H.; Leboeuf, M.; Cavé, A. Aporphinoid Alkaloids, III. *J. Nat. Prod.* **1983**, *46*, 761–835.
38. Guinaudeau, H.; Leboeuf, M.; Cavé, A. Aporphinoid alkaloids. *J. Nat. Prod.* **1994**, *57*, 1033–1135.
39. Castedo, L.; Tojo, G. Phenantrene Alkaloids. In *The Alkaloid*; Brossi, A., Ed.; Academic: New York, NY, USA, 1990; Volume 39, pp. 99–138.

40. Chen, Z.-L.; Zhu, D.-Y. Aristolochia Alkaloids. In *The Alkaloids: Chemistry and Pharmacology*; Brossi, A., Ed.; New York, NY, USA, 1987; Volume 31, pp. 29–65.
41. Kumar, V.; Poonam, P.A.K.; Parmar, V.S. Naturally occurring aristolactams, aristolochic acids and dioxoaporphines and their biological activities. *Nat. Prod. Rep.* **2003**, *20*, 565–583.
42. Bentley, K.W. β-Phenylethylamines and the isoquinoline alkaloids. *Nat. Prod. Rep.* **2006**, *23*, 444–463.
43. Zhang, Y.-N.; Zhong, X.-G.; Zheng, Z.-P.; Hu, X.-D.; Zuo, J.-P.; Hu, L.-H. Discovery and synthesis of new immunosuppressive alkaloids from the stem of *Fissistigma oldhamii* (Hemsl.) Merr. *Bioorg. Med. Chem.* **2007**, *15*, 988–996.
44. Choi, Y.L.; Kim, J.K.; Choi, S.-U.; Min, Y.-K.; Bae, M.-A.; Kim, B.T.; Heo, J.-N. Synthesis of aristolactam analogues and evaluation of their antitumor activity. *Bioorg. Med. Chem. Lett.* **2009**, *19*, 3036–3040.
45. Hedge, V.R.; Borges, S.; Pu, H.; Patel, M.; Gullo, V.P.; Wu, B.; Kirchmeier, P.; William, M.J.; Madison, V.; Fischmann, T.; Chan, T.-Z. Semi-synthetic aristolactams—inhibitors of CDK2 enzyme. *Bioorg. Med. Chem. Lett.* **2010**, *20*, 1384–1387.
46. Li, L.; Wang, X.; Chen, J.; Ding, H.; Zhang, Y.; Hu, T.-C. Hu, L.-H.; Jiang, H.-L.; Shen, X. The natural product Aristolactam AIIIa as a new ligand targeting the polo-box domain of polo-like kinase 1 potently inhibits cancer cell proliferation. *Acta Pharm. Sin.* **2009**, *30*, 1443–1453.
47. Swaffar, D.S.; Holley, C.J.; Fitch, R.W.; Elkin, K.R.; Zhang, C.; Sturgill, J.P.; Menachery, M.D. Phytochemical investigation and *in vitro* cytotoxic evaluation of alkaloids from *Abuta rufescens*. *Planta Med.* **2012**, *78*, 230–232.
48. Costa, E.V.; Pinheiro, M.L.B.; Barison, A.; Maia, B.H.L.N.S.; Campos, F.R.; Salvador, M.J.; Cabral, E.C.; Eberlin, M.N. Alkaloids from the bark of *Guatteria hispida* and their evaluation as antioxidant and antimicrobial agents. *J. Nat. Prod.* **2010**, *73*, 1180–1183.
49. Chang, H.-C.; Chang, F.-R.; Wu, Y.-C.; Lai, Y.-H. Anti-cancer effect of liriodenine on human lung cancer cells. *Kaohsiung J. Med. Sci.* **2004**, *20*, 365–371.
50. Chen, C.-Y.; Chen, S.-Y.; Chen, C.-H. Liriodenine induces G1/S cell cycle arrest in human colon cancer cells via nitric oxide- and p53-mediated pathway. *Process Biochem.* **2012**, *47*, 1460–1468.
51. Kim, K.H.; Moon, E.; Choi, S.U.; Kim, S.Y.; Lee, K.R. Biological evaluation of phenolic constituents from the trunk of *Berberis koreana*. *Bioorg. Med. Chem. Lett.* **2011**, *21*, 2270–2273.
52. Li, F.; Awale, S.; Tezuka, Y.; Kadota, S. Cytotoxic constituents from Brazilian red propolis and their structure-activity relationship. *Bioorg. Med. Chem.* **2008**, *16*, 5434–5440.
53. Beauchard, A.; Ferandin, Y.; Frère, S.; Lozach, O.; Blairvacq, M.; Meijer, L.; Thiéry, V.; Besson, T. Synthesis of novel 5-substituted indirubins as potential inhibitor of protein kinases. *Bioorg. Med. Chem.* **2006**, *14*, 6434–6443.

Sample Availability: Samples of the compounds **1**–**15** are available from the authors.

Food Chemistry 150 (2014) 235–245

Contents lists available at ScienceDirect

Food Chemistry

journal homepage: www.elsevier.com/locate/foodchem

Analytical Methods

Comprehensive profiling and marker identification in non-volatile citrus oil residues by mass spectrometry and nuclear magnetic resonance

Guillaume Marti [a], Julien Boccard [a], Florence Mehl [a], Benjamin Debrus [a], Laurence Marcourt [a], Philippe Merle [b], Estelle Delort [b], Lucie Baroux [b], Horst Sommer [b], Serge Rudaz [a], Jean-Luc Wolfender [a,*]

[a] School of Pharmaceutical Sciences, EPGL, University of Geneva, University of Lausanne, 30 Quai Ernest-Ansermet, 1211 Geneva 4, Switzerland
[b] Firmenich SA, Corporate R&D Division, Geneva, CH-1211, Switzerland

ARTICLE INFO

Article history:
Received 24 February 2013
Received in revised form 15 October 2013
Accepted 24 October 2013
Available online 1 November 2013

Keywords:
Cold-pressed lemon oils
Classification
Geographic origin
^{1}H NMR
UHPLC–TOF-MS
Chemometrics
MS–NMR correlation analysis

ABSTRACT

The detailed characterization of cold-pressed lemon oils (CPLOs) is of great importance for the flavor and fragrance (F&F) industry. Since a control of authenticity by standard analytical techniques can be bypassed using elaborated adulterated oils to pretend a higher quality, a combination of advanced orthogonal methods has been developed. The present study describes a combined metabolomic approach based on UHPLC–TOF-MS profiling and ^{1}H NMR fingerprinting to highlight metabolite differences on a set of representative samples used in the F&F industry. A new protocol was set up and adapted to the use of CPLO residues. Multivariate analysis based on both fingerprinting methods showed significant chemical variations between Argentinian and Italian samples. Discriminating markers identified in mixtures belong to furocoumarins, flavonoids, terpenoids and fatty acids. Quantitative NMR revealed low citropten and high bergamottin content in Italian samples. The developed metabolomic approach applied to CPLO residues gives some new perspectives for authenticity assessment.

1. Introduction

Citrus (Rutaceae) essential oils are widely employed in the perfumery, cosmetic, pharmaceutical and food industries for their rich olfactory profiles as well as for their fixative and antimicrobial properties (Maffei, Gertsch, & Appendino, 2011). Cold-pressed lemon oils (CPLOs) are obtained by industrial extraction from the peels of citrus fruits, resulting in a complex product of several hundred compounds (Dugo & Mondello, 2010). CPLOs are typically composed of a mixture of a significant volatile fraction (85–99%) that can be further processed by distillation and the remaining 1–15% non-volatile residue (CPLOR) (Dugo & Mondello, 2010). The volatile fraction is predominantly composed of mono- and sesquiterpene hydrocarbons and their oxygenated derivatives, aliphatic aldehydes, alcohols and esters, whereas the non-volatile fraction contains hydrocarbons, sterols, fatty acids, waxes, non-volatile terpenes, carotenoids and flavonoids, as well as coumarins and furocoumarins that are largely found in the Rutaceae family (Groppo, Pirani, Salatino, Blanco, & Kallunki, 2008).

The complex composition of CPLOs influences their organoleptic qualities and varies according to the geographic origin of the plant, the types of cultivars used, the extraction process and the harvest date (Costa et al., 2010; Luro et al., 2012). Because of the importance of CPLOs in many industrial fields, a precise assessment of the composition of this product is necessary for quality assurance, authenticity evaluation, adulteration detection and classification according to phenotypic criteria. In this context, numerous analytical methods based mainly on hyphenated chromatographic techniques have been developed (Tranchida, Bonaccorsi, Dugo, Mondello, & Dugo, 2012). Most of the studies have been focused on the analysis of the volatile fraction of CPLOs that can efficiently be investigated by gas chromatography–mass spectrometry–flame ionisation detector (GC–MS–FID) (Tranchida et al., 2012). Although the non-volatile fraction represents a relatively small part of the total citrus oil content, its analysis is also highly important, as key compounds may act as odour fixatives and can strongly influence the organoleptic properties of the entire oil (Sommer et al., 2003). Additionally, it can constitute a relevant source of information for authentication purposes. Several studies have highlighted a typical pattern of the non-volatile fraction with regard to the genuineness of citrus essential oils (Dugo & Mondello, 2010). Moreover, the high levels of oxygen heterocyclic compounds in the non-volatile fraction is of significant interest with regard to their pharmacological properties (Santana, Uriarte, Roleira, Milhazes, & Borges, 2004). The toxicological effects of these compounds are also of concern because photo-toxicity has previously been highlighted by several reports (Placzek, Fromel,

* Corresponding author. Tel.: +41 223793385; fax: +41 223793399.
E-mail address: Jean-Luc.Wolfender@unige.ch (J.-L. Wolfender).

0308-8146/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.foodchem.2013.10.103

Eberlein, Gilbertz, & Przybilla, 2007). Because of the considerable diversity of their constituents, which subsequently results in different physicochemical properties, CPLOs are challenging to analyse with conventional reverse phase (RP) chromatography. However, most of the liquid chromatography (LC) based methods dedicated to the analysis of the non-volatile fraction involve RP-based C18 columns to achieve high-performance liquid chromatography (HPLC), with either UV (LC-UV) or MS (LC–MS) detection (Frerot & Decorzant, 2004; Tranchida et al., 2012).These methods have mainly been developed to characterise and quantify coumarins and furocoumarins. These constituents possess characteristic chromophores and are well detected with LC-UV, while they are known to be unstable under typical GC inlet conditions (Peroutka, Schulzova, Botek, & Hajslova, 2007). The sensitivity response function and accuracy of reversed-phase HPLC-UV and HPLC–MS was previously compared (Frerot & Decorzant, 2004). The results obtained with HPLC-UV at 310 nm within 40 min were found to be promising, but several co-elutions occurred during the analysis of the citrus oil samples, leading to a problem of overall selectivity. The same method was also used to assess the limits of quantification (LOQ) based on HPLC-UV for complex fragrance mixtures analyses (Macmaster et al., 2012). More comprehensive studies were achieved using MS- and NMR-based methods. For example, eleven new psoralens and coumarins were identified in CPLOs by combining HPLC–MS, GC–MS and NMR data (Ziegler & Spiteller, 1992). CPLOR analyses by LC–NMR resulted in the identification of 3 coumarins and 13 psoralens (Sommer et al., 2003). Fast HPLC coupled to a diode array detector (DAD) allowed the separation of 21 oxygen heterocyclic compounds in citrus oils from different species within eight minutes (Bonaccorsi, McNair, Brunner, Dugo, & Dugo, 1999). Comprehensive LC × LC, have also been used for the analysis of psoralens and carotenoids in CPLOs (Dugo et al., 2006; Tranchida et al., 2012).

To date, very few studies have used either direct or hyphenated fingerprinting spectroscopic method with multivariate data analysis (MVA) for the sample classification or biomarker identification of CPLOs. Near-infrared spectroscopy (NIR) was previously used to discriminate citrus species and to quantify key components of citrus oils (Steuer, Schulz, & Lager, 2001). GC–MS was successfully applied to unravel the chemical polymorphisms of several citrus oils from different taxa (Lota, de Rocca Serra, Tomi, Jacquemond, & Casanova, 2002). In a recent study, we reported the classification of industrial citrus oils by means of spectroscopic or separative methods and MVA (Mehl et al., 2014).

The present study investigates methods that provide a rapid fingerprinting of CPLORs based on NMR fingerprinting and UHPLC–TOF-MS rapid profiling. These methods were applied to the study of industrial batches of CPLOs from two different geographic origins. Chemometric tools were used to assess the classification potential of the approach and to identify and quantify significant markers. A covariance analysis of the orthogonal LC–MS and NMR fingerprints was also performed to investigate the potential of such multiplatform approaches for the identification of key components that were detected with both methods.

2. Experimental section

2.1. Description of the sample set

In an industrial context, CPLOs may come from suppliers that blend both cultivars as well as extraction processes. To evaluate the magnitude of changes in composition mainly due to origins and cultivars, a selected set of oils from Italy and Argentina was used as a method development support. Supplier samples were used as such to focus on the origin only. Eight independent production batches were obtained from the same supplier for each country and were composed of 16 CPLOs from 2 different geographic origins (Italy ($n = 8$) and Argentina ($n = 8$)). All samples were produced by the BOE (Brown Oil Extractor) industrial extraction process. Italian oils were obtained from the Feminello cultivar and extracted between January and February 2011. Oils originating from Argentina were of the Eureka cultivar and extracted between June and July 2011.

2.2. Sample preparation

To remove the major volatile constituents from the CPLOs, 1 mL of $CHCl_3$ was added to 400 μL (~340 mg) of each sample and evaporated under vacuum at 40 °C for 6 h (Genevac HT-4X, Ipswich, UK). The remaining residue in each flask, containing only the non-volatile fraction of the CPLOs (i.e., the CPLOR) was weighed (~10–16 mg) and used for further analyses.

2.3. NMR spectra acquisition

The CPLORs were dissolved in deuterated chloroform at 10 mg/mL and transferred to 5 mm NMR tubes. ^{1}H NMR spectra were recorded on a Bruker® AVANCE I 500 NMR spectrometer at 500.13 MHz (Bruker BioSpin GmbH, Fällanden, Switzerland) equipped with a BBO probe. The ^{1}H NMR acquisition was also optimised for quantification purposes using the PULCON method developed by Wider and Dreier (2006). The 360° RF pulse was calibrated with a sufficient relaxation delay, minimising the influence of different relaxation times between external standard samples (bergapten) and analyte samples. For each sample, 64 K data points were collected using a spectral width of 8000 Hz with a relaxation delay of 20 s, an acquisition time of 4.09 s and a pulse width of 11 μs (90° pulse). All spectra were recorded at a constant temperature of 298 K.

Each spectrum was imported into MNova 7.1 (Mestrelab Research, Santiago, Spain) for spectra alignment and peak detection. The CPLOR ^{1}H NMR fingerprints were aligned in reference to the $CDCl_3$ solvent residual signal. The region from 7.25 to 7.28 ppm was removed prior to data reduction using a bin width of 0.02 ppm. The spectra were then normalised to total area and exported as ASCII files. 2D-NMR spectra were recorded from a pooled sample at a concentration of 30 mg/mL in $CDCl_3$ on the same instrument. COSY, HSQC and HMBC spectra were measured using the standard pulse sequences of the Bruker pulse library.

2.4. UHPLC–TOF-MS rapid profiling

MS fingerprints were acquired on a Micromass LCT Premier time-of-flight (TOF) mass spectrometer (Waters, MA, USA) equipped with an electrospray interface and coupled with an Acquity UPLC system (Waters, MA, USA). The ESI conditions were as follows: capillary voltage, 2800 V; cone voltage, 40 V; MCP detector voltage, 2400 V; source temperature, 120 °C; desolvation temperature, 300 °C; cone gas flow, 20 L/h; and desolvation gas flow, 800 L/h. Detection was performed in positive ion (PI) mode with a m/z range of 100–1000 Da and a scan time of 0.25 s in "W" mode. The mass spectrometer was calibrated using sodium formate, and leucine enkephalin was used as the internal standard at 2 μg/mL and infused through the Lock Spray™ probe at a flow rate of 10 μL/min using a second LC pump (Shimadzu LC-10ADvp, Duisburg, Germany).

Three UHPLC columns, C4 (Waters Acquity BEH300 C4, 50 × 2.1 mm i.d., 1.7 μm), C8 (Waters Acquity BEH C8, 50 × 1 mm i.d., 1.7 μm) and phenyl (Waters Acquity BEH Phenyl, 50 × 1 mm i.d., 1.7 μm), were assessed using a gradient beginning at 95% water acidified with 0.1% formic acid (FA) that increased to

95% acetonitrile acidified with 0.1% FA or methanol with 0.1% FA in 5 min while maintaining a constant column temperature of 40 °C.

The column temperature was changed from 40 °C to 70 °C while adapting the flow rate to preserve the optimal theoretical peak capacity that was calculated with a freely available HPLC calculator software.

To compare the CPLOs and CPLORs, a gradient was used that progressed from 95% of water acidified with 0.1% FA (phase A) to 95% methanol acidified with 0.1% FA (phase B) in 3 min using the phenyl column at a constant temperature of 60 °C. After peak picking (see below), the spectra were computed to build a data table of 28 observations × 680 variables for further multivariate data analysis.

For CPLOR analyses, an aliquot of each sample prepared for NMR spectroscopy was dissolved in methanol at 1 mg/mL to achieve UHPLC–UV-DAD-TOF-MS profiles. The same mobile phases were used with a linear gradient as follows: 65% A to 95% B over 4.0 min, held at 95% B for a further 1.1 min, then returned to initial conditions (65% A) in 0.1 min for 1.1 min of equilibration before subsequent analysis. The flow rate was set to 0.4 mL/min, and the column temperature was kept at 60 °C.

The UV spectra were recorded from 200 to 450 nm with a resolution of 1.2 nm. All samples were injected in a random order, and blank solvent containing only the dilution solvent (90/10 MeOH/$CDCl_3$, v/v) and a quality control (QC) sample containing an aliquot of all the samples were injected every four samples. This ensured the reproducibility of data acquisition and the absence of carryover during the analyses (Ranjbar, Wang, & Ressom, 2011).

UHPLC–TOF-MS fingerprints were processed with MZmine 2.9.1 for mass signal extraction and alignment from 0 to 5 min with *m/z* values from 100 to 1000 Da with the following parameters: baseline correction was applied to all chromatograms, and the chromatogram builder was set to a minimum time span of 0.06 min, a minimum height of 10 and a *m/z* tolerance of 10 ppm. The local minimum search algorithm was applied for chromatogram deconvolution. Each peak list was de-isotoped and aligned using the RANSAC alignment method. The resulting peak matrix, containing areas of aligned peaks characterised by retention time and *m/z* ratio, from each sample was gap-filled and exported into .csv files prior to multivariate data analysis.

2.5. Multivariate data analysis

For each analytical technique used in this study, the data were directly imported into SIMCA-P© software v12.1 (Umetrics, Sweden) to perform multivariate data analysis. The data were Pareto scaled, and a 7-fold procedure was applied to cross-validate the PCA and OPLS-DA models. NMR–MS correlations were computed using MatLab® version R2012a (The MathWorks, Natick, MA, USA).

2.6. Chemical standards

All furocoumarin standards used in the study of Frerot et al. were provided by Firmenich SA (Frerot & Decorzant, 2004).

3. Results and discussion

Two complementary metabolite fingerprinting methods were considered for the rapid high-throughput analysis of the non-volatile constituents of the CPLOs: ^{1}H-NMR and LC–MS. The reproducibility, high throughput and resolution of the metabolite features were evaluated during the development of the methods. The methods were developed in the context of their application to a high number of industrial batch samples over a large time scale.

3.1. High-throughput UHPLC–TOF-MS method development

For LC–MS, the development of a rapid profiling method was necessary for the high-throughput monitoring of industrial citrus oils in order to be in-line with the requirement of MS-based metabolomics (Wolfender, Rudaz, Choi, & Kim, 2012). Ultra-high-pressure liquid chromatography coupled to time-of-flight mass spectrometry (UHPLC–TOF-MS) was chosen as an analytical platform, which presents numerous advantages for metabolomic studies. In UHPLC, the use of columns packed with sub-2 μm particles yields the same peak capacity compared to standard HPLC columns with gradient times shorter than 10 min for complex plant extracts (Wolfender, Marti, & Ferreira Queiroz, 2010). In TOF-MS, the high resolving power of the detector increases the number of detectable features and provides high mass accuracy data for molecular formula assignments and further peak annotation (Wolfender et al., 2012).

As a first step of the analytical development, several chromatographic conditions were assessed (e.g., solvent, column phase chemistry, temperatures and flow rate). Due to the high levels of apolar compounds with high log *P* in the citrus oil blends, reversed phases based on C18 were rapidly rejected to avoid carryover issues and to limit irreversible adsorption on the stationary phase. Indeed, preliminary assays on sub-2 μm particle size C18 columns with ESI-TOF-MS detection revealed that hydrophobic peaks with strong intensity continued to elute with 100% organic phase and that the column pressure increased after a few injections and needed to be washed with a strong apolar solvent.

Sub-2 μm columns with lower hydrophobicity properties (e.g., C4, C8 and Phenyl) were therefore chosen for evaluation. To assess the quality of the short LC–MS profiling, the number of features detected in ESI positive ionisation mode was measured first. A representative pool of CPLOs was analysed in triplicate with the three types of columns using two organic solvents (see supplementary S1A). In all cases, methanol provided a better detection and yielded roughly two times more features than acetonitrile. The C4-grafted stationary phase showed the lowest number of detected features, while the C8 and phenyl-grafted particles resulted in approximately the same number of detectable compounds.

Second, the selectivity was assessed by checking the separation of selected furocoumarin isomer standards, as they are key components of CPLOs (e.g., heraclenin and oxypeucedanin, 8-geranyloxypsoralen and bergamottin, imperatorin and isoimperatorin). As shown in supplementary S1B, the phenyl-grafted stationary phase demonstrated a better peak shape for the furocoumarin standards compared to the C8 column, while the observed selectivity with the C4 column was clearly insufficient for these compounds. The high selectivity of the phenyl column can be attributed to the π–π interactions between the aromatic compounds and the phenyl-grafted phase (Yang, Fazio, Munch, & Drumm, 2005). To increase the throughput and to further improve the elution of the apolar constituents, the influence of column temperature was investigated. The temperature (40–70 °C) and flow rate (0.3–0.6 mL/min) conditions were optimised to maintain the highest possible peak capacity for a given gradient (<5 min) (Guillarme et al., 2009). An increase in temperature reduced the analysis time for a similar peak capacity and the gradient time was reduced by a factor of two in increasing the temperature from 40 °C to 70 °C (data not shown). A temperature of 60 °C was finally selected because no degradation could be detected for any of the standards, and it permitted a rapid gradient time of 3 min.

To estimate the carryover effects, 8 successive injections of pooled CPLOs dissolved at 1% v/v in methanol with interleaved methanol blank samples every two analyses were performed. The percentage of common detected features in the blank and CPLO samples was determined. After 8 injections, 65% of the detected

features in the CPLO samples were also detected in the blank samples. These results highlighted a strong carryover issue, which was of concern for the reliability of further data analyses. In the optimised short profiling obtained on the phenyl column, a close inspection of the UHPLC–TOF-MS chromatogram of the pooled CPLOs (see Fig. 1A-top) showed many peaks eluting at the end of the gradient between two and three minutes. These apolar compounds were suspected to increase the carryover and consequently affect the instrument repeatability. Because the CPLOs contained a very large proportion of volatile lipophilic constituents, it was assumed that a portion of these late-eluting compounds was responsible for the strong carryover measured. To remove the volatile compounds from the citrus oil blend, a method that provide efficient evaporation and yielded a stable non-volatile residue (CPLOR) for further analyses was investigated. Therefore, enriched CPLORs were obtained after vacuum-assisted centrifugal evaporation of the volatile fraction under mild conditions (see experimental).

The effect of the evaporation procedure was further characterised by means of UHPLC–TOF-MS profiling. For this purpose, 7 independent CPLOs were evaporated, and the CPLORs were dissolved at a concentration of 1 mg/mL in MeOH. Similarly, the same CPLO samples were directly dissolved in 1% v/v in methanol to compare their profiles. As shown in Fig. 1A, the procedure removed the very lipophilic constituents and predominantly highlighted compounds of medium polarity. To further evaluate the stability of the CPLOs and CPLORs, each sample was injected just after dilution in MeOH and then again 24 h after the first analysis. Principal component analysis (PCA) was performed after extracting mass features from the 28 recorded UHPLC–TOF-MS profiles (see Fig. 1B). As expected, the CPLOs and CPLORs were separated into two distinct groups along the first principal component (PC) axis, which comprised 66% of the initial variance, indicating the main variability originated from the evaporation, as this was also clearly observed in the individual profiles (Fig. 1A). A closer examination of the groupings revealed that the CPLO samples were tightly clustered just after dilution but showed a dispersion effect along the second PC after 24 h, while the CPLOR groups were clustered uniformly. The variability observed for the CPLO samples after 24 h was most likely due to degradation products in the organic solvent, while the good clustering of the CPLORs indicate good overall stability after evaporation. Furthermore, the analysis of the interleaved blank samples of the CPLOR series showed no significant difference between the blanks prior to injection of the samples and the interleaved blanks indicating no detectable carryover effect.

Based on these results, the best method for profiling the non-volatile constituents of CPLOs was the analysis of the corresponding CPLORs after evaporation using a phenyl sub-2 μm column (see experimental). Because only CPLORs were considered for further analysis, the gradient slope used for the rapid UHPLC–TOF-MS profiling was re-optimised to better separate the detected features throughout the chromatogram (see experimental).

3.1.1. ^{1}H-NMR fingerprinting

NMR fingerprinting was also considered as an orthogonal method for assessing metabolomic variations within the CPLO batches, and the ^{1}H NMR spectra of the CPLOs and CPLORs were compared. ^{1}H NMR is indeed a robust high-throughput and non-destructive

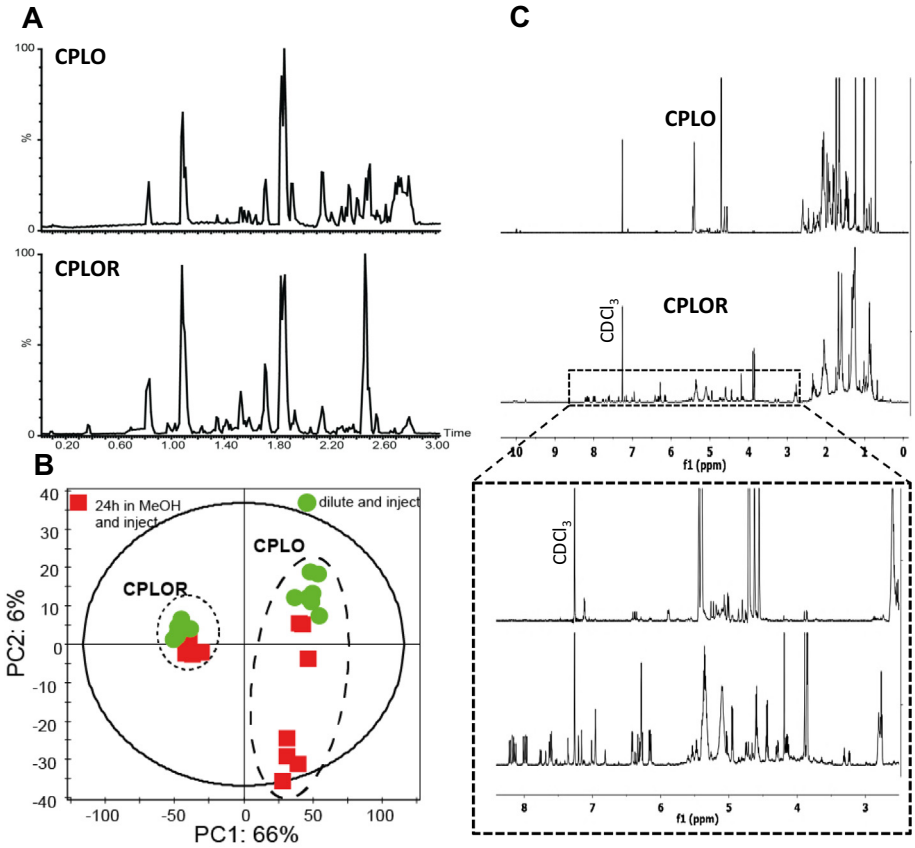


Fig. 1. Analysis of CPLOs and CPLORs. (A) UHPLC–TOF-MS base peak intensity chromatogram of pooled CPLOs and CPLORs in ESI PI mode. (B) Principal component analysis (PCA) score plot from UHPLC–TOF-MS analysis of seven citrus oil samples diluted at 1% in methanol (CPLOs) or dried and diluted at 1 mg/mL in methanol (CPLORs). The sample set was injected just after dilution (green dots) and 24 h later (red squares). (C) ^{1}H NMR spectra of pooled CPLOs and CPLORs in $CDCl_3$. The inset focuses on the NMR domain ranging from 2.5 to 8.5 ppm. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

method, and it provides reproducible fingerprints of the main constituents of natural extracts that are compatible with further MVA for either classification or biomarker identification (Wolfender et al., 2012). Contrary to MS, this approach yields direct quantitative relationships between the features, but it lacks sensitivity. It has also frequently been used for quality control and authenticity studies (Cevallos-Cevallos, Reyes-De-Corcuera, Etxeberria, Danyluk, & Rodrick, 2009).

To investigate whether relevant information could be obtained on the non-volatile constituents of the CPLOs, the samples were dissolved in $CDCl_3$ and analysed. As expected, a visual inspection of the CPLO features showed that the samples were strongly dominated by the signals of the volatile constituents (see Fig. 1C-top), and complementary 2D-NMR analysis indicated that the most intense signals were from limonene, which is well known as the main component of the citrus oils. In these spectra, the signals of aromatics that could be ascribed to the non-volatiles constituents (e.g., furocoumarins) were not detected most likely because of the restricted dynamic range of NMR spectroscopy and their presence at low concentration. As for LC–MS, the analysis of the residue (CPLORs) highlighted features that were characteristic of the non-volatile constituents between 2.5 and 8.5 ppm and that were hidden in the CPLOs due to the intense aliphatic proton signals (inset Fig. 1C). The superimposed NMR spectra (see Fig. 1C) of the CPLOs and CPLORs showed a completely different pattern.

Thus, the analysis of CPLORs instead of the CPLOs was found to be much more efficient for both NMR fingerprinting and LC–MS profiling. These orthogonal analytical methods were further evaluated for their classification potential, marker identification and quantification between CPLO batches of different origins.

3.2. Analysis of Argentinian and Italian industrial samples

To evaluate the potential of the developed high-throughput metabolomic methods, independent batches of CPLOs from Italy and Argentina, all extracted by brown oil extraction (BOE) industrial processes, were analysed. The sample set considered consisted of 8 samples of each origin that were analysed in three independent series.

3.2.1. Multivariate data analysis of the UHPLC–TOF-MS CPLOR fingerprints

As described, all CPLOs were evaporated, and the corresponding CPLORs were analysed using the optimised UHPLC–TOF-MS rapid profiling method. The samples were injected randomly, and interleaved blanks and quality control (QC) samples corresponding to a pool of the whole sample set were simultaneously analysed to further evaluate the overall variability of the sample set.

The UHPLC–TOF-MS 3D ion map of the QC samples (see Fig. 2A) (RT 0–5 min × *m/z* 100–1000 × intensity) illustrated the very good orthogonal resolution obtained for all features detected with this high-throughput method. This method allowed the detection of more than 700 features that were characterised by their retention time (RT), high resolution *m/z* and their corresponding area (RT × *m/z* × area). The features of all samples and QC samples were exported after peak picking, deconvolution and alignment procedures into a 21 samples × 731 features matrix. PCA was then performed to assess the consistency of the LC–MS preprocessing and detect putative outliers (see Fig. 2B). QC samples were clustered in the centre of the score plot, confirming the stability of the UHPLC–TOF-MS analysis (Ranjbar et al., 2011). Two partially over-

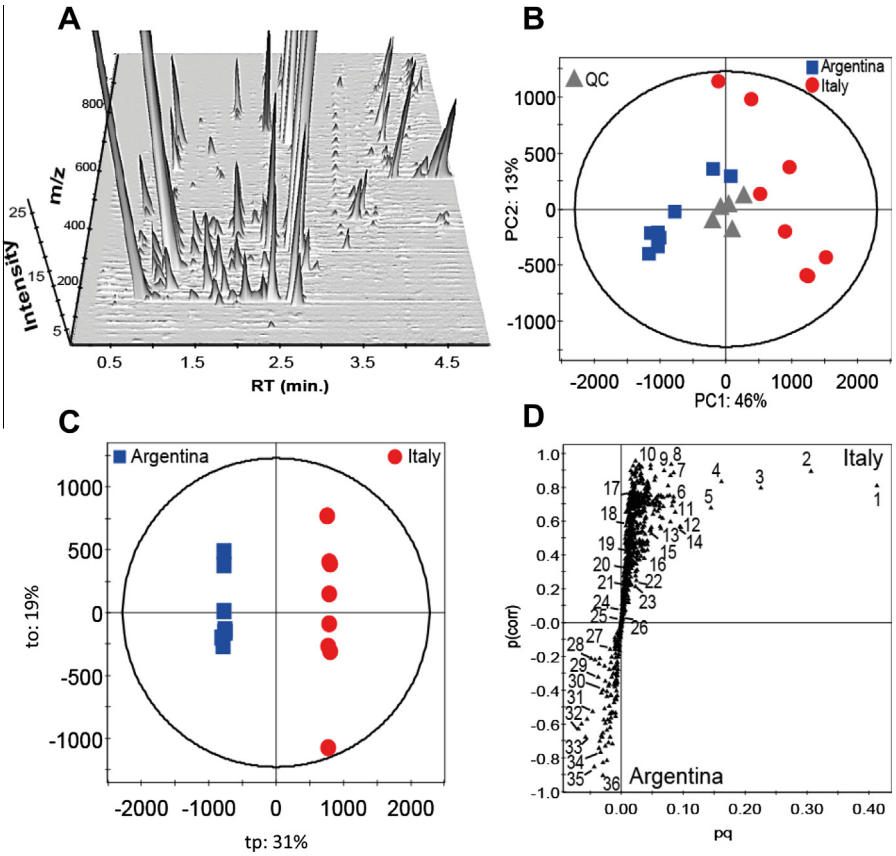


Fig. 2. UHPLC–TOF-MS profiling of CPLOR. (A) UHPLC–TOF-MS 3D ion map (RT × *m/z* × intensity) of pooled CPLORs in ESI PI mode. (B) Pareto-scaled PCA score plot of UHPLC–TOF-MS profiling of eight Argentinian CPLORs (blue squares), eight Italian CPLORs (red dots) and QC samples (grey triangles). (C) OPLS-DA score plot of the same series. (D) S-plot calculated from OPLS-DA model. Loadings were numbered according to Table 1. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

lapped groups that corresponded to the Italian and Argentinian CPLORs were highlighted along the first PC, which accounted for 46% of the total variance. Interestingly, this preliminary trend indicated that difference related to sample origin might exist. To further verify that markers of CPLO origin (discrimination between Italian and Argentinian samples) could be identified in these groups, a supervised OPLS-DA model was computed. This model was characterised by a high predictive ability ($Q^2cum = 0.93$) (OPLS-DA score plot in Fig. 2C). The loadings were plotted according to S-plot projection to highlight discriminating features by combining the amplitudes (covariance loadings) and the reliabilities (correlation loadings) of the variables (see Fig. 2D). Consequently, the most discriminating features were located in the top right corner for the Italian samples and the bottom left corner for the Argentinian samples. The investigation of the feature characteristics, either of Italian or Argentinian origin, was further conducted by inspection of the corresponding high resolution TOF-MS mass spectra to determine molecular ion species and assignment of molecular formulas. This was completed semi-automatically by the molecular formula prediction toolbox of MZmine, which comprises several heuristic filters for molecular formula validation (Pluskal, Uehara, & Yanagida, 2012).

The putative molecular formulas that were retained in this case fulfilled the element count heuristic rules, an isotopic pattern matching score greater than 95% and a mass error less than 2 mDa. A cross search of this information with entries from the Dictionary of Natural Products (DNP, CRC Press) corresponding either to the citrus genus or coumarin- and psoralen-related compounds revealed putative structures for 36 markers. The majority of these compounds were identified at the chemical class level (coumarins, polymethoxyflavones, furocoumarins, unsaturated fatty acids and terpenoids) and have previously been reported in *Citrus* species. The identities of ten compounds were confirmed by comparisons with commercial standards (see Table 1).

The most discriminating marker samples of Italian origin (upper right area of the S-plot in Fig. 2D) were coumarins and furocoumarins (compounds **1**–**4** in Table 1), which were identified as bergamottin (**1**), 5-geranyloxy-7-methoxycoumarin (**2**), 8-geranyloxypsoralen (**3**) and isoimperatorin (**4**). Interestingly, methoxylated flavones (**7**–**8**) and hydroxylated fatty acids (**9**–**10**) were also detected in the upper part of the S-plot, indicating the importance of those chemical classes in the CPLOR classification. Other features had lower contributions to the discriminant model. As an example, heraclenin (**26**), positioned in the centre of the S-plot, is supposed to be found in equal amount in the two classes. Additionally, the Argentinian samples were characterised by higher concentrations of citropten (**31**), an unidentified coumarin (**33**), a trimethoxyflavone (**32**), an unsaturated fatty acid (**34**) and byakangelicol (**35**). Interestingly, the most significant feature contributing to the characterisation of the Argentinian samples was attributed to a terpene derivative (**36**) according to its calculated molecular formula ($C_{10}H_{16}O$). This peak matched with fourteen compounds in the Dictionary of Natural Product (DNP) database within the *Citrus* genus and could not be readily identified based only on high-resolution mass spectral data.

Table 1
Identified features in cold-pressed oil residues.

No	HR-MS $[M+H]^+$	RT (min.)	MF	Chemical class	DNP hit[a]	Putative ID[b]	Error (mDa)	Isotope pattern score (%)
1	339.1620	2.40	$C_{21}H_{22}O_4$	Furocoumarin	–	Bergamottin[1,2]	2	98
2	329.1787	2.46	$C_{20}H_{24}O_4$	Coumarin	–	5-Geranyloxy-7-methoxycoumarin[2]	1.3	97
3	339.1619	2.16	$C_{21}H_{22}O_4$	Furocoumarin	–	8-Geranyloxypsoralen[1]	1.8	96
4	271.0963	1.48	$C_{16}H_{14}O_4$	Furocoumarin	–	Isoimperatorin[1]	0.2	99
5	355.1547	2.09	$C_{21}H_{22}O_5$	Furocoumarin	2[1]	–	0.7	96
6	203.0348	2.42	$C_{11}H_6O_4$	Furocoumarin	1[1]	Bergaptol	0.9	98
7	373.1315	1.50	$C_{20}H_{20}O_7$	Pentamethoxyflavone	3[1]	–	1.5	95
8	403.1413	1.33	$C_{21}H_{22}O_8$	Hexamethoxyflavone	4[1]	–	0.8	98
9	337.2388	2.59	$C_{20}H_{32}O_4$	Hydroxylated fatty acid	–	–	1.5	96
10	359.2200	2.57	$C_{20}H_{30}O_4$	Hydroxylated fatty acid	–	–	2	96
11	193.0505	2.47	$C_{10}H_8O_4$	Coumarin	9[2]	–	1	98
12	305.1031	0.51	$C_{16}H_{16}O_6$	Furocoumarin	–	Oxypeucedanin hydrate[1]	1	97
13	219.1737	1.57	$C_{15}H_{22}O$	Sesquiterpene	6[1]	–	0.7	95
14	374.1327	1.51	$C_{20}H_{20}O_7$	Pentamethoxyflavone	3[1]	–	1.5	95
15	335.1145	0.56	$C_{17}H_{18}O_7$	Furocoumarin	–	Byakangelicin[1]	2	96
16	371.1493	1.75	$C_{21}H_{22}O_6$	Furocoumarin	5[2]	–	0.4	95
17	271.0963	1.20	$C_{16}H_{14}O_4$	Furocoumarin	–	Imperatorin[1]	0.2	96
18	315.1604	2.11	$C_{19}H_{22}O_4$	Coumarin	20[2]	–	1.2	95
19	269.2472	3.64	$C_{17}H_{32}O_2$	Unsaturated fatty acid	–	–	0.2	96
20	355.1542	1.70	$C_{21}H_{22}O_5$	Furocoumarin	–	Epoxybergamottin[1]	0.2	97
21	231.1013	1.18	$C_{14}H_{14}O_3$	Coumarin	7[2]	–	0.2	98
22	301.1071	1.39	$C_{17}H_{16}O_5$	Furocoumarin	5[2]	–	0.1	96
23	261.1118	1.52	$C_{15}H_{16}O_4$	Coumarin	5[1]	–	0.4	97
24	247.0586	0.59	$C_{13}H_{10}O_5$	Furocoumarin	–	Isopimpinellin[1]	1.5	96
25	221.1898	1.02	$C_{15}H_{24}O$	Sesquiterpene	–	–	0.2	99
26	287.0915	0.74	$C_{16}H_{14}O_5$	Furocoumarin	–	Heraclenin[1,2]	0.4	97
27	287.0920	0.90	$C_{16}H_{14}O_5$	Furocoumarin	–	Oxypeucedanin[1,2]	0.1	96
28	277.1077	0.86	$C_{15}H_{16}O_5$	Coumarin	2[1]	–	0.6	96
29	313.2741	2.76	$C_{19}H_{36}O_3$	Hydroxylated fatty acid	–	–	0.3	97
30	203.0335	1.69	$C_{11}H_6O_4$	Furocoumarin	2[2]	–	0.6	96
31	207.0667	0.58	$C_{11}H_{10}O_4$	Coumarin	2[1]	Citropten[2]	1.5	99
32	313.1077	1.69	$C_{18}H_{16}O_5$	Trimethoxyflavone	1[1]	4′,5,7-trimethoxyflavone	0.7	96
33	361.1655	1.78	$C_{20}H_{24}O_6$	Coumarin	4[1]	–	0.9	96
34	279.2321	2.61	$C_{18}H_{30}O_2$	Unsaturated fatty acid	–	–	0.2	98
35	317.1035	0.94	$C_{17}H_{16}O_6$	Furocoumarin	–	Byakangelicol[1,2]	1.5	97
36	153.1288	0.92	$C_{10}H_{16}O$	Monoterpene	14[1]	–	1	95

[a] Type of filter used: [1]keyword: *Citrus*; [2]skeletons: coumarin or furocoumarin sub-substructure.
[b] Type of metabolite assignment: [1]comparison with commercial standard, [2]identification by 2D-NMR of pooled CPLORs.

3.2.2. Multivariate data analysis of the ^{1}H-NMR CPLOR fingerprints

The same sample set was also analysed by ^{1}H NMR fingerprinting to verify if the main markers highlighted by MS could also be detected, to further confirm their identities and to establish the quantitative relationships between the most abundant markers. The 16 CPLOR samples were dissolved at 10 mg/mL in $CDCl_3$, and their ^{1}H NMR fingerprints were recorded according to the optimised methods described in the experimental section. As observed for the pooled CPLOR spectrum (see Fig. 1C), the signals related to the non-volatile constituents were recorded in all the fingerprints. As for MS, the preprocessing of the data was necessary to build a matrix for further MVA. NMR spectrum reduction was achieved by binning. The goal of such data binning was to reduce small variations in the resonance positions of individual peaks across samples. The best results were obtained by reducing the spectra to 522 bins with a bin size of 0.02 ppm, which provided satisfactory peak alignment without critical losses in spectral resolution (see Fig. 3A). This binning resulted in a matrix of 16 samples × 522 features for the whole sample set. As for MS, the resulting PCA separated the samples into two clusters that could be easily related to sample origin (see Fig. 3B). To identify the signals (i.e., chemical shifts) responsible for the group separation, a highly predictive OPLS-DA model was computed (Q^2cum = 0.98) (see Fig. 3C), and the resulting S-plot was used to highlight the most relevant signals (see Fig. 3D). As shown in Fig. 3D, the most discriminating NMR chemical shifts that were characteristic of Italian samples were 1.26, 1.28, 1.6, 4.96 and 7.6 ppm, while signals at 1.7, 2.18, 2.24, 5.9 and 10.0 ppm were typical for Argentinian samples. To provide an overview of the main spectral differences between groups, all loadings were visualised by generating pseudo-spectra in which positive or negative correlations to either the Italian or the Argentinian group were displayed. The signal intensities were also partially taken into account via Pareto scaling (see Fig. 3E). This pseudo-spectrum revealed various discriminating chemical shift domains of interest. In the aliphatic region between 1.0 and 2.5 ppm, the Italian CPLORs were characterised by intense signals from 1.26 to 1.7 ppm, while Argentinian samples showed intense signals between 1.7 and 2.25 ppm. These aliphatic protons were most likely related to fatty acids or terpenoid derivatives. The 4.0–8.0 ppm region displayed mainly signals that were detected

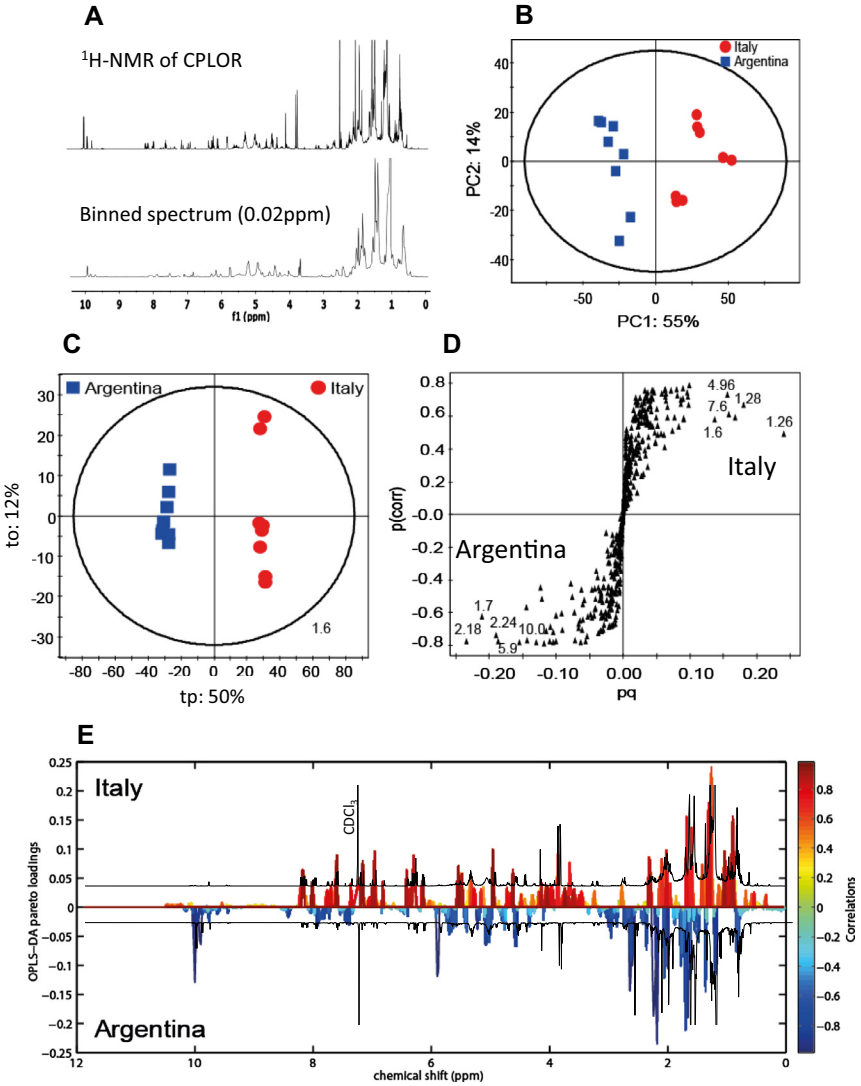


Fig. 3. ^{1}H-NMR fingerprints of CPLORs from Italy and Argentina. (A) Top: ^{1}H NMR spectrum of a representative CPLOR in $CDCl_3$. Bottom: reduced ^{1}H-NMR spectrum using bin size of 0.02 ppm. (B) Pareto-scaled PCA score plot of ^{1}H NMR binned spectra from eight Italian samples of the same series as Fig. 3B. (C) OPLS-DA score plot of the same series. (D) S-plot calculated from OPLS-DA model. Numbers represent significant chemical shifts in ppm for each class. (E) OPLS-DA Pareto-scaled loadings coloured according to their correlations: red, positively correlated signal of Italian samples; blue, positively correlated signals of Argentinian samples. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

in the Italian samples. This region is representative of heterocyclic oxygen compounds present in the citrus oil (Sommer et al., 2003). Interestingly, characteristic signals were observed in the Argentinian samples approximately 10 ppm. The complexity of the ^{1}H NMR CPLOR fingerprints did not allow the identification of its main components. To associate the main discriminating ^{1}H NMR signals to structural features, homo- and heterocorrelated 2D-NMR spectra were acquired directly from the concentrated pooled samples of each CPLOR origin group. With the help of COSY, HSQC and HMBC spectra, for example, the main discriminating signals of the Italian samples in the aromatic region at 7.6 ppm were attributed to the proton in the 7 position of bergamottin, indicating a higher content of this compound in samples from Italy. These 2D-NMR data also provided the unambiguous identifications of 4 furocoumarins (bergamottin, heraclenin, byakangelicol and oxypeucedanin) and 2 coumarins (5-geranyloxy-7-methoxycoumarin and citropten) (see supplementary S2). All identified compounds were also detected in the UHPLC–TOF-MS rapid profiling results. Bergamottin, which was the most discriminating compound in the Italian sample with NMR, was also the most discriminating compound in the S-plot of the MS data (see Table 1 and Fig. 2D).

3.2.3. NMR–MS correlation for identifying key markers

The NMR- and MS-based results analysed individually showed evidence of discriminating markers of origin in the CPLORs. After identification, some of them were found to be discriminating in

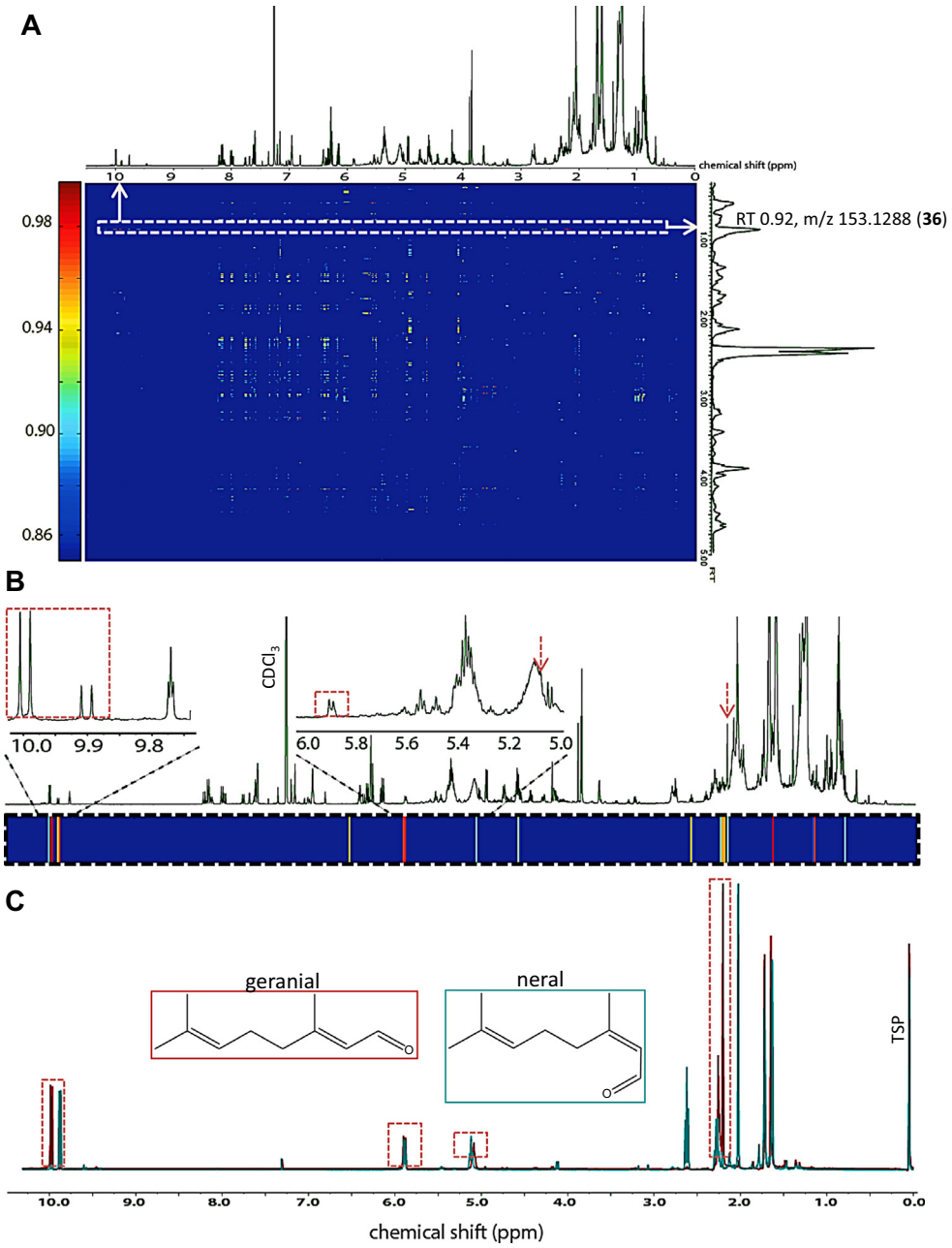


Fig. 4. NMR–MS correlation. (A) NMR–MS correlation map (cutoff: 0.85). Top: ^{1}H NMR spectrum of pooled CPLORs in $CDCl_3$. Right: UHPLC–TOF-MS chromatogram of pooled CPLORs. The dashed rectangle highlights the correlation line for the NMR signal at 10 ppm corresponding to compound **36** in Table 1. (B) ^{1}H NMR of CPLOR, dashed square: enlargement of the NMR–MS correlation map of compound **36**. Top: ^{1}H NMR spectrum of pooled CPLORs with focus on area from 9.6 to 10.2 ppm and from 5 to 6 ppm. (C) Superimposed ^{1}H NMR spectra of pure geranial (red) and neral (blue) in $CDCl_3$. Dashed red squares and arrows highlight common signals of geranial found in the pooled CPLOR spectra. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

both models. To comprehensively investigate which markers could be linked between the MS and NMR data matrices, a correlation analysis was undertaken.

Indeed, while the complementarity of the MS and NMR methods is usually applied to single component identification strategies (Wolfender et al., 2012), multiple sample analyses on MS and NMR platforms in a parallel approach provides the opportunity to apply a statistical analysis of signal amplitude covariance between the technologies and to direct cross-correlations of data for assignment purposes (Crockford et al., 2006). This approach can be particularly useful to improve molecular biomarker identifications by highlighting high correlations between a given *m/z* feature and several ^{1}H NMR chemical shifts.

The NMR and MS data matrices used for individual MVA (see above) were directly centred and scaled to unit variance using MATLAB prior to cross-correlation calculations. A correlation cutoff value of 0.85 was then applied to the resulting correlation map to select the most significant correlation patterns (see Fig. 4A). Interestingly, the range from 5.0 to 8.0 ppm in the ^{1}H NMR spectra showed strong correlations with LC–MS features eluted from 1.5 to 3 min. This region corresponded to the oxygen heterocyclic compounds annotated on the basis of LC–MS data (see Table 1). As an example, the LC–MS peak of bergamottin (RT 2.4 min, *m/z* 339.1620) showed a good correlation with the ^{1}H NMR signals at 6.26, 8.16, 6.96, 7.60 and 7.16 ppm, corresponding to protons in positions 3, 4, 6 7 and 8 of the furo-benzopyranone moiety, respectively, and matched with the chemical shifts of a pure standard. Interestingly, the characteristic discriminating signal of Argentinian CPLORs at 10.0 ppm (see above) was correlated with the LC–MS peak at 0.92 min, corresponding to compound 36 (see Fig. 4A) with a molecular formula of $C_{10}H_{16}O$. As mentioned above, the LC–MS information led to 14 possibilities after cross searching the DNP database. However, with the help of the 2D-NMR data, this signal was attributed to an aldehyde functional group though an HSQC correlation with the corresponding carbonyl (^{13}C 191.2 ppm). It was also linked to an alkene chain according to its COSY correlation with an alkene proton (^{1}H 5.98/^{13}C 127.1 ppm). Further identification in mixture was not possible due to overlapping signals, but cross searching the DNP database reduced the possibilities to geranial or neral. To verify this hypothesis, ^{1}H NMR spectra of both pure isomer standards were recorded to compare the chemical shifts of both isomers and NMR/MS correlations (see Fig. 4C). A deeper analysis of the NMR–MS correlations of compound 36 displayed strong correlation for the doublet at 10.0 and 9.9 ppm (J = 8.1 Hz) along with the signal at 5.9 ppm (dd, J = 8.1; 1.3 Hz). Additionally, other significant correlations were also observed at 5.1 ppm and 2.18 ppm (see Fig. 4B). As shown in Fig. 4B, the superimposed spectra of neral and geranial displayed many overlapped signals, but they could be readily distinguished by chemical shifts of the aldehyde proton (9.9 and 10.0 ppm, respectively), the methyl in *cis* or *trans* position (1.99 and 2.18 ppm, respectively) and the overlapping signals observed for CH_2 at 2.23 ppm in geranial. Regarding the NMR/MS correlations, intense signals were observed in the overlapped region at 5.1 ppm, and an intense correlation was displayed at 2.18 ppm, confirming the methyl group in the *trans* position along with the overlapped CH_2 signals characteristic of geranial. Moreover, doublets at 10.0 and 9.9 ppm in the CPLOR spectra suggested the presence of both isomers, with geranial as major compound.

3.3. Quantitative NMR applied to identified component in CPLORs

One of the main advantages of ^{1}H NMR fingerprints is the possibility to obtain absolute quantitative information directly from mixtures, as the proton signal area is proportional to the molar concentration (Wishart, 2008). LC–MS provides sensitive detection of many features, but it has limited quantification capabilities because MS ionisation is compound-dependent, and matrix effects can reduce the ionisation efficiency of a given feature. As an example, the observed UHPLC–TOF-MS ion trace of geranial was just above the noise level due to its limited ionisation in positive ESI mode, although it was readily detected by ^{1}H NMR in the CPLOR mixture. In the present work, the markers were quantified based on their characteristic signals in the mixture. To avoid the use of an internal standard (IS), the pulse length-based concentration (PULCON) qNMR method was used (Wider & Dreier, 2006).

This method is based on the comparison of the absolute integral value of one proton from an external standard recorded at a known concentration and the absolute integral of the targeted proton within the mixture. In the present study, bergapten was used as an external standard for its close structural similarity to the other well-known furocoumarins found in *Citrus*. For each identified structure (see above), a non-overlapping signal corresponding to a single proton was selected and carefully integrated in all spectra and was compared to the external standard. Because all data were acquired under the same conditions and the concentration of the external standard was controlled, the molar concentration of each compound in the CPLOR could be readily calculated (see Fig. 5). Overall, the quantification results confirmed the relatively high amounts of furocoumarins and coumarins in the Italian samples with significantly higher concentrations of bergamottin (**1**) (70 mg/g in CPLOR) and 5-geranyloxy-7-methoxycoumarin (**2**) (48 mg/g in CPLOR). Heraclenin (**26**) and oxypeucedanin (**27**) were found at comparable quantities in all samples studied. Finally, the concentration of citropten (**31**) was significantly higher in the

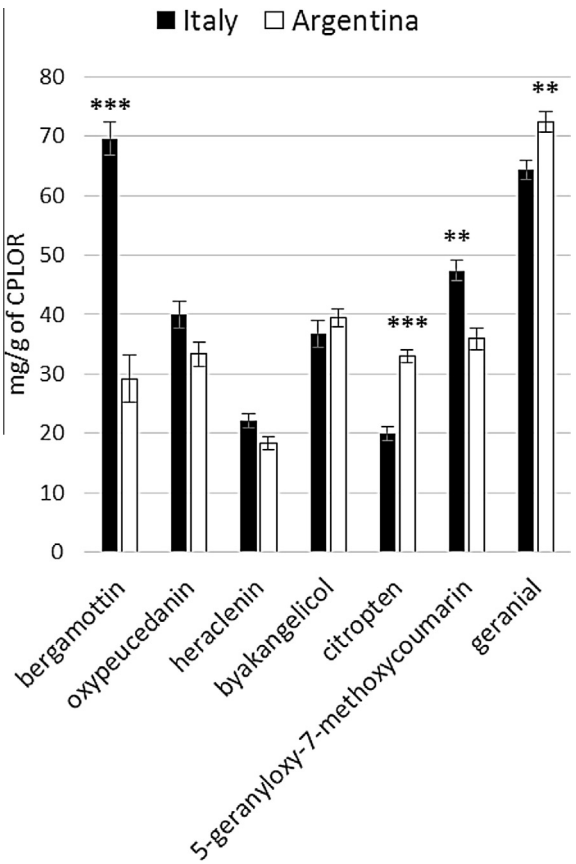


Fig. 5. quantitative NMR: Bar plot of concentrations in mg/g of identified compounds in CPLORs deduced from quantitative NMR method ± SEM. Stars indicate significant differences calculated by impaired *t*-test (Italian samples, n = 8; Argentinian samples, n = 8: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$).

Argentinian CPLORs. The quantification of geranial (**36**) based on the aldehyde proton signal at 10.0 ppm confirmed the relatively high levels of this terpene in the Argentinian CPLOR samples (72 mg/g in CPLOR) compared to the Italian CPLOR samples (64 mg/g in CPLOR). In the sample set studied the difference observed for geranial, although small, was found to be statistically significant ($p < 0.01$). This also demonstrates the potential of NMR to highlight subtle changes with high confidence level in this context. To confirm the overall differences observed between batches from different origins used in this study (furocoumarins being the most noticeable ones), a larger sample set would however be mandatory since the ratio of natural products can be sensitive to many factors.

4. Conclusions

The goal of this study was to evaluate if relevant information on the composition could be extracted from the non-volatile fraction of CPLOs from an unbiased metabolomic approach and reveal marker compounds associated with the sample origin. Orthogonal NMR and LC–MS profiling methods were found to be suitable, provided that the correct sample preparation procedure was applied. In particular this requires an evaporative step prior to analysis, which directly enriches the non-volatile constituents and maintains instrument repeatability as well as sample stability. As expected, NMR fingerprinting mainly revealed quantitative changes of the main constituents responsible for discrimination. Overall, seven major compounds were identified in mixture. On the other hand, the short UHPLC–TOFMS profiling highlighted discriminating markers including minor components independently from their absolute abundance. A total of 721 features were detected in the 5 min gradient and several chemical classes were revealed such as coumarins, furocoumarins, flavonoids, fatty acids and terpenoids.

Both types of data were found to be complementary, and correlation analysis of the NMR and MS dataset matrices allowed in particular the identification of geranial. This combined approach provides molecular formula assignments of constituents detected in mixtures by NMR without the need for further separation. Moreover, qNMR spectroscopy provided a good estimation of the ratio of the identified markers related to the sample origin. Italian samples were characterised by their high content in bergamottin and 5-geranyloxy-7-methoxycoumarin while citropten was mostly present in Argentinian CPLORs.

The possibility to highlight markers in the non-volatile residues of CPLOs gives a new perspective for authenticity assessment because it can widen the scope and the power of classical methods. This is of particular importance for the F&F industry since it can help to differentiate between oils of different qualities and prices and identify possible adulterations.

The information yielded by LC–MS and NMR is complementary to that obtained by classical GC–MS and LC-UV analysis in this field. Because of the sophistication level of adulteration, combined techniques are necessary to support quality control in their day-today work. While being advanced, the methods proposed can be integrated for the survey of selected batches and since the sample preparation was kept simple, analyses can be easily externalised, especially in the case of ^{1}H NMR.

The metabolomic approach presented here gives a holistic view of the changes that can be expected in CPLORs. With this strategy more precise trends in CPLORs content related to either cultivar, origin, processes and suppliers can be monitored over years. Metabolomic profiling has thus a high potential to be implemented as QC support of CPLOs in a long term perspective.

Acknowledgements

The authors would like to thank Robert Brauchli and Sandy Frank of the Firmenich Spectroscopy group for performing the NMR measurements on the citrus oil samples, Eric Frerot for providing the standards, André Pinto of Geneva University for his help in NMR data acquisition and Rolland Thoos for technical support.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.foodchem.2013.10.103.

References

Bonaccorsi, I. L., McNair, H. M., Brunner, L. A., Dugo, P., & Dugo, G. (1999). Fast HPLC for the analysis of oxygen heterocyclic compounds of citrus essential oils. *Journal of Agricultural and Food Chemistry, 47*(10), 4237–4239.

Cevallos-Cevallos, J. M., Reyes-De-Corcuera, J. I., Etxeberria, E., Danyluk, M. D., & Rodrick, G. E. (2009). Metabolomic analysis in food science: A review. *Trends in Food Science & Technology, 20*(11–12), 557–566.

Costa, R., Dugo, P., Navarra, M., Raymo, V., Dugo, G., & Mondello, L. (2010). Study on the chemical composition variability of some processed bergamot (*Citrus bergamia*) essential oils. *Flavour and Fragrance Journal, 25*(1), 4–12.

Crockford, D. J., Holmes, E., Lindon, J. C., Plumb, R. S., Zirah, S., Bruce, S. J., et al. (2006). Statistical heterospectroscopy, an approach to the integrated analysis of NMR and UPLC–MS data sets: Application in metabonomic toxicology studies. *Analytical Chemistry, 78*(2), 363–371.

Dugo, G., & Mondello, L. (2010). *Citrus oils: Composition, advanced analytical techniques, contaminants, and biological activity*. Boca Raton: CRC Press.

Dugo, P., Skerikova, V., Kumm, T., Trozzi, A., Jandera, P., & Mondello, L. (2006). Elucidation of carotenoid patterns in citrus products by means of comprehensive normal-phase × reversed-phase liquid chromatography. *Analytical Chemistry, 78*(22), 7743–7750.

Frerot, E., & Decorzant, E. (2004). Quantification of total furocoumarins in citrus oils by HPLC coupled with UV, fluorescence, and mass detection. *Journal of Agricultural and Food Chemistry, 52*(23), 6879–6886.

Groppo, M., Pirani, J. R., Salatino, M. L., Blanco, S. R., & Kallunki, J. A. (2008). Phylogeny of Rutaceae based on twononcoding regions from cpDNA. *American Journal of Botany, 95*(8), 985–1005.

Guillarme, D., Grata, E., Glauser, G., Wolfender, J. L., Veuthey, J. L., & Rudaz, S. (2009). Some solutions to obtain very efficient separations in isocratic and gradient modes using small particles size and ultra-high pressure. *Journal of Chromatography A, 1216*(15), 3232–3243.

Lota, M. L., de Rocca Serra, D., Tomi, F., Jacquemond, C., & Casanova, J. (2002). Volatile components of peel and leaf oils of lemon and lime species. *Journal of Agricultural and Food Chemistry, 50*(4), 796–805.

Luro, F., Venturini, N., Costantino, G., Paolini, J., Ollitrault, P., & Costa, J. (2012). Genetic and chemical diversity of citron (*Citrus medica* L.) based on nuclear and cytoplasmic markers and leaf essential oil composition. *Phytochemistry, 77*, 186–196.

Macmaster, A. P., Owen, N., Brussaux, S., Brevard, H., Hiserodt, R., Leijs, H., et al. (2012). Quantification of selected furocoumarins by high-performance liquid chromatography and UV-detection: Capabilities and limits. *Journal of Chromatography A, 1257*, 34–40.

Maffei, M. E., Gertsch, J., & Appendino, G. (2011). Plant volatiles: Production, function and pharmacology. *Natural Product Reports, 28*(8), 1359–1380.

Mehl, F., Marti, G., Boccard, J., Debrus, B., Merle, P., Delort, E., et al. (2014). Differentiation of lemon essential oil based on volatile and non-volatile fractions with various analytical techniques: A metabolomic approach. *Food Chemistry, 143*, 325–335.

Peroutka, R., Schulzova, V., Botek, P., & Hajslova, J. (2007). Analysis of furanocoumarins in vegetables (*Apiaceae*) anal citrus fruits (*Rutaceae*). *Journal of the Science of Food and Agriculture, 87*(11), 2152–2163.

Placzek, M., Fromel, W., Eberlein, B., Gilbertz, K. P., & Przybilla, B. (2007). Evaluation of phototoxic properties of fragrances. *Acta Dermato-Venereologica, 87*(4), 312–316.

Pluskal, T., Uehara, T., & Yanagida, M. (2012). Highly accurate chemical formula prediction tool utilizing high-resolution mass spectra, MS/MS fragmentation, heuristic rules, and isotope pattern matching. *Analytical Chemistry, 84*(10), 4396–4403.

Ranjbar, M. R. N., Wang, Y., & Ressom, H. W. (2011). Quality assessment of LC–MS metabolomic data. In *Bioinformatics and biomedicine workshops (BIBMW), 2011 IEEE international conference on bioinformatics and biomedicine* (pp. 1034–1036).

Santana, L., Uriarte, E., Roleira, F., Milhazes, N., & Borges, F. (2004). Furocoumarins in medicinal chemistry. Synthesis, natural occurrence and biological activity. *Current Medicinal Chemistry, 11*(24), 3239–3261.

Sommer, H. B., Krammer, G., Kindel, G., Kuhnle, T., Reinders, G., Reiss, I., et al. (2003). HPLC–NMR a powerful tool for the identification of non-volatiles in lemon peel oils. *Perfumer & Flavorist, 28*(1).

Steuer, B., Schulz, H., & Lager, E. (2001). Classification and analysis of citrus oils by NIR spectroscopy. *Food Chemistry, 72*(1), 113–117.
Tranchida, P. Q., Bonaccorsi, I., Dugo, P., Mondello, L., & Dugo, G. (2012). Analysis of Citrus essential oils: State of the art and future perspectives. A review. *Flavour and Fragrance Journal, 27*(2), 98–123.
Wider, G., & Dreier, L. (2006). Measuring protein concentrations by NMR spectroscopy. *Journal of the American Chemical Society, 128*(8), 2571–2576.
Wishart, D. S. (2008). Quantitative metabolomics using NMR. *Trac-Trends in Analytical Chemistry, 27*(3), 228–237.
Wolfender, J. L., Marti, G., & Ferreira Queiroz, E. (2010). Advances in techniques for profiling crude extracts and for the rapid identificationof natural products: Dereplication, quality control and metabolomics. *Current Organic Chemistry, 14*(16), 1808–1832.
Wolfender, J. L., Rudaz, S., Choi, Y. H., & Kim, H. K. (2012). Plant metabolomics: From holistic data to relevant biomarkers. *Current Medicinal Chemistry*.
Yang, M., Fazio, S., Munch, D., & Drumm, P. (2005). Impact of methanol and acetonitrile on separations based on pi–pi interactions with a reversed-phase phenyl column. *Journal of Chromatography A, 1097*(1–2), 124–129.
Ziegler, H., & Spiteller, G. (1992). Coumarins and psoralens from sicilian lemon oil (*Citrus limon* (L.) Burm. f.). *Flavour and Fragrance Journal, 7*(3), 129–139.

Journal of Chromatography A, 1382 (2015) 136–164

Contents lists available at ScienceDirect

Journal of Chromatography A

journal homepage: www.elsevier.com/locate/chroma

Review

Current approaches and challenges for the metabolite profiling of complex natural extracts

Jean-Luc Wolfender[a,*], Guillaume Marti[a,b], Aurélien Thomas[c], Samuel Bertrand[a,d]

[a] School of Pharmaceutical Sciences, EPGL, University of Geneva, University of Lausanne, quai Ernest-Ansermet 30, CH-1211 Geneva 4, Switzerland
[b] Université de Toulouse, UPS, UMR 152 Pharma-DEV, Institut de Recherche et Développement (IRD), Université Toulouse 3, Faculté des Sciences Pharmaceutiques, F-31062 Toulouse cedex 09, France
[c] Unit of Toxicology, CURML, University of Lausanne, Bugnon Street 21, 1011 Lausanne, Switzerland
[d] Groupe Mer, Molécules, Santé-EA 2160, UFR des Sciences Pharmaceutiques et Biologiques, Université de Nantes, 9 rue Bias, BP 53508, F-44035 Nantes Cedex 01, France

ARTICLE INFO

Article history:
Received 25 July 2014
Received in revised form 23 October 2014
Accepted 26 October 2014
Available online 31 October 2014

Keywords:
Metabolite profiling and fingerprinting
Metabolite annotation and identification
Natural products
Liquid chromatography
Mass spectrometry and NMR
Metabolomics

ABSTRACT

Metabolite profiling is critical in many aspects of the life sciences, particularly natural product research. Obtaining precise information on the chemical composition of complex natural extracts (metabolomes) that are primarily obtained from plants or microorganisms is a challenging task that requires sophisticated, advanced analytical methods. In this respect, significant advances in hyphenated chromatographic techniques (LC–MS, GC–MS and LC–NMR in particular), as well as data mining and processing methods, have occurred over the last decade. Together, these tools, in combination with bioassay profiling methods, serve an important role in metabolomics for the purposes of both peak annotation and dereplication in natural product research. In this review, a survey of the techniques that are used for generic and comprehensive profiling of secondary metabolites in natural extracts is provided. The various approaches (chromatographic methods: LC–MS, GC–MS, and LC–NMR and direct spectroscopic methods: NMR and DIMS) are discussed with respect to their resolution and sensitivity for extract profiling. In addition the structural information that can be generated through these techniques or in combination, is compared in relation to the identification of metabolites in complex mixtures. Analytical strategies with applications to natural extracts and novel methods that have strong potential, regardless of how often they are used, are discussed with respect to their potential applications and future trends.

Contents

* Corresponding author. Tel.: +41 223793385; fax: +41 223793399.
E-mail address: jean-luc.wolfender@unige.ch (J.-L. Wolfender).

http://dx.doi.org/10.1016/j.chroma.2014.10.091
0021-9673/

1. Introduction

Metabolite analysis, particularly metabolite profiling in complex biological matrices, is essential in many fields of the life sciences. The concept of **metabolite profiling** is not new [1]. In 1971, Horning and Horning first introduced the concept of metabolic profiling using mass spectrometry (MS) [2], and in that same year, Pauling et al. reported analyses of urine vapour and breath by gas chromatography (GC) and related the obtained profiles to the effects of a defined diet. Since then, many powerful analytical methods, both chromatographic and spectroscopic, have been developed or improved [3], and metabolite analyses in various types of matrices, has become necessary in various research fields. While many methods have reached a high level of resolution and sensitivity for such analyses, the unambiguous identification of metabolites remains challenging [4].

This review is primarily focused on crude extracts of natural origins as matrices of interest (primarily plant and microbial extracts that are defined as "natural extracts"). Such biological matrices are challenging to analyse due to the concomitant presence of metabolites with high chemodiversity [5–7], which makes them particularly interesting to study. In natural product research, the term "**metabolite**" is widespread and refers to compounds (small molecules with molecular weights (MW) of <1000 Da) that are either essential to sustain the life of a given organism via normal metabolic processes ("**primary metabolites**") or non-essential but necessary for survival in a given environment ("**secondary metabolite**"). Primary metabolites include amino acids, lipids, and carbohydrates; secondary metabolites are related to defence and signalling mechanisms and include polyphenols, alkaloids, terpenes, polyketides, and hormones [8]. These latter compounds are also referred to as natural products and have historically been a significant source for lead compounds for drug discovery due to their action on pharmacological targets [9]. The profiling of such extracts in classical natural product research has been underway for years, especially for dereplication purposes [10]. **Dereplication** is the process of testing sample mixtures that are active in screening in order to differentiate the novel compounds from active substances that have already been studied [11]. With the significant development that is occurring in metabolomics for biology and natural product research, this type of analysis is gaining more and more importance from both targeted and untargeted analytical perspectives [12].

Metabolomics is defined as a non-selective, universally applicable, comprehensive analytical approach for the identification and quantification of metabolites in a biological system. This area of research strives to obtain complete metabolite fingerprints, detect differences between metabolites and generate hypotheses to explain these differences. In natural product research, metabolomics is considered the large-scale analysis of metabolites of a given organism during various physiological states [13]. The advent of powerful modern analytical methods generate large data set which can be Interpreted from a holistic manner only by reducing their dimensionality using multivariate data analysis (MVDA) approaches. This highlights the significant variations that occurs at the metabolome level. The results can then be correlated with new biological knowledge. To achieve this goal, the metabolites of complex natural extracts must be comprehensively analysed by metabolite profiling and fingerprinting, which is a significant analytical challenge, as several extraction methods [14] and complementary sophisticated analytical platforms must be used [15].

In metabolomics, different analytical strategies are used to determine the chemical composition of a given biological matrix or extract. The profiling of different constituents can be achieved on several levels, such as "metabolite (or metabolic) fingerprinting," "metabolite profiling" and "metabolite target analysis." The term "**metabolic**" is more often used in drug research; it is frequently used to describe the metabolic fate of an administered drug [16]. Alternatively, in the analysis of complex mixtures, these approaches are defined according to Fiehn [16]:

(i) "**Metabolite fingerprinting**" is the rapid classification of samples. In these high-throughput analyses, extensive metabolite identification and quantitation are generally not used. MVDA is applied in fingerprinting to determine differences and classify samples. The purpose of this method is not to identify each individual metabolite, but to compare patterns or "fingerprints" of metabolites that change in a given biological system. In the frame of metabolomics, it is aims to be untargeted and is used as an hypothesis-generating approach.

(ii) "**Metabolite profiling**" focuses on the analysis of a large group of metabolites that is either related to a specific metabolic pathway or a class of compounds [17]. In most cases, metabolite profiling is more targeted than metabolite fingerprinting and follows specific hypotheses in the approach. Therefore specific analytical methods are developed for analyte determination. Metabolite profiling is the oldest and most established approach and is considered the precursor of metabolomics.

(iii) "**Metabolite target analysis**" focuses on the investigation of metabolites that are related to a particular metabolic pathway. The purpose of this strategy is to observe the specific metabolic modifications that may be related to a particular hypothesis.

When such analyses are performed on a sufficient number of biological replicates, they enable researchers to discriminate and classify samples into groups and monitor changes in metabolome composition related to factors such as a given physiological status, a stress or stimulus, genetic modification, or interaction with other organisms, among others. The use of biomarkers found in a fingerprint is not discussed herein; however, mining of these data is discussed in many other reviews (e.g., [18,19]).

In the field of plant science, the size of the metabolome (the array of all metabolites in a given organism) is very difficult to estimate because metabolome size is highly dependent on the specific organism under study. Some reports have estimated that up to 15,000 distinct compounds are present within a given plant species [20,21], and more than 200,000 natural products have been reported [6]. This extensive chemical diversity can be quantified or visualised by assessing the chemical space occupied by natural products [5,22]. Furthermore, this chemical diversity is also directly associated with the high variability in the intrinsic physicochemical properties of natural products, which make separation, detection and identification from their natural matrices extremely challenging. Therefore, a single analytical technique is not sufficient for a comprehensive analysis of a complex metabolome, and the utilisation of multiple technologies is necessary. In this review, analytical methods for rapid metabolite identification of natural extracts is described with respect to dereplication and metabolomics natural product research, particularly LC–MS profiling (Fig. 1C). Nevertheless, fingerprinting methods, such as direct MS (Fig. 1B) or NMR spectroscopy (Fig. 1A) are also presented. In many cases, these approaches are utilised for metabolite identification. Additionally, methods that link the secondary metabolite composition of an extract with their bioactivities, for example, HPLC, are discussed.

Various excellent reviews have dealt with different aspects of natural product analyses [23,24] focussing on metabolite profiling by LC [25–27], GC [28,29], MS [26,30,31], NMR [32] or bioassays [33,34]. The present review discussed the wide range of analytical methods that are used to analyse metabolites, with a specific emphasis on the possibilities and limitations of the various approaches, or a combination of them, for deciphering the composition of natural extracts. A particular focus is made on the dereplication strategies in combination with biological profiling. In addition, future trends are discussed.

2. Recent chromatographic improvements in metabolite profiling

Because natural extracts may have complex compositions, the chromatographic separation of metabolites based on differences in their physicochemical properties is required prior to detection. Profiling methods should involve high chromatographic resolution to meet the standards for data acquisition and dereplication (see Section 4). HPLC has been recognised since the early 1980s as the most versatile technique for the efficient and direct separation of natural products from crude mixtures without requiring complex sample preparation [35]. Since then, separation techniques for natural products have significantly improved in terms of convenience, speed, choice of column stationary phases, sensitivity, applicability to a variety of sample matrices and the use of spectroscopic detectors [36,37]. In HPLC, the development of columns with different phase chemistries that are stable (especially

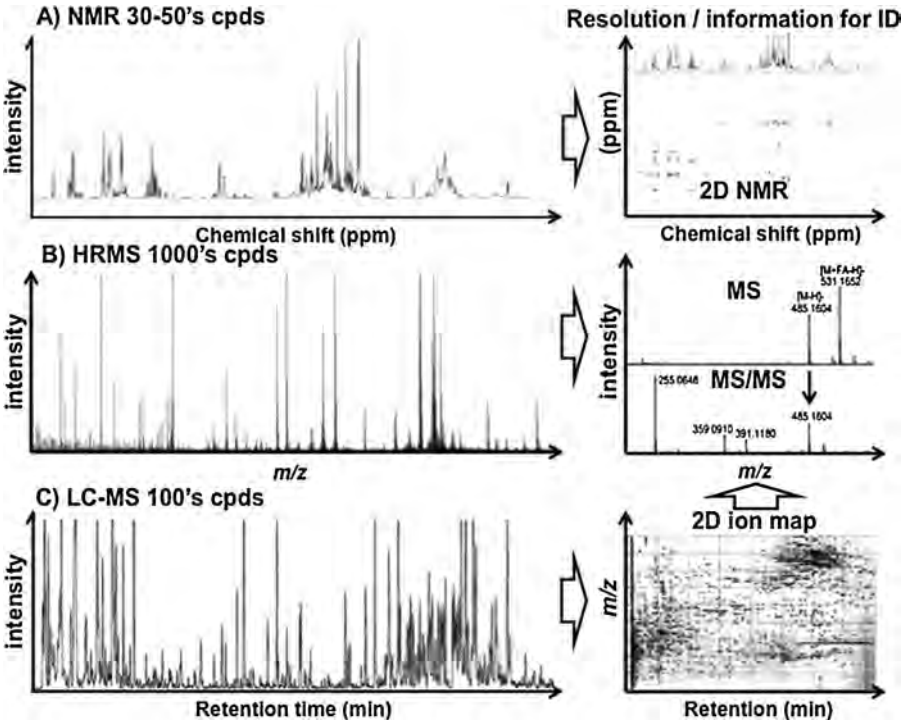


Fig. 1. Various methods of metabolite profiling or fingerprinting of crude extracts. Lefts panel: (A) NMR (^{1}H NMR), (B) DIMS (TOFMS) and (C) LC–MS (UHPLC–TOFMS) profiling. For each method a rough estimate of the number of compounds that can be detected in a plant extract are indicated, this range from a few tens to about one thousand metabolites according to the technique used. Right panel: (A) 2D NMR of the extract profiled by ^{1}H NMR for enhanced resolution of the features, (B) HR-MS and corresponding MS/MS spectra of s single compounds for metabolite identification, (C) 2D ion map of the HR UHPLC–TOFMS showing all features detected. The number of detected metabolites represents a rough estimation and usually many more features than real metabolites (as resolved by UHPLC) are reported as is the case with DIMS approach.

Adapted from [15] with permission.

reversed-phase) has enabled the separation of almost any type of natural product [38–40]. Crude extracts of natural origins can be separated either by using raw mixtures or by enriching the mixtures through solid-phase extraction (SPE) or liquid–liquid extraction (LLE). These separations are generally performed using reversed-phase chromatography on C_{18} material eluting with a gradient of acetonitrile–water (ACN–H_2O) or methanol–water (MeOH–H_2O) solvent system, supplemented with 0.1% formic acid. To improve separation efficiency, various modifiers or buffers can be added to the mobile phase to tune the separation selectivity or detection sensitivity [41].

In addition to HPLC, other separation methods, such as GC, play an important role in both the analysis of volatile compounds found in the essential oils of plant extracts and profiling primary metabolites after derivatisation in metabolomics studies [42]. To some extent, capillary electrophoresis (CE) [43] can also be used, especially for charged metabolites. The use of CE for the generic profiling of crude natural matrices is infrequent and is not discussed in this review; however, it has been previously reviewed by Tubaon et al. [44]. CE also shows potential for the analysis of proteins, peptides, carbohydrates and small biomolecules [45].

In metabolite profiling studies, the detectors hyphenated to these chromatographic methods, which are necessary for further peak annotation or dereplication, must provide on-line structural information regarding the separated compounds. For liquid-based separation techniques, photodiode array (PDA), MS, NMR or a combination of these approaches is used. For GC, electron impact (EI) MS is generally utilised [46], and additional information is obtained from infrared (IR) spectroscopy [47].

With respect to conventional HPLC separation, metabolite profiling of natural products in plant or microbial extracts has primarily been based on LC–MS and LC-PDA analyses. These processes have been summarised by Steinmann et al. [26]. In these profiling methods, a reversed-phase gradient from 30 to 120 min generally provides satisfactory resolution of the extract constituents. Since 2000, LC has evolved rapidly towards improving separation performance [48]. In addition, innovations in the stationary phases of HPLC platforms have significantly improved throughput and resolution (see Sections 2.2 and 2.3). A major advancement was the introduction of ultra-high pressure liquid chromatography (UHPLC) technology, which allows both higher throughput and improved chromatographic resolution and has now been largely adopted in the laboratory.

2.1. Ultra-high pressure liquid chromatography (UHPLC)

In metabolite profiling, there is a strong need for high chromatographic resolution in order to deconvolute a metabolite from the associated complex matrix. Reduction of the packing particle size in chromatography allows higher separation efficiency (higher plate count) and a shorter analysis time [49]. However, the main limitation is the backpressure that is generated. To overcome this, dedicated instrumentation that withstands pressures of >400 bar is necessary when using very small particles (i.e., sub-2 μm). With the introduction of UHPLC operating at very high pressures with sub-2 μm-packed columns, shorter analysis times and increases in peak capacity, sensitivity and reproducibility have been attained [48]. Thus, interest in UHPLC for natural product research with respect to metabolite profiling and dereplication, as well as for fast fingerprinting and metabolomics, has increased rapidly [27].

From a theoretical viewpoint, it is possible to increase the throughput by a factor of 9 compared to conventional HPLC by selecting the appropriate column length. Such ultra-fast separations have been experimentally demonstrated for extracts of natural origins, with analysis times ranging from 1 to 5 min [50–53]. However, for high resolution, increasing the plate count by a factor of 3 is possible by utilising identical column lengths for both HPLC and UHPLC [27]. In contrast, using optimal flow rate conditions with long columns is difficult due to the significant backpressure generated when they are packed with sub-2 μm particles. To compensate, higher temperatures may be used to allow higher flow rates [54]. Separations involving 150 mm or even longer UHPLC columns have been reported and have demonstrated higher efficiencies [55].

Thus, reproducible high-throughput (Fig. 2A) or high-resolution (Fig. 2B and C) UHPLC methods have been developed using concentration sensitive detection methods, such as electrospray (ESI) MS, and these result in narrow LC peaks for highly sensitive detection [56]. When UHPLC is coupled to HRMS analysers the data obtained results in ion maps (m/z vs RT) that can be used for fingerprinting many biological replicates for MVDA analysis in metabolomics (See ion maps in Fig. 2A) or for high resolutions ion maps (see HR ion maps in Fig. 2B) that provide a precise localisation of any feature and MS information on base line separated metabolites. In addition, ion suppression problems related to metabolite co-elution are less common with UHPLC [57].

2.2. Alternative LC-based separation approaches

Another important innovation in the application of chromatography to metabolite profiling is the use of monolithic columns [48,60]. Compared to standard HPLC columns, monolithic columns contain a single solid polymer as the stationary phase (consisting of a network of polymethacrylate or polystyrene copolymers or bonded silica) [61]. These latter stationary phases exhibit increased external porosity and flow-through pore size, which lead to low intrinsic flow resistances and thus enhanced separation speeds [62]. Such high-throughput separations are beneficial, for example, in analysing natural extracts (e.g., phenols in apple extracts [63], bacosides and apigenin in *Bacopa monnieri* [64]). The primary disadvantages of monolithic columns are the lack of variety phase chemistries (only C_{18} and C_8 columns are available; see Section 2.4) and the restricted column geometries (usually <10 cm, based on availability at column suppliers) due to the heterogeneity of longer columns, which reduces elution efficiency. A few papers have shown high efficiencies for long capillary columns but, to our knowledge, such columns were specifically produced for the application by the authors [65]. Therefore, as an alternative to monolithic columns, superficially porous particles (i.e., shell, fused-core™, core-shell™, partially porous or pellicular) were developed. These particles are made of a solid, nonporous core surrounded by a shell of porous material [48,60] and exhibit high kinetic performance with reasonable backpressures. These types of columns are widely used with particle sizes of 2.5–2.7 μm and provide high separation efficiency on standard HPLC systems without the need for an ultra-high pressure pump. However, such columns do not fully achieve the performance of UHPLC technology when used in traditional HPLC systems because of instrumental constraints (e.g., dwell volume and extra-column volume, etc.) [66]. These columns suffer from lower loading capacity compare to standard HPLC columns [67], however for sensitive LC–MS applications the amounts of extract loaded are usually below the loading capacities. Core shell columns have recently been made available for semi-preparative separation but still suffer from low loading capacities.

As an alternative to HPLC, supercritical fluid chromatography (SFC) has been applied to natural product analysis [68]. This technique uses CO_2, which possesses a modest supercritical point (31 °C and 74 bar), as the main mobile phase eluent [48]. The low mobile phase viscosity and increased molecular diffusion of compounds allow the use of high flow rates and/or longer columns for either rapid or highly efficient separations, respectively. CO_2 can be modified using a polar co-solvent, such as methanol or ethanol to

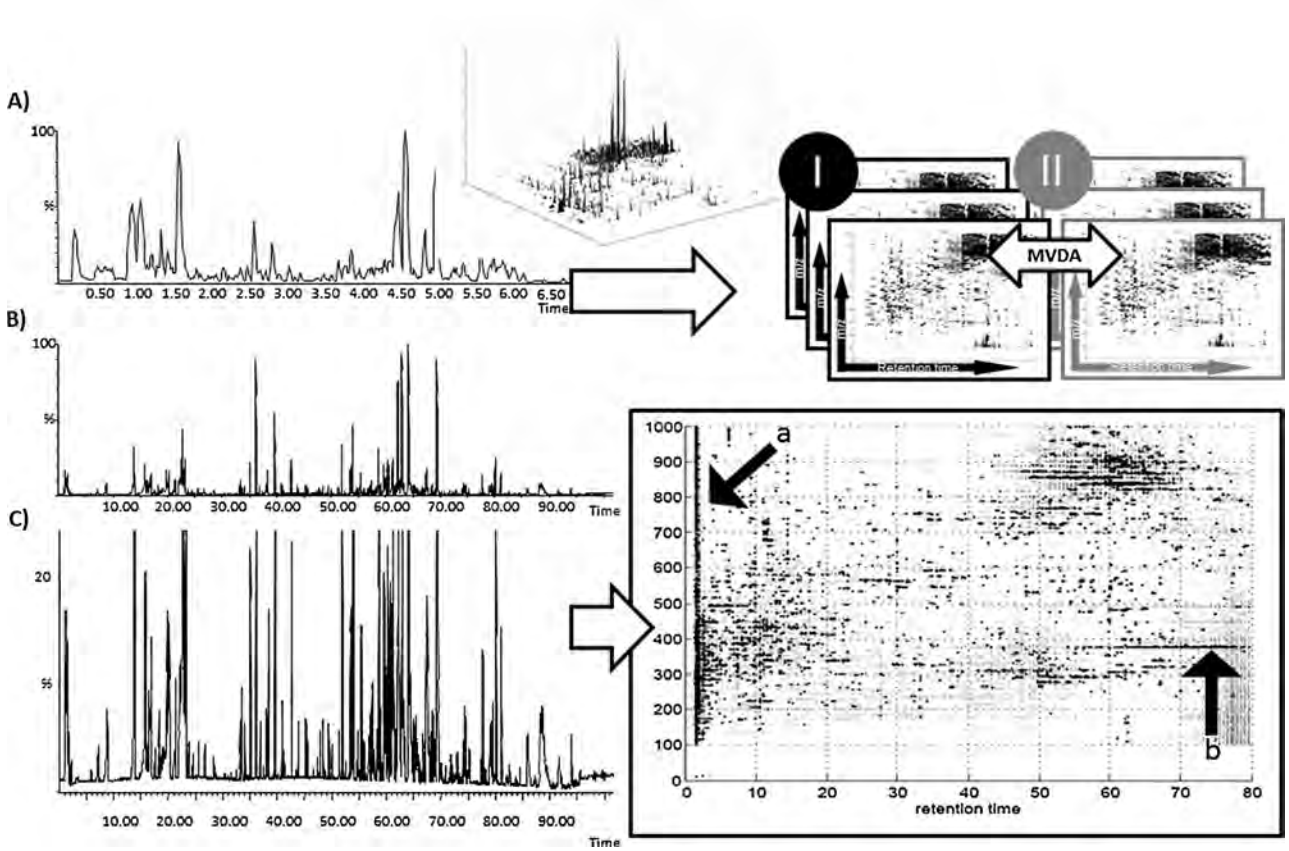


Fig. 2. Different LC fingerprinting and profiling approaches by NI ESI UHPLC–TOFMS with the corresponding ion maps obtained. Analyses with high and low peak capacities were performed. (A) High throughput fingerprints obtained using a 1.0 mm × 50 mm column for a metabolomics study where control and stressed *Arabidopsis thaliana* plants were compared. These rapid fingerprints, 7 min runs enable the analysis of multiple replicates that allows multivariate data analysis (MVDA) for highlighting features (*m*/*z* vs RT) that were induced in the wound response of the plant (see e.g., [58]). (B and C) High-resolution LC–MS metabolite profiling using a 2.1 mm × 150 mm column that enable a base line of many constituents. Such a detailed profiling is obtained by applying a chromatographic gradient transfers from (A) to (B) that ensure the same selectivity and enable a precise localisation of the feature of interest (biomarkers) for metabolite ID and or subsequent microisolation (see e.g., [3]). (C) Expansion of (B) at the 20% intensity threshold to highlight the very detailed composition information provided by TOF-MS detection. On the high-resolution 2D ion map "a" represents the highly polar compounds not retained by reverse phase separation on the C_{18} column, and "b" represents lipophilic compounds that are tailing in such conditions. Such types of features are generally discarded form the peak picking procedures used for MVDA in metabolomics.
Adapted from [58], [59] and [24] with permission.

increase the eluent strength of the mobile phase. Despite the fact that the SFC apparatus are more expensive (about 20–30% more), this complementary approach to HPLC shows potential for the profiling of natural extracts because it is well suited for analysing non-polar constituents strongly retained in reversed-phase (RP) HPLC. However, because the SFC stationary phase is very often polar (i.e., bare silica or silica modified with a polar group such as diol or amide), the elution of analytes possessing strong hydrogen bonding capabilities may be difficult. Detection is not a critical issue in SFC because a large array of detectors can be used including UV, ELSD or MS [69]. SFC–MS coupling need more complicated interfacing than RP LC–MS but, this is far less difficult than normal phase (NP) LC. Although this approach is not yet commonly used in natural product analyses, it has been successfully applied to various types of compounds, including artemisinine [70], ginkgolides [71], alkaloids [72] and lipids [73]. However, with the exception of lipids, the possibility to do generic gradients to profile a very large number of metabolites in a complex natural extract still need to be exemplified. In addition the capacity of SFC to profile polar constituents has to be better documented. With the recent introduction of more reliable SFC systems compatible with columns packed with small particles [74,75], applications for the efficient metabolite profiling of highly non-polar or even volatile compounds, are expected to be developed. Finally, SFC has significant advantages over HPLC for metabolite isolation because it uses CO_2 which is cost effective, do not need evaporation steps and is more environmentally friendly.

2.3. Nano/capillary liquid chromatography

Because electrospray ionisation (ESI) is concentration dependent, the development of miniaturised LC strategies have been studied to improve the sensitivity and increase the number of detected compounds, especially endogenous compounds that are present in small amounts [76,77]. When using the same amount of injected material, the sensitivity enhancement is the square of the standard diameter and nanocolumn ratio. For example, a theoretical increase in sensitivity of 110-fold may be obtained through the reduction of the i.d. of the column below 300 μm. Additionally, for some analyses of natural matrices, a higher sensitivity might be crucial for the extremely small amounts of sample in some studies. In addition to improve sensitivity, miniaturisation allows a reduction in the amount of solvent that is used, which also reduces costs [78]. Some disadvantages are however link to miniaturisation, such as lower robustness (primarily for nanoLC systems), increased care in sample preparation and particle filtering, extensive re-equilibration time and peak enlargement with high dead-volume systems. Altogether this may impair analysis throughput and repeatability.

Different flow rate segments can be applied from the nano- (<1 μL/min) to capillary (1–20 μL/min) and micro- (30–100 μL/min) scales. Although nanoscale LC systems that are combined with a nanospray ion source have been routinely used in proteomics, very few applications to the analyses of metabolites in natural product research have been reported [79,80]. The lack of studies in this area is most likely due to the fact that natural product extracts can generally be obtained in sufficient amounts. However, with microfractionation studies in HPLC activity-based profiling studies (see Section 6) [33], the amount collected can be extremely small [81], and nanoLC–MS may be useful. Indeed, in these cases, while the main fraction of the μg collected must be used for the assay, an extremely small fraction remains for further LC–MS profiling. In other instances, only a small part of a herbarium sample may be available for profiling, and the nanoscale approach can be extremely useful. The potential of nanoLC–MS methods has been demonstrated for crude plant extract analyses. For example, Fanali et al. developed a nanoLC-based method that showed complex profiles for different aloe extracts [82]. Additionally, the applicability of nanoLC–MS has been evaluated for the analysis of phenolic compounds in olive oil. Compared to conventional LC, the method demonstrated better sensitivity with reduced consumption of mobile phase, although a lower reproducibility was achieved to obtain the same quantitative performance [83]. The recent development of both monolith and sub-2 μm capillary columns allows significant miniaturisation of the LC scale, while using flow rates compatible with a conventional ESI source [78,84]. Due to the continual breakthrough in the field of LC-miniaturisation allowing to have improved robustness and ease-of-use instruments for routinely application, nano, capillary and microLC systems are expected to play increasingly important roles, particularly in metabolomics [85–87].

2.4. Selectivity and phase chemistries in liquid chromatography

As mentioned, natural products are very diverse [88,89] and show a broad range of polarities [22]. Therefore, a large number of protocols are used for LC separations. To improve peak separation, the selectivity can be optimised by modifying the stationary phase chemistry or the HPLC conditions (i.e., gradient profile, temperature and nature of organic modifier). For given selectivities, as mentioned above (see Sections 2.2 and 2.3), the column geometry (i.e., length and particle size) can be adapted to afford higher peak capacities and/or higher throughput. For the metabolite profiling of crude natural extracts, the large majority of applications utilise RP LC. The stationary phases of choice in RP LC for natural product profiling are octadecyl group-bonded silica gel (C_{18}) phases [41]. This phase is largely used with gradient elution using a water and methanol/acetonitrile solvent system. In some cases, none of the traditional reversed stationary phases (C_{18} or C_8) result in satisfactory separation. Generic RP gradients are generally preferred for profiling because they are fully compatible with MS detection using atmospheric pressure ionisation sources (ESI or APCI).

For the separation of compound classes that are difficult to resolve on C_{18}/C_8, alternate RP stationary phases have been developed, such as polar embedded alkyl, polar end-capped alkyl, fluorinated, aromatic-based and alkyl C_{30} [90]. For example, a C_{30} phase is very efficient for the separation of carotenoids and is often used for the separation of lipophilic plant constituents [91].

Even with the large panel of C_{18} RP stationary phases (differentiated by bonding density or/and by end-capping treatments) available, significant differences can be observed under RP LC [92,93] or SFC [94] conditions.

The profiling of lipophilic constituents in extracts can be efficiently performed using NP LC [95,96], which can, to some extent, be compatible with MS [95]. However, under such separation conditions, MS and UV-diode array detection (DAD) techniques are difficult to implement for retrieving the structural information necessary for LC-peak annotation, depending on the solvent (e.g., solvent cut-off for UV detection). A relatively low number of LC–MS applications for natural extracts have been reported to date. For example, procyanidins from various fruits have been quantified and partially identified using NP LC–MS with fluorescent detection [95].

Thus, lipophilic natural products are still most commonly profiled under RP conditions, which often compromises their elution. In C_{18} RP, natural products can be eluted using an eluent with increased eluting strength, such as isopropanol [97] or tetrahydrofuran [98]; however, the latter solvent is not as compatible with MS, and high concentrations should be avoided (technical data). Additionally, high temperature conditions that improve the eluting strength of the mobile phase may be used. For example, comparable retention times have been obtained under isocratic conditions using 8:2 methanol/water at 25 °C and 1:9 methanol/water at 185 °C [99]. A few procedures using non-aqueous reverse phase (NARP) LC–MS have been very efficient for profiling highly non-polar triacylglycerols in plant oils [100]. Alternatively, the use of less retentive columns (C_8, C_4 and cyano) also reduces the number of interactions between non-polar compounds and the stationary phase, thus facilitating their analysis [101].

As mentioned, alternative solutions can also be envisioned, such as the use of modern SFC–MS with columns packed with sub-2 μm, fully porous particles [69,74]. Some reports have already demonstrated efficient chemical profiling of triacylglycerols and diacylglycerols using this approach [102].

The profiling of highly polar compounds also requires an approach other than RP on C_{18}. They are poorly retained in LC–MS with a C_{18} material (the elution of polar metabolites is shown on the ion maps displayed in Fig. 2B (arrow "a")). Hydrophilic interaction chromatography (HILIC) provides one such solution [103,104]. In HILIC, the analytes interact with the polar stationary phase, which is composed of silica gel, chemically modified silica (i.e., diol, amide and aminopropyl groups or zwitterionic phases) or polymers with polar functional groups [105]. HILIC utilises a large diversity of stationary phases that possess distinct characteristics. In all cases, compounds are eluted with an aprotic solvent mixed with water, which is the strongest eluting solvent. This approach has been successfully applied to the separation of peptides, oligosaccharides, glycosides, amino sugars, amino acids, sugar nucleotides and phosphorylated metabolites [104–106]. Interestingly, HILIC was also successfully utilised for the separation of non-polar compounds such as lipids [101]. It should be mentioned that in HILIC high throughput may be impaired by lack of reproducibility usually related to a slow equilibration of the column, particularly with mobile phases containing buffers [107].

Importantly, in metabolomics studies, no single method is sufficiently comprehensive to provide a complete view of all metabolites present in a given extract [15]. A recent comparison of analyses of the urine metabolome by RP, HILIC and a pentafluorophenylpropyl (PFPP) stationary phase showed that <12% of the annotated compounds were adequately analysed in all three LC systems [108]. This result clearly highlights the need for complementary methods.

2.5. Multiple dimensions in liquid chromatography

To obtain a comprehensive view of the metabolome, emphasis should on the resolution of a greater number of compounds as well as the investigation of less abundant species with off-line or on-line multidimensional LC–MS. This goal can be achieved by combining two or more separation dimensions prior to MS detection in order to increase the peak capacity (i.e., the number of peaks that can be theoretically separated with a resolution of 1) [109]. Independent of

the exact strategy, orthogonal separation modes that involve different physicochemical interaction are generally combined to increase the resolving power for complex biological extracts [110].

Unlike proteomics, the use of multidimensional LC is not widespread in metabolomics. Some studies have recently been developed based on combinations of normal- and reversed-phases, especially in off-line setups [111]. In such investigations, the metabolites were first fractionated according to their hydrophilic affinity in one dimension prior to the injection of all fractions and subsequent separation according to their hydrophobicity. To increase the throughput of these processes, an on-line approach using a solvent-evaporating interface that overcomes the solvent immiscibility issues of mobile phases has been demonstrated [111].

Alternatively, the combination of HILIC and RP LC is attractive because both techniques use the same type of eluent, a mixture of acetonitrile and water with a volatile buffer [112]. Unlike RP, this aqueous phase is the strongest eluent in HILIC [113,114]. Although an off-line combination is straightforward, the on-line hyphenation of these two modes has also been demonstrated. In the comprehensive LC × LC setup, the mobile phase at the outlet of the first column was adjusted to ensure that the analyte was focused at the second column head [84,115]. This strategy has been successfully applied for the analysis of metabolites with a wide polarity range using a single experiment [116,117]. Bleeder et al. successfully demonstrated the effectiveness of comprehensive HILIC × RP LC for the analyses of rooibos (*Aspalathus linearis*) phenolic compounds [118]. Although 16-fold faster than the off-line mode, their comprehensive HILIC × RP setup suffered from a significant loss of peak capacity in the second dimension compared to the off-line mode, due to the shorter analysis time. However, the gain in both on- and off-line modes with respect to the one-dimensional analysis allowed broader coverage of the complex phenolic compositions of the rooibos samples.

In addition to hydrophobic or hydrophilic modes of separation, orthogonal separation according to the charge state of molecules can be used. For example, the use of strong or weak anion exchange columns in the first dimension may be promising for the LC × LC–MS analysis of metabolites [119]. In this manner, metabolites can be first separated according to their charge using an ion exchange column. The metabolites can then be focused on-line in the second dimension using a trap column for desalting before being separated by RP LC according to their hydrophobicity on a C_{18} column [120]. Because ion exchange chromatography use high water content eluent, its use as first dimension leads to a straightforward hyphenation with RP LC. Similarly, the HILIC × RP LC combination approach affords orthogonal separation with a eluent composition (usually aqueous salt gradient) compatible with RP LC [77].

Ion mobility (IMS) can also be regarded as an additional orthogonal dimensions to chromatography for the separation of isomeric ions based on the differences in cross sections (CSS) rather than differences in their *m*/*z* values (see Section 3.3). In such cases, the LC × IMS combination can afford a highly resolved 2D chromatogram, even when isomeric pairs co-elute in the first HPLC dimension [121]. For example, UHPLC-ESI-IMS–MS has successfully been applied for the rapid identification and structural characterisation of flavonoid glycosides from cauliflower waste [122].

Generally, LC × LC separation methods show great potential for the analysis of natural extracts, especially for very detailed metabolome analysis in which all metabolites should ideally be separated in the chromatographic dimension. However, unlike comprehensive GC (GC × GC) [123], the application of such methods remains relatively uncommon. Despite the complexity of generated data, a tremendous increase of the use of LC × LC in the future is expected in regard of the potential of this technique for deciphering complex biological matrices.

2.6. Gas chromatography coupled to mass spectrometry

In natural product research, GC–MS techniques have historically enabled both the separation and identification of natural compounds in complex mixtures. Capillary GC demonstrates very high peak capacities, and electronic impact (EI) MS, even with nominal mass resolution, affords spectral patterns that can be easily searched against databases, as the ionisation energy is fixed at 70 eV. The primary restriction of the method is that the analytes must be volatile; thus, the technique is restricted to the analysis of volatile plant fractions (e.g., headspace, essential oils, flavours, fragrances, aromas and extracts prepared through specific techniques) [28]. The technique is routinely and successfully used for the analysis of volatiles [29].

However, GC–MS is also useful for the analysis of non-volatile constituents, provided that sample enrichment steps and derivatisation steps are performed [124,125]. Thus, despite a derivatisation step not being necessary for LC–MS, GC–MS remains one of the most widely used analytical techniques. The very high chromatographic resolution of capillary GC or GC × GC and the performances of both EIMS and HRMS instruments makes the techniques particularly useful for the analysis of complex biological extracts [126]. Thus, GC–MS has been extensively applied to metabolomics and, to a broader extent, to natural product metabolomics, especially primary metabolite analysis [127,128]. To increase the volatility and thermal stability of molecules, numerous derivatisations can be used to "protect" functional groups, such as —OH, —COOH, —NH and —SH; for example, alkylation, silylation and acylation reactions have been utilised [124,125]. Among these derivatisations, a two-step procedure involving the combination of methoximation and trimethylsilylation has successfully enabled large-scale metabolomics studies with GC–MS [129,130]. However, the derivatisation steps may lead to increased complexity due to incomplete reaction and biased the analyses [131].

Due to the highly reproducible mass fragmentation observed in EI, the strength of GC-EI–MS is the possibility of inter-instrument, library-based searching for the identification of compounds [132]. However, compound identification can be ineffective when insufficient chromatographic resolution results in the contamination of the mass spectrum of interest [133]. To improve identification, software, including AMDIS [134], has been developed that allows for the deconvolution of complex spectra and the matching of a purified spectrum to a reference spectrum [135–137].

To increase both peak capacity and sensitivity, the use of comprehensive GC (GC × GC) is increasing in metabolomics through the development of robust commercial devices [138–140]. In contrast to conventional GC, compounds separated on a first column are transferred to a second column with different physicochemical properties using a modulator. This topic has been reviewed in previous reports that provide extensive details of the technical aspects [138,141,142]. In a recent feasibility study, GC × GC was compared to conventional GC in the profiling of flavonoids from both standards and complex natural product matrices, including propolis and *Chrysanthemum* [143]. The study confirmed the capability of this technique for the qualitative and quantitative analyses of this class of plant secondary metabolites.

In another study, Gao et al. assessed the profiles of volatiles using GC × GC–MS to determine the biomarkers in *Eucalyptus globulus* leaves infected by the necrotrophic fungus *Teratosphaeria nubilosa* [144]. Utilising the developed strategy, more than 40 biomarkers were highlighted, providing novel insights into plant defence mechanisms against this infection.

Although MS was naturally hyphenated with GC through EI more than 40 years ago, the recent emergence of atmospheric pressure chemical ionisation (APCI) as a commercial ionisation source is another area of growing interest [145,146]. Interestingly, APCI

overcomes some common disadvantages of EI and chemical ionisation (CI) and is complementary to both ion sources through its use of a more "universal" and soft ionisation process, respectively [145,147,148]. In a recent work, a GC comparison between APCI and EI ionisation sources was performed for the characterisation of avocado metabolites [149]. Interestingly, the complementary capabilities of this technique were demonstrated with respect to EI. With the on-going development of spectral libraries, GC-APCI–MS will continue to develop and likely become a primary technique in the metabolomics toolbox [150]. When no matches are present in the database identification is rather complicated due to the high fragmentation observed in GC-EI–MS, therefore the low fragmentation observed in GC-APCI–MS could be particularly interesting to help the identification of unknown metabolites. Thus, GC-EI–MS in combination with GC–APCI(+)-QTOFMS has proven to be highly successful for the identification of unknown metabolites present in *Arabidopsis thaliana* [151]. In this case, the APCI source reduces fragmentation, and thus, the protonated molecular ions are more easily observed. If completely unknown metabolites are highlighted they can be subsequently isolated by preparative GC and identified de novo by NMR [152]. This is feasible but remains more complex than with LC approaches.

To improve separation of compounds within a single analysis, strategies coupling LC and GC were also developed [153]. Obviously natural products need to be either directly suitable for GC analysis, or derivatised prior to the LC × GC separation. Despite its high separation efficiency LC × GC was rarely applied to natural extracts but was successful for the analysis of sterols in edible oils [154].

3. Strategies without chromatographic separation

Different spectroscopic techniques can be used to profile metabolites from crude natural extracts. Direct MS and NMR (as detailed below) as well as infrared (IR) approaches have been successfully used. All of these methods provide spectroscopic fingerprints of the complex metabolite mixture that can be treated using MVDA in metabolomics studies for discrimination purposes [15]. However, some methods, including MS (through MS/MS spectra) and NMR (through 2D NMR), can also provide important information for the direct identification of metabolites in crude mixtures (Fig. 1).

Fourier transform-middle IR spectrometry (FT-MIR) is commonly used for the fingerprinting of different natural matrices, especially for quality control of medicinal plants in powder form [155] or plant extracts [156], and for the rapid qualitative evaluation of sample preparations in metabolomics [157]. However, the method is strictly used with MVDA to classify samples in metabolomics and does not provide useful data for the metabolite identification of specific markers.

In the case of direct profiling of extracts, very complex convoluted data are obtained and compound identification may be impaired. Therefore the careful optimisation of the extraction steps may be necessary to improve the quality profiles [50,158,159]. The extract can also be enriched in the metabolites of interest prior to analysis. For small sample quantities, LLE or SPE are used for a rapid enrichment. LLE and vacuum liquid chromatography (VLC) can be employed for larger samples that are sometimes needed, for example complex NMR profiling. These latter techniques are also generally used as preliminary fractionation steps for natural product isolation [160].

Because NMR and direct MS analyses not only provide fingerprints but also structural information for analyte identification, these approaches are discussed in more detail as complementary or orthogonal approaches to the chromatographic MS methods presented in Section 2.

3.1. NMR profiling

NMR is widely used for metabolite fingerprinting of crude natural extracts and does not require prior chromatographic separation [15]. The method is simple and reproducible; thus, it does not require specific sample preparation techniques [127]. In the case of ^{1}H NMR acquisition, this approach has high throughput capacity and spectra can be acquired in a few tens of scans (ca 1 min acquisition), this is however depending on the sensitivity of the instrument used and the amount of sample available. A combination of the chemical shifts (the nature of the chemical environment in which a particular nucleus is located), spin–spin coupling (the number and nature of nearby nuclei, which provides connectivity information) and peak intensity (concentration of protons) provides important structural information for the identification of any compounds of interest [161]. NMR analyses also provide direct quantitative information because the intensity of the proton signal is proportional to the molar concentration of the metabolites [162–164]. Through 2D NMR experiments, much information regarding the atom connectivities of the metabolites within a complex mixture can also be obtained [96,165,166] and in some cases, complete de novo metabolite identifications can be achieved [32]. Because NMR chemical shifts are highly reproducible using a specific solvent or buffer conditions, comparison of the ^{1}H NMR chemical shifts with those reported in databases results in unambiguous identification, provided that standard protocols for sample preparation are used [15]. With natural extracts, this is particularly useful for the identification of primary metabolites (e.g., amino acids and sugars), and secondary metabolites may be identified based on their chemical shifts and signal multiplicities, if ^{1}H NMR databases that contain the compounds under study are available, as is the case for flavonoids [167]. While the technique is extensively used for the profiling of biological fluids in metabolomics for medical applications [168], standard protocols are also used for NMR profiling of natural extracts [169] and has led to many applications [170,171].

The resolution of ^{1}H NMR spectra is dependent on the field strength of the magnet used, and profiling methods for instruments up to 1 GHz have been reported. Furthermore, unlike MS, metabolites produce signals that are convoluted and superimposed on each other when mixtures are profiled using NMR. Additionally, many of the signals occur as multiplets due to spin-spin coupling. One method for obtaining higher resolution is to perform 2D ^{1}H–^{1}H J-resolved experiments. These experiments provide spectra that exhibit a coupling constant in a dimension orthogonal to that of the chemical shift axis. A projection of the obtained 2D plot onto the chemical-shift axis yields a fingerprint of peaks without multiplicities and thus provides better resolution and avoids some overlapping [172].

Several types of NMR databases are available [15], and many studies dedicated to establishing these databases have recently been published. For example, MetaboMiner automates the identification and quantification of metabolites using ^{1}H NMR data of biological origin [173]. Additionally, NMR spectra can be simulated by different types of software to predict chemical shifts and signal multiplicities; these greatly support the metabolite identification process [167].

The low sensitivity of NMR is considered the major limitation of this application. Low sensitivity is intrinsic to the method of detection in NMR. The signal in an NMR experiment results from the transition between low and high nuclear spin energy states, and the low sensitivity is due to the small differences in the Zeeman energy level populations. State-of-the-art NMR probes using high field magnets (e.g., cryoprobes on NMR magnets >600 MHz) have enabled the detection of single constituents on the low microgram (nanomole scale) scale when solubilised in microtubes [174]. Such sensitivities are high for NMR applications, but are still

significantly lower than what can be achieved through MS profiling methods. With natural extracts, milligram amounts are typically required for profiling studies in conventional 5 mm i.d. tubes. The detection of particular metabolites in complex mixtures by NMR can be enhanced by statistical correlation analyses between LC–MS and NMR datasets in metabolomics studies that enable the specific detection of a specific set of metabolites, based on the covariance of MS and NMR signals, as exemplified in Section 4.3.5.

Additionally, NMR is a very useful tool for the structural elucidation of metabolites that cannot be dereplicated by MS profiling methods. As discussed in Section 4.3.1, microNMR methods directly hyphenated to HPLC or used after the targeted MS microisolation of given metabolites is the ultimate tool for unambiguous identification [175].

3.2. Direct infusion MS (DIMS)

Deciphering the complex composition of biological samples based only on MS without prior chromatographic analysis has developed, especially with the recent advancement in HRMS and ultra-HRMS and in robots used for automated nano-infusion. The direct infusion of samples in the MS interface (DIMS) can significantly increase the throughput compared to LC–MS. Similar to NMR, DIMS has primarily been used to record fingerprints for differential metabolomics studies rather than for the dereplication and comprehensive identification of compounds present within an extract. Another advantage with MVDA is the simplification of data pre-processing steps using dedicated toolboxes [176] compared to LC–MS pre-processing treatments for metabolomics. DIMS is usually employed with API sources, such as ESI or APCI, and samples can be either infused through a syringe pump or directly injected into the LC–MS flow (plug flow) using short LC runs (2–3 min). Samples may also be continuously injected using nano-infusion devices, such as the chip-based Nanomate™ [177]. Because no dimension orthogonal to MS is used for m/z separation, the acquisition of total MS spectra fingerprints (TMS) must preferably be made on HRMS instruments [24]. HRMS and ultra-HRMS instruments are considered the state-of-the-art for numerous DIMS-based studies. Such mass analysers include time-of-flight (TOF) (resolution of 50,000, accuracy of 1 ppm), Orbitrap (resolution of 100,000, accuracy <1 ppm) and Fourier transform ion cyclotron resonance (FTICR) (resolution >500,000, accuracy <1 ppm). These instruments provide feature deconvolution and metabolite annotation in a single run based on molecular formula determination with high accuracy. HRMS/MS experiments give additional structural information. However, while the resolution of isobars is possible, isomers, which are likely to be present in natural extracts, may not be separated from each other [15].

Fingerprints recorded at nominal mass can be sufficient for sample classification or quality control purposes. For example, wine made from a combination of different grape varieties and ageing have been separated using the MVDA of DIMS fingerprints (1/100, v/v) with a methanol/water solution [178]. DIMS is also useful for certifying the origin and detecting the falsification or adulteration of complex products of natural origin [179–181]. The use of HRMS DIMS provides improved information because more features can be detected. For example, an interesting study of wine ageing in oak barrels that was based on HRMS fingerprints acquired using FT-ICR-MS allowed the discrimination of wines according to their age and variety and provided the metabolite signature of the forest location in which the oaks used to manufacture the barrels were grown (Fig. 3) [182].

In HRMS spectra, chemical formula assignments are confirmed by matching mass accuracies and by comparing the observed and modelled isotopic distributions (see Section 4.1.1). Access to ultra-high HRMS is sometimes necessary to avoid false positive interpretations (Fig. 3). In the study of Gougeon et al., database cross searching, limited to *Vitis vinifera* metabolites, allowed the putative annotation of discriminant signals and led to an interesting hypothesis relating the geographic biomarkers of barrels and the organoleptic properties of wine. Another strategy currently used for feature annotation involves targeted HRMS/MS experiments on relevant features to obtain fragmentation patterns that can be readily compared to the literature data or dedicated databases [183,184]. For example, direct flow injection FT-ICR-MS was used to assess metabolite modification in grapevine leaves upon downy mildew infestation [185]. Control and infested leaves were clearly segregated, and targeted MS/MS experiments allowed the identification of 19 biomarkers involved in the grapevine leaf response. Another original approach developed by Giavalisco et al. takes advantage of HRMS fingerprints combined with ^{13}C isotope labelling of the entire metabolome from *A. thaliana* and demonstrated substantial improvements in the detection of true metabolites (compared to background ions) and their subsequent assignments [186].

Although DIMS is easy to set up and provides much information on natural extracts and other biological matrices, the studied metabolites suffer from ionisation suppression, which arises from competitive ionisation with other components in the matrix [187]. ESI is particularly sensitive to these matrix effects, which leads to minor compounds being overlooked, the decrease or absence of signals from major constituents, and bias in the accuracy of quantified results. However, ion suppression can be reduced by dilution assays to obtain a broad coverage of signals across the whole spectrum. For example, to develop a high-throughput metabolome fingerprint platform, Fuhrer et al. recorded high linear responses at low concentrations for 73 metabolites spiked in yeast extracts [188]. Additionally, to assess the quality of measured data, quality control samples (QC) representative of the biological matrix under study and infused periodically throughout the experiment can be incorporated into the experimental design. Thus, the within- and between-batch variations can be measured and even corrected using QC samples as reference spectra [189]. Overall, DIMS-based analytical methods represent a valuable high-throughput strategy for the examination of complex extracts or the monitoring of compounds that would otherwise be outside the usual scope of RP chromatography [177]. This fact is especially true for compounds that cannot easily be profiled under standard RP conditions, such as highly polar or lipophilic compounds. In some cases, including lipidomics, DIMS represents the method of choice because better results for either untargeted lipid profiling or targeted quantification have been obtained compared to LC–MS based methods [190,191].

3.3. A new dimension in MS: ion mobility spectrometry (IMS)

One of the major drawbacks of DIMS-based methods is the inability to segregate isobaric or isomeric compounds in complex mixtures using nominal or HRMS spectra, respectively. However, IMS represents a valuable tool for addressing this issue. IMS separates ions on the millisecond timescale according to their shape, size and charge using a chamber filled with an inert gas, thus adding an additional dimension to DIMS. From the time required for a given ion to cross the chamber, the collision cross-section value (CCS), which appears to be a unique physicochemical property of the compound, can be derived [192]. The IMS apparatus is generally hyphenated to TOF-MS (IMS-TOF-MS) because of its fast acquisition rate. Using such a configuration, >1100 ions were detected in human blood samples, and 300 isomeric compounds were separated [193]. Interestingly, each class of compounds was detected in a characteristic area of the 2D IMS–MS spectrum (drift time $\times$ m/z), thus facilitating feature annotation. Both the

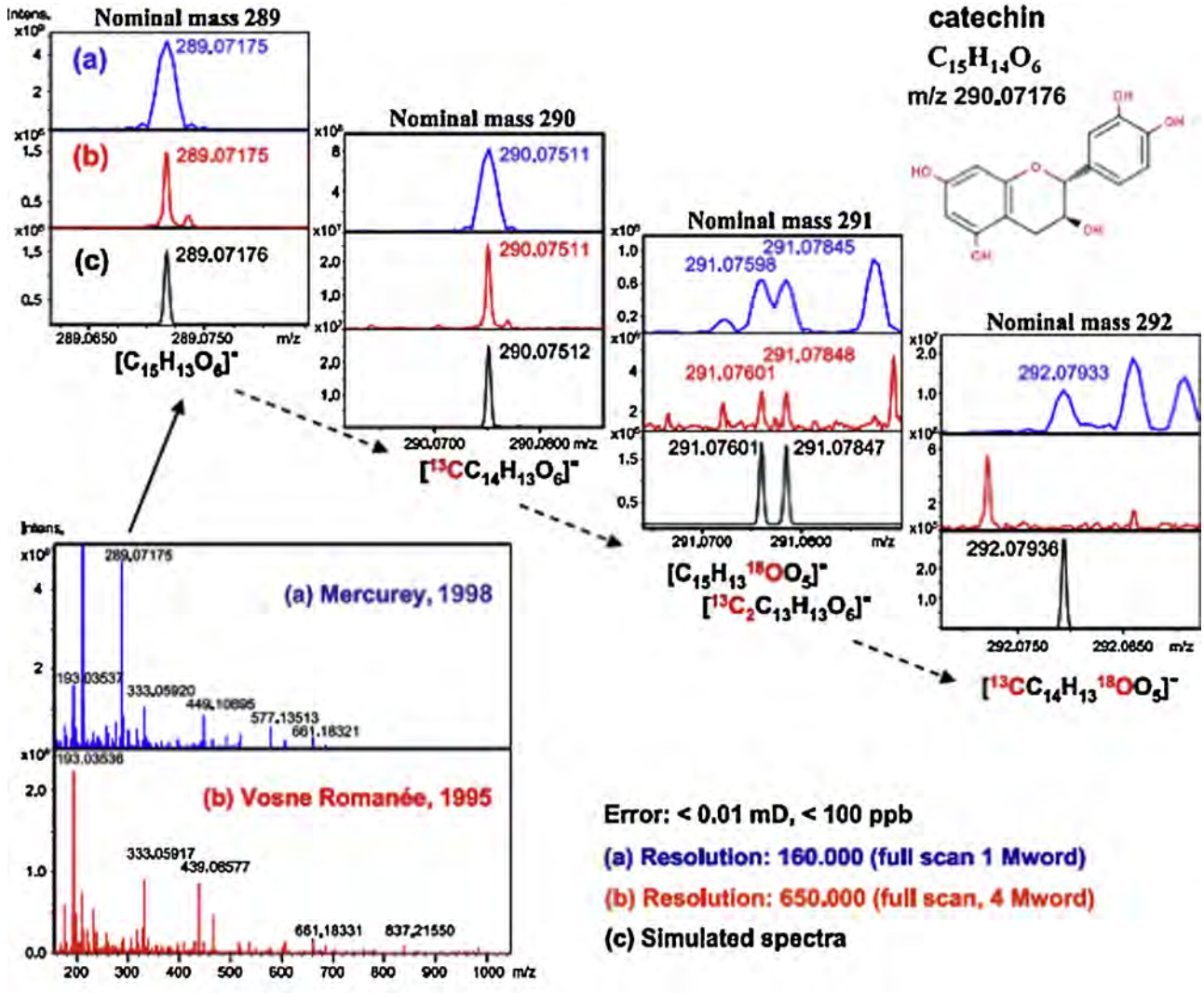


Fig. 3. DIMS FT-ICR-MS spectra of wine samples directly infused in the high resolution MS instrument without any sample preparation. The DIMS fingerprints of (A) Mercurey 1998 and (B) Vosne Romanée 1995 acquired by infusing the sample for 30 s in the NI ESI mode from *m/z* 200 to 1000. The mass peak 289.07178 typically dominant in the Mercurey is represented in detail with corresponding resolutions of around 160.000 (blue) and 650.000 (red) for the Mercurey and Vosne Romanée wines, respectively, and compared to the simulated spectrum, showing the presence of corresponding isotopologues at nominal mass *m/z* 290, 291, and 292. The presence of the isotopologue in the mass spectra is used to confirm the assigned elemental compositions of the signals in the spectra. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
Adapted from [182] with permission.

signal-to-noise ratio (S/N) and the separation of isobaric drugs were significantly improved using an IMS interfaced between a DESI source and a TOF-MS [194]. Isobaric flavonoids frequently encountered in natural extracts were resolved more effectively using IMS-TOF-MS compared to rapid profiling by UHPLC–TOF-MS [195]. IMS is also frequently associated with LC to improve the peak capacity of the system. Paglia et al. investigated the potential of UPLC-IMS–MS as a metabolomic platform for reproducible compound detection and identification. Interestingly, the measured CCS values were highly reproducible, and a comparison between the modelled and measured CCS greatly improved the identification of metabolites from human cell extracts [196].

3.4. Desorption based methods: MALDI, DESI and DART

Other direct techniques for metabolite fingerprinting complementary to DIMS are matrix-assisted laser desorption/ionisation (MALDI) and ambient MS.

MALDI is primarily applied to the study of large biomolecules (e.g., proteins, nucleotides and biopolymers), as conventional matrices generate numerous background ions and limit the detection of small compounds (<500 Da). However, low MW compounds can also be profiled to some extent using this technique. Ion suppression and matrix interferences can be reduced by the use of modified matrices and adapted protocols [197,198]. Graphene was used as a matrix for the first time, and the fragmentation of analytes was not observed; in addition, good reproducibility was observed for the detection of amino acids, polyamines and anticancer drugs [199]. A growing area of research based on MALDI approaches is the direct study of the spatial distribution of metabolites in tissues [200].

Ambient ionisation techniques have emerged for surface analysis in the open air without sample preparation and thus allow the analysis of samples in their native form by substantially reducing the risk of artefact formation. Ions are generated under an ambient environment using either a spray of charged micro-droplets (desorption electrospray ionisation, DESI) or a gas stream activated by a plasma discharge (direct analysis in real time, DART) [201]. Key advantages of ambient MS compared to other metabolite profiling methods are the lack of sample preparation (solid-phase extraction,

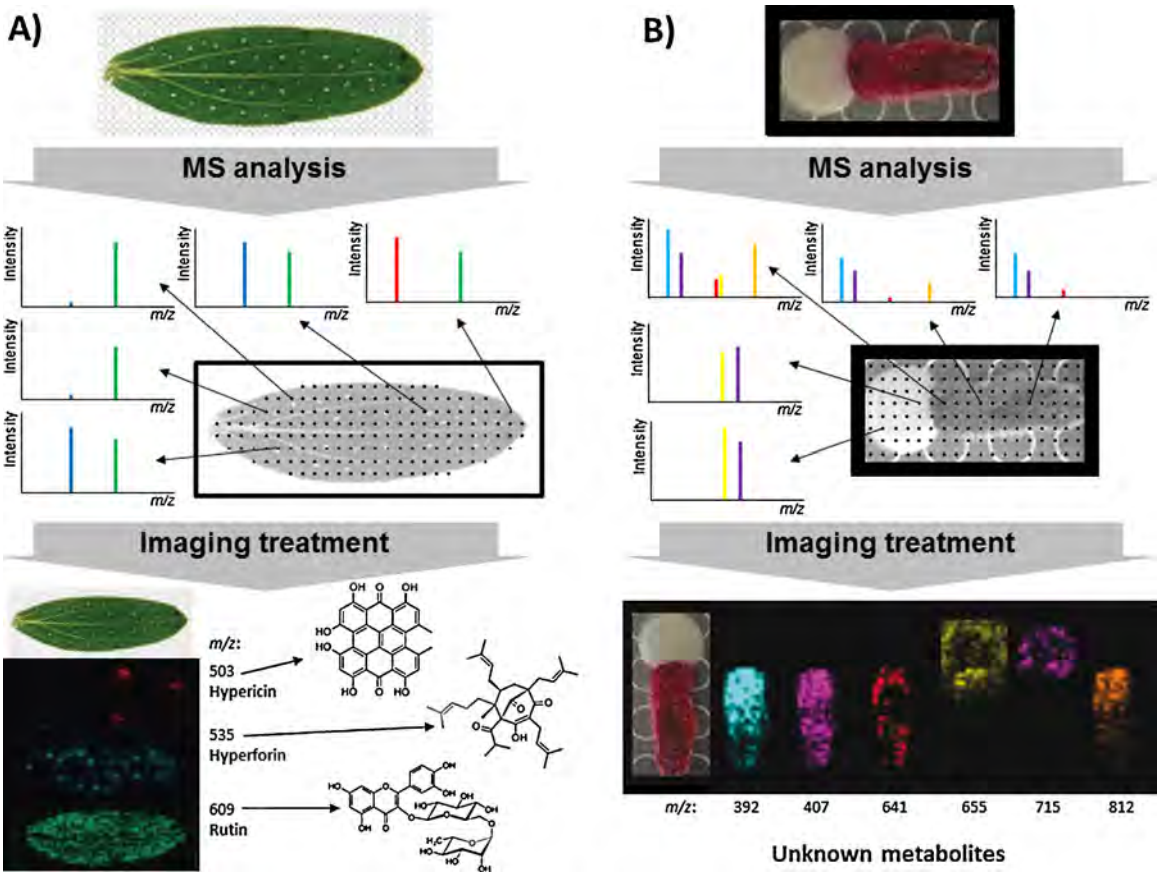


Fig. 4. Imaging MS strategy applied to natural matrices for (A) plant leaf surface analysis or (B) microbial interaction at a surface of growing colony or solid media culture. The sample surfaces are sequentially rastered to acquire a MS spectrum at every location. Then, images are generated as heat maps for each ion detected to highlight their location. (A) Analysis of a *Hypericum perforatum* leaf. Heat maps images of the selected ions (*m*/*z*) reveals the presence of hypericin, hyperforin and rutin on the leaves. (B) Analysis of the confrontation zone of the microbial co-culture (*Bacillus subtilis* and *Streptomyces coelicolor*). The images show the distributions of unknown metabolites as a result of the stress caused by the confrontation.
Adapted from [214,215] with permission.

concentration, chromatographic separations, etc.) and its ability to obtain MS fingerprints of the surface in a condition close to its native state. These emerging tools find interesting applications in the rapid analysis of pharmaceutical tablets [202], surface explosive residues [203], and volatiles [204], among others [205]. The immediacy of the results obtained is beneficial for the monitoring of industrial processes or in the area of QC. In natural product research, untargeted fingerprints acquired using DART-MS have been used to monitor the tea fermentation process. Key components were identified by the statistical analyses of fingerprints; biomarkers were then putatively identified by MS^n experiments [206]. A high-throughput assessment of beer was conducted using the DART-TOF-MS analysis of crude samples after degassing [207]. Additional applications of such simple and direct MS methods are expected to be applied increasingly often, for example, in plant QC. DESI was successfully used to screen camptothecin and the derivative 9-methoxycamptothecin in *Nothapodytes nimmoniana*. Results obtained after incision of the leaves and stem to access to the underlying structures allowed the detection of both targeted compounds with higher intensity arising from the stem [208].

3.5. Imaging mass spectrometry

Desorption and ambient ionisation methods can not only obtain direct MS spectra from a sample but also monitor these MS spectra at different locations on the surface of a given sample. Imaging MS is an innovative approach for obtaining information regarding the spatial distribution of analytes on a tissue [209,210] or microbial colony [211] surface. Remarkably, approximately thousands of unlabelled molecules, including proteins, peptides, and lipids, can be selectively mapped with a high correlation to the histological details of the tissue surface [212,213]. The imaging MS workflow is illustrated in Fig. 4.

Using either a laser or an ion beam, an imaging MS experiment is performed by rastering the entire surface according to a user-defined array of fixed spatial resolution and dimension. A mass spectrum is acquired at each coordinate (or pixel) with a lateral resolution that generally ranges from 5 to 300 μm [216]. Ion images are reconstructed with dedicated software to integrate the intensity of selected *m*/*z* features with respect to their coordinates of origin [217]. In plant sciences, the Imaging MS proof-of-concept has been demonstrated in recent years [215]. Interestingly, Ye et al. detected a large array of metabolites, including organic acids, amino acids, sugars, lipids, flavonoids and their conjugates, which allowed the numerous regioselective distributions between the roots and nodules of the *Medicago truncatula* plant to be evaluated [218]. Enriched by a decade of remarkable developments, Imaging MS is currently mature enough to be considered a basic technique with the ability to drive the discovery of novel biological insights [219,220]. For example, Shroff et al. used Imaging MS to measure the spatial distribution of glucosinolates in *A. thaliana* leaves [221] (e.g., as in Fig. 4A). The study revealed that the distribution of glucosinolates has a selective pattern across the leaves and may influence the feeding behaviour of *Helicoverpa armigera* larvae, thus playing a pivotal role in the defence of leaves against chewing herbivores by creating a barrier to feeding.

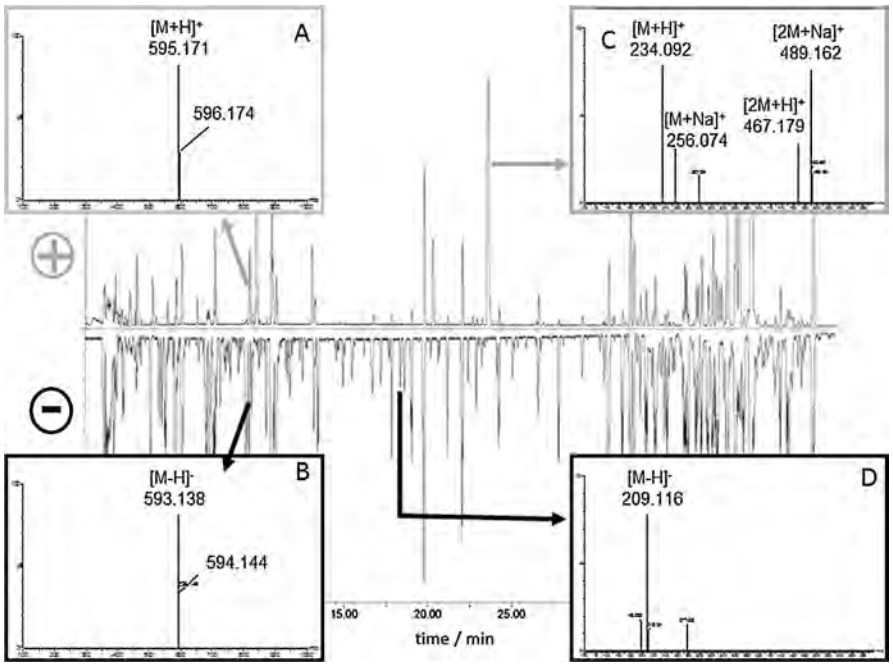


Fig. 5. UHPLC–TOFMS metabolite profiling of a crude *Arabidopsis thaliana* extract obtained in PI (upper trace) and NI (bottom trace) ESI modes. The chromatograms show the complementarity of both ion polarities for the detection of a maximal number of metabolites. (A and B) PI and NI TOFMS spectra obtained for a flavonoid diglycoside. The combination of both spectra enabled to confirm that M is 594.158 and the molecular formula $C_{27}H_{30}O_{15}$. (C) Example of a metabolite strongly ionised in PI mode. The presence of multiple ion adducts in the form of protonated and sodium adducts as well as dimers complicated the spectra but enable the attribution of M with need for the corresponding NI data. (D) Example of a metabolite (Jasmonic acid) giving rises only to a deprotonated molecule in NI mode and no ionisation in PI mode. In this case and without standard or data base information it is difficult to ascertain the M unambiguously.
Adapted from [24] with permission.

MALDI imaging has also been successfully used to highlight metabolite induction in microorganism interactions [222] (e.g., as in Fig. 4B). Using this method, a large variety of natural products (desferioxamines, prodiginine or actinorhodin) were induced at the interaction zone of *Streptomyces coelicolor* with other actinomycetes [223]. NanoDESI has been found to be an efficient method for detailed imaging of small molecules [224].

4. Metabolite identification strategies

After the metabolite profiles are acquired at the highest possible resolution, with a high degree of separation of the chemical constituents of an extract, the identification of the eluting compounds becomes the next challenge. MS [27] and NMR [225] have most often been used for this purpose. Due to its higher sensitivity, MS is primarily used for dereplication purposes.

The general strategy for natural product identification or LC peak annotation is based on database searches. In these searches, MS, HRMS, MS/MS or NMR spectra are compared to a large number of existing databases. With NMR, the searches are limited to particular solvent compositions because the chemical shifts of signals may vary significantly in different solvents. However, achieving annotation with strong evidence remains rather complicated, even though large structural databases of >200,000 natural compounds exist [226]. Therefore, other strategies are necessary for the identification of unknown or unreported metabolites, which is a challenge for metabolite profiling.

4.1. Metabolite identification based on MS

Due to its high sensitivity, MS detection is commonly used for metabolite profiling. Low-resolution mass spectrometers were initially used for the identification of the natural products in a given extract [227,228]. Success was largely based to the existence of a local database [228]. Historically, the first databases were developed for EI-MS because this technique results in reproducible spectral patterns on different mass spectrometers [128]. More recently, the use of HR instruments has significantly improved dereplication procedures, allowing the use of generic databases through molecular formula searches. Various strategies have been employed regarding detection using MS, HRMS and MS/MS [15,229–232].

4.1.1. HRMS

In most LC–HRMS applications, the metabolites are ionised using ESI, a soft ionisation technique. In comparison to EI which generates fragments, ESI has the great advantage of producing mostly molecular ion species that appear in the form of single or multiple adducts, such as $[M+H]^+$, $[M+Na]^+$, $[M+H+CH_3CN]^+$ (if acetonitrile is used as the solvent), and $[M+H-H_2O]^+$ in positive ion mode (PI) or $[M-H]^-$, $[M+HCO_2]^-$ and $[M-H+CO_2]^-$ in negative ion mode (NI) [233]. Furthermore, dimers, which complicate interpretation, may also be formed in the ion source (e.g., as in Fig. 5C). Therefore, prior to identification, the correct MW should be defined through adduct recognition [234]. Additionally, comparison of the different ionisation modes (PI or NI) may help to unambiguously show the molecular ions [235] (Fig. 5A and B). Once the molecular ion is carefully selected, the molecular formula (MF) is determined based on the accuracy of the mass and spectrum [232] (Fig. 6B and C). However, definitive MF determination remains a difficult task, even when mass accuracies <1 ppm are obtained [236]. This fact is particularly true for compounds with a high MW (>500 Da). To reduce the number of possible MFs, different heuristic filters can be applied [237], including: restrictions on the number of elements; LEWIS and SENIOR chemical rules; isotopic patterns; hydrogen/carbon ratios; the elemental ratios of nitrogen, oxygen, phosphorus and sulphur versus carbon; and elemental ratio probabilities (Fig. 6B). Based on the putative MFs,

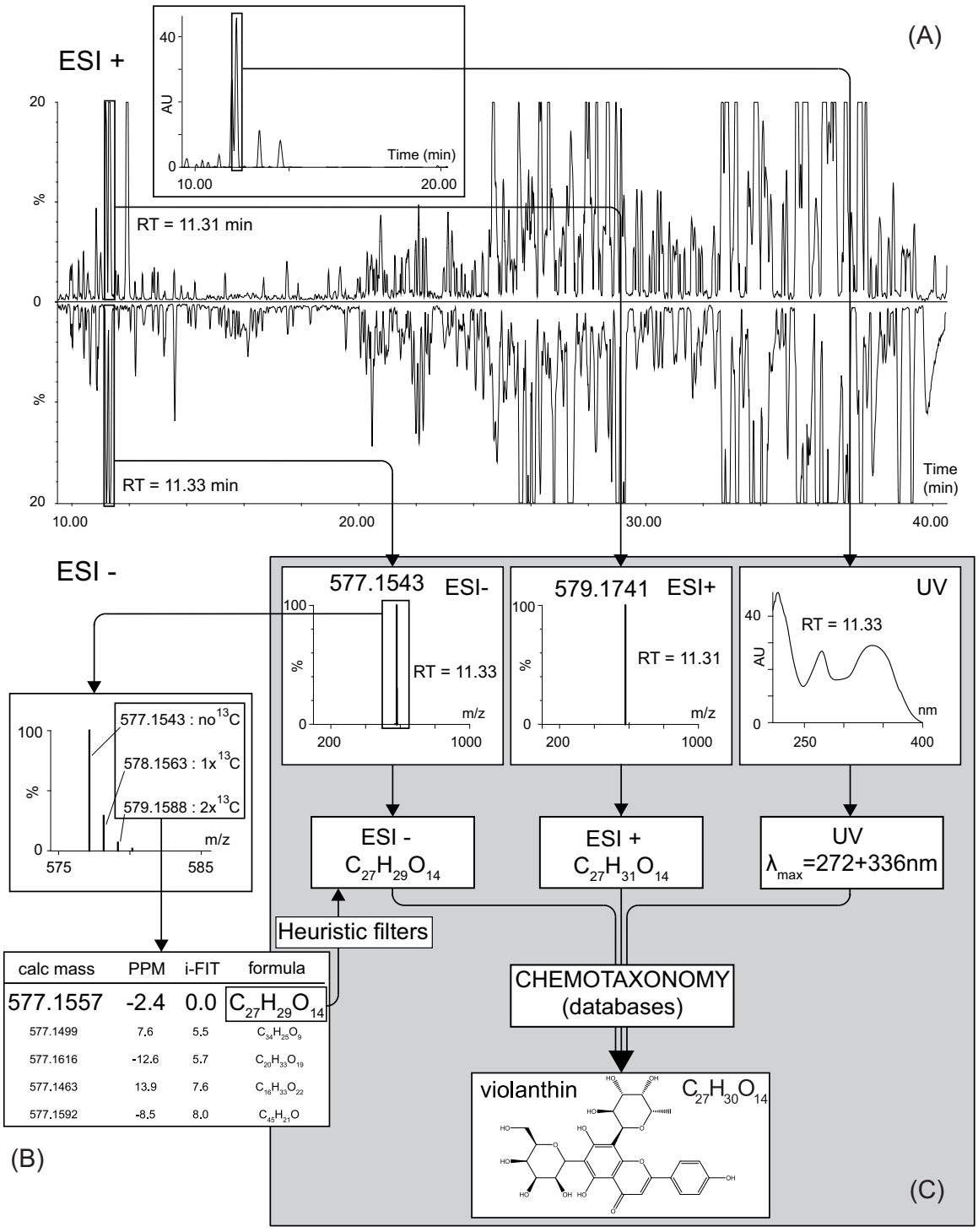


Fig. 6. Example of LC–MS peak annotation based on a high resolution *Viola tricolor* profiling on a C_{18} UHPLC column (150 mm × 2.1 mm; 1.7 μm) obtained with a slow gradient (5–95% ACN in 50 min). (A) PI (upper trace) and NI (lower trace) ESI-TOFMS BPI chromatograms and a UV trace (366 nm) is displayed in the inset. (B) Putative molecular formulas assignment based on the 15 ppm precision and isotopic pattern (iFIT) obtained from the NI ESI-TOFMS spectrum of the LC peak at RT 11.33 min. Application of heuristic filtering enable to ascertain the molecular formula assignment [237] (C) The LC peak at RT 11.33 min is annotated based on PI and NI molecular formula assignment and the UV PDA spectrum. Final structural assignment is based on a cross search with chemotaxonomic information which considerably reduces the number of possibilities. Such an approach for metabolite identification is still ambiguous (level of ID 2 according to MSI [4]) and assume that the metabolite has been previously characterised. Adapted from [41] with permission of The Royal Society of Chemistry.

the identification of metabolites can be achieved using various natural product databases (Table S1). However, this strategy typically leads to multiple putative identities for each compound detected.

Therefore, various strategies have been used to reduce, as much as possible, the number of possible structures that correspond to a given MF. Because UV-visible spectral data can be very informative, depending on the characteristic chromophores of a class of constituents, such data are often used to enhance dereplication [230,238,239]. Additionally, the phylogenetic information of the analysed organism also reduces the number of possible structures and improves the accuracy of peak annotation [240,241]. However, the success of dereplication is related to the presence

of the compounds in the databases. Therefore, the hit rate is substantially reduced for poorly studied organisms. When metabolic pathway data are available, selection based on pathway relationships may also improve annotation [242]. These workflows are highly time-consuming but may be substantially improved by process automation [235,243]. Marti et al. have evaluated a large number of metabolomics studies in the crop sciences and found that only a limited number of metabolites were identified [244].

An example of full annotation based on the described procedure is provided in Fig. 6 using the UHPLC-HRMS profile of an extract from *Viola tricolor*. The peak at a RT of 11.33 min displayed a molecular ion at *m/z* 577.1543 $[M-H]^-$ in NI mode and at *m/z* 579.1741 $[M+H]^+$ in PI mode (Fig. 6A). The presence of both molecular ions confirmed that the MW was 578 Da. In the 15 ppm tolerance range, the HR mass in NI mode allowed five possible MFs (Fig. 6B). Based on a heuristic filter, one valid MF was finally determined, C27H30O14. A database search [226] for this specific formula combined with the chemotaxonomic information of the *Viola* genus revealed that the only possibility was violanthin (Fig. 6C). This annotation was confirmed by the UV-PDA spectrum (Fig. 6C), in which both maxima were well matched with those of this flavonoid diglycoside.

The major drawback of this strategy is that the success of the annotation is directly related to the presence of the compounds in the databases. Therefore, various strategies based on compound-to-compound structural connections have been explored to further improve annotation based on direct HR-MS data from compounds not reported in any databases. These approaches consist of revealing structurally related compounds [245,246] that exhibit small changes, such as the addition of an alkyl/alkyloxy chain, hydroxy, or amino group or the reduction of a double bond. This can be achieved by obtaining accurate relative mass differences. These results can be confirmed by the changes observed in the relative RTs of the two compounds, which are related to structural modifications that affect the lipophilicity. However, this approach suffers from its targeted nature and is limited to specific structural modifications. Therefore, highly modified analogues may not be unambiguously annotated using this approach. To overcome such limitations, strategies based on mass defect filtering allow the identification of structurally related compounds because they possess similar mass defects (the difference between the accurate mass of a compound and its nominal mass) [247,248]. This approach considers the fact that substituents impart relatively low and defined effects on the mass defect of their core substructure. This approach has been successfully applied to the characterisation of 50 ophiopogonins and 27 ophiopogonones from *Ophiopogon japonicus* extracts [249]. In this case, all compounds differed only in the presence of hydroxy, methoxy, methyl, acetyl and sugar moieties. Complete elucidation of the substitution patterns was achieved through careful study of the MS^n spectra. These approaches enable annotation based on the unambiguous full de novo identification of related natural products present in an extract.

4.1.2. HR-MS/MS and $HRMS^n$

To improve annotation, additional information is required, including that provided by HR tandem MS (HR-MS/MS) and multistage MS ($HRMS^n$) spectra [30,231].

The fragmentation pattern obtained by HRMS has important advantages over nominal mass measurements. For example, the HR fragmentation pattern can be employed to reduce the number of MFs based on the identity of the fragment MFs [250,251]. This process is efficient for unambiguous MF attribution, although it is generally insufficient for identification of all compounds present in an extract [4].

In many cases, the fragment information generated by CID MS, MS/MS or MS^n can be simply interpreted from the acquired spectra. In plant extracts, in particular, many metabolites are widespread as glycosides. O- or C-glycosides, such as flavonoids, which are substituted with mono- or oligosaccharide moieties. In MS/MS, the fragmentation of such compounds leads to characteristic neutral losses, specifically, 162 amu for hexose, 146 amu for deoxyhexose, 132 amu for pentose and 176 amu for uronic acid moieties, when these glycosides are linked through an acetal bond to their aglycones. In contrast, the spectra of C-glycosides are dominated by cross-ring cleavages of the saccharide moiety (X ions); losses of 90, 120 and 150 amu are observed. The presence of such fragments is diagnostic. Monitoring the ion traces of a given aglycone versus its glycosides also enables localisation of the various glycosylated forms of a given metabolite in an extract. However, spectral interpretation is limited, and the substituent position cannot generally be ascertained (rules exist for specific types of flavonoids [252]), and the exact nature of the sugars cannot be determined. Even using less studied natural product skeletons, this approach can lead to the identification of large numbers of compounds [253]. Similar approach leads to the identification of series of compounds in *Penicilium expansum* extracts. Based on the fragmentation mechanisms of four known communesins, seven unknown compounds were characterised without further purification [254]. However, in the large majority of cases, the spectral interpretation for dereplication is very limited, and the process is time consuming.

Thus, the main strategy for automatic identification based on MS^n spectra consists of comparing the MS^2 (MS/MS) spectra of particular LC peaks to those present in a database however, this approach is instrument-specific and the database need to be built by the user based on his own set of available standards [230,239]. This approach improves confidence in the identification but suffers from a lack of reproducibility in fragmentation spectra taken on different MS platforms [232,255]. In addition, databases did not contain an exhaustive list of the secondary metabolites [256] (Table S1). To overcome this issue, approaches based on simulation of the various MS fragmentation spectra have been successfully applied. Simulation helps to discriminate between multiple potential structures. Approaches based on computational fragmentation in mass spectra have been successfully used to discriminate between candidate structures and were thus able to help identify the compounds under study [232]. With specific types of compounds, such as lipids, it is possible to simulate large libraries of fragmentation spectra, which can be used as databases [257].

Depending on the compound type, however, such predictions are not always so successful, and fragment information from a very significant proportion of the metabolites present in a natural extract must be collected for successful dereplication. Therefore, an intermediate approach consists of utilising the fragmentation mechanisms of known compounds in a given extract to simulate a larger MS/MS spectral library of analogous compounds. This approach can be automated when searching for specific compounds based on selected structural modifications. For example, Ridder et al. identified 25 novel green tea derivatives in urinary samples using traditional intestinal degradation and human biotransformation pathways for 75 known metabolites [258]. To improve this approach, the generation of large databases in silico based on known fragmentation mechanisms can be used to cover a wide chemical space (Table S1) [232,259], as in the case of lipids in the LipidBlast database [257]. For less studied fragmentation patterns, which is common with natural products, in silico fragmentation based on basic fragmentation rules or on combinatorial bond fragmentation can be used [232]. The latter approach was developed to generate simulated MS^n spectra; although the approach is not based on any reported fragmentation rules, it has proven to be highly successful [232,259].

When annotation is unsuccessful, the method of choice remains manual inspection of the MS^n spectra. This approach is usually successful for specific compound classes and is highly time

consuming. To improve this process, an untargeted MS^2 spectral similarity search has been applied to the identification of a large variety of natural products using similarity scores [224,260]. The systematic comparison of spectra allowed the generation of a metabolic network, which aided in the identification of structural analogues. Additionally, this approach enabled the identification of redundant substructures and was able to partially characterise specific metabolites. For example, a large series of desferrioxamines were identified in co-cultures in which *Streptomyces coelicolor* were combined with other bacteria [223]. This approach required the manual inspection of all similar MS^n spectra to identify the unknown compounds based on structural differences.

4.1.3. MS/MS acquisition modes for dereplication purposes

The emergence of hybrid MS, including LIT-Orbitrap Q-Exactive and Q-TOF systems, has promoted the development of automated data acquisition in the MS/MS mode to maximise throughput capabilities and the information obtained from a single analytical run [37,217,261]. Of the many approaches, the primary strategies are data-dependent acquisition (DDA) and data-independent acquisition (DIA) [262]. In DDA, MS/MS spectra are acquired if selected criteria (e.g., threshold, charge of the detected compound, and dynamic exclusion list) are met in a survey scan, which is generally based on a full scan to ensure that the process is non-discriminative [263]. For example, this approach has enabled the systematic selection and fragmentation of the base peak ion of the full scan spectrum for the acquisition of specific MS/MS spectra. The same approach can then be extended to the second and third most abundant ions. Although beneficial, this approach is limited by the duty cycle that is necessary to acquire the MS/MS dependent scan [264]. Because of this requirement, only a handful of the hundreds of molecular species contained in a full scan can be fragmented, thus compromising the MS/MS acquisition of less abundant analytes of interest. For example, in peptide digest analysis, <20% of peptides can be targeted for MS/MS [265]. Unlike DDA, DIA is not biased towards the detection of the most abundant ions in a full scan spectrum because it does not use a selection step prior to fragmentation, which results in the acquisition of all ions present at any time in the chromatographic separation in the MS/MS spectrum [266]. For this purpose, any m/z contained in the range of interest are fragmented either by simultaneously broadbanding all ions (entering the MS at a single chromatographic time point) or by multiplexing the full m/z range into smaller m/z isolation windows [262]. Regarding instrumental capabilities, MS manufacturers have developed different strategies, including "MS to the E" (MS^E, simultaneous acquisition of MS spectra at low and high collision energies), all-ion fragmentation (AIF), multiplexed MS/MS data-independent acquisition (MSX-DIA) and Sequential Window Acquisition of all THeoretical mass spectra (SWATH) [267]. For example, MS^E combines fast consecutive scanning at low and high collision energies which allows to correlate pseudo MS/MS spectra (high energy ESI-MS) to their precursor ions (low energy ESI-MS), through deconvolution in the time domain of all ions with overlapping elution profiles at a given RT. This mode has been successfully applied to plant metabolomics studies [268]. Recently, this approach was used to assess local and systemic herbivore-induced changes in maize leaves, sap, roots and root exudates without any prior assumptions regarding their function [269]. A systematic comparison of Q-TOF and Orbitrap-type instruments in MS^E and AIF modes, respectively, resulted in similar efficiencies for untargeted plant metabolomics studies [270]. As a multiplex DIA approach, both SWATH and MSX-DIA are based on the fragmentation of all ions entering the MS by stepping the fixed m/z window increments across the mass range of interest [271]. Although this approach results in a more complex dataset, multiple advantages have been achieved with these strategies, including comprehensive qualitative/quantitative enquiries of samples with high specificity and the possibility of retrospectively mining data because all m/z are fragmented within the LC time frame. To date, such acquisition strategies have primarily been applied to proteomics; however, their numerous benefits are expected to result in applications to comprehensive investigations of complex natural extracts in the near future [272].

4.1.4. Retention time prediction

Even when combined with chemotaxonomic information [240,241], the various LC–MS-based approaches often remain insufficient for unambiguous metabolite identification. This statement is especially true for secondary metabolites known to possess highly diverse structures that are species specific, whereas primary metabolites are more ubiquitous. Generally, only putative identification is achieved unless pure standards are available or MS/MS matches under similar conditions are possible, provided that databases exist. Therefore, additional information is necessary to improve metabolite annotation. Essentially, a comparison of the relative RT of analogous compounds can significantly aid with the discrimination between various possible structures. Because the retention time in RP chromatography is known to correlate with the log P [273], calculated log P values of specific analytes can provide valuable orthogonal information to assist in the identification of annotated metabolites [241,274]. Therefore, more generic approaches based on RT prediction and the physicochemical properties of compounds may be a valuable approach discriminating between multiple possible structures. This approach provides valuable results for analogous compounds with larger structural variations, such as small neutral compounds [275], phenolic compounds [276], peptides [277] and steroids [278]. Generic models, calculated using artificial neural networks with a large number of descriptors, have also been developed to aid in the dereplication of compounds present in human serum samples [279].

To improve RT prediction, models were built that depending on the phase chemistry used for compound elution, and these were limited to natural products. In the case of RP LC, models were constructed based on hydrogen bond acidity (α^H) and basicity (β^H), the dipolarity/polarisability parameter (π^*), the McGowan volume, the octanol/water partition coefficient (log P, calculated as the S + log P), the MW, the topological polar surface area and the number of rotatable bonds; these models allow for RT predictions with high accuracy (approximately 3.27 min for 90% of the tested metabolites in a 30 min gradient UHPLC analyses) [280]. Similar models were also developed for HILIC that were based on the calculated octanol/water partition coefficient (log D) at pH 3.5, negative charge at pH 3.5, positive charge at pH 3.5, number of rotatable bonds, number of phosphate groups, number of hydrogen bond donors and MW [281]. The actual RTs for 93% of the metabolite standards were within 35% of the predicted RTs.

4.2. Annotation standards

All of these approaches have resulted in significant reductions in the number of possible annotations for a given metabolite; however, completely unambiguous identifications using generic MS-based methods without access to instrument-dedicated MS/MS databases remain unattainable at present. Thus, the comprehensive data necessary for annotation remain under debate [4]. In a community effort to build a general consensus, the recommended reporting standards for compound identification have been described as follows [282]: "Authors should clearly differentiate and report the level of identification rigour for all metabolites reported based on a four-level system ranging from Level 1 (identified compound) via Levels 2 and 3 (putatively annotated compounds and compound classes) to Level 4 (unidentified or

unclassified metabolites which nevertheless can be differentiated based upon spectral data)" [4]. The systematic usage of such levels of identification is critical because reporting putative annotations might bias readers, who may over-interpret the data published in metabolomics studies by considering all peak annotations to be completely unambiguous. This is far from reality, especially for the profiling of secondary metabolites in natural extracts.

If doubt exists in compound annotation, the only solutions remaining are purification and de novo identification by NMR or synthesis of the putative structure of a given metabolite for further confirmation. Recently, LC–MS-targeted isolation protocols based on software-driven protocol optimisation were used to accelerate purification procedures and efficiently isolate compounds of interest for NMR identification [3,283].

4.3. Complementary on-line and at-line NMR for metabolite identification

NMR is a mandatory complement to MS-based dereplication strategies when an unknown metabolite must be identified or when dereplication remains ambiguous or speculative. NMR is applied to the analysis of new plant and microorganism species that have not been previously investigated.

For dereplication purposes, 1D proton NMR (^{1}H NMR) spectra are used to complement the MS data obtained by metabolite profiling; this approach is often sufficient for unambiguous identification. Such complementary data can be obtained using either on-line or at-line NMR.

Because ^{1}H NMR spectroscopy detects any hydrogen-containing compound, this universal mode of detection has an advantage over other methods, which are usually limited to a specific range of compounds with particular physicochemical properties. Another advantage of NMR spectroscopy is its capacity to quantify mixtures in a non-destructive manner using adapted protocols [164]. In this respect, NMR spectroscopy can be considered the ideal method for hyphenation with HPLC when used for the separation and identification of compounds in complex mixtures [284].

However, LC–NMR hyphenation remains challenging due to the very low intrinsic sensitivity of NMR relative to the small absolute quantity of raw samples that are usually injected in HPLC [285]. On-flow LC–NMR as a metabolite profiling technique is not extensively discussed here due to these sensitivity limitations. Thus, the use of LC–NMR, LC-SPE–NMR or other microflow NMR methods for the targeted identification of specific metabolites within HPLC metabolite profiles is often crucial for metabolite identification. The on-line and at-line HPLC approaches that are used for this purpose, either for 1D NMR or for comprehensive 2D NMR acquisition, are summarised below. The sensitivities currently attainable by different methods are also provided.

4.3.1. NMR sensitivity and probe designs

NMR probe design is probably the most important parameter with respect to S/N and overall performance. A major improvement in recent years is the reduction of coil diameter, which has led to a net increase in sensitivity. The direct consequence of this reduction is the smaller sample volume necessary, which has decreased from approximately 150 μL for 3 mm tubes to 40 μL for 1.7 mm tubes and to a few microlitres for capillary flow probe NMR (CapNMR^{TM}) [174]. The development of cryoprobes, in which the coils and preamplifier are chilled to 20 K with He, thus resulting in the elimination of electronic noise [286], is a major advancement [286]. The recent design of coils with high-temperature superconducting materials (HTSs), in contrast to conventional copper-based conductors, has also enhanced mass sensitivity. For example, a standard 5 mm room temperature probe provides good results with 600 μL of a 1 mM solution (e.g., MW = 500, 300 μg of sample). An S/N increase of up to four-fold should be obtained using a comparable cryogenic probe with the same volume. An HTS cryogenic probe will enhance the sensitivity by a factor of 5, thus being equivalent to a 5 mm room temperature probe with only 15 μg of sample [284]. Overall, the improvement in S/N ratio between conventional and dedicated sensitive probes can reach a factor of 20.

4.3.2. Direct LC–NMR coupling

The hyphenation of LC and NMR (LC–NMR) usually involves an HPLC–UV or –MS system coupled to an NMR instrument equipped with a flow probe. The flow cell of such a probe classically involves detection volumes ranging from 40 to 120 μL to obtain symmetric line shapes and sufficient sensitivity. Direct LC–NMR coupling is the simplest approach for recording ^{1}H NMR spectra directly from the LC flow. This on-flow mode involves the use of a LC column with a high loading capacity (>10 mm i.d. or >250 mm long) in combination with deuterated water and a common organic phase (typically CH_3CN in RP mode). Strong signals resulting from the LC solvents are usually suppressed through dedicated pulse sequences, such as water suppression enhanced through T1 effects [287]. A typical on-flow run results in a 2D plot, displaying NMR frequencies on one axis and chromatographic separation time on the other axis. This strategy is primarily used for the dereplication of complex mixtures. For example, anti-inflammatory furanocoumarins were fully dereplicated on-line from the crude extract of *Angelica dahurica* Fisch. ex Hoffm. (Apiaceae) using LC–MS-NMR [288]. An ethanolic extract (2.5 mg) was injected and separated in isocratic mode (D_2O/CH_3CN mixture) on a C_{30} column (250 mm × 4.6 mm, 3 μm) at a flow rate of 0.5 mL/min; the spectra were recorded on-flow using a cryo LC–NMR flow probe (60 μL flow cell). The same approach allowed for the on-line identification of various flavonoids from the roots of *Sophora flavescens* Ait. (Fabaceae) [289].

Higher sensitivities can be accomplished using LC–NMR in the stop-flow mode. This hyphenation setup involves an UV or MS detector-triggered valve to stop LC flow when peaks of interest are detected in order to park it in the centre of the LC–NMR probe. In this mode, improved line shape quality and efficient solvent suppression signals afford additional structural information for a given LC peak, and 2D spectra can be recorded for the most abundant compounds. This method enables the dereplication of the minor and major components of a complex extract, as exemplified by the work of Acevedo et al. on various *Vitis sp*. grapes. In this study, 33 anthocyanins were identified using a combination of MS/MS, LC–NMR and 2D NMR data. To obtain sufficient structural information for the minor components of the extract, a fraction collector was installed after the on-flow LC–NMR apparatus, thus enabling further NMR analyses on the fraction of interest. The major disadvantage of LC–NMR in the stop-flow mode is the low quality of the LC separation, which result from the necessary incorporation of multiple stop-flow events during elution and column overloading to obtain sufficient S/N of the detected peaks.

4.3.3. Indirect coupling through LC-SPE–NMR

To overcome the major drawbacks of direct LC–NMR hyphenation, indirect coupling through LC-SPE–NMR is a valuable strategy. In this configuration, a solid-phase extraction system is interfaced between sensitive detectors and the NMR flow probe. The LC peaks triggered by MS or UV are automatically trapped on SPE cartridges pre-conditioned with a weak solvent (usually water) using a divert valve. Then, the SPE is dried under a nitrogen flow, and the retained compound is eluted with an adequate mixture and volume of deuterated solvents for direct transfer to the NMR flow cell. This method presents several advantages compared to direct hyphenation, including multiple trappings of the same peak, which results from repeated injections of the same mixture and thus avoids column overloading while promoting optimal LC

separation. Another advantage of this approach is the use of deuterated solvents for SPE elution alone, which limits the use of solvent signal suppression sequences. The sensitivity improvement achieved by LC-SPE–NMR is suitable for the de novo compound identification of complex extracts, and 2D NMR spectra can be efficiently recorded. For example, 6 novel inositol derivatives were purified from the roots of *Taraxacum officinale* (Asteraceae) and characterised using LC-SPE–NMR [290]. The resolved peaks of interest were triggered using a DAD detector on polydivinylbenzene resin-filled SPE cartridges, followed by drying with nitrogen gas. The compounds were then eluted with CD_3CN into NMR tubes, enabling the acquisition of 1D and 2D NMR spectra.

Fully integrated systems allow for the automation of all steps from separation and targeted MS trapping to SPE multiple trapping enrichment procedure to the NMR spectra acquisition of the isolated LC peaks. Such an approach was used for the de novo characterisation of the metabolites of *Carthamus oxyacantha* M. Bieb. (Asteraceae) [291]. A fully automated HPLC-PDA-HRMS-SPE–NMR strategy focused on novel compounds and allowed the identification of 15 metabolites, including 4 novel spiro compounds in only a few weeks. This approach was compared with a classical isolation procedure that used vacuum liquid chromatography (VLC) and semi-preparative HPLC, which required nearly 2 months, consumed >20 L of organic solvents and afforded primarily known metabolites and only two novel spiro compounds.

Schoonen et al. recently developed a novel strategy that involves the substitution of SPE cartridges with a solvent exchange module [292]. The peaks of interest were first concentrated by evaporating 90% of the LC solvent before dilution in a deuterated solvent for NMR analysis. The entire evaporation process was monitored using the machine's vision software platform to maintain a stable droplet volume. Overall, recoveries of >90% were obtained for volatile, polar, non-polar and thermosensitive compounds.

4.3.4. Microfractionation and microNMR

Another alternative for the indirect coupling of LC to NMR is the collection of all fractions triggered by either a sensitive detector or at a constant time interval into vials or deep 96-well plates. Next, the fractions are dried and dissolved in appropriate volumes of deuterated solvents, followed by filtering and loading into tubes for a top-loading probe or direct injection into a flow probe. This at-line strategy has the advantage of bypassing the SPE development steps and can easily be coupled to post-chromatographic biological assays to rapidly focus on the bioactive fractions (see Section 6). For example, an efficient dereplication strategy for the identification of compounds with anti-fungal activity against *Candida albicans* was developed based on semi-preparative HPLC separation followed by a 96-well microdilution assay [293]. Crude extracts (50 mg) of *Alphitonia zizyphoides* (Spreng.) Gray. (Rhamnaceae) were sufficient to conduct biological assays using 40% of the amount in the wells, while the remaining fractions were used for the identification of active components by UHPLC–TOFMS and microflow NMR (CapNMR with an active sample volume of 1.5 μL). This strategy allowed the dereplication of betulinic acid, a triterpene known to possess anti-*Candida* activity.

An alternative for obtaining the same sensitivity as microflow NMR is the use of microtubes ranging from 3 to 1 mm in size, thus avoiding the possibility of sample precipitation and clogging of the microflow probe. HPLC time-dependant separation followed by activity profiling and subsequent dereplication by micro-NMR was used to investigate the anti-plasmodial activity of *Artemisia persica* Boiss. (Asteraceae) [294]. First, 350 μg of extract was separated for activity profiling and dereplication by LC–MS^n, and at-line microNMR analysis revealed the array of unknown compounds in the active fractions. Targeted purification by medium pressure liquid chromatography (MPLC) and semi-preparative HPLC was used to identify five new bisabololoxide sesquiterpene diesters with anti-plasmodial activities in the μM range.

Cryoprobe techniques for 1.7 mm tubes have recently been introduced. The concomitant use of small volumes with electronic noise reduction from cryogenation provides important benefits for sensitivity. With such probes, the ^{1}H NMR spectrum of low MW indole alkaloids can be acquired from <2 μg within approximately 1 min, yielding very good S/N ratios, and sub-μg quantities can be analysed with longer acquisition times. The corresponding HSQC was obtained in 1 h with 10 μg [175]. These types of probes have recently been used for the analysis of triterpenoids from *Ganoderma lucidum* (Ganodermataceae) extracts [295].

This result demonstrates that ^{1}H NMR spectra can be obtained for specific HPLC peaks using a standard 4 mm i.d. column in a single analytical run with only a few tens of μg of crude extract injected. The sensitivity of NMR compared to MS is at least one order of magnitude less. However, the rapid and MS targeted micro-isolation [3,283] or SPE trapping [290,291] of given LC peaks for unambiguous identification is quite effective and can be critical for identifying metabolites in profiling studies.

4.3.5. Statistical spectroscopic tools

The NMR information of a single component can also be extracted from the NMR profiles of complex mixtures (see Section 3.1) without micro-isolation strategies. In the past decade, the development of covariance NMR to identify correlations between spectroscopic signals recorded from 2D NMR experiments has simplified the chemical characterisation of compounds in mixtures [166,296]. This strategy was extended by Nicholson et al., who developed correlations of variable intensities across multiple independent samples, thus providing a covariance matrix of the entire dataset under study (statistical total correlation spectroscopy, STOCSY) [297]. The STOCSY approach generates a pseudo-2D NMR spectrum from multiple ^{1}H NMR experiments for each sample recorded, leading to improvements in resolution. This technique has been applied to metabolomics studies based on ^{1}H NMR. Biomarker identification and relationships between metabolite can be achieved with a combined strategy using MVDA to highlight proton NMR signals related to a given biological state complemented by signals-signals correlation revealed by STOCSY approach. By virtually enhancing the overall resolution, STOCSY is also applicable to LC–NMR experiments for signal deconvolution of co-eluting compounds [298]. This flexible technique is also appropriate for identifying the associations between NMR and MS data (statistical heterospectroscopy, SHY) acquired independently from the same samples or through hyphenation techniques, such as LC–MS-NMR [299]. For example, biomarkers of geographic origin from industrial citrus oils were identified using the SHY approach [300]. The citrus oils were first evaporated to remove highly volatile compounds before UHPLC–TOFMS profiling and ^{1}H NMR fingerprinting were performed on samples with Italian and Argentinean origins. An orthogonal projection on latent structure-discriminant analysis (OPLS-DA) model was set up from the LC–MS data to sort the most discriminating biomarkers of origin (Fig. 7A). Features at m/z 339.1599 at 2.63 min and m/z 153.1286 at 0.90 min were strongly involved in citrus oil origin separation, according to the OPLS-DA model. The covariance of the NMR and MS data matrices provided a NMR-MS map, which displayed correlations between a given m/z feature and several ^{1}H NMR chemical shifts (Fig. 7B). The molecular formula $C_{21}H_{22}O_4$ was calculated based on HRMS, the isotope pattern and heuristic filters from the LC–MS peak at m/z 339.1620. A cross-research investigation with a dictionary of natural products (DNP) using lemon and citrus as filters displayed two putative hits. The NMR-MS maps showed good correlations with the ^{1}H NMR signals at 6.26, 8.16, 6.96, 7.60 and 7.16 ppm, corresponding to the protons at positions 3, 4, 6, 7 and 8 of the furobenzopyranone

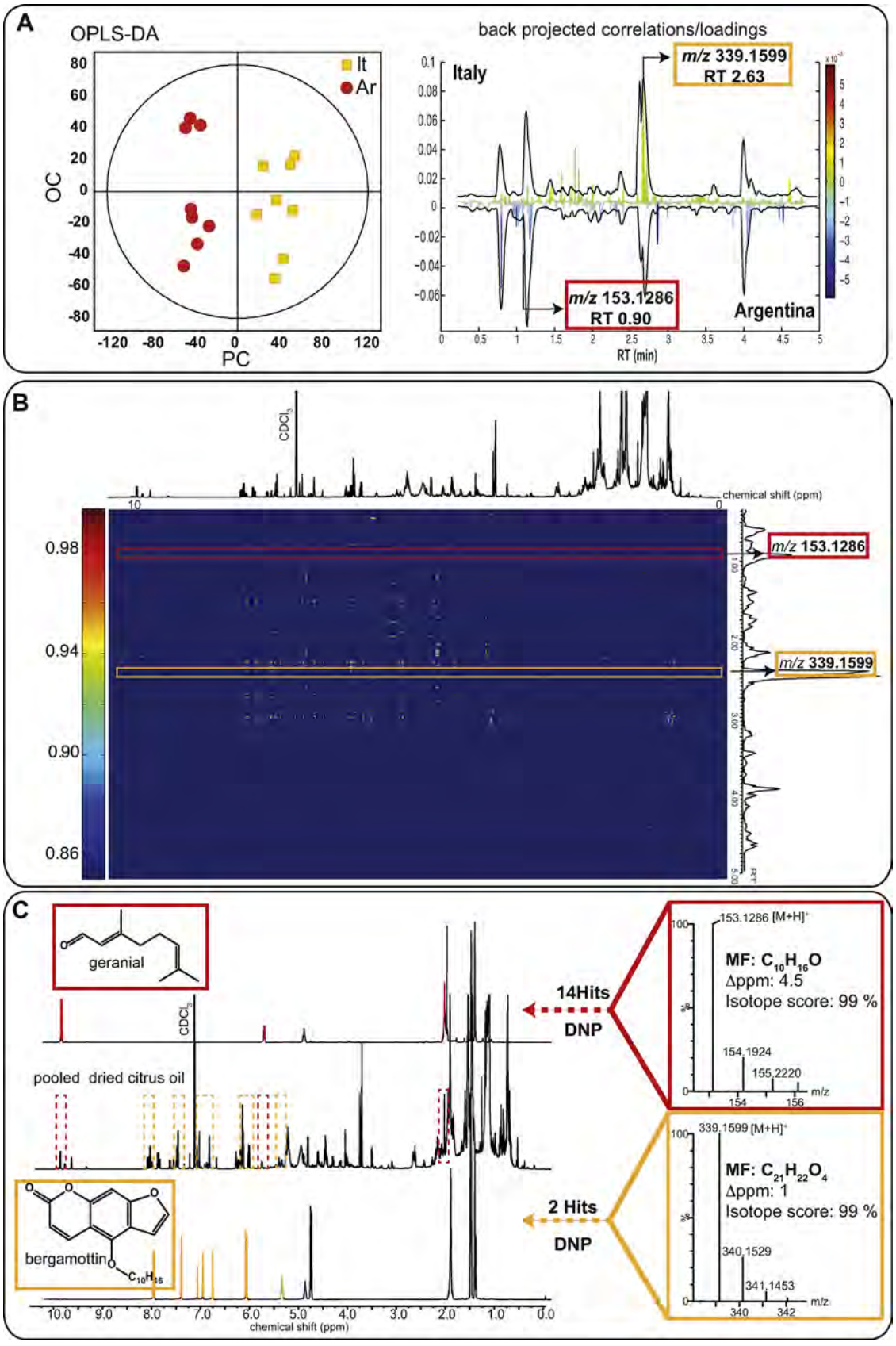


Fig. 7. NMR-MS dereplication strategy using covariance between sets of LC–MS and NMR metabolite profiling data to highlight NMR signals correspond to feature changes in the LC–MS analyses after MVDA. Metabolite profiling was performed on dried citrus oil from different origins. (A) OPLS-DA of LC–MS dataset from dried citrus oil of Italy and Argentina (right). The back projected correlation/loadings plot displays the LC–MS chromatogram for pooled extract from Italy (up) and Argentina (bottom) superimposed to the pareto scaled correlation of all detected features (left) according to the OPLS-DA model. The insets highlight two important discriminant loadings involved in origin separation. (B) NMR-MS correlation map (cutoff: 0.85). Top: ^{1}H NMR spectrum of pooled dried citrus oil in $CDCl_3$. Right: UHPLC–TOFMS chromatogram of the same extract. The red and orange rectangles highlights the correlation line of features at *m*/*z* 153.1286 and *m*/*z* 339.1599 respectively. (C) The squares to the left display the isotope pattern

moiety of berganottin (Fig. 7C bottom). Using the same approach, the MF of $C_{10}H_{16}O$ was deduced from the LC–MS peak at 0.90 min (m/z 153.1288), matching 14 hits in the DNP. The NMR-MS map revealed a high correlation between the LC–MS peak and the signals at 10.0, 5.6 and 2.4 ppm, allowing the unambiguous identification of geranial, which was confirmed by comparison with a pure standard (Fig. 7C top).

5. Multiple profiling platform strategies to enlarge metabolome coverage

Comprehensive coverage of the metabolome requires the combination of multiple orthogonal and complementary analysis platforms [129,130]. While natural extracts are non-exhaustively analysed using single profiling such as NMR, GC–MS, LC–MS, DIMS and CE–MS [124,129]; several workflows for metabolomics involving various combinations of these techniques have been described [130,301–306]. For example, GC– and LC–MS analyses based on full scan acquisition modes are highly complementary because each provides data for metabolites with different physicochemical properties, and generally, only a limited number of compounds are simultaneously detected by both platforms [130,304]. In serum plasma, for example, the unique combination of both GC–MS and UHPLC-HRMS has routinely enabled the simultaneous acquisition of >4500 metabolic features, and 3000 metabolite peaks that have been putatively annotated [130]. GC can efficiently separate volatiles and relatively polar metabolites of low MW analytes (<350 Da), including amino, organic and fatty acids and carbohydrates after derivatisation. In contrast, LC–MS detects higher MW compounds [307]. In a multiple platform strategy, genetically modified tomatoes overexpressing the taste-modifying protein miraculin were assessed using GC–, LC– and CE–TOFMS. This combination enabled the coverage of 86% of the chemical diversity listed in the LycoCyc database [303]. Schweiger et al. combined targeted GC–MS and UHPLC–TOFMS profiling to assess the specificity of the foliar metabolic responses of five plant species to symbiosis with arbuscular mycorrhizal fungus [302]. In addition, Tugizimana et al. have developed a multiple platform-based analysis that included GC-FID, GC–MS, GC × GC–TOFMS, UPLC–MS and ^{1}H NMR to assess the effect of ergosterols on the metabolome of cultured tobacco cells [305]. Their study revealed that external ergosterol stimuli triggered adaptive changes in the cell metabolome, notably through modifications in the biosynthesis of secondary metabolites. In another study, Mari et al. have established an LC–MS and GC–MS approach for comprehensively analysing and comparing a reference extract of *Potentilla anserina* with those obtained from various manufacturers [306]. The integrated approach allowed the characterisation of a large number of metabolites and the detection of the isoflavone genistein, which is considered as an active compound for oestrogen therapy.

In the frame of LC–MS, the use of various interfaces in both positive and negative ionisation modes also provides very complementary data sets. In the LC–MS profiling of body fluids, separations are routinely performed with a combination of both PI and NI ESI detections [308]. In addition, the use of different MS ionisation sources has also demonstrated that ESI and APCI ionisation techniques are highly complementary; ideally, both should be used to increase the number of metabolites detected for a given sample [56].

As shown, the metabolite profiling of a given organism or extract remains far from comprehensive. Multiple analytical approaches can be used to generate a high number of features, some of which are redundant between datasets, while others are unique. Software to mine all of the obtained data into meaningful data sets that are fully annotated is required to further obtain a comprehensive metabolome survey [309].

6. HPLC-based activity profiling

In natural product research, a primary objective is the discovery of novel bioactive compounds useful for drug discovery. To do so, the early identification of bioactive compounds in complex matrices is essential in order to isolate only the natural products of interest. This approach would be useful for avoiding the rediscovery of known compounds [30,310]. Various approaches can be employed to highlight the biologically active compounds in extracts. In addition to bioautography (bioassay directly on TLC plates) [311], which is not included here due to difficulties when applied to metabolite identification, such approaches are referred to as HPLC-based activity profiling [33,34]. Accordingly, LC peaks correlating to a specific bioactivity (e.g., enzyme inhibition, antifungal activity or in vivo activity in small model organisms) can be efficiently highlighted in the crude extract profiles, and these compounds can be fully or partially identified through metabolite profiling methods (see Section 4) [253,293]. The entire approach can be performed with only a few mg of crude extract, and a specially equipped platform can enable miniaturisation of the drug discovery process to the level of a few μg of active compounds. In addition to accelerating the bioactivity-guided isolation process at a large scale, this approach also allows limited amounts of a sample (e.g., plant herbarium samples, microorganisms cultured at small scale, etc.) to be assayed and can provide significant data regarding the bioactive natural products. Two different approaches are predominantly used, the direct on-line coupling of HPLC with bioassays and at-line post chromatographic bioactivity evaluation after microfractionation of an extract, which involves drying and re-dissolving the fractions in a suitable biocompatible media.

6.1. On-line biological evaluation

This approach is based on the direct biological evaluation of the eluted compound using on-line post-column reactions [33,312]. The generated biological profiles can be directly compared to the recorded chromatograms, and metabolite identification can also be performed on-line [313]. For enzymatic assays, the post-column setup generally consists of post-column addition of the enzyme, followed by a second addition of the enzymatic substrate, which allows the enzymatic activity to be monitored. Between the two post-column injections, the enzyme may react with the eluted compounds, which indicates the bioactive compounds [33]. For example, on-line antioxidant activity is rather straightforward to measure because only chemical reagents are involved. Radical scavengers in complex mixtures can be tracked by adding a coloured radical such as the purple DPPH$^\bullet$, after DAD, which is only reduced to non-coloured DPPH when radical scavengers are eluted from the column. Thus, a negative peak indicates radical scavenging properties [314,315]. Enzymatic activities can also be assessed on-line; this strategy was successfully applied for the identification

of each discriminant feature. The number of hits in Dictionary of Natural Products (DNP) [226] was reduced by analysis of the correlations between LC–MS and ^{1}H NMR. The dashed squares highlight the most important correlations observed for m/z 153.1286 (red) and m/z 339.1599 (orange) in the ^{1}H NMR spectra of the pooled citrus oils (middle). Comparison with ^{1}H NMR of pure standards of geranial (up) and bergamottin (bottom) achieve the identification with important MS-NMR correlated signals of each feature coloured in red or yellow respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
Adapted from [300] with permission.

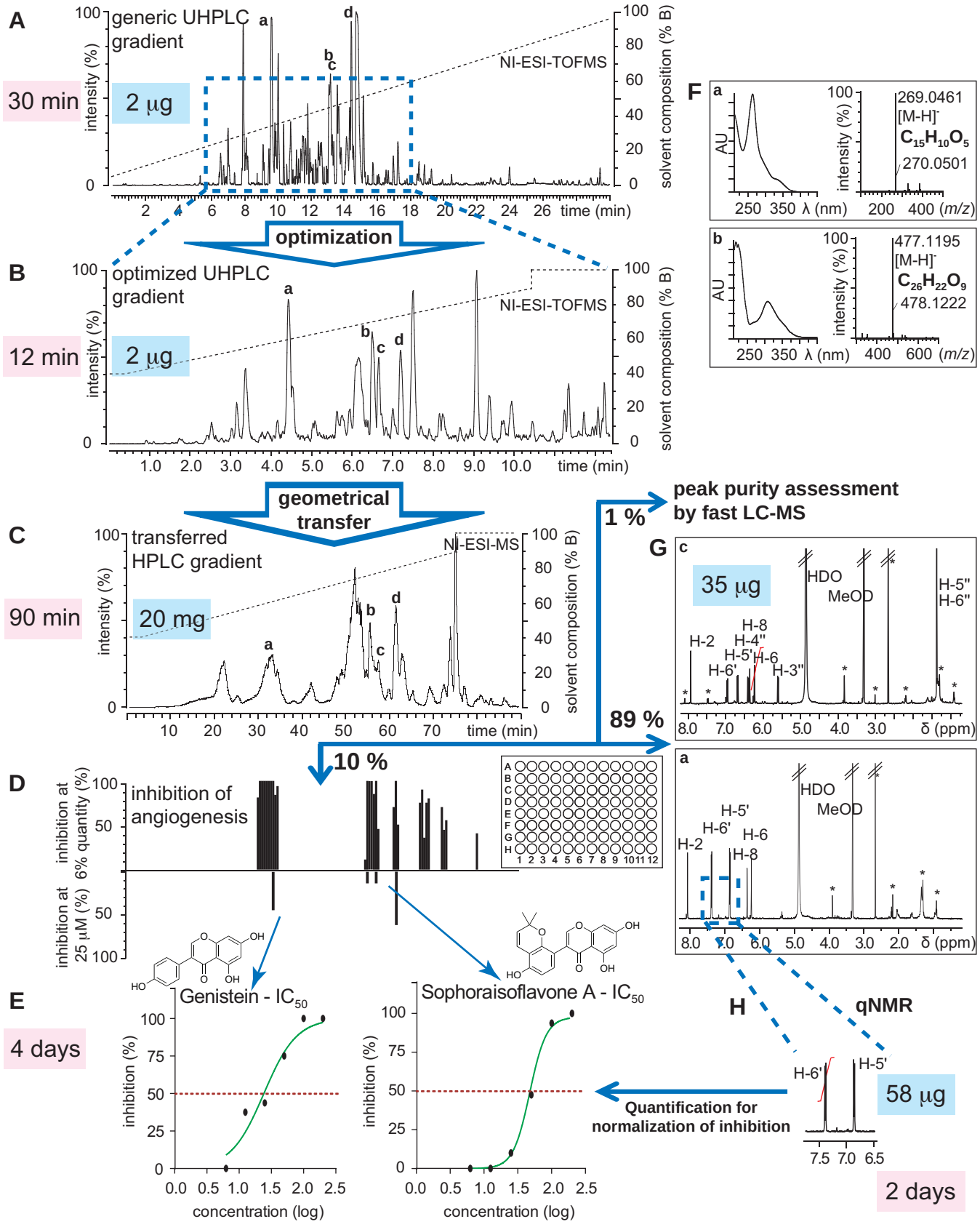


Fig. 8. HPLC based activity at line profiling of bioactive natural products with in vivo anti-inflammatory and anti-angiogenic assays in 96 well plates using Zebrafish larvae as a model. The generic procedure including the biological assay and dereplication part as well de novo identification by at-line microNMR and subsequent quantification for assessing IC50 value as the µg level is displayed. (A) Generic ultra-high pressure liquid chromatography–photo diode array–time of flight mass spectrometry (UHPLC-PDA–TOFMS) chromatogram; (B) Optimised UHPLC-PDA–TOFMS chromatogram for methanolic extract of *R. viscose*; (C) Semi-preparative high performance liquid chromatography (HPLC) chromatogram for the microfractionation of the enriched extract of *R. viscosa*. The chromatographic gradient was geometrically transferred using

of natural products with inhibitory activities (Table S2). In this case, simultaneously identifying metabolites while determining the cathepsin B inhibitors is possible using high temperature LC–MS with a post-column continuous-flow enzymatic assay [99]. This high temperature LC (208 °C) allows the methanol content of the mobile phase to be significantly reduced, thereby improving the compatibility of the solvent with enzymatic bioassays.

6.2. At-line biological evaluation

While efficient for the rapid identification of bioactive constituents, on-line setups are not compatible with all types of assays due to the compromise that must be made between the HPLC analytical conditions and assay requirements. Furthermore, on-line biological evaluation is not compatible with in vitro assays, including anti-microbial, anti-parasitic and cytotoxicity assays. Thus, at-line methods have been developed. Such approaches primarily consist of the post-column collection of microfractions, drying of the collected fraction for solvent removal to avoid solvent interference with the bioassay, and biological evaluation of the microfractions (Fig. 8) [33]. Finally, the biological profile, the activity measured for each microfraction, is compared to the HPLC profile to localise the LC peaks that are responsible for the activity. In these profiles, the extract amounts (few mg) are greater than those for typical analytical profiling; therefore, the approach provides the possibility for both MS [316] and analysis of the content of the microfraction using at-line micro-NMR methods for dereplication and de novo identification of the bioactive compounds [175,225]. The full process is exemplified in Fig. 8. Microfractionation is typically performed in deep 96-well plates, which are commonly used for bioassays; however, smaller fractions collected in 1536 wells can ameliorate the loss of resolution during the collection process [81]. Thus, even a relatively low rate of fractionation can effectively indicate the bioactive compounds [238].

At-line HPLC bioassays are relatively easy to implement and have been used for several applications, including assays that can be directly performed by dilution in wells or using TLC by bioautography after spotting the various microfractions collected. Examples of bioautography assays after HPLC microfractionation are shown in Table S2. Other assays have also been reported to be effective in dilutions, including simple chemical assays for antioxidant activity [317]. Similarly, enzymatic assays have also been extensively explored (Table S2). This robust approach is highly compatible with cell-based bioassays (Table S2). In vivo assays can also be performed directly at-line with miniaturised assays, particularly those involving zebrafish (Fig. 8) [163,318].

Due to the relatively low chromatographic resolution of the microfractionation, co-elution of compounds may occur in the wells. This problem can be solved by analysing the microfractions near the bioactive fraction, followed by profiling. Thus, correlation of the activity with the presence of signals in the active and non-active microfractions may highlight active metabolites [319,320]. Correlating the activity to metabolite profiles through statistical methods is referred to as biochemometrics [319]. This approach has been successfully applied for determining the 100 most active compounds of a highly complex extract of *Oplopanax horridus* [319]. In this study, fractionation was performed by countercurrent chromatography (CCC), and each fraction was analysed by GC–MS. However, similar results can also be obtained using HPLC microfractionation.

7. Conclusions

As demonstrated in this review, based on significant advances in analytical techniques, complex matrices, such as crude natural extracts, can be profiled effectively. The tremendous developments in HRMS, as well as data mining tools, particularly over the last 10 years, have been crucial for enhancing the throughput and the quality of data that can be acquired for peak annotation and dereplication.

Currently, MS not only allows chemical information on the content of extracts to be profiled by either LC–MS or DIMS approaches, but using imaging methods, intact tissues and solid culture media can also be analysed to provide spatial information for the location of given metabolites, which can aid in interpreting data from a biological viewpoint. With the development of instruments that provide increasingly higher resolution and higher acquisition frequencies, different MS^n experiments and full scan spectra can be acquired. The option of acquiring HRMS and HRMS/MS spectra in an untargeted manner in a single analytical run greatly aids in helping peak annotation for metabolite profiling. In addition to this important development in MS, the separation sciences have also evolved extensively. In particular, the introduction of sub-2 μm particle technology for both HPLC and SFC has significantly improved the resolution and throughput of chromatographic separation processes. Using these methods, very high peak capacity values can be obtained with gradient separations in HPLC, and the enhanced resolution of all constituents improves their detection and the quality of the spectra obtained. In addition, the separation of isomers that are likely to occur in natural extract matrices is possible. As a result of the orthogonality delivered in the second dimension, the development of comprehensive chromatography methods, including LC × LC and LC × IMS, shows great potential in deciphering the complex composition of extracts, even though such methods are still not used frequently in natural product research.

Together, these MS-based methods provide very detailed profiles of both the primary and minor constituents in extracts, as the response of MS is compound dependant. Thus, in non-targeted approaches, deducing quantitative information from the profiles recorded is not currently possible. Thus, NMR, which intrinsically provides a quantitative response, is an efficient and simple orthogonal method for profiling extracts. NMR offers an unbiased view of the main metabolites of an extract and their actual concentrations. While the method is intrinsically insensitive, NMR remains essential for determining the major constituents, particularly constituents that are not easily ionised by MS. The metabolite profiles recorded are either not influenced or are only slightly influenced by the type of instrument used for acquisition; thus, the technique can be universally utilised for comparison between laboratories. With continued improvements in deconvolution software, the sensitivity and the high reproducibility of NMR methods will remain critical for metabolomics profiling as a mandatory complement to MS.

mathematical models to obtain a comparable elution of extract constituents. Fractions were collected every 30 s directly into 96-deepwell plates. The so generated microfractions were aliquoted for anti-angiogenic screening (10% aliquot A), for fast LC–MS analysis (1%, aliquot B), and for NMR analysis (89%, aliquot C); (D) Anti-angiogenic screen on 180 microfractions generated by microfractionation. Positive bars show inhibition of angiogenesis of microfractions tested at high concentration; negative bars show inhibition of angiogenesis of selected microfractions at 25 mM. The concentration was determined by quantitative NMR (qNMR) (H); (E) Determination of IC50 using the quantitative information obtained by qNMR (H); (F) On-line PDA and high-resolution MS information from (A) for the dereplication of plant constituents; (G) ^{1}H NMR spectra using the CapNMR™ probe for structure confirmation of bioactive constituents; (H) Integration of well resolved aromatic protons for quantification of bioactive constituents to establish the potency of the anti-angiogenic and anti-inflammatory activity of the targeted compounds genistein sophoraisoflavone A.

Adapted from [163] with permission.

An important bottleneck in the metabolite profiling process remains the unambiguous identification of metabolites based on MS data alone. Indeed, no central spectroscopic database for secondary metabolites that encompasses all natural products is currently available in the dictionary of natural products [226], which is a good compilation of reported compounds from natural sources (>200,000 entries). While dereplication based on MS/MS can be effective and robust on a given instrument [321], the need for generic methods that allow comparison of MS/MS patterns, HRMS data and molecular formula assignments, to confirm metabolite identification remains. Often, this process is performed manually, and simple questions, such as unambiguously determining the MW of an unknown from the adduct ions or (de)protonated molecules recorded in both PI and NI modes can be difficult; however, software solutions are starting to become available [235]. Recent developments in this field, such as MS^n spectral trees [322] and the use of neural networks [323] to assess the similarities between MS/MS, are of great interest for future development. Orthogonal approaches, such as RT prediction in RP HPLC and/or HILIC, are also necessary. Contrarily to LC–MS, GC–MS are very reproducible and both RT indexes and MS spectra acquired in the EI mode significantly contribute to unambiguous peak annotation. This allows the exchange of databases between instruments. Because GC–MS is a classic and reproducible profiling method, it is gaining increasing interest in metabolomics, especially for profiling primary metabolites after derivatisation [324]. In addition, open access databases are available for profiling [325].

When analysing unknown compounds or confirming dereplication results, NMR remains the technique of choice for structural identification. The latest developments with respect to sensitivity, the introduction of cryogenated micro-NMR probes and automated LC-SPE-NMR–MS platforms, have provided methods for way to rapidly targeting the isolation of specific compounds for de novo identification. With the best technology, only a few μg are required for 2D NMR, and ^{1}H NMR spectra can be recorded in the sub-μg range. This has significantly facilitated the isolation process, which is otherwise complex and time-consuming for natural product research. In this respect, the use of chromatography software to optimise separation and facilitate gradients on the LC–MS semi-preparative scale renders the isolation process rapid, efficient and rational. The development of NMR databases combined with prediction software has greatly improved the amount of structural information that can be extracted from ^{1}H NMR spectra alone and accelerated the dereplication process [167].

The microfractionation of crude natural extracts also introduces the possibility of directly assessing the bioactivities of LC-peaks within their biological matrices. With the miniaturisation of enzymatic, cell and in vivo assays (e.g., zebrafish) in 96-well plates, HPLC-based biological activity profiling has become available for many types of assays and is a complement to metabolite profiling, allowing bioactive compounds to easily be localised and identified partly. In particular, these methods should accelerate the discovery of compounds from natural extracts and assist in rationalisation of the classical bioactivity-guided approach, which was regarded as too slow and complex for the pharmaceutical and agricultural industries [326].

To render the dereplication of secondary metabolites more efficient and unambiguous, the construction of an open database that merges taxonomic and chemical information with spectroscopic and chromatographic data is paramount. With advances in expert systems, spectral simulation tools and multivariate methods, taking advantage of these data should be possible, even if they are obtained using different analytical platforms. Natural product chemists that are trained to identify secondary metabolites should be aware of the need for rapid and unambiguous annotation in the metabolomics community. With these tools available and with scientist have common platforms [327] for collecting data [328], advances in this field will continue to improve the accuracy and effectiveness of metabolite profiling. With respect to the application of generic methods to complex natural matrices, the possibility of simply clicking a button to achieve annotation and a solution, as seen on "CSI Miami" for peak annotation in forensics, remains far from a reality. However, researchers have the tools that are required to successfully investigate nature's chemodiversity and elucidate critical biological issues.

Acknowledgments

JLW is thankful to the Swiss National Science Foundation for supporting the activity concerning metabolite profiling through grants CRSII3_127187, 205320/135 190-1, CR23I3_143733/1 and 200020_146200. We gratefully thank Adlin Afzan, Davy Guillarme and Mariam Mnatsakanyan for revising the manuscript and Yann Guitton for its help to find the natural products databases.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at http://dx.doi.org/10.1016/j.chroma.2014.10.091.

References

[1] J. van der Greef, H. van Wietmarschen, B. van Ommen, E. Verheij, Looking back into the future: 30 years of metabolomics at TNO, Mass Spectrom. Rev. 32 (2013) 399–415.
[2] E.C. Horning, M.G. Horning, Human metabolic profiles obtained by GC and GC/MS, J. Chromatogr. Sci. 9 (1971) 129–140.
[3] G. Glauser, D. Guillarme, E. Grata, J. Boccard, A. Thiocone, P.-A. Carrupt, J.-L. Veuthey, S. Rudaz, J.-L. Wolfender, Optimized liquid chromatography–mass spectrometry approach for the isolation of minor stress biomarkers in plant extracts and their identification by capillary nuclear magnetic resonance, J. Chromatogr. A 1180 (2008) 90–98.
[4] D.J. Creek, W.B. Dunn, O. Fiehn, J.L. Griffin, R.D. Hall, Z. Lei, R. Mistrik, S. Neumann, E.L. Schymanski, L.W. Sumner, R. Trengove, J.-L. Wolfender, Metabolite identification: are you sure? And how do your peers gauge your confidence? Metabolomics 10 (2014) 350–353.
[5] J. Larsson, J. Gottfries, L. Bohlin, A. Backlund, Expanding the ChemGPS chemical space with natural products, J. Nat. Prod. 68 (2005) 985–991.
[6] R.A. Dixon, D. Strack, Phytochemistry meets genome analysis, and beyond, Phytochemistry 62 (2003) 815–816.
[7] W. Schwab, Metabolome diversity: too few genes, too many metabolites? Phytochemistry 62 (2003) 837–849.
[8] E. Grotewold, Plant metabolic diversity: a regulatory perspective, Trends Plant Sci. 10 (2005) 57–62.
[9] D.G.I. Kingston, Modern natural products drug discovery and its relevance to biodiversity conservation, J. Nat. Prod. 74 (2010) 496–511.
[10] K. Hostettmann, J.-L. Wolfender, C. Terreaux, Modern screening techniques for plant extracts, Pharm. Biol. 39 (2001) 18–32.
[11] F.E. Koehn, G.T. Carter, The evolving role of natural products in drug discovery, Nat. Rev. Drug Discov. 4 (2005) 206–220.
[12] C. Guy, J. Kopka, T. Moritz, Plant metabolomics coming of age, Physiol. Plant 132 (2008) 113–116.
[13] L.W. Sumner, P. Mendes, R.A. Dixon, Plant metabolomics: large-scale phytochemistry in the functional genomics era, Phytochemistry 62 (2003) 817–836.
[14] Y.H. Choi, R. Verpoorte, Metabolomics what you see is what you extract, Phytochem. Anal. 25 (2014) 289–290.
[15] J.-L. Wolfender, S. Rudaz, Y. Hae Choi, H. Kyong Kim, Plant metabolomics: from holistic data to relevant biomarkers, Curr. Med. Chem. 20 (2013) 1056–1090.
[16] O. Fiehn, Combining genomics, metabolome analysis, and biochemical modelling to understand metabolic networks, Comp. Funct. Genome 2 (2001) 155–168.
[17] J.L. Wolfender, G. Glauser, J. Boccard, S. Rudaz, MS-based plant metabolomic approaches for biomarker discovery, Nat. Prod. Commun. 4 (2009) 1417–1430.
[18] K.H. Liland, Multivariate methods in metabolomics-from pre-processing to dimension reduction and statistical analysis, Trends Anal. Chem. 30 (2011) 827–841.
[19] E. Saccenti, H.C.J. Hoefsloot, A.K. Smilde, J.A. Westerhuis, M.M.W.B. Hendriks, Reflections on univariate and multivariate analysis of metabolomics data, Metabolomics 10 (2014) 361–374.
[20] J.C. D'Auria, J. Gershenzon, The secondary metabolism of *Arabidopsis thaliana*: growing like a weed, Curr. Opin. Plant Biol. 8 (2005) 308–316.

[21] R.A. Dixon, Natural products and plant disease resistance, Nature 411 (2001) 843–847.
[22] C.M. Dobson, Chemical space and biology, Nature 432 (2004) 824–828.
[23] X.M. Liang, Y. Jin, Y.P. Wang, G.W. Jin, Q. Fu, Y.S. Xiao, Qualitative and quantitative analysis in quality control of traditional Chinese medicines, J. Chromatogr. A 1216 (2009) 2033–2044.
[24] J.-L. Wolfender, G. Marti, E.F. Queiroz, Advances in techniques for profiling crude extracts and for the rapid identification of natural products: dereplication, quality control and metabolomics, Curr. Org. Chem. 14 (2010) 1808–1832.
[25] R.I. Boysen, M.T.W. Hearn, 9.02 – High performance liquid chromatographic separation methods, in: M. Lew, L. Hung-Wen (Eds.), Comprehensive Natural Products II, Elsevier, Oxford, 2010, pp. 5–49.
[26] D. Steinmann, M. Ganzera, Recent advances on HPLC/MS in medicinal plant analysis, J. Pharm. Biomed. Anal. 55 (2011) 744–757.
[27] P.J. Eugster, D. Guillarme, S. Rudaz, J.L. Veuthey, P.A. Carrupt, J.L. Wolfender, Ultra high pressure liquid chromatography for crude plant extract profiling, J. AOAC Int. 94 (2011) 51–70.
[28] C. Bicchi, C. Cagliero, P. Rubiolo, New trends in the analysis of the volatile fraction of matrices of vegetable origin: a short overview. A review, Flavour Frag. J. 26 (2011) 321–325.
[29] P. Rubiolo, B. Sgorbini, E. Liberto, C. Cordero, C. Bicchi, Essential oils and volatiles: sample preparation and analysis. A review, Flavour Frag. J. 25 (2010) 282–290.
[30] D. Krug, R. Muller, Secondary metabolomics: the impact of mass spectrometry-based approaches on the discovery and characterization of microbial natural products, Nat. Prod. Rep. 31 (2014) 768–783.
[31] M. Ernst, D.B. Silva, R.R. Silva, R.Z.N. Vencio, N.P. Lopes, Mass spectrometry in plant metabolomics strategies: from analytical platforms to data acquisition and processing, Nat. Prod. Rep. 31 (2014) 784–806.
[32] R.C. Breton, W.F. Reynolds, Using NMR to identify and characterize natural products, Nat. Prod. Rep. 30 (2013) 501–524.
[33] O. Potterat, M. Hamburger, Concepts and technologies for tracking bioactive compounds in natural product extracts: generation of libraries and hyphenation of analytical processes with bioassays, Nat. Prod. Rep. 30 (2013) 546–564.
[34] O. Potterat, M. Hamburger, Combined use of extract libraries and HPLC-based activity profiling for lead discovery: potential challenges and practical considerations, Planta Med. 80 (2014) 1171–1181.
[35] D.G.I. Kingston, High performance liquid chromatography of natural products, J. Nat. Prod. 42 (1979) 237–260.
[36] T.K. Natishan, Recent developments of achiral HPLC methods in pharmaceuticals using various detection modes, J. Liq. Chromatogr. Relat. Technol. 27 (2005) 1237–1316.
[37] J.W. Allwood, R. Goodacre, An introduction to liquid chromatography–mass spectrometry instrumentation applied in plant metabolomic analyses, Phytochem. Anal. 21 (2010) 33–47.
[38] H. Wu, J. Guo, S. Chen, X. Liu, Y. Zhou, X. Zhang, X. Xu, Recent developments in qualitative and quantitative analysis of phytochemical constituents and their metabolites using liquid chromatography–mass spectrometry, J. Pharm. Biomed. Anal. 72 (2013) 267–291.
[39] Z. Liu, S. Rochfort, Recent progress in polar metabolite quantification in plants using liquid chromatography–mass spectrometry, J. Integr. Plant Biol. 56 (2014) 816–825.
[40] K. Sandra, P. Sandra, Lipidomics from an analytical perspective, Curr. Opin. Chem. Biol. 17 (2013) 847–853.
[41] P.J. Eugster, J.-L. Wolfender, UHPLC in natural products analysis, in: D Guillarme, J.-L. Veuthey (Eds.), UHPLC in Life Sciences, Royal Society of Chemistry, Cambridge, UK, 2012, pp. 354–386.
[42] S. Wiklund, E. Johansson, L. Sjostrom, E.J. Mellerowicz, U. Edlund, J.P. Shockcor, J. Gottfries, T. Moritz, J. Trygg, Visualization of GC/TOF-MS-based metabolomics data for identification of biochemically interesting compounds using OPLS class models, Anal. Chem. 80 (2008) 115–122.
[43] G. Bonvin, J. Schappler, S. Rudaz, Capillary electrophoresis–electrospray ionization-mass spectrometry interfaces: fundamental concepts and technical developments, J. Chromatogr. A 1267 (2012) 17–31.
[44] R.M.S. Tubaon, H. Rabanes, P.R. Haddad, J.P. Quirino, Capillary electrophoresis of natural products: 2011–2012, Electrophoresis 35 (2014) 190–204.
[45] C. Ibáñez, C. Simó, V. García-Cañas, A. Cifuentes, M. Castro-Puyana, Metabolomics, peptidomics and proteomics applications of capillary electrophoresis–mass spectrometry in Foodomics: a review, Anal. Chim. Acta 802 (2013) 1–13.
[46] C.B. Hill, U. Roessner, Metabolic Profiling of Plants by GC–MS, The Handbook of Plant Metabolomics, Wiley-VCH Verlag GmbH & Co. KGaA, 2013, pp. 1–23.
[47] P.R. Griffiths, D.A. Heaps, P. Brejna, The gas chromatography/infrared interface: past, present, and future, Appl. Spectrosc. 62 (2008) 259A–270A.
[48] S. Fekete, A. Grand-Guillaume Perrenoud, D. Guillarme, Evolution and current trends in liquid and supercritical fluid chromatography, Curr. Chromatogr. 1 (2014) 15–40.
[49] J.H. Knox, Practical aspects of LC theory, J. Chromatogr. Sci. 15 (1977) 352–364.
[50] S. Bertrand, O. Schumpp, N. Bohni, A. Bujard, A. Azzollini, M. Monod, K. Gindro, J.-L. Wolfender, Detection of metabolite induction in fungal co-cultures on solid media by high-throughput differential ultra-high pressure liquid chromatography–time-of-flight mass spectrometry fingerprinting, J. Chromatogr. A 1292 (2013) 219–228.
[51] T. Ito, T. Odake, H. Katoh, Y. Yamaguchi, M. Aoki, High-throughput profiling of microbial extracts, J. Nat. Prod. 74 (2011) 983–988.
[52] M. Muller, S. Munne-Bosch, Rapid and sensitive hormonal profiling of complex plant samples by liquid chromatography coupled to electrospray ionization tandem mass spectrometry, Plant Methods 7 (2011) 37.
[53] D. Guillarme, C. Casetta, C. Bicchi, J.-L. Veuthey, High throughput qualitative analysis of polyphenols in tea samples by ultra-high pressure liquid chromatography coupled to UV and mass spectrometry detectors, J. Chromatogr. A 1217 (2010) 6882–6890.
[54] D. Guillarme, E. Grata, G. Glauser, J.-L. Wolfender, J.-L. Veuthey, S. Rudaz, Some solutions to obtain very efficient separations in isocratic and gradient modes using small particles size and ultra-high pressure, J. Chromatogr. A 1216 (2009) 3232–3243.
[55] P.J. Eugster, D. Biass, D. Guillarme, P. Favreau, R. Stöcklin, J.-L. Wolfender, Peak capacity optimisation for high resolution peptide profiling in complex mixtures by Liquid Chromatography coupled to Time-of-Flight Mass Spectrometry: application to the *Conus consors* cone snail venom, J. Chromatogr. A 1259 (2012) 187–199.
[56] G. Theodoridis, H.G. Gika, I.D. Wilson, Mass spectrometry-based holistic analytical approaches for metabolite profiling in systems biology studies, Mass Spectrom. Rev. 30 (2011) 884–906.
[57] D. Guillarme, J. Schappler, S. Rudaz, J.-L. Veuthey, Coupling ultra-high-pressure liquid chromatography with mass spectrometry, Trends Anal. Chem. 29 (2010) 15–27.
[58] E. Grata, J. Boccard, D. Guillarme, G. Glauser, P.-A. Carrupt, E.E. Farmer, J.-L. Wolfender, S. Rudaz, UPLC–TOF-MS for plant metabolomics: a sequential approach for wound marker analysis in *Arabidopsis thaliana*, J. Chromatogr. B 871 (2008) 261–270.
[59] E. Grata, D. Guillarme, G. Glauser, J. Boccard, P.-A. Carrupt, J.-L. Veuthey, S. Rudaz, J.-L. Wolfender, Metabolite profiling of plant extracts by ultra-high-pressure liquid chromatography at elevated temperature coupled to time-of-flight mass spectrometry, J. Chromatogr. A 1216 (2009) 5660–5668.
[60] T.L. Chester, Recent developments in high-performance liquid chromatography stationary phases, Anal. Chem. 85 (2012) 579–589.
[61] A. Marston, Role of advances in chromatographic techniques in phytochemistry, Phytochemistry 68 (2007) 2786–2798.
[62] K. Cabrera, Applications of silica-based monolithic HPLC columns, J. Sep. Sci. 27 (2004) 843–852.
[63] F. Chinnici, A. Gaiani, N. Natali, C. Riponi, S. Galassi, Improved HPLC determination of phenolic compounds in Cv. golden delicious apples using a monolithic column, J. Agric. Food Chem. 52 (2003) 3–7.
[64] P. Bhandari, N. Kumar, B. Singh, V. Singh, I. Kaur, Silica-based monolithic column with evaporative light scattering detector for HPLC analysis of bacosides and apigenin in *Bacopa monnieri*, J. Sep. Sci. 32 (2009) 2812–2818.
[65] K. Miyamoto, T. Hara, H. Kobayashi, H. Morisaka, D. Tokuda, K. Horie, K. Koduki, S. Makino, O. Núñez, C. Yang, T. Kawabe, T. Ikegami, H. Takubo, Y. Ishihama, N. Tanaka, High-efficiency liquid chromatographic separation utilizing long monolithic silica capillary columns, Anal. Chem. 80 (2008) 8741–8750.
[66] M.W. Dong, S. Fekete, D. Guillarme, Superficially porous particles: perspectives, practices, and trends, LC–GC N. Am. 32 (2014) 420–433.
[67] D. Guillarme, J. Ruta, S. Rudaz, J.L. Veuthey, New trends in fast and high-resolution liquid chromatography: a critical comparison of existing approaches, Anal. Bioanal. Chem. 397 (2010) 1069–1082.
[68] Y. Wang, Supercritical fluid chromatography of natural products, in: G.K. Webster (Ed.), Supercritical Fluid Chromatography: Advances and Applications in Pharmaceutical Analysis, CRC, Press, Boca Raton, FL, USA, 2014, p. 297.
[69] A. Grand-Guillaume Perrenoud, J.-L. Veuthey, D. Guillarme, Coupling state-of-the-art supercritical fluid chromatography and mass spectrometry: from hyphenation interface optimization to high-sensitivity analysis of pharmaceutical compounds, J. Chromatogr. A 1339 (2014) 174–184.
[70] M. Köhler, W. Haerdi, P. Christen, J.-L. Veuthey, Determination of artemisinin and artemisinic acid by capillary and packed supercritical fluid chromatography, J. High Resolut. Chromatogr. 20 (1997) 62–66.
[71] J. Thompson, B. Strode III, L.T. Taylor, T.A. van Beek, Supercritical fluid chromatography of ginkgolides A, B, C and J and bilobalide, J. Chromatogr. A 738 (1996) 115–122.
[72] J. Balsevich, L.R. Hogge, A.J. Berry, D.E. Games, I.C. Mylchreest, Analysis of indole alkaloids from leaves of *Catharanthus roseus* by means of supercritical fluid chromatography/mass spectrometry, J. Nat. Prod. 51 (1988) 1173–1177.
[73] T. Yamada, T. Uchikata, S. Sakamoto, Y. Yokoi, S. Nishiumi, M. Yoshida, E. Fukusaki, T. Bamba, Supercritical fluid chromatography/Orbitrap mass spectrometry based lipidomics platform coupled with automated lipid identification software for accurate lipid profiling, J. Chromatogr. A 1301 (2013) 237–242.
[74] A. Grand-Guillaume Perrenoud, J.-L. Veuthey, D. Guillarme, Comparison of ultra-high performance supercritical fluid chromatography and ultra-high performance liquid chromatography for the analysis of pharmaceutical compounds, J. Chromatogr. A 1266 (2012) 158–167.
[75] A. Grand-Guillaume Perrenoud, J. Boccard, J.-L. Veuthey, D. Guillarme, Analysis of basic compounds by supercritical fluid chromatography: attempts to improve peak shape and maintain mass spectrometry compatibility, J. Chromatogr. A 1262 (2012) 205–213.
[76] M. Wilm, M. Mann, Analytical properties of the nanoelectrospray ion source, Anal. Chem. 68 (1996) 1–8.

[77] J.R. Yates, C.I. Ruse, A. Nakorchevsky, Proteomics by mass spectrometry: approaches, advances, and applications, Annu. Rev. Biomed. Eng. 11 (2009) 49–79.
[78] P.A. Levkin, S. Eeltink, T.R. Stratton, R. Brennen, K. Robotti, H. Yin, K. Killeen, F. Svec, J.M. Frechet, Monolithic porous polymer stationary phases in polyimide chips for the fast high-performance liquid chromatography separation of proteins and peptides, J. Chromatogr. A 1200 (2008) 55–61.
[79] S. Fanali, E. Camera, B. Chankvetadze, G. D'Orazio, M.G. Quaglia, Separation of tocopherols by nano-liquid chromatography, J. Pharm. Biomed. Anal. 35 (2004) 331–337.
[80] G. Merola, Z. Aturki, G. D'Orazio, R. Gottardo, T. Macchia, F. Tagliaro, S. Fanali, Analysis of synthetic cannabinoids in herbal blends by means of nano-liquid chromatography, J. Pharm. Biomed. Anal. 71 (2012) 45–53.
[81] M. Giera, F. Heus, L. Janssen, J. Kool, H. Lingeman, H. Irth, Microfractionation revisited: a 1536 well high resolution screening assay, Anal. Chem. 81 (2009) 5460–5466.
[82] S. Fanali, Z. Aturki, G. D'Orazio, A. Rocco, A. Ferranti, L. Mercolini, M.A. Raggi, Analysis of Aloe-based phytotherapeutic products by using nano-LC–MS, J. Sep. Sci. 33 (2010) 2663–2670.
[83] R. Garcia-Villalba, A. Carrasco-Pancorbo, G. Zurek, M. Behrens, C. Bassmann, A. Segura-Carretero, A. Fernandez-Gutierrez, Nano and rapid resolution liquid chromatography–electrospray ionization-time of flight mass spectrometry to identify and quantify phenolic compounds in olive oil, J. Sep. Sci. 33 (2010) 2069–2078.
[84] X. Gao, Q. Zhang, D. Meng, G. Isaac, R. Zhao, T.L. Fillmore, R.K. Chu, J. Zhou, K. Tang, Z. Hu, R.J. Moore, R.D. Smith, M.G. Katze, T.O. Metz, A reversed-phase capillary ultra-performance liquid chromatography–mass spectrometry (UPLC–MS) method for comprehensive top-down/bottom-up lipid profiling, Anal. Bioanal. Chem. 402 (2012) 2923–2933.
[85] Q. Ni, K.R. Reid, C.F. Burant, R.T. Kennedy, Capillary LC–MS for high sensitivity metabolomic analysis of single islets of Langerhans, Anal. Chem. 80 (2008) 3539–3546.
[86] X. He, M. Kozak, S. Nimkar, Ultra-sensitive measurements of 11-Nor-delta(9)-tetrahydrocannabinol-9-carboxylic acid in oral fluid by microflow liquid chromatography–tandem mass spectrometry using a benchtop quadrupole/orbitrap mass spectrometer, Anal. Chem. 84 (2012) 7643–7647.
[87] R. Ramautar, G.J. de Jong, Recent developments in liquid-phase separation techniques for metabolomics, Bioanalysis 6 (2014) 1011–1026.
[88] T. El-Elimat, X. Zhang, D. Jarjoura, F.J. Moy, J. Orjala, A.D. Kinghorn, C.J. Pearce, N.H. Oberlies, Chemical diversity of metabolites from fungi, cyanobacteria, and plants relative to FDA-approved anticancer agents, ACS Med. Chem. Lett. 3 (2012) 645–649.
[89] M. Feher, J.M. Schmidt, Property distributions: differences between drugs, natural products, and molecules from combinatorial chemistry, J. Chem. Inf. Comput. Sci. 43 (2002) 218–227.
[90] M. Przybyciel, Novel phases for HPLC separations, LC–GC LC Column Technol. Suppl. 4 (2006) 49–52.
[91] H. Matsumoto, Y. Ikoma, M. Kato, T. Kuniga, N. Nakajima, T. Yoshida, Quantification of carotenoids in citrus fruit by LC–MS and comparison of patterns of seasonal changes for carotenoids among citrus varieties, J. Agric. Food Chem. 55 (2007) 2356–2368.
[92] J. Layne, Characterization and comparison of the chromatographic performance of conventional, polar-embedded, and polar-endcapped reversed-phase liquid chromatography stationary phases, J. Chromatogr. A 957 (2002) 149–164.
[93] T. Németh, E. Haghedooren, B. Noszál, J. Hoogmartens, E. Adams, Three methods to characterize reversed phase liquid chromatographic columns applied to pharmaceutical separations, J. Chemomet. 22 (2008) 178–185.
[94] E. Lesellier, C. West, A. Tchapla, Classification of special octadecyl-bonded phases by the carotenoid test, J. Chromatogr. A 1111 (2006) 62–70.
[95] L. Gu, M. Kelm, J.F. Hammerstone, G. Beecher, D. Cunningham, S. Vannozzi, R.L. Prior, Fractionation of polymeric procyanidins from lowbush blueberry and quantification of procyanidins in selected foods with an optimized normal-phase hpLC–MS fluorescent detection method, J. Agric. Food Chem. 50 (2002) 4852–4860.
[96] E. Giacomelli, S. Bertrand, A. Nievergelt, V. Zwick, C. Simoes-Pires, L. Marcourt, E. Rivara-Minten, M. Cuendet, A. Bisio, J.-L. Wolfender, Cancer chemopreventive diterpenes from *Salvia corrugata*, Phytochemistry 96 (2013) 257–264.
[97] O.L. Knittelfelder, B.P. Weberhofer, T.O. Eichmann, S.D. Kohlwein, G.N. Rechberger, A versatile ultra-high performance LC–MS method for lipid profiling, J. Chromatogr. B 951–952 (2014) 119–128.
[98] H. Chen, M. Angiuli, C. Ferrari, E. Tombari, G. Salvetti, E. Bramanti, Tocopherol speciation as first screening for the assessment of extra virgin olive oil quality by reversed-phase high-performance liquid chromatography/fluorescence detector, Food Chem. 125 (2011) 1423–1429.
[99] A.R. de Boer, J.M. Alcaide-Hidalgo, J.G. Krabbe, J. Kolkman, C.N. van Emde Boas, W.M.A. Niessen, H. Lingeman, H. Irth, High-temperature liquid chromatography coupled on-line to a continuous-flow biochemical screening assay with electrospray ionization mass spectrometric detection, Anal. Chem. 77 (2005) 7894–7900.
[100] M. Lísa, M. Holčapek, Triacylglycerols profiling in plant oils important in food industry dietetics and cosmetics using high-performance liquid chromatography–atmospheric pressure chemical ionization mass spectrometry, J. Chromatogr. A 1198–1199 (2008) 115–130.
[101] L. Zheng, R. T'Kind, S. Decuypere, S.J. von Freyend, G.H. Coombs, D.G. Watson, Profiling of lipids in *Leishmania donovani* using hydrophilic interaction chromatography in combination with Fourier transform mass spectrometry, Rapid Commun. Mass Spectrom. 24 (2010) 2074–2082.
[102] Q. Zhou, B. Gao, X. Zhang, Y. Xu, H. Shi, L. Yu, Chemical profiling of triacylglycerols and diacylglycerols in cow milk fat by ultra-performance convergence chromatography combined with a quadrupole time-of-flight mass spectrometry, Food Chem. 143 (2014) 199–204.
[103] M.A. Strege, Hydrophilic interaction chromatography–electrospray mass spectrometry analysis of polar compounds for natural product drug discovery, Anal. Chem. 70 (1998) 2439–2445.
[104] V.V. Tolstikov, O. Fiehn, Analysis of highly polar compounds of plant origin: combination of hydrophilic interaction chromatography and electrospray ion trap mass spectrometry, Anal. Biochem. 301 (2002) 298–307.
[105] A. Petruczynik, M. Waksmundzka-Hajnos, Application of hydrophilic interaction chromatography in phytochemical analysis, Acta Chromatogr. 25 (2013) 1–25.
[106] Z. Li, T.D. Sharkey, Metabolic profiling of the methylerythritol phosphate pathway reveals the source of post-illumination isoprene burst from leaves, Plant Cell Environ. 36 (2013) 429–437.
[107] L.R. Snyder, J.J. Kirkland, J.W. Dolan, Introduction to Modern Liquid Chromatography, Wiley, 2009.
[108] S. Boudah, M.-F. Olivier, S. Aros-Calt, L. Oliveira, F. Fenaille, J.-C. Tabet, C. Junot, Annotation of the human serum metabolome by coupling three liquid chromatography methods to high-resolution mass spectrometry, J. Chromatogr. B 966 (2014) 34–47.
[109] P. Horvatovich, B. Hoekman, N. Govorukhina, R. Bischoff, Multidimensional chromatography coupled to mass spectrometry in analysing complex proteomics samples, J. Sep. Sci. 33 (2010) 1421–1437.
[110] E.I. Chen, J.R. Yates 3rd, Cancer proteomics by quantitative shotgun proteomics, Mol. Oncol. 1 (2007) 144–159.
[111] H. Nie, R. Liu, Y. Yang, Y. Bai, Y. Guan, D. Qian, T. Wang, H. Liu, Lipid profiling of rat peritoneal surface layers by online normal- and reversed-phase 2D LC QToF-MS, J. Lipid Res. 51 (2010) 2833–2844.
[112] D.V. McCalley, Is hydrophilic interaction chromatography with silica columns a viable alternative to reversed-phase liquid chromatography for the analysis of ionisable compounds? J. Chromatogr. A 1171 (2007) 46–55.
[113] R. Tuytten, F. Lemiere, W. Van Dongen, E. Witters, E.L. Esmans, R.P. Newton, E. Dudley, Development of an on-line SPE-LC-ESI–MS method for urinary nucleosides: hyphenation of aprotic boronic acid chromatography with hydrophilic interaction LC-ESI–MS, Anal. Chem. 80 (2008) 1263–1271.
[114] P. Jandera, Stationary phases for hydrophilic interaction chromatography, their characterization and implementation into multidimensional chromatography concepts, J. Sep. Sci. 31 (2008) 1421–1437.
[115] K.M. Kalili, A. de Villiers, Systematic optimisation and evaluation of on-line, off-line and stop-flow comprehensive hydrophilic interaction chromatographyxreversed phase liquid chromatographic analysis of procyanidins, Part I: Theoretical considerations, J. Chromatogr. A 1289 (2013) 58–68.
[116] Y. Wang, R. Lehmann, X. Lu, X. Zhao, G. Xu, Novel, fully automatic hydrophilic interaction/reversed-phase column-switching high-performance liquid chromatographic system for the complementary analysis of polar and apolar compounds in complex samples, J. Chromatogr. A 1204 (2008) 28–34.
[117] A. Thomas, J. Deglon, T. Steimer, P. Mangin, Y. Daali, C. Staub, On-line desorption of dried blood spots coupled to hydrophilic interaction/reversed-phase LC/MS/MS system for the simultaneous analysis of drugs and their polar metabolites, J. Sep. Sci. 33 (2010) 873–879.
[118] T. Beelders, K.M. Kalili, E. Joubert, D. de Beer, A. de Villiers, Comprehensive two-dimensional liquid chromatographic analysis of rooibos (*Aspalathus linearis*) phenolics, J. Sep. Sci. 35 (2012) 1808–1820.
[119] D.Y. Bang, M.H. Moon, On-line two-dimensional capillary strong anion exchange/reversed phase liquid chromatography–tandem mass spectrometry for comprehensive lipid analysis, J. Chromatogr. A 1310 (2013) 82–90.
[120] T. Kato, M.C. Kasuya, K. Hatanaka, Rapid separation of gangliosides using strong anion exchanger cartridges, J. Oleo Sci. 57 (2008) 397–400.
[121] C.W.N. Damen, G. Isaac, J. Langridge, T. Hankemeier, R.J. Vreeken, Enhanced lipid isomer separation in human plasma using reversed-phase UPLC with ion-mobility/high-resolution MS detection, J. Lipid Res. 55 (2014) 1772–1783.
[122] G.B. Gonzales, K. Raes, S. Coelus, K. Struijs, G. Smagghe, J. Van Camp, Ultra(high)-pressure liquid chromatography–electrospray ionization-time-of-flight-ion mobility-high definition mass spectrometry for the rapid identification and structural characterization of flavonoid glycosides from cauliflower waste, J. Chromatogr. A 1323 (2014) 39–48.
[123] G. Purcaro, C. Cordero, E. Liberto, C. Bicchi, L.S. Conte, Toward a definition of blueprint of virgin olive oil by comprehensive two-dimensional gas chromatography, J. Chromatogr. A 1334 (2014) 101–111.
[124] K. Dettmer, P.A. Aronov, B.D. Hammock, Mass spectrometry-based metabolomics, Mass Spectrom. Rev. 26 (2007) 51–78.
[125] O. Fiehn, Metabolomics – the link between genotypes and phenotypes, Plant Mol. Biol. 48 (2002) 155–171.
[126] P. Begley, S. Francis-McIntyre, W.B. Dunn, D.I. Broadhurst, A. Halsall, A. Tseng, J. Knowles, H. Consortium, R. Goodacre, D.B. Kell, Development and performance of a gas chromatography–time-of-flight mass spectrometry analysis for large-scale nontargeted metabolomic studies of human serum, Anal. Chem. 81 (2009) 7038–7046.

[127] W.B. Dunn, Current trends and future requirements for the mass spectrometric investigation of microbial, mammalian and plant metabolomes, Phys. Biol. 5 (2008) 011001.
[128] T. Kind, G. Wohlgemuth, D.Y. Lee, Y. Lu, M. Palazoglu, S. Shahbaz, O. Fiehn, FiehnLib: mass spectral and retention index libraries for metabolomics based on quadrupole and time-of-flight gas chromatography/mass spectrometry, Anal. Chem. 81 (2009) 10038–10048.
[129] W.B. Dunn, N.J. Bailey, H.E. Johnson, Measuring the metabolome: current analytical technologies, Analyst 130 (2005) 606–625.
[130] W.B. Dunn, D. Broadhurst, P. Begley, E. Zelena, S. Francis-McIntyre, N. Anderson, M. Brown, J.D. Knowles, A. Halsall, J.N. Haselden, A.W. Nicholls, I.D. Wilson, D.B. Kell, C. Goodacre, Human Serum Metabolome (HUSERMET) Consortium, Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry, Nat. Protoc. 6 (2011) 1060–1083.
[131] S. Rodríguez-Sánchez, O. Hernández-Hernández, A.I. Ruiz-Matute, M.L. Sanz, A derivatization procedure for the simultaneous analysis of iminosugars and other low molecular weight carbohydrates by GC–MS in mulberry (*Morus* sp.), Food Chem. 126 (2011) 353–359.
[132] P. Ausloos, C.L. Clifton, S.G. Lias, A.I. Mikaya, S.E. Stein, D.V. Tchekhovskoi, O.D. Sparkman, V. Zaikin, D. Zhu, The critical evaluation of a comprehensive mass spectral library, J. Am. Soc. Mass Spectrom. 10 (1999) 287–299.
[133] F. Versace, F. Sporkert, P. Mangin, C. Staub, Rapid sample pre-treatment prior to GC–MS and GC–MS/MS urinary toxicological screening, Talanta 101 (2012) 299–306.
[134] T. Davies, The new automated mass spectrometry deconvolution and identification system (AMDIS), Spectrosc. Eur. 10 (1998) 24–27 http://chemdata.nist.gov/mass-spc/amdis/
[135] S.E. Stein, An integrated method for spectrum extraction and compound identification from gas chromatography/mass spectrometry data, J. Am. Soc. Mass Spectrom. 10 (1999) 770–781.
[136] R. Aggio, S.G. Villas-Boas, K. Ruggiero, Metab: an R package for high-throughput analysis of metabolomics data generated by GC–MS, Bioinformatics 27 (2011) 2316–2318.
[137] W.B. Dunn, D.I. Ellis, Metabolomics current analytical platforms and methodologies, Trends Anal. Chem. 24 (2005) 285–294.
[138] S.T. Chin, P.J. Marriott, Multidimensional gas chromatography beyond simple volatiles separation, Chem. Commun. 50 (2014) 8819–8833.
[139] Y. Nolvachai, P.J. Marriott, GC for flavonoids analysis: past, current, and prospective trends, J. Sep. Sci. 36 (2013) 20–36.
[140] A. Mostafa, M. Edwards, T. Gorecki, Optimization aspects of comprehensive two-dimensional gas chromatography, J. Chromatogr. A 1255 (2012) 38–55.
[141] K.M. Pierce, J.C. Hoggard, R.E. Mohler, R.E. Synovec, Recent advancements in comprehensive two-dimensional separations with chemometrics, J. Chromatogr. A 1184 (2008) 341–352.
[142] R.C. Ong, P.J. Marriott, A review of basic concepts in comprehensive two-dimensional gas chromatography, J. Chromatogr. Sci. 40 (2002) 276–291.
[143] X. Gao, S.J. Williams, O.L. Woodman, P.J. Marriott, Comprehensive two-dimensional gas chromatography, retention indices and time-of-flight mass spectra of flavonoids and chalcones, J. Chromatogr. A 1217 (2010) 8317–8326.
[144] L.W. Hantao, H.G. Aleme, M.M. Passador, E.L. Furtado, F.A. Ribeiro, R.J. Poppi, F. Augusto, Determination of disease biomarkers in *Eucalyptus* by comprehensive two-dimensional gas chromatography and multivariate data analysis, J. Chromatogr. A 1279 (2013) 86–91.
[145] C.N. McEwen, R.G. McKay, A combination atmospheric pressure LC/MS:GC/MS ion source: advantages of dual AP-LC/MS:GC/MS instrumentation, J. Am. Soc. Mass Spectrom. 16 (2005) 1730–1738.
[146] R. Schiewek, M. Lorenz, R. Giese, K. Brockmann, T. Benter, S. Gab, O.J. Schmitz, Development of a multipurpose ion source for LC–MS and GC–API MS, Anal. Bioanal. Chem. 392 (2008) 87–96.
[147] T. Portoles, L. Cherta, J. Beltran, F. Hernandez, Improved gas chromatography–tandem mass spectrometry determination of pesticide residues making use of atmospheric pressure chemical ionization, J. Chromatogr. A 1260 (2012) 183–192.
[148] C.J. Wachsmuth, M.F. Almstetter, M.C. Waldhier, M.A. Gruber, N. Nurnberger, P.J. Oefner, K. Dettmer, Performance evaluation of gas chromatography–atmospheric pressure chemical ionization-time-of-flight mass spectrometry for metabolic fingerprinting and profiling, Anal. Chem. 83 (2011) 7514–7522.
[149] E. Hurtado-Fernandez, T. Pacchiarotta, E. Longueira-Suarez, O.A. Mayboroda, A. Fernandez-Gutierrez, A. Carrasco-Pancorbo, Evaluation of gas chromatography–atmospheric pressure chemical ionization–mass spectrometry as an alternative to gas chromatography–electron ionization–mass spectrometry: avocado fruit as example, J. Chromatogr. A 1313 (2013) 228–244.
[150] T. Pacchiarotta, R.J. Derks, E. Hurtado-Fernandez, P. van Bezooijen, A. Henneman, R. Schiewek, A. Fernandez-Gutierrez, A. Carrasco-Pancorbo, A.M. Deelder, O.A. Mayboroda, Online spectral library for GC-atmospheric pressure chemical ionization-ToF MS, Bioanalysis 5 (2013) 1515–1525.
[151] N. Strehmel, J. Kopka, D. Scheel, C. Böttcher, Annotating unknown components from GC/EI-MS-based metabolite profiling experiments using GC/APCI(+)-QTOFMS, Metabolomics 10 (2013) 324–336.
[152] D. Sciarrone, S. Pantò, A. Rotondo, L. Tedone, P.Q. Tranchida, P. Dugo, L. Mondello, Rapid collection and identification of a novel component from *Clausena lansium* Skeels leaves by means of three-dimensional preparative gas chromatography and nuclear magnetic resonance/infrared/mass spectrometric analysis, Anal. Chim. Acta 785 (2013) 119–125.
[153] G. Purcaro, S. Moret, L. Conte, Hyphenated liquid chromatography–gas chromatography technique: recent evolution and applications, J. Chromatogr. A 1255 (2012) 100–111.
[154] M. Biedermann, K. Grob, C. Mariani, Transesterification and on-line lc-gc for determining the sum of free and esterified sterols in edible oils and fats, Lipid/Fett 95 (1993) 127–133.
[155] N. Mncwangi, I. Vermaak, A.M. Viljoen, Mid-infrared spectroscopy and short wave infrared hyperspectral imaging—a novel approach in the qualitative assessment of *Harpagophytum procumbens* and *H. zeyheri* (Devil's Claw), Phytochem. Lett. 7 (2014) 143–149.
[156] F. Mehl, G. Marti, J. Boccard, B. Debrus, P. Merle, E. Delort, L. Baroux, V. Raymo, M.I. Velazco, H. Sommer, J.-L. Wolfender, S. Rudaz, Differentiation of lemon essential oil based on volatile and non-volatile fractions with various analytical techniques: a metabolomic approach, Food Chem. 143 (2014) 325–335.
[157] J. Kuligowski, D. Pérez-Guaita, J. Escobar, I. Lliso, M. de la Guardia, B. Lendl, M. Vento, G. Quintás, Infrared biospectroscopy for a fast qualitative evaluation of sample preparation in metabolomics, Talanta 127 (2014) 181–190.
[158] M.Y. Mushtaq, Y.H. Choi, R. Verpoorte, E.G. Wilson, Extraction for metabolomics: access to the metabolome, Phytochem. Anal. 25 (2014) 291–306.
[159] M. Halabalaki, S. Bertrand, A. Stefanou, K. Gindro, S. Kostidis, E. Mikros, A.L. Skaltsounis, J.-L. Wolfender, Sample preparation issues in NMR-based plant metabolomics: optimisation for *Vitis* woods samples, Phytochem. Anal. 25 (2014) 350–356.
[160] F. Bucar, A. Wube, M. Schmid, Natural product isolation – how to get from biological material to pure compounds, Nat. Prod. Rep. 30 (2013) 525–545.
[161] G.F. Pauli, S.-N. Chen, D.C. Lankin, J. Bisson, R.J. Case, L.R. Chadwick, T. Gödecke, T. Inui, A. Krunic, B.U. Jaki, J.B. McAlpine, S. Mo, J.G. Napolitano, J. Orjala, J. Lehtivarjo, S.-P. Korhonen, M. Niemitz, Essential parameters for structural analysis and dereplication by 1H NMR spectroscopy, J. Nat. Prod. 77 (2014) 1473–1487.
[162] D.S. Wishart, Quantitative metabolomics using NMR, Trends Anal. Chem. 27 (2008) 228–237.
[163] N. Bohni, M.L. Cordero-Maldonado, J. Maes, D. Siverio-Mota, L. Marcourt, S. Munck, A.R. Kamuhabwa, M.J. Moshi, C.V. Esguerra, P.A.M. de Witte, A.D. Crawford, J.-L. Wolfender, Integration of microfractionation, qNMR and zebrafish screening for the *in vivo* bioassay-guided isolation and quantitative bioactivity analysis of natural products, PLoS ONE 8 (2013) e64006.
[164] C. Simmler, J.G. Napolitano, J.B. McAlpine, S.-N. Chen, G.F. Pauli, Universal quantitative NMR analysis of complex natural samples, Curr. Opin. Biotechnol. 25 (2014) 51–59.
[165] S.M. Al-Massarani, S. Bertrand, A. Nievergelt, A. Mohammed El-Shafae, T. Abdullah Al-Howiriny, N. Musayeib Al-Musayeib, M. Cuendet, J.-L. Wolfender, Acylated pregnane glycosides from *Caralluma sinaica*, Phytochemistry 79 (2012) 129–140.
[166] S.L. Robinette, R. Brüschweiler, F.C. Schroeder, A.S. Edison, NMR in metabolomics and natural products research: two sides of the same coin, Acc. Chem. Res. 45 (2011) 288–297.
[167] V.V. Mihaleva, T.A. te Beek, F. van Zimmeren, S. Moco, R. Laatikainen, M. Niemitz, S.P. Korhonen, M.A. van Driel, J. Vervoort, MetIDB: a publicly accessible database of predicted and experimental 1H NMR spectra of flavonoids, Anal. Chem. 85 (2013) 8700–8707.
[168] J.K. Nicholson, J.C. Lindon, Systems biology – metabonomics, Nature 455 (2008) 1054–1056.
[169] H.K. Kim, Y.H. Choi, R. Verpoorte, NMR-based metabolomic analysis of plants, Nat. Protoc. 5 (2010) 536–549.
[170] H.K. Kim, Y.H. Choi, R. Verpoorte, NMR-based plant metabolomics: where do we stand, where do we go? Trends Biotechnol. 29 (2011) 267–275.
[171] L.S. Pimenta, H. Kim, R. Verpoorte, Y. Choi, NMR-based metabolomics: a probe to utilize biodiversity, in: U. Roessner, D.A. Dias (Eds.), Metabolomics Tools for Natural Product Discovery, Humana Press, 2013, pp. 117–127.
[172] K. Ali, F. Maltese, A. Figueiredo, M. Rex, A.M. Fortes, E. Zyprian, M.S. Pais, R. Verpoorte, Y.H. Choi, Alterations in grapevine leaf metabolism upon inoculation with *Plasmopara viticola* in different time-points, Plant Sci. 191 (2012) 100–107.
[173] E. Alm, T. Slagbrand, K. Åberg, E. Wahlström, I. Gustafsson, J. Lindberg, Automated annotation and quantification of metabolites in ^{1}H NMR data of biological origin, Anal. Bioanal. Chem. 403 (2012) 443–455.
[174] T.F. Molinski, NMR of natural products at the 'nanomole-scale', Nat. Prod. Rep. 27 (2010) 321–329.
[175] N. Bohni, E.F. Queiroz, J.-L. Wolfender, On-line and at-line liquid chromatography nuclear magnetic resonance and related micro-nuclear magnetic resonance methods in natural product analysis, in: K. Hostettmann, H. Stuppner (Eds.), Encyclopedia of Analytical Chemistry (Plant Analysis: Chemical and Biological), John Wiley & Sons, Chichester, UK, 2014, pp. 1–31.
[176] X. Wei, W. Sun, X. Shi, I. Koo, B. Wang, J. Zhang, X. Yin, Y. Tang, B. Bogdanov, S. Kim, Z. Zhou, C. McClain, X. Zhang, MetSign. A computational platform for high-resolution mass spectrometry-based metabolomics, Anal. Chem. 83 (2011) 7668–7675.
[177] J. Draper, A. Lloyd, R. Goodacre, M. Beckmann, Flow infusion electrospray ionisation mass spectrometry for high throughput, non-targeted metabolite fingerprinting: a review, Metabolomics 9 (2012) 4–29.

[178] A.C.H.F. Sawaya, R.R. Catharino, E.M.P. Facco, A. Fogaça, H.T. Godoy, C.E. Daudt, M.N. Eberlin, Monitoring of wine aging process by electrospray ionization mass spectrometry, Food Sci. Technol. 31 (2011) 730–734.
[179] J. de, O. Alves, W.B. Neto, H. Mitsutake, P.S.P. Alves, R. Augusti, Extra virgin (EV) and ordinary (ON) olive oils: distinction and detection of adulteration (EV with ON) as determined by direct infusion electrospray ionization mass spectrometry and chemometric approaches, Rapid Commun. Mass Spectrom. 24 (2010) 1875–1880.
[180] H.-H. Gan, C. Soukoulis, I. Fisk, Atmospheric pressure chemical ionisation mass spectrometry analysis linked with chemometrics for food classification—a case study: geographical provenance and cultivar classification of monovarietal clarified apple juices, Food Chem. 146 (2014) 149–156.
[181] L. Mattoli, F. Cangi, C. Ghiara, M. Burico, A. Maidecchi, E. Bianchi, E. Ragazzi, L. Bellotto, R. Seraglia, P. Traldi, A metabolite fingerprinting for the characterization of commercial botanical dietary supplements, Metabolomics 7 (2011) 437–445.
[182] R.D. Gougeon, M. Lucio, M. Frommberger, D. Peyron, D. Chassagne, H. Alexandre, F. Feuillat, A. Voilley, P. Cayot, I. Gebefügi, N. Hertkorn, P. Schmitt-Kopplin, The chemodiversity of wines can reveal a metabologeography expression of cooperage oak wood, Proc. Natl. Acad. Sci. U.S.A. 106 (2009) 9174–9179.
[183] M. Rojas-Cherto, J.E. Peironcely, P.T. Kasper, J.J. van der Hooft, R.C. de Vos, R. Vreeken, T. Hankemeier, T. Reijmers, Metabolite identification using automated comparison of high-resolution multistage mass spectral trees, Anal. Chem. 84 (2012) 5524–5534.
[184] Y. Sawada, R. Nakabayashi, Y. Yamada, M. Suzuki, M. Sato, A. Sakata, K. Akiyama, T. Sakurai, F. Matsuda, T. Aoki, M.Y. Hirai, K. Saito, RIKEN tandem mass spectral database (ReSpect) for phytochemicals: a plant-specific MS/MS-based data resource and database, Phytochemistry 82 (2012) 38–45 http://spectra.psc.riken.jp/
[185] L. Becker, A. Poutaraud, G. Hamm, J.-F. Muller, D. Merdinoglu, V. Carré, P. Chaimbault, Metabolic study of grapevine leaves infected by downy mildew using Negative Ion Electrospray–Fourier Transform Ion Cyclotron Resonance Mass Spectrometry, Anal. Chim. Acta 795 (2013) 44–51.
[186] P. Giavalisco, J. Hummel, J. Lisec, A.C. Inostroza, G. Catchpole, L. Willmitzer, High-resolution direct infusion-based mass spectrometry in combination with whole ^{13}C metabolome isotope labeling allows unambiguous assignment of chemical sum formulas, Anal. Chem. 80 (2008) 9417–9425.
[187] C. Böttcher, E.v. Roepenack-Lahaye, E. Willscher, D. Scheel, S. Clemens, Evaluation of matrix effects in metabolite profiling based on capillary liquid chromatography electrospray ionization quadrupole time-of-flight mass spectrometry, Anal. Chem. 79 (2007) 1507–1513.
[188] T. Fuhrer, D. Heer, B. Begemann, N. Zamboni, High-throughput accurate mass metabolome profiling of cellular extracts by flow injection–time-of-flight mass spectrometry, Anal. Chem. 83 (2011) 7074–7080.
[189] J.A. Kirwan, D.I. Broadhurst, R.L. Davidson, M.R. Viant, Characterising and correcting batch variation in an automated direct infusion mass spectrometry (DIMS) metabolomics workflow, Anal. Bioanal. Chem. 405 (2013) 5147–5157.
[190] X.L. Guan, I. Riezman, M.R. Wenk, H. Riezman, Chapter 15 – Yeast lipid analysis and quantification by mass spectrometry, in: W. Jonathan, G. Christine, R.F. Gerald (Eds.), Methods Enzymology, Academic Press, 2010, pp. 369–391.
[191] H.R. Jung, T. Sylvänne, K.M. Koistinen, K. Tarasov, D. Kauhanen, K. Ekroos, High throughput quantitative molecular lipidomics, Biochim. Biophys. Acta 1811 (2011) 925–934.
[192] C. Lapthorn, F. Pullen, B.Z. Chowdhry, Ion mobility spectrometry–mass spectrometry (IMS–MS) of small molecules: separating and assigning structures to ions, Mass Spectrom. Rev. 32 (2013) 43–71.
[193] P. Dwivedi, A.J. Schultz, H.H. Hill, Metabolic profiling of human blood by high-resolution ion mobility mass spectrometry (IM-MS), Int. J. Mass Spectrom. 298 (2010) 78–90.
[194] A.S. Galhena, G.A. Harris, M. Kwasnik, F.M. Fernández, Enhanced direct ambient analysis by differential mobility-filtered desorption electrospray ionization–mass spectrometry, Anal. Chem. 82 (2010) 9159–9163.
[195] M. Groessl, A. Azzollini, P.J. Eugster, B. Plet, J.-L. Wolfender, R. Knochenmuss, Comparison of UHPLC-ESI–MS and Hadamard transform atmospheric pressure ion mobility-ESI-MS for rapid profiling of isomeric flavonoids, CHIMIA 68 (2014) 135–139.
[196] G. Paglia, J.P. Williams, L.C. Menikarachchi, J.W. Thompson, R. Tyldesley-Worster, S. Halldórsson, O. Rolfsson, M.A. Moseley, D.F. Grant, J. Langridge, B.Ø. Palsson, G. Astarita, Ion mobility-derived collision cross-sections to support metabolomics applications, Anal. Chem. 86 (2014) 3985–3993.
[197] H.-Y. Wang, X. Chu, Z.-X. Zhao, X.-S. He, Y.-L. Guo, Analysis of low molecular weight compounds by MALDI-FTICR-MS, J. Chromatogr. B 879 (2011) 1166–1179.
[198] R. Shroff, A. Svatoš, Proton sponge: a novel and versatile MALDI matrix for the analysis of metabolites using mass spectrometry, Anal. Chem. 81 (2009) 7954–7959.
[199] X. Dong, J. Cheng, J. Li, Y. Wang, Graphene as a novel matrix for the analysis of small molecules by MALDI-TOF MS, Anal. Chem. 82 (2010) 6208–6214.
[200] C. Eriksson, N. Masaki, I. Yao, T. Hayasaka, M. Setou, MALDI imaging mass spectrometry – a mini review of methods and recent developments, Mass Spectrom. 2 (2013) S0022.
[201] D.J. Weston, Ambient ionization mass spectrometry: current understanding of mechanistic theory; analytical performance and application areas, Analyst 135 (2010) 661–668.
[202] F.M. Fernández, R.B. Cody, M.D. Green, C.Y. Hampton, R. McGready, S. Sengaloundeth, N.J. White, P.N. Newton, Characterization of solid counterfeit drug samples by desorption electrospray ionization and direct-analysis-in-real-time coupled to time-of-flight mass spectrometry, ChemMedChem 1 (2006) 702–705.
[203] F. Rowell, J. Seviour, A.Y. Lim, C.G. Elumbaring-Salazar, J. Loke, J. Ma, Detection of nitro-organic and peroxide explosives in latent fingermarks by DART- and SALDI-TOF–mass spectrometry, Forensic Sci. Int. 221 (2012) 84–91.
[204] S.D. Maleknia, T.M. Vail, R.B. Cody, D.O. Sparkman, T.L. Bell, M.A. Adams, Temperature-dependent release of volatile organic compounds of eucalypts by direct analysis in real time (DART) mass spectrometry, Rapid Commun. Mass Spectrom. 23 (2009) 2241–2246.
[205] M.E. Monge, G.A. Harris, P. Dwivedi, F.M. Fernández, Mass spectrometry recent advances in direct open air surface sampling/ionization, Chem. Rev. 113 (2013) 2269–2308.
[206] K. Fraser, G.A. Lane, D.E. Otter, S.J. Harrison, S.-Y. Quek, Y. Hemar, S. Rasmussen, Monitoring tea fermentation/manufacturing by direct analysis in real time (DART) mass spectrometry, Food Chem. 141 (2013) 2060–2065.
[207] T. Cajka, K. Riddellova, M. Tomaniova, J. Hajslova, Ambient mass spectrometry employing a DART ion source for metabolomic fingerprinting/profiling: a powerful tool for beer origin recognition, Metabolomics 7 (2011) 500–508.
[208] A. Srimany, D.R. Ifa, H.R. Naik, V. Bhat, R.G. Cooks, T. Pradeep, Direct analysis of camptothecin from Nothapodytes nimmoniana by desorption electrospray ionization mass spectrometry (DESI-MS), Analyst 136 (2011) 3066–3068.
[209] P. Chaurand, S.A. Schwartz, M.L. Reyzer, R.M. Caprioli, Imaging mass spectrometry: principles and potentials, Toxicol. Pathol. 33 (2005) 92–101.
[210] A. Thomas, P. Chaurand, Advances in tissue section preparation for MALDI imaging MS, Bioanalysis 6 (2014) 967–982.
[211] C.-J. Shih, P.-Y. Chen, C.-C. Liaw, Y.-M. Lai, Y.-L. Yang, Bringing microbial interactions to light using imaging mass spectrometry, Nat. Prod. Rep. 31 (2014) 739–755.
[212] P. Chaurand, S.A. Schwartz, D. Billheimer, B.J. Xu, A. Crecelius, R.M. Caprioli, Integrating histology and imaging mass spectrometry, Anal. Chem. 76 (2004) 1145–1155.
[213] P. Chaurand, M.E. Sanders, R.A. Jensen, R.M. Caprioli, Proteomics in diagnostic pathology: profiling and imaging proteins directly in tissue sections, Am. J. Pathol. 165 (2004) 1057–1068.
[214] Y.-L. Yang, Y. Xu, P. Straight, P.C. Dorrestein, Translating metabolic exchange with imaging mass spectrometry, Nat. Chem. Biol. 5 (2009) 885–887.
[215] N. Bjarnholt, B. Li, J. D'Alvise, C. Janfelt, Mass spectrometry imaging of plant metabolites – principles and possibilities, Nat. Prod. Rep. 31 (2014) 818–837.
[216] A. Rompp, B. Spengler, Mass spectrometry imaging with high resolution in mass and space, Histochem. Cell Biol. 139 (2013) 759–783.
[217] A. Thomas, S. Lenglet, P. Chaurand, J. Deglon, P. Mangin, F. Mach, S. Steffens, J.L. Wolfender, C. Staub, Mass spectrometry for the evaluation of cardiovascular diseases based on proteomics and lipidomics, Thromb. Haemost. 106 (2011) 20–33.
[218] H. Ye, E. Gemperline, M. Venkateshwaran, R. Chen, P.M. Delaux, M. Howes-Podoll, J.M. Ane, L. Li, MALDI mass spectrometry-assisted molecular imaging of metabolites during nitrogen fixation in the Medicago truncatula-Sinorhizobium meliloti symbiosis, Plant J. 75 (2013) 130–145.
[219] J.L. Norris, R.M. Caprioli, Analysis of tissue specimens by matrix-assisted laser desorption/ionization imaging mass spectrometry in biological and clinical research, Chem. Rev. 113 (2013) 2309–2342.
[220] P. Neubert, A. Walch, Current frontiers in clinical research application of MALDI imaging mass spectrometry, Expert Rev. Proteome 10 (2013) 259–273.
[221] R. Shroff, F. Vergara, A. Muck, A. Svatos, J. Gershenzon, Nonuniform distribution of glucosinolates in Arabidopsis thaliana leaves has important consequences for plant defense, Proc. Natl. Acad. Sci. U. S. A. 105 (2008) 6196–6201.
[222] S. Bertrand, N. Bohni, S. Schnee, O. Schumpp, K. Gindro, J.-L. Wolfender, Metabolite induction via microorganism co-culture: a potential way to enhance chemical diversity for drug discovery, Biotechnol. Adv. 32 (2014) 1180–1204.
[223] M.F. Traxler, J.D. Watrous, T. Alexandrov, P.C. Dorrestein, R. Kolter, Interspecies interactions stimulate diversification of the *Streptomyces coelicolor* secreted metabolome, mBio 4 (2013) e00459–e00513.
[224] J. Watrous, P. Roach, T. Alexandrov, B.S. Heath, J.Y. Yang, R.D. Kersten, M. van der Voort, K. Pogliano, H. Gross, J.M. Raaijmakers, B.S. Moore, J. Laskin, N. Bandeira, P.C. Dorrestein, Mass spectral molecular networking of living microbial colonies, Proc. Natl. Acad. Sci. U. S. A. 109 (2012) E1743–E1752.
[225] K.T. Johansen, S. Wubshet, N.T. Nyberg, HPLC–NMR revisited: using time-slice HPLC–SPE–NMR with database assisted dereplication, Anal. Chem. 85 (2013) 3183–3189.
[226] Chapman, Hall, Dictionary of Natural Products on DVD (23:1), CRC Press, Taylor & Francis Group, 2014 http://dnp.chemnetbase.com/
[227] A. Fredenhagen, C. Derrien, E. Gassmann, An MS/MS library on an ion-trap instrument for efficient dereplication of natural products. Different fragmentation patterns for $[M+H]^+$ and $[M+Na]^+$ ions, J. Nat. Prod. 68 (2005) 385–391.
[228] K.F. Nielsen, J. Smedsgaard, Fungal metabolite screening: database of 474 mycotoxins and fungal metabolites for dereplication by standardised liquid chromatography–UV–mass spectrometry methodology, J. Chromatogr. A 1002 (2003) 111–136.
[229] T. El-Elimat, M. Figueroa, B.M. Ehrmann, N.B. Cech, C.J. Pearce, N.H. Oberlies, High-resolution MS, MS/MS, and UV database of fungal secondary metabolites as a dereplication protocol for bioactive natural products, J. Nat. Prod. 76 (2013) 1709–1716.

[230] A. Klitgaard, A. Iversen, M.R. Andersen, T.O. Larsen, J.C. Frisvad, K.F. Nielsen, Aggressive dereplication using UHPLC–DAD–QTOF: screening extracts for up to 3000 fungal secondary metabolites, Anal. Bioanal. Chem. 406 (2014) 1933–1943.
[231] T. Kind, O. Fiehn, Advances in structure elucidation of small molecules using mass spectrometry, Bioanal. Rev. 2 (2010) 23–60.
[232] K. Scheubert, F. Hufsky, S. Böcker, Computational mass spectrometry for small molecules, J. Cheminform. 5 (2013) 12.
[233] K.F. Nielsen, M. Månsson, C. Rank, J.C. Frisvad, T.O. Larsen, Dereplication of microbial natural products by LC–DAD-TOFMS, J. Nat. Prod. 74 (2011) 2338–2348.
[234] C. Kuhl, R. Tautenhahn, C. Böttcher, T.R. Larson, S. Neumann, CAMERA. An integrated strategy for compound spectra extraction and annotation of liquid chromatography/mass spectrometry data sets, Anal. Chem. 84 (2012) 283–289.
[235] D.J. Creek, A. Jankevics, K.E.V. Burgess, R. Breitling, M.P. Barrett, IDEOM. An Excel interface for analysis of LC–MS based metabolomics data, Bioinformatics 28 (2012) 1048–1049.
[236] T. Kind, O. Fiehn, Metabolomic database annotations via query of elemental compositions: mass accuracy is insufficient even at less than 1 ppm, BMC Bioinform. 7 (2006) 234.
[237] T. Kind, O. Fiehn, Seven golden rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry, BMC Bioinform. 8 (2007) 105.
[238] J. Yang, Q. Liang, M. Wang, C. Jeffries, D. Smithson, Y. Tu, N. Boulos, M.R. Jacob, A.A. Shelat, Y. Wu, R.R. Ravu, R. Gilbertson, M.A. Avery, I.A. Khan, L.A. Walker, R.K. Guy, X.-C. Li, UPLC–MS-ELSD-PDA as a powerful dereplication tool to facilitate compound identification from small-molecule natural product libraries, J. Nat. Prod. 77 (2014) 902–909.
[239] S. Kildgaard, M. Mansson, I. Dosen, A. Klitgaard, J. Frisvad, T. Larsen, K. Nielsen, Accurate dereplication of bioactive secondary metabolites from marine-derived fungi by UHPLC–DAD-QTOFMS and a MS/HRMS library, Mar. Drugs 12 (2014) 3681–3705.
[240] S. Bertrand, A. Azzollini, O. Schumpp, N. Bohni, J. Schrenzel, M. Monod, K. Gindro, J.-L. Wolfender, Multi-well fungal co-culture for *de novo* metabolite-induction in time series studies based on untargeted metabolomics, Mol. BioSyst. 10 (2014) 2289–2298.
[241] C.S. Funari, P.J. Eugster, S. Martel, P.-A. Carrupt, J.-L. Wolfender, D.H.S. Silva, High resolution ultra high pressure liquid chromatography–time-of-flight mass spectrometry dereplication strategy for the metabolite profiling of Brazilian Lippia species, J. Chromatogr. A 1259 (2012) 167–178.
[242] R.R. Silva, F. Jourdan, D.M. Salvanha, F. Letisse, E.L. Jamin, S. Guidetti-Gonzalez, C.A. Labate, R.Z.N. Vêncio, ProbMetab: an R package for Bayesian probabilistic annotation of LC–MS-based metabolomics, Bioinformatics 30 (2014) 1336–1337.
[243] M. Brown, D.C. Wedge, R. Goodacre, D.B. Kell, P.N. Baker, L.C. Kenny, M.A. Mamas, L. Neyses, W.B. Dunn, Automated workflows for accurate mass-based putative metabolite identification in LC/MS-derived metabolomic datasets, Bioinformatics 27 (2011) 1108–1112.
[244] G. Marti, M. Erb, S. Rudaz, T. Turlings, J.-L. Wolfender, Search for low-molecular-weight biomarkers in plant tissues and seeds using metabolomics: tools, strategies, and applications, in: G.K. Agrawal, R. Rakwal (Eds.), Seed Development: OMICS Technologies toward Improvement of Seed Quality and Crop Yield, Springer, Netherlands, 2012, pp. 305–341.
[245] H. Doerfler, X. Sun, L. Wang, D. Engelmeier, D. Lyon, W. Weckwerth, *mzGroupAnalyzer* – predicting pathways and novel chemical structures from untargeted high-throughput metabolomics data, PLoS ONE 9 (2014) e96188.
[246] S. Rogers, R.A. Scheltema, M. Girolami, R. Breitling, Probabilistic assignment of formulas to mass peaks in metabolomics experiments, Bioinformatics 25 (2009) 512–518.
[247] H. Zhang, D. Zhang, K. Ray, M. Zhu, Mass defect filter technique and its applications to drug metabolite identification by high-resolution mass spectrometry, J. Mass Spectrom. 44 (2009) 999–1016.
[248] C.A. Hughey, C.L. Hendrickson, R.P. Rodgers, A.G. Marshall, K. Qian, Kendrick mass defect spectrum: a compact visual analysis for ultrahigh-resolution broadband mass spectra, Anal. Chem. 73 (2001) 4676–4681.
[249] T. Xie, Y. Liang, H.J.A. Hao, L. Xie, P. Gong, C. Dai, L. Liu, A. Kang, X. Zheng, G. Wang, Rapid identification of ophiopogonins and ophiopogonones in Ophiopogon japonicus extract with a practical technique of mass defect filtering based on high resolution mass spectrometry, J. Chromatogr. A 1227 (2012) 234–244.
[250] Y. Konishi, T. Kiyota, C. Draghici, J.-M. Gao, F. Yeboah, S. Acoca, S. Jarussophon, E. Purisima, Molecular formula analysis by an MS/MS/MS technique to expedite dereplication of natural products, Anal. Chem. 79 (2006) 1187–1197.
[251] T. Pluskal, T. Uehara, M. Yanagida, Highly accurate chemical formula prediction tool utilizing high-resolution mass spectra, MS/MS fragmentation, heuristic rules, and isotope pattern matching, Anal. Chem. 84 (2012) 4396–4403.
[252] V. Vukics, A. Guttman, Structural characterization of flavonoid glycosides by multi-stage mass spectrometry, Mass Spectrom. Rev. 29 (2010) 1–16.
[253] W.F. Smyth, T.J.P. Smyth, V.N. Ramachandran, F. O'Donnell, P. Brooks, Dereplication of phytochemicals in plants by LC-ESI–MS and ESI-MSn, Trends Anal. Chem. 33 (2012) 46–54.
[254] I. Kerzaon, Y.F. Pouchus, F. Monteau, B. Le Bizec, M.-R. Nourrisson, J.-F. Biard, O. Grovel, Structural investigation and elucidation of new communesins from a marine-derived Penicillium expansum Link by liquid chromatography/electrospray ionization mass spectrometry, Rapid Commun. Mass Spectrom. 23 (2009) 3928–3938.
[255] X. Yang, P. Neta, S.E. Stein, Quality control for building libraries from electrospray ionization tandem mass spectra, Anal. Chem. 86 (2014) 6393–6400.
[256] R. Tautenhahn, K. Cho, W. Uritboonthai, Z. Zhu, G.J. Patti, G. Siuzdak, An accelerated workflow for untargeted metabolomics using the METLIN database, Nat. Biotechnol. 30 (2012) 826–828.
[257] T. Kind, K.-H. Liu, D.Y. Lee, B. DeFelice, J.K. Meissen, O. Fiehn, LipidBlast *in silico* tandem mass spectrometry database for lipid identification, Nat. Methods 10 (2013) 755–758.
[258] L. Ridder, J.J.J. van der Hooft, S. Verhoeven, R.C.H. de Vos, J. Vervoort, R.J. Bino, In silico prediction and automatic LC–MSn annotation of green tea metabolites in urine, Anal. Chem. 86 (2014) 4767–4774.
[259] F. Hufsky, K. Scheubert, S. Bocker, New kids on the block: novel informatics methods for natural product discovery, Nat. Prod. Rep. 31 (2014) 807–817.
[260] J.Y. Yang, L.M. Sanchez, C.M. Rath, X. Liu, P.D. Boudreau, N. Bruns, E. Glukhov, A. Wodtke, R. de Felicio, A. Fenner, W.R. Wong, R.G. Linington, L. Zhang, H.M. Debonsi, W.H. Gerwick, P.C. Dorrestein, Molecular networking as a dereplication strategy, J. Nat. Prod. 76 (2013) 1686–1699.
[261] K.P. Law, Y.P. Lim, Recent advances in mass spectrometry: data independent analysis and hyper reaction monitoring, Expert Rev. Proteome 10 (2013) 551–566.
[262] J.D. Chapman, D.R. Goodlett, C.D. Masselon, Multiplexed and data-independent tandem mass spectrometry for global proteome profiling, Mass Spectrom. Rev. 33 (2014) 452–470.
[263] A. Thomas, J. Déglon, S. Lenglet, F. Mach, P. Mangin, J.-L. Wolfender, S. Steffens, C. Staub, High-throughput phospholipidic fingerprinting by online desorption of dried spots and quadrupole-linear ion trap mass spectrometry: evaluation of atherosclerosis biomarkers in mouse plasma, Anal. Chem. 82 (2010) 6687–6694.
[264] J.D. Egertson, A. Kuehn, G.E. Merrihew, N.W. Bateman, B.X. MacLean, Y.S. Ting, J.D. Canterbury, D.M. Marsh, M. Kellmann, V. Zabrouskov, C.C. Wu, M.J. MacCoss, Multiplexed MS/MS for improved data-independent acquisition, Nat. Methods 10 (2013) 744–746.
[265] A. Michalski, J. Cox, M. Mann, More than 100,000 detectable peptide species elute in single shotgun proteomics runs but the majority is inaccessible to data-dependent LC–MS/MS, J. Proteome Res. 10 (2011) 1785–1793.
[266] K. Blackburn, M.B. Goshe, Challenges and strategies for targeted phosphorylation site identification and quantification using mass spectrometry analysis, Brief. Funct. Genome Proteome 8 (2009) 90–103.
[267] T. Geiger, J. Cox, M. Mann, Proteomics on an Orbitrap Benchtop mass spectrometer using all-ion fragmentation, Mol. Cell. Proteomics 9 (2010) 2252–2261.
[268] P. Le, M. McCooeye, A. Windust, Application of UPLC-QTOF–MS in MSE mode for the rapid and precise identification of alkaloids in goldenseal (*Hydrastis canadensis*), Anal. Bioanal. Chem. 406 (2014) 1739–1749.
[269] G. Marti, M. Erb, J. Boccard, G. Glauser, G.R. Doyen, N. Villard, C.A.M. Robert, T.C.J. Turlings, S. Rudaz, J.-L. Wolfender, Metabolomics reveals herbivore-induced metabolites of resistance and susceptibility in maize leaves and roots, Plant Cell Environ. 36 (2013) 621–639.
[270] G. Glauser, N. Veyrat, B. Rochat, J.-L. Wolfender, T.C.J. Turlings, Ultra-high pressure liquid chromatography–mass spectrometry for plant metabolomics: a systematic comparison of high-resolution quadrupole-time-of-flight and single stage Orbitrap mass spectrometers, J. Chromatogr. A 1292 (2013) 151–159.
[271] L.C. Gillet, P. Navarro, S. Tate, H. Rost, N. Selevsek, L. Reiter, R. Bonner, R. Aebersold, Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis, Mol. Cell. Proteomics 11 (2012), O111.016717.
[272] A. Bouslimani, L.M. Sanchez, N. Garg, P.C. Dorrestein, Mass spectrometry of natural products: current, emerging and future technologies, Nat. Prod. Rep. 31 (2014) 718–729.
[273] Y. Henchoz, D. Guillarme, S. Martel, S. Rudaz, J.-L. Veuthey, P.-A. Carrupt, Fast log P determination by ultra-high-pressure liquid chromatography coupled with UV and mass spectrometry detections, Anal. Bioanal. Chem. 394 (2009) 1919–1930.
[274] J. Stanstrup, M. Gerlich, L. Dragsted, S. Neumann, Metabolite profiling and beyond: approaches for the rapid processing and annotation of human blood serum mass spectrometry data, Anal. Bioanal. Chem. 405 (2013) 5037–5048.
[275] A. Téllez, M. Rosés, E. Bosch, Modeling the retention of neutral compounds in gradient elution RP-HPLC by means of polarity parameter models, Anal. Chem. 81 (2009) 9135–9145.
[276] J. Akbar, S. Iqbal, F. Batool, A. Karim, K. Chan, Predicting retention times of naturally occurring phenolic compounds in reversed-phase liquid chromatography: a quantitative structure-retention relationship (QSRR) approach, Int. J. Mol. Sci. 13 (2012) 15387–15400.
[277] T. Bączek, R. Kaliszan, Predictions of peptides' retention times in reversed-phase liquid chromatography as a new supportive tool to improve protein identification in proteomics, Proteomics 9 (2009) 835–847.
[278] M. Salo, H. Sirén, P. Volin, S. Wiedmer, H. Vuorela, Structure-retention relationships of steroid hormones in reversed-phase liquid chromatography and micellar electrokinetic capillary chromatography, J. Chromatogr. A 728 (1996) 83–88.
[279] L.M. Hall, L.H. Hall, T.M. Kertesz, D.W. Hill, T.R. Sharp, E.Z. Oblak, Y.W. Dong, D.S. Wishart, M.-H. Chen, D.F. Grant, Development of Ecom50 and retention index models for nontargeted metabolomics: identification of

1,3-dicyclohexylurea in human serum by HPLC/mass spectrometry, J. Chem. Inform. Model. 52 (2012) 1222–1237.
[280] P.J. Eugster, J. Boccard, B. Debrus, L. Bréant, J.-L. Wolfender, S. Martel, P.-A. Carrupt, Retention time prediction for dereplication of natural products (CxHyOz) in LC–MS metabolite profiling, Phytochemistry 108 (2014) 196–207.
[281] D.J. Creek, A. Jankevics, R. Breitling, D.G. Watson, M.P. Barrett, K.E.V. Burgess, Toward global metabolomics analysis with hydrophilic interaction liquid chromatography–mass spectrometry: improved metabolite identification by retention time prediction, Anal. Chem. 83 (2011) 8703–8710.
[282] L. Sumner, A. Amberg, D. Barrett, M. Beale, R. Beger, C. Daykin, T.M. Fan, O. Fiehn, R. Goodacre, J. Griffin, T. Hankemeier, N. Hardy, J. Harnly, R. Higashi, J. Kopka, A. Lane, J. Lindon, P. Marriott, A. Nicholls, M. Reily, J. Thaden, M. Viant, Proposed minimum reporting standards for chemical analysis, Metabolomics 3 (2007) 211–221.
[283] S. Bertrand, O. Schumpp, N. Bohni, M. Monod, K. Gindro, J.-L. Wolfender, *De novo* production of metabolites by fungal co-culture of *Trichophyton rubrum* and *Bionectria ochroleuca*, J. Nat. Prod. 76 (2013) 1157–1165.
[284] J.L. Wolfender, N. Bohni, K. Ndjoko-Ioset, A.S. Edison, Chapter 16 – Advanced spectroscopic detectors for identification and quantification: nuclear magnetic resonance, in: S. Fanali, P.R. Haddad, C.F. Poole, P. Schoenmakers, D. Lloyd (Eds.), Liquid Chromatography, Elsevier, Amsterdam, 2013, pp. 349–384.
[285] J.-L. Wolfender, E.F. Queiroz, K. Hostettmann, Phytochemistry in the microgram domain—a LC–NMR perspective, Magn. Reson. Chem. 43 (2005) 697–709.
[286] H. Kovacs, D. Moskau, M. Spraul, Cryogenically cooled probes—a leap in NMR technology, Prog. Nucl. Magn. Reson. Spectrosc. 46 (2005) 131–155.
[287] R.T. McKay, How the 1D-NOESY suppresses solvent signal in metabonomics NMR spectroscopy: an examination of the pulse sequence components and evolution, Concepts Magn. Reson. A 38A (2011) 197–220.
[288] S.W. Kang, C.Y. Kim, D.G. Song, C.H. Pan, K.H. Cha, D.U. Lee, B.H. Um, Rapid identification of furanocoumarins in *Angelica dahurica* using the online LC-MMR–MS and their nitric oxide inhibitory activity in RAW 264.7 cells, Phytochem. Anal. 21 (2010) 322–327.
[289] S.J. Kim, S.M. Kim, M.C. Kim, S.W. Kang, B.H. Um, Fast identification of flavonoids in the roots of *Sophora flavescens* by on-flow LC–NMR, J. Med. Plant. Res. 4 (2010) 2452–2459.
[290] O. Kenny, T.J. Smyth, C.M. Hewage, N.P. Brunton, P. McLoughlin, 4-Hydroxyphenylacetic acid derivatives of inositol from dandelion (*Taraxacum officinale*) root characterised using LC-SPE–NMR and LC–MS techniques, Phytochemistry 98 (2014) 197–203.
[291] K.T. Johansen, S.G. Wubshet, N.T. Nyberg, J.W. Jaroszewski, From retrospective assessment to prospective decisions in natural product isolation: HPLC-SPE–NMR analysis of *Carthamus oxyacantha*, J. Nat. Prod. 74 (2011) 2454–2461.
[292] J.W. Schoonen, P. Vulto, N. de Roo, J. van Duynhoven, H. van der Linden, T. Hankemeier, Solvent exchange module for LC–NMR hyphenation using machine vision-controlled droplet evaporation, Anal. Chem. 85 (2013) 5734–5739.
[293] S. Bertrand, C. Petit, L. Marcourt, R. Ho, K. Gindro, M. Monod, J.-L. Wolfender, HPLC profiling with at-line microdilution assay for the early identification of antifungal compounds in plants from French Polynesia, Phytochem. Anal. 25 (2014) 106–112.
[294] F. Moradi-Afrapoli, S.N. Ebrahimi, M. Smiesko, M. Raith, S. Zimmermann, F. Nadjafi, R. Brun, M. Hamburger, Bisabololoxide derivatives from *Artemisia persica*, and determination of their absolute configurations by ECD, Phytochemistry 85 (2013) 143–152.
[295] S.G. Wubshet, K.T. Johansen, N.T. Nyberg, J.W. Jaroszewski, Direct ^{13}C NMR detection in HPLC hyphenation mode: analysis of *Ganoderma lucidum* terpenoids, J. Nat. Prod. 75 (2012) 876–882.
[296] R. Bruschweiler, F. Zhang, Covariance nuclear magnetic resonance spectroscopy, J. Chem. Phys. 120 (2004) 5253–5260.
[297] O. Cloarec, M.E. Dumas, A. Craig, R.H. Barton, J. Trygg, J. Hudson, C. Blancher, D. Gauguier, J.C. Lindon, E. Holmes, J. Nicholson, Statistical total correlation spectroscopy: an exploratory approach for latent biomarker identification from metabolic 1H NMR data sets, Anal. Chem. 77 (2005) 1282–1289.
[298] O. Cloarec, A. Campbell, L.H. Tseng, U. Braumann, M. Spraul, G. Scarfe, R. Weaver, J.K. Nicholson, Virtual chromatographic resolution enhancement in cryoflow LC–NMR experiments via statistical total correlation spectroscopy, Anal. Chem. 79 (2007) 3304–3311.
[299] D.J. Crockford, E. Holmes, J.C. Lindon, R.S. Plumb, S. Zirah, S.J. Bruce, P. Rainville, C.L. Stumpf, J.K. Nicholson, Statistical heterospectroscopy an approach to the integrated analysis of NMR and UPLC–MS data sets: application in metabonomic toxicology studies, Anal. Chem. 78 (2006) 363–371.
[300] G. Marti, J. Boccard, F. Mehl, B. Debrus, L. Marcourt, P. Merle, E. Delort, L. Baroux, H. Sommer, S. Rudaz, J.-L. Wolfender, Comprehensive profiling and marker identification in non-volatile citrus oil residues by mass spectrometry and nuclear magnetic resonance, Food Chem. 150 (2014) 235–245.
[301] J.W. Allwood, D.I. Ellis, R. Goodacre, Metabolomic technologies and their application to the study of plants and plant-host interactions, Physiol. Plant 132 (2008) 117–135.
[302] R. Schweiger, M.C. Baier, M. Persicke, C. Muller, High specificity in plant leaf metabolic responses to arbuscular mycorrhiza, Nat. Commun. 5 (2014) 3886.
[303] M. Kusano, H. Redestig, T. Hirai, A. Oikawa, F. Matsuda, A. Fukushima, M. Arita, S. Watanabe, M. Yano, K. Hiwasa-Tanase, H. Ezura, K. Saito, Covering chemical diversity of genetically-modified tomatoes using metabolomics for objective substantial equivalence assessment, PLoS ONE 6 (2011) e16989.
[304] Z. Lei, D.V. Huhman, L.W. Sumner, Mass spectrometry strategies in metabolomics, J. Biol. Chem. 286 (2011) 25435–25442.
[305] F. Tugizimana, P.A. Steenkamp, L.A. Piater, I.A. Dubery, Multi-platform metabolomic analyses of ergosterol-induced dynamic changes in *Nicotiana tabacum* cells, PLoS ONE 9 (2014) e87846.
[306] A. Mari, D. Lyon, L. Fragner, P. Montoro, S. Piacente, S. Wienkoop, V. Egelhofer, W. Weckwerth, Phytochemical composition of *Potentilla anserina* L. analyzed by an integrative GC–MS and LC–MS metabolomics platform, Metabolomics 9 (2013) 599–607.
[307] D.K. Lee, M.H. Yoon, Y.P. Kang, J. Yu, J.H. Park, J. Lee, S.W. Kwon, Comparison of primary and secondary metabolites for suitability to discriminate the origins of Schisandra chinensis by GC/MS and LC/MS, Food Chem. 141 (2013) 3931–3937.
[308] J. Ivanisevic, Z.-J. Zhu, L. Plate, R. Tautenhahn, S. Chen, P.J. O'Brien, C.H. Johnson, M.A. Marletta, G.J. Patti, G. Siuzdak, Toward 'omic scale metabolite profiling: a dual separation–mass spectrometry approach for coverage of lipid and central carbon metabolism, Anal. Chem. 85 (2013) 6876–6884.
[309] J. Boccard, S. Rudaz, Harnessing the complexity of metabolomic data with chemometrics, J. Chemomet. 28 (2014) 1–9.
[310] T. Roemer, D. Xu, S.B. Singh, C.A. Parish, G. Harris, H. Wang, J.E. Davies, G.F. Bills, Confronting the challenges of natural product-based antifungal discovery, Chem. Biol. 18 (2011) 148–164.
[311] A. Marston, Thin-layer chromatography with biological detection in phytochemistry, J. Chromatogr. A 1218 (2011) 2676–2683.
[312] J. Kool, M. Giera, H. Irth, W.A. Niessen, Advances in mass spectrometry-based post-column bioaffinity profiling of mixtures, Anal. Bioanal. Chem. 399 (2011) 2655–2668.
[313] H.A.G. Niederländer, T.A. van Beek, A. Bartasiute, I.I. Koleva, Antioxidant activity assays on-line with liquid chromatography, J. Chromatogr. A 1210 (2008) 121–134.
[314] T. van Beek, K.R. Tetala, I. Koleva, A. Dapkevicius, V. Exarchou, S.F. Jeurissen, F. Claassen, E.C. van der Klift, Recent developments in the rapid analysis of plants and tracking their bioactive constituents, Phytochem. Rev. 8 (2009) 387–399.
[315] M. Mnatsakanyan, T.A. Goodie, X.A. Conlan, P.S. Francis, G.P. McDermott, N.W. Barnett, D. Shock, F. Gritti, G. Guiochon, R.A. Shalliker, High performance liquid chromatography with two simultaneous on-line antioxidant assays: evaluation and comparison of espresso coffees, Talanta 81 (2010) 837–842.
[316] G. Lang, N.A. Mayhudin, M.I. Mitova, L. Sun, S. van der Sar, J.W. Blunt, A.L.J. Cole, G. Ellis, H. Laatsch, M.H.G. Munro, Evolving trends in the dereplication of natural product extracts: new methodology for rapid small-scale investigation of natural product extracts, J. Nat. Prod. 71 (2008) 1595–1599.
[317] S.G. Wubshet, N.T. Nyberg, M.V. Tejesvi, A.M. Pirttilä, M. Kajula, S. Mattila, D. Staerk, Targeting high-performance liquid chromatography–high-resolution mass spectrometry-solid-phase extraction-nuclear magnetic resonance analysis with high-resolution radical scavenging profiles-bioactive secondary metabolites from the endophytic fungus *Penicillium namyslowskii*, J. Chromatogr. A 1302 (2013) 34–39.
[318] S. Challal, N. Bohni, O.E. Buenafe, C.V. Esguerra, P.A.M. de Witte, J.-L. Wolfender, A.D. Crawford, Zebrafish bioassay-guided microfractionation for the rapid *in vivo* identification of pharmacologically active natural products, CHIMIA 66 (2012) 229–232.
[319] T. Inui, Y. Wang, S.M. Pro, S.G. Franzblau, G.F. Pauli, Unbiased evaluation of bioactive secondary metabolites in complex matrices, Fitoterapia 83 (2012) 1218–1225.
[320] A. Azzollini, S. Bertrand, A. Nievergelt, J. Boccard, S. Rudaz, M. Cuendet, J.-L. Wolfender, Integrative approach in natural product research: at-line coupling of micro-fractionation with NMR, LC–MS and bioassays, CHIMIA 66 (2012) 487.
[321] J.J.J. van der Hooft, J. Vervoort, R.J. Bino, J. Beekwilder, R.C.H. de Vos, Polyphenol identification based on systematic and robust high-resolution accurate mass spectrometry fragmentation, Anal. Chem. 83 (2011) 409–416.
[322] J.J.J. van der Hooft, J. Vervoort, R.J. Bino, R.C.H. de Vos, Spectral trees as a robust annotation tool in LC–MS based metabolomics, Metabolomics 8 (2012) 691–703.
[323] F. Allen, R. Greiner, D. Wishart, Competitive fragmentation modeling of ESI-MS/MS spectra for putative metabolite identification, Metabolomics (2014), http://dx.doi.org/10.1007/s11306-014-0676-4.
[324] F. Dethloff, A. Erban, I. Orf, J. Alpers, I. Fehrle, O. Beine-Golovchuk, S. Schmidt, J. Schwachtje, J. Kopka, Profiling methods to identify cold-regulated primary metabolites using gas chromatography coupled to mass spectrometry, Methods Mol. Biol. 1166 (2014) 171–197.
[325] T. Kind, G. Wohlgemuth, Y. Lee do, Y. Lu, M. Palazoglu, S. Shahbaz, O. Fiehn, FiehnLib: mass spectral and retention index libraries for metabolomics based on quadrupole and time-of-flight gas chromatography/mass spectrometry, Anal. Chem. 81 (2009) 10038–10048.
[326] B. David, J.-L. Wolfender, D. Dias, The pharmaceutical industry and natural products: historical status and new trends, Phytochem. Rev. (2014), http://dx.doi.org/10.1007/s11101-014-9367-z.
[327] K. Haug, R.M. Salek, P. Conesa, J. Hastings, P. de Matos, M. Rijnbeek, T. Mahendraker, M. Williams, S. Neumann, P. Rocca-Serra, E. Maguire, A. González-Beltrán, S.-A. Sansone, J.L. Griffin, C. Steinbeck, MetaboLights—an open-access general-purpose repository for metabolomics studies and associated meta-data, Nucleic Acids Res. 41 (2013) D781–D786.
[328] S.A. Sansone, P. Rocca-Serra, D. Field, E. Maguire, C. Taylor, O. Hofmann, H. Fang, S. Neumann, W.D. Tong, L. Amaral-Zettler, K. Begley, T. Booth, L.

Bougueleret, G. Burns, B. Chapman, T. Clark, L.A. Coleman, J. Copeland, S. Das, A. de Daruvar, P. de Matos, I. Dix, S. Edmunds, C.T. Evelo, M.J. Forster, P. Gaudet, J. Gilbert, C. Goble, J.L. Griffin, D. Jacob, J. Kleinjans, L. Harland, K. Haug, H. Hermjakob, S.J.H. Sui, A. Laederach, S.G. Liang, S. Marshall, A. McGrath, E. Merrill, D. Reilly, M. Roux, C.E. Shamu, C.A. Shang, C. Steinbeck, A. Trefethen, B. Williams-Jones, K. Wolstencroft, I. Xenarios, W. Hide, Toward interoperable bioscience data, Nat. Genet. 44 (2012) 121–126.

Glossary

ACN: acetonitrile
AIF: all-ion fragmentation
APCI: atmospheric pressure chemical ionisation
API: atmospheric pressure ionisation
CCC: counter current chromatography
CI: chemical ionisation
CSS: collision cross-section value
DDA: data-dependent acquisition
DIA: data-independent acquisition
DIMS: direct infusion mass spectrometry
EI: electronic impact
ESI: electrospray ionisation
GC: gas chromatography
HILIC: hydrophilic interaction liquid chromatography
HPLC: high-performance liquid chromatography
HR: high resolution
Imaging MS: imaging MS is sometime referred as IMS
IMS: ion mobility
IPC: ion pair chromatography
IR: infrared
LLE: liquid–liquid extraction
MALDI: matrix-assisted laser desorption/ionisation
MPLC: medium pressure liquid chromatography
MS^E: "MS to the E", simultaneous acquisition of MS spectra at low and high collision energy
MSX-DIA: multiplexed MS/MS data-independent acquisition
MS: mass spectrometry
MVDA: multivariate data analysis
MW: molecular weight
NMR: nuclear magnetic resonance
NARP: non-aqueous reverse phase
NI: negative ionisation
NP: normal phase
PDA: photodiode array
PI: positive ionisation
QC: quality control
RP: reversed-phase
RT: retention time
SCX: strong cation exchange
SFC: supercritical fluid chromatography
SPE: solid phase extraction
SWATH: Sequential Window Acquisition of all THeoretical mass spectra
TMS: total MS spectra fingerprints
UHPLC: ultra-high pressure liquid chromatography
VLC: vacuum liquid chromatography

analytical chemistry

pubs.acs.org/ac Article

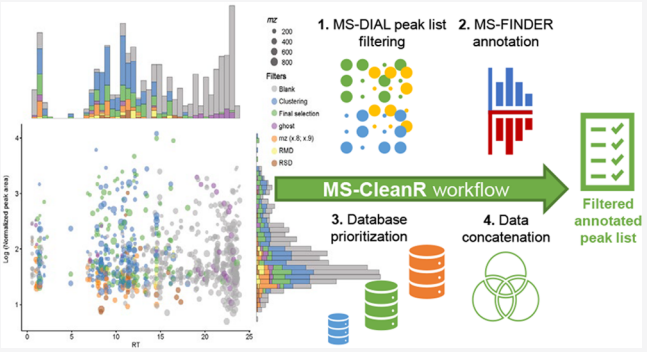
MS-CleanR: A Feature-Filtering Workflow for Untargeted LC−MS Based Metabolomics

Ophélie Fraisier-Vannier, Justine Chervin, Guillaume Cabanac, Virginie Puech, Sylvie Fournier, Virginie Durand, Aurélien Amiel, Olivier André, Omar Abdelaziz Benamar, Bernard Dumas, Hiroshi Tsugawa, and Guillaume Marti*

Cite This: *Anal. Chem.* 2020, 92, 9971−9981 Read Online

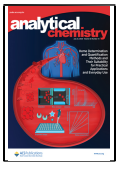
ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Untargeted metabolomics using liquid chromatography−mass spectrometry (LC−MS) is currently the gold-standard technique to determine the full chemical diversity in biological samples. However, this approach still has many limitations; notably, the difficulty of accurately estimating the number of unique metabolites profiled among the thousands of MS ion signals arising from chromatograms. Here, we describe a new workflow, MS-CleanR, based on the MS-DIAL/MS-FINDER suite, which tackles feature degeneracy and improves annotation rates. We show that implementation of MS-CleanR reduces the number of signals by nearly 80% while retaining 95% of unique metabolite features. Moreover, the annotation results from MS-FINDER can be ranked according to the database chosen by the user, which enhance identification accuracy. Application of MS-CleanR to the analysis of *Arabidopsis thaliana* grown in three different conditions fostered class separation resulting from multivariate data analysis and led to annotation of 75% of the final features. The full workflow was applied to metabolomic profiles from three strains of the leguminous plant *Medicago truncatula* that have different susceptibilities to the oomycete pathogen *Aphanomyces euteiches*. A group of glycosylated triterpenoids overrepresented in resistant lines were identified as candidate compounds conferring pathogen resistance. MS-CleanR is implemented through a Shiny interface for intuitive use by end-users (available at https://github.com/eMetaboHUB/MS-CleanR).



Untargeted or discovery-based metabolomics have become an essential tool in all biological sciences including clinical research,[1,2] plant science,[3] and natural product mining,[4] among many other applications. Living organisms are estimated to contain more than one million distinct compounds.[5] According to the MetaboLights database (DB), 80% of untargeted metabolomics workflows rely on liquid chromatography−mass spectrometry (LC−MS) (https://www.ebi.ac.uk/metabolights/). Due to its broad coverage of metabolites, LC−MS based metabolomics has become the preferred tool to detect hundreds of compounds encountered in complex biological materials. Many software programs have been developed to turn features (m/z × retention time (RT) pairs) extracted from LC−MS raw data into chromatographic peak lists, including web-based interfaces such as XCMS,[6] Workflow4Metabolomics,[7] local GUI with MZmine,[8] and MS-DIAL.[9] Despite significant progress in feature extraction, it is challenging to accurately estimate the number of unique metabolites in a crude extract profiled by LC−MS.[10] On average, untargeted LC−MS yields hundred to thousands of signals, which may be attributed to either isotopes, contaminants, adducts, dimers, multimers, and heteromeric complexes or artifacts. The feature attribution processing which aims to decipher ion linkages is an essential step prior to metabolite annotation, which refers to tentative metabolite assignment to a given feature. Following these steps, the annotation rate can be calculated as the number of unique annotated metabolites over total features counts. Patti and colleagues[11] used the term "degenerate features" to describe feature relationships between multiple m/z signals arising from in-source phenomena and derived from the same metabolite. Their study demonstrated that feature inflation is highly underestimated in untargeted LC−MS based metabolomics. Additionally, this redundancy trend may have important consequences on metabolite annotation by increasing both false positive results and the number of "unknown" arising from wrongly attributed signals. This is especially true when

Received: April 13, 2020
Accepted: June 26, 2020
Published: June 26, 2020



https://dx.doi.org/10.1021/acs.analchem.0c01594
Anal. Chem. 2020, 92, 9971−9981

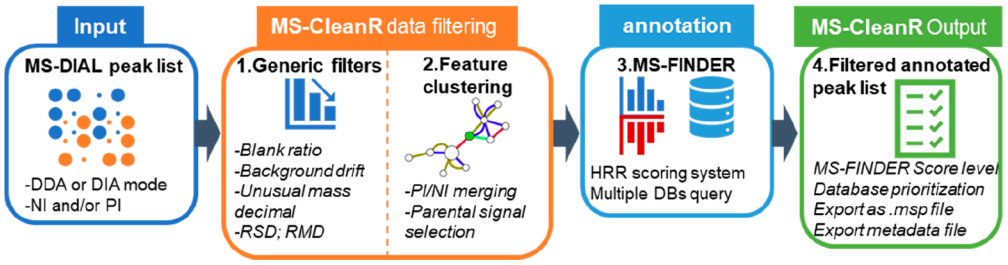


Figure 1. Schematic representation of MS-CleanR workflow.

the metabolite annotation process is based on *in silico* modeling of fragmentation patterns, as are Sirius,[12] MS-FINDER,[13] MetFrag,[14] or CFM-ID,[15] since tandem mass spectrometry (MS/MS) spectra are processed without taking into account feature relationships. As a result, most untargeted metabolomics studies focus on a subset of identified metabolites for which spectral data are easily accessible from public repositories or in-house DBs.

Feature attribution processing has been tackled by several packages using either correlation approach across multiple samples like MSClust,[16] RAMClust,[17] MS-FLO,[18] with additional MS/MS similarity filtering implemented in compMS2Miner,[19] or inferring feature attribution and annotation using metabolic pathway associations in xMSannotator.[20] Other approaches driven by peak shape similarity in a user defined retention time window are applicable to one single chromatogram. These include CAMERA[21] based on Pearson correlation, MetAssign[22] constructed through a Bayesian clustering approach to sort related m/z signals, and CliqueMS[23] which added an estimated adducts frequency to peak similarity for feature relationship deciphering. Within the all-in-one program MS-DIAL, a combination of approaches is gathered by peak character estimation algorithm[24] (MS-DIAL-PCE) providing the ion linkages based on (i) adducts mass differences calculation, (ii) unexpected in-source phenomena using correlated chromatograms within the sample, and (iii) putative ion source fragments candidates based on MS/MS similarity and analogous metabolite profiles among samples. MS-DIAL can handle one single chromatogram or a set of samples providing an aligned peak list (n samples × peak area of detected features) with putative ion linkages data. MS-FINDER is a partner program of MS-DIAL, in which unknown structures can be elucidated from MS/MS spectra by the hydrogen rearrangement rules-(HRR) based scoring system.[13]

Here, we describe a third tool interfaced between these two programs, called MS-CleanR, which aims to obtain an annotated and clean peak list by embedding the results from MS-DIAL and MS-FINDER. Starting from the aligned peak list determined by the MS-DIAL deconvolution process, our R package first removes noise signals by using several generic filters. In the second step, each feature is clustered to construct a graph based on the results of MS-DIAL-PCE and optionally extended by Pearson correlation links. At this point, each graph is submitted to the multilevel optimization of modularity algorithm[25] to extract the most probable parental mass. Optionally, clustered ion features can be merged between positive ionization (PI) and negative ionization (NI) modes and the adduct relationships are corrected accordingly. The cleaned-up feature list is then exported to MS-FINDER for annotation purposes. Finally, the package merges the MS-FINDER annotation output with the cleaned-up peak list and includes the possibility to prioritize identification according to the DBs used for MS-FINDER interrogation. The whole MS-CleanR workflow (Figure 1) is easily accessible through a shiny user interface, and it is available as open source code.

METHODS

Standards. Individual solutions of natural products (NPs) compounds (Metasci, Toronto, Canada) were prepared at 100 μg/mL in H_2O or MeOH according to the supplier's recommendations. We selected 51 NPs eluting from 2 to 18 min as a first test mixture to construct DB-level 1 annotation. The second test mixture was set up using 167 standards from the IROA Mass Spectrometry Library of Standards (Sigma-Aldrich, Darmstadt, Germany) (see Supplementary Text S1 for details).

Plant Material. *A. thaliana* (wild-type Col-O) were grown either in hydroponic culture, in plastic pots (high density), or in Jiffy pots. For each growing condition, 200 mg of plant material per sample were collected, placed in a FastPrep tube (MP Biomedicals Lysing Matrix D, Illkirch, France), and frozen in liquid nitrogen. For extraction, each sample was ground with a Mixer Mill MM 400 grinder (Retsh, Eragny sur Oise, France) by applying two cycles of 30 s at 30 m/s. Biphasic sample extraction was adapted from Salem et al., 2016.[26] Samples were filtered through 0.2 μm PTFE filters (Thermo Scientific) and transferred to vials. An extraction blank (without plant material) and a QC (Quality Control) sample (aliquot of all samples) were also prepared to validate the LC−MS profiles.

Germinated seedlings of *Medicago truncatula* were transferred onto M medium[24] and then placed in a growth chamber at 22 °C and 50% humidity with cycles of 16 h light−8 h dark for 14 days. The roots were ground with a Mixer Mill MM 400 grinder by applying two cycles of 30 s at 300 Hz. A total of 100 mg of ground tissue was placed in 2 mL FastPrep tubes containing 1.4 mm ceramic spheres (Lysing Matrix D) and extracted with 1 mL of acidified aqueous solution of methanol ($MeOH/H_2O/HCOOH$, 80:19:1). After two cycles of 20 s at 6 m/s in the FastPrep-24 (MP Biomedicals), the samples were centrifuged at 4 °C and 10 000 rpm for 10 min. The supernatants were transferred into vials. An extraction blank and quality control (QC) were also done for extraction and analytical validation (see Supplementary Text S1 for details).

UHPLC−HRMS Profiling. Ultrahigh-performance liquid chromatography−high-resolution MS (UHPLC−HRMS) analyses were performed on a Q Exactive Plus quadrupole mass spectrometer, equipped with a heated electrospray probe (HESI II) coupled to an U-HPLC Ultimate 3000 RSLC system (Thermo Fisher Scientific, Hemel Hempstead, U.K.). Samples were separated on a Luna Omega Polar C18 column (150 mm × 2.1 mm i.d., 1.6 μm, Phenomenex, Sartrouville,

https://dx.doi.org/10.1021/acs.analchem.0c01594
Anal. Chem. 2020, 92, 9971−9981

France) equipped with a guard column. Mobile phase A (MPA) was water with 0.05% formic acid (FA), and mobile phase B (MPB) was acetonitrile with 0.05% FA. The solvent gradient was 0 min, 100% MPA; 1 min, 100% MPA; 22 min, 100% MPB; 25 min, 100% MPB; 25.5 min, 100% MPA; 28 min, 100% MPA. The flow rate was 0.3 mL/min, and the column temperature was set to 40 °C; the autosampler temperature was set to 10 °C, and the injection volume was fixed to 2 μL for standard mixes and plant extracts. Mass detection was performed in positive ionization (PI) and negative ionization (NI) modes at 30 000 resolving power [full width at half-maximum (fwhm) at 400 m/z] for MS1 and 17 500 for MS2 with an automatic gain control (AGC) target of 1×10^6 for full scan MS1 and 1×10^5 for MS2. Ionization spray voltages were set to 3.5 kV (for PI) and 2.5 kV (for NI), and the capillary temperature was set to 256 °C for both modes. The mass scanning range was m/z 70−1050 for standards and m/z 100−1500 for plant extracts. Each full MS scan was followed by data-dependent acquisition of MS/MS spectra for the six most intense ions using stepped normalized collision energy of 20, 40, and 60 eV.

Data Processing. LC−MS data were first processed with MS-DIAL version 4.12. MS1, and MS2 tolerances were set to 0.01 and 0.05 Da, respectively, in centroid mode for each data set. Peaks were aligned on a quality control (QC) reference file with a RT tolerance of 0.1 min and a mass tolerance of 0.015 Da. Minimum peak height was set to 70% below the observed total ion chromatogram (TIC) baseline for a blank injection. MS-DIAL data was cleaned with MS-CleanR by selecting all filters with a minimum blank ratio set to 0.8, a maximum relative standard deviation (RSD) set to 30, and a relative mass defect (RMD) ranging from 50 to 3 000. The maximum mass difference for feature relationships detection was set to 0.005 Da, and the maximum RT difference was set to 0.025 min. The Pearson correlation links were considered only for biological data sets with correlation ≥0.8 and statistically significant α = 0.05. Two peaks were kept in each cluster: the most intense and the most connected. The kept features were annotated with MS-FINDER version 3.26. The MS1 and MS2 tolerances were set to 5 and 15 ppm, respectively. Formula finder were exclusively processed with C, H, O, N, P, and S atoms. DBs based on the genus and the family of the plant species (Supplementary Table S3, Supplementary Table S4, Supplementary Table S7, Supplementary Table S8) being investigated were constituted with the dictionary of natural product (DNP-CRC press, DNP on DVD v. 28.2) and the internal generic databases used were KNApSAcK, PlantCyc, HMDB, LipidMaps, NANPDB, and UNPD. Annotation prioritization was done by ranking genus DB followed by Family DB and then generic DB (internal DB from MS-FINDER). Statistical analysis and mass spectral similarity network were detailed in Supplementary Text S1.

RESULTS AND DISCUSSION

MS-CleanR Workflow and Implementation. MS-CleanR uses as input a MS-DIAL peak list processed in data dependent analysis (DDA) or data independent analysis (DIA) using either positive ionization mode (PI) or negative ionization mode (NI) or both. First, MS-CleanR applies generic filters encompassing blank injection signal subtraction starting from feature height ratio between QC and blank detected signals. A second generic filter, called "ghost blank peaks" is based on the high-background ion drift removal. Optionally, an unusual mass defect can be filtered out. A fourth generic filtering approach is the application of a "relative standard deviation" (RSD) threshold among sample classes. Finally, we introduced a fifth filter based on the "relative mass defect" (RMD) calculation. The RMD is calculated in ppm as [(mass defect/measured monoisotopic mass) × 10^6)]. It can be used to filter compound classes[27] and may remove signals with unexpected RMD values. All these filters are tunable by the user.

The second step involves a feature clustering method based on MS-DIAL peak character estimation algorithm (MS-DIAL-PCE), which aggregates several possible relationships at the same RT range: ion correlation among samples, MS/MS fragments in higher m/z, possible adducts and chromatogram correlations.[24] Optionally, Pearson's correlation between features located in the same RT window (typically 0.025 min) can be added during the clustering process. At this point, each feature is seen as a node within a graph. First, all neutral losses and dimers/heteromers are detected and taken out of consideration. Adduct attribution is based on MS-DIAL-PCE and extended by calculating m/z differences using the adduct list provided within the MS-CleanR package (tunable by the user). To filter out ambiguous adduct relationships, the frequency of apparition of each identified adduct was computed based on all adducts detected within the data set. Each adduct relationship is represented as an edge linking two nodes in our graph, and the weight is given as the product of its nodes' frequencies. The resulting graph allows one to discard misidentified relationships by keeping only the most probable adduct type for each feature. During this step, MS-CleanR can merge PI and NI mode if both acquisition modes are detected. Among each cluster, one to n features (tunable by the user) can be selected for further annotation: the most intense, the most connected, or both. During this study, we obtained the best results by keeping 2 features per cluster with selection based on both intensity and degree.

Then, all selected features are exported to MS-FINDER program for *in silico*-based annotation using hydrogen rearrangement rules (HRR) scoring system. During this step, multiple databases can be queried and each annotation result will be handled by MS-CleanR (see Supplementary Table S3 and Supplementary Table S4 for user-defined DB models).

The final step will merge annotation results with the filtered peak list by prioritizing database annotation depending on user choice. This latter function can greatly improve the annotation accuracy particularly when dealing with taxonomically defined extracts.[28] Optionally, all results can be exported as .msp files for mass spectral similarity networking purposes. For the detailed workflow, see Supplementary Text S1.

Workflow Benchmarking on Pooled Standards. To validate our approach, we benchmarked the MS-CleanR workflow by using a mixture of 51 NPs standards profiled in NI and PI modes with a reverse phase column and a 25 min gradient. The resulting data were compiled in an in-house DB comprising RT, HRMS, and MS/MS fragmentation patterns (DB-level 1 annotation according to the Metabolomics Standards Initiative-MSI[29]). To test whether the workflow retained features arising from unique metabolites and removed useless signals, we compared the results obtained by using a combination of MS-DIAL and MS-FINDER and DB-level 1 annotation to those obtained by using MS-CleanR. For the latter, we created another DB of the same metabolite set including accurate mass, molecular formula, and SMILES

https://dx.doi.org/10.1021/acs.analchem.0c01594
Anal. Chem. 2020, 92, 9971−9981

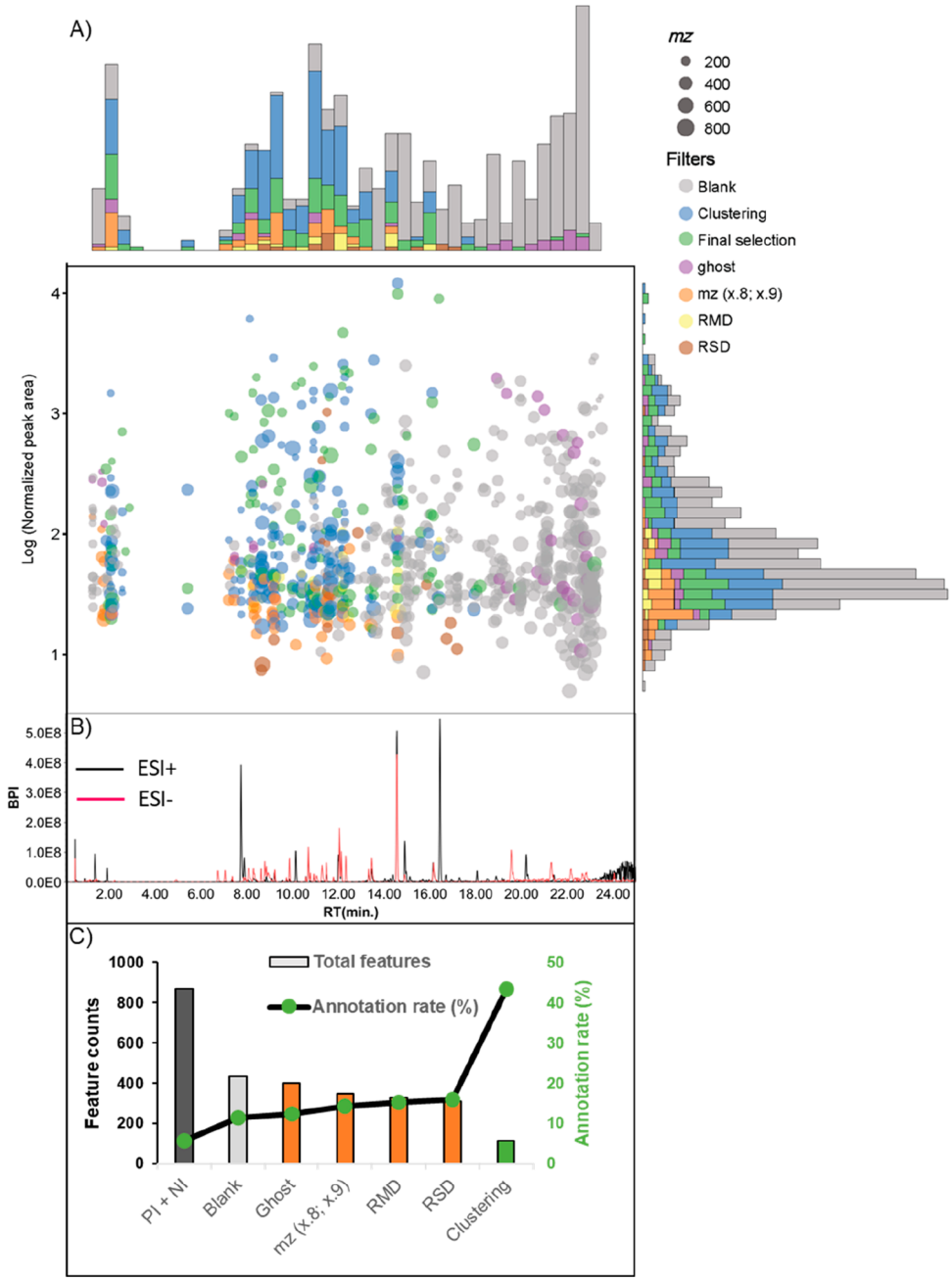


Figure 2. Feature filtering of the LC−MS data set from 51 NPs standards. Generic filters and the feature clustering algorithm were applied to the initial PI + NI mode data set. (A) Scatter plot of LC−MS features (x = RT in minute, y = log of normalized peak area, size = m/z value) colored according to filtering steps. Margin histogram plots display feature count along RT (x) and peak area (y). (B) Base peak chromatogram of LC−MS from 51 NPs standards compounds. (C) Bar plot displays feature counts after successive filters. The line plot displays annotation rate (unique metabolites/feature counts in %).

strings (DB-level 2 annotation according to the MSI) to reproduce real-case annotation processing. All generic filters were used, and the first two features within each cluster (n = 2), either most intense or most connected, were exported for annotation using the "formula prediction and structure elucidation by *in silico* fragmentation tool" in MS-FINDER (Supplementary Table S1).

The behavior of MS-CleanR filtering approach was benchmarked using a combination of both ionization modes acquired from a mixture of 51 NPs standards. The base peak intensity (BPI) chromatogram of the PI and NI modes displayed little overlap as well as a higher signal thickness between 6 and 18 min (Figure 2B). As shown in Figure 2A, the blank filtering process was the most effective after 16 min, within a sprinkling peak area. This result could be explained by lower competition for ion currents which induces higher signals for blank samples. The background ion drift signal filter (namely, "ghost") followed the behavior of blank filters. Unexpected mass signals (namely, "mz (x.8; x.9)") were mainly spread over low intensity signals, e.g., with m/z repeating signal at 158.964 (TFA contaminant) in PI or 112.982 in NI (formic acid dimer contaminant). This filter is particularly useful in

https://dx.doi.org/10.1021/acs.analchem.0c01594
Anal. Chem. 2020, 92, 9971−9981

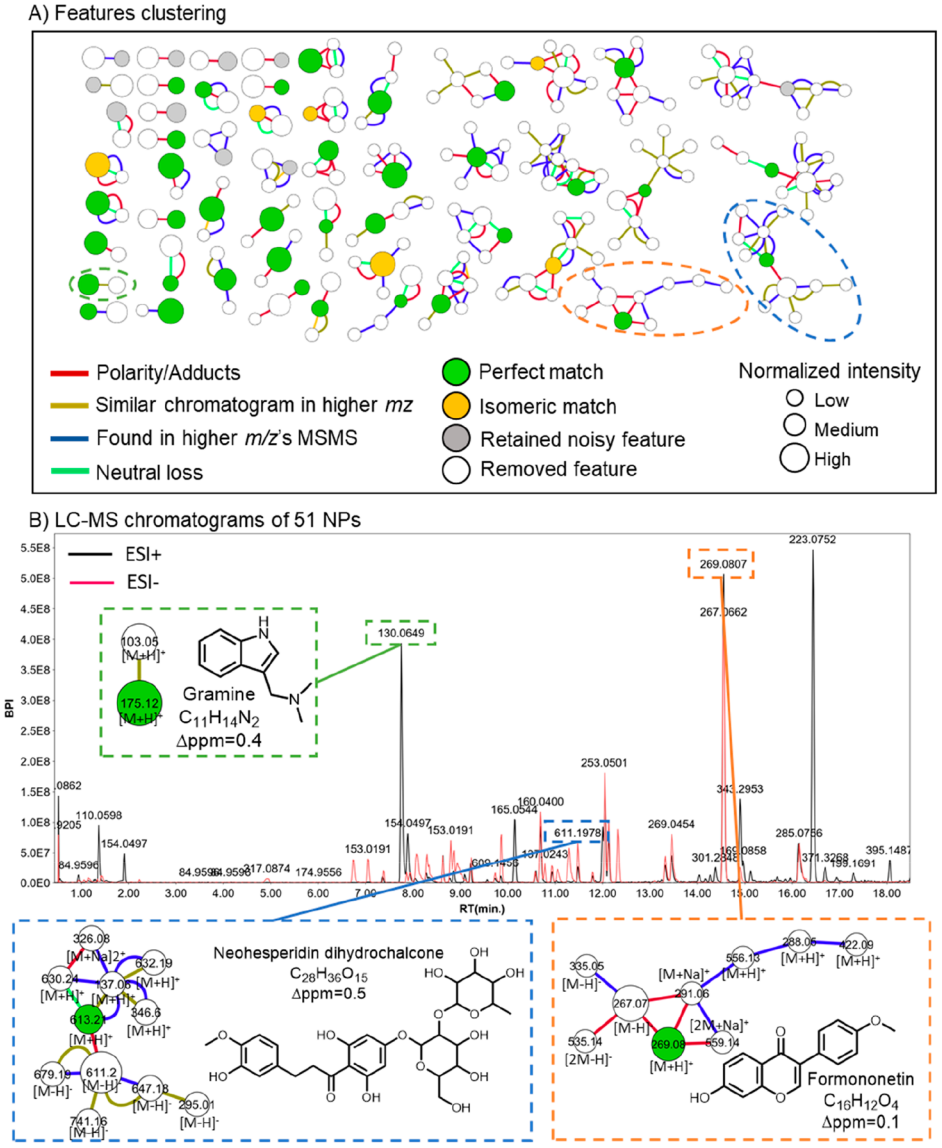


Figure 3. MS-CleanR feature clustering of 51 NPs. Clustering was based on the peak character estimation and multilevel optimization of modularity algorithms. (A) Cluster plot of the whole data set excluding size one clusters. (B) UHPLC−HRMS base peak intensity (BPI) chromatogram of the standard mixture containing 51 NPs. Three representative compounds and their respective clusters are indicated.

signal-rich areas such as the injection peak observed between 1 and 2 min. The RMD filter was kept with default parameter of MS-CleanR which encompass most NPs mass defect values (between 50 and 3000 ppm). The RSD filter was also kept with a default value of 30%, which is generally the deviation value accepted between two consecutive injections of the same sample. Both RMD and RSD were spread between medium to low intensity peaks and removed a low number of signals. Interestingly, the final clustering-based feature selection displayed a repartition from low to high intensity signals all along the chromatogram. According to our observations, the parental feature arising from a unique metabolite signal among each cluster was highlighted by either a PI/NI adduct link in the case where both ionization modes were used (e.g., $[M + H]^+/[M - H]^-$, $[M + Na]^+/[M + FA - H]^-$); the most intense peak of the cluster and/or the peak with the most relationships to other features (i.e., the highest "degree" of connection). A selection based on the most intense and connected feature per cluster avoids the risk of retaining only the most intense signals of each cluster (as shown by the peak pattern at 5 min on Figure 2A). We advise combining both filtering options for optimal feature selection.

As anticipated, we observed a significant feature inflation in this mixture of 51 NPs standards: 869 signals from the PI and NI acquisition modes were detected (Figure 2C). This approximately 95% feature inflation is consistent with a previous report of 10 000−30 000 features detected after injection of 900 unique metabolites,[30] and with a study that used isotope labeling as a feature filtering approach.[11] Blank ratio filtering deleted 50% of the features, and the other generic filters described above removed 15% of the remaining features. Feature clustering caused a further reduction of 18%, resulting in a total of 115 features retained. Overall, the workflow filtered out 80% of all detected signals. By using this approach, there was a remarkable improvement in the annotation rate

https://dx.doi.org/10.1021/acs.analchem.0c01594
Anal. Chem. 2020, 92, 9971−9981

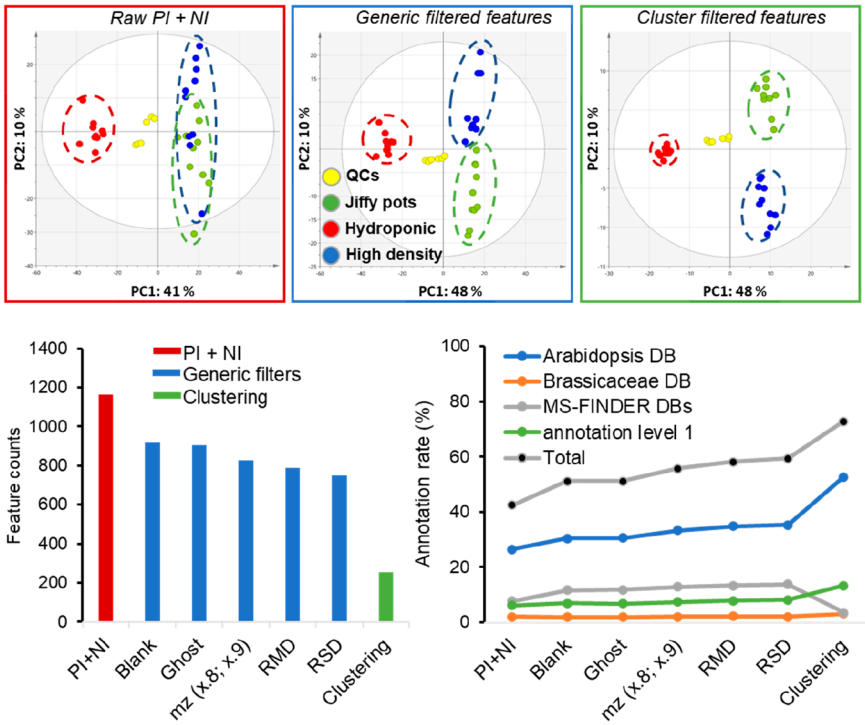


Figure 4. LC−MS data set processing of the metabolomes of *A. thaliana* plants growing in different conditions. Top: Sequential PCA score plots of raw PI and NI modes data set and after applying generic filters and feature clustering. Dotted circles indicate biological sample type distribution (yellow, QC injections; green, plants growing in Jiffy pots at low density; blue, plants growing in plastic pots at high density; red plants in hydroponic culture). Bottom: The bar plot shows the feature counts after successive filtering steps. The line plot displays the annotation rate (unique metabolites/feature counts expressed as %) after successive filtering steps using annotation DBs prioritization.

(unique metabolites/detected features) from 5% to 45% (Figure 2C).

Consequently, 10 metabolites exhibited an isolated *mz*-RT signal whereas 41 metabolites displayed multiple signals grouped into clusters of the same RT range (±0.05 min, depending on MS-DIAL parameters used for RT windows) from 2 to 11 features (Figure 3A). The parental feature arising from a unique metabolite (green dots, Figure 3A) displayed either the most intense signal and/or highest degree of connection to other features among each cluster. It was also noticed that if a PI/NI linkage between two pseudomolecular ions was detected, the parental signal was one of them. The application of these rules allowed the annotation of 50 metabolites, 44 of which matched perfectly with level 1 annotation DB (Supplementary Table S1). The remaining ones were annotated as an isobaric/isomeric match due to prioritization of the highest MS-FINDER scoring value (e.g., 4-aminosalicylic acid and 5-aminosalicylic acid). In the case of gramine, for example, the major pseudomolecular ion had an m/z value of 130.0649 at RT 7.75 min (Figure 3B). By applying feature clustering, we detected an in-source fragment corresponding to the neutral loss of the dimethylamine group at m/z 130.0649. This feature was removed and only the signal at m/z 103.054 and m/z 175.1228 were retained. Since the signal at m/z 103.054 displayed a similar chromatogram to the most intense signal at m/z 175.1228, this last was exported and annotated as gramine (Δppm = 0.4) with a perfect match. The detected signals at 11.47 min; m/z 611.1983 $[M - H]^-$ and 11.48 min; m/z 613.2122 $[M + H]^+$ in NI and PI modes, respectively, were grouped in a cluster of 11 features. Feature relationships among this cluster were mainly related to similar MS/MS patterns among multiple features at the same RT range in the PI mode while similar chromatograms were mainly detected in NI mode (Figure 3B). In this case, feature selection was driven by the PI/NI adduct relationship and the feature with the highest MS-Finder annotation score was retained in the final peak list and identified as neohesperidin dihydrochalcone (Δppm = 0.4). Formononetin displayed a pseudomolecular ion $[M + H]^+$ in PI (RT = 14.55; m/z at 269.0806) and $[M - H]^-$ in NI (RT = 14.55; m/z at 267.0662) modes. The formononetin cluster encompassed complex adduct relationships in the NI mode and multiple similar MS/MS patterns in the PI mode. The detection of the most intense and connected features among this cluster added to the PI/NI relationship has conducted the selection of signals at m/z 269.0806 for annotation, which provided a perfect match with level 1 annotation DB. The only mismatch was encountered for phloridzin due to the neutral loss of a glucose moiety in both PI and NI modes. Only genine was detected in PI mode, resulting in selection of this signal in the final peak list.

To more closely model a real biological sample, we standardized our workflow by using a mixture of 167 standard compounds from the IROA Mass Spectrometry library (Supplementary Table S2). As above, we found significant feature inflation: 6732 signals after concatenation of PI and NI data sets (Figure S2). Unlike the standardization with NPs, above, the generic filters removed only 15% of features. The most important improvement was obtained by feature clustering, which filtered out 90% of the detected features leaving 611 signals. Among these, 127 features were identified as a perfect match compared to Level-1 annotation DB, and 21 were annotated as an isomeric match (Supplementary Table S2). Twelve features were removed due to their coelution with

https://dx.doi.org/10.1021/acs.analchem.0c01594
Anal. Chem. 2020, 92, 9971−9981

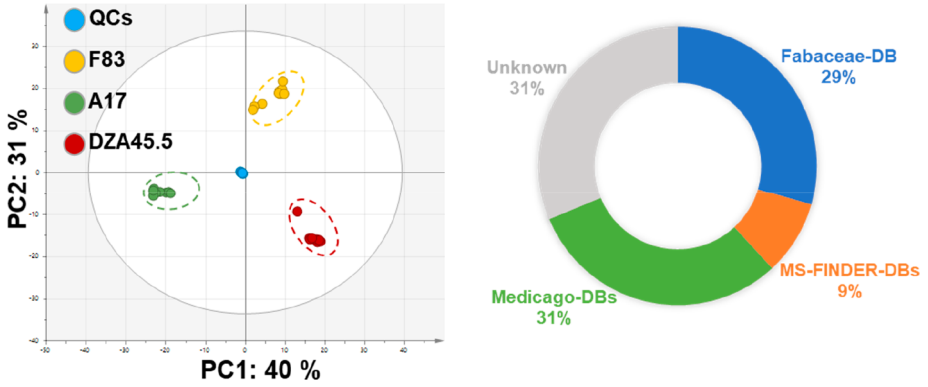


Figure 5. LC−MS NI data set processing of the metabolome of roots from three strains of *M. truncatula*. Left: PCA score plot after applying the MS-CleanR workflow. Dotted circles enclose samples from each plant strain. Right: Circular plot of the proportions of features annotated with reference to the indicated databases (DB).

other compounds, and four had a significant RT shift due to their poor peak shapes. The final three compounds were not annotated because of neutral loss of the same moiety in PI and NI modes, which led to their misidentification. Overall, the annotation rate with this workflow was 27% (Figure S2), and 90% of unique metabolites were retained.

Evaluation of MS-CleanR on Biological Samples. To evaluate the utility of the workflow on a real data set, we set up an experiment to compare metabolome changes in *Arabidopsis thaliana* plants due to different culture conditions and the age of the plants. Three cultural conditions were assessed (low-density growth in Jiffy pots for 32 days, high-density growth in plastic pots for 21 days, and hydroponic culture in liquid MS medium for 14 days), and 10 biological replicates were analyzed per culture condition. At harvest time, 4 leaves (2 cotyledons and 2 leaves) were observed for hydroponic plants, the densely seeding plants showed not more than two small, but completed, developed leaves, while the jiffy growing plants harbored large and well-developed rosette leaves. Extracts were made from the aerial parts of the plants grown in pots and from the roots and green tissues of plants in hydroponic culture, and the extracts were profiled by LC−MS. The data sets acquired in PI and NI modes were treated using the MS-CleanR workflow with default parameters (see Methods). Sequential principal component analysis (PCA) was used to provide an unsupervised overview of the LC−MS fingerprints resulting from the generic filters and feature clustering (Figure 4). The PCA score plot of raw PI and NI mode data displayed 51% of total explained variance using the first two principal components. QC samples appeared in the center of the PCA score plot, demonstrating the reproducibility of the LC−MS analysis. As expected, the youngest plants growing hydroponically were completely separate on the first principal component (PC1) axis from the older plants growing in pots. The plants growing in Jiffy pots and plastic pots could not be distinguished in the raw data set, and after the generic filter step, the data formed more distinct clusters; the total explained variance was slightly improved (58%), and the number of features decreased by 35% (Figure 4). After the feature clustering step, the number of features was reduced by 80%. All data sets were annotated with in-lab DB (level 1) and with MS-FINDER (level 2) by reference to external DBs of *Arabidopsis* (Supplementary Table S3) and Brassicaceae compounds (Supplementary Table S4) and an internal MS-FINDER plant-related DB (comprising PlantCyc, KNApSAcK, HMDB, LIPID MAPS, and UNPD). In the raw PI and NI data set exported from MS-DIAL (1163 features), 42% of all features were annotated, 26% of them appeared in the *Arabidopsis* DB, 2% in the Brassicaceae DB, 7% in the internal MS-FINDER DBs, and 6% with in-lab DB (Figure 4); 58% of all features were unidentified. The generic filters removed 15% of all features and increased the annotation rate to 59%. Feature clustering drastically reduced the number of features (254 m/z × RT pairs) and increased the annotation rate to 73%. Using annotation DB prioritization, 53% of retained features were annotated in the *Arabidopsis* genus and 13% at level 1 with in-lab DB; only 27% remained unidentified. Orthogonal projections to latent structures-discriminant analysis (OPLS-DA) of the most highly ranked features identified three amino acids (oxoproline, citrulline, and glutamine) that discriminate between growth in pots and hydroponic growth (Supplementary Table S5). This may be related to differences in nitrogen availability in the hydroponics medium and in potting soil.

Metabolic Profiling with MS-CleanR. Untargeted metabolomic profiling has emerged as the method of choice to identify metabolic markers associated with beneficial traits in plants, such as resistance to biotic stresses. In this context, the MS-CleanR workflow could greatly improve the results of untargeted metabolomics. To illustrate this point, we used as models the legume *Medicago truncatula* and the pathogenic oomycete *Aphanomyces euteiches*, a major pathogen of several legume species.[31] Genome-wide association studies of 179 lines of *M. truncatula* have identified major loci involved in the resistance of the plant to *A. euteiches*.[32] Moreover, genes encoding enzymes involved in the synthesis of antimicrobial metabolites are expressed in uninfected plants.[33] This suggests that antimicrobial metabolites in uninfected plants may be useful biomarkers with which to select legume lines resistant to *A. euteiches*. To identify these metabolites, we applied the MS-CleanR workflow to analyze the metabolomes of roots from three different strains of *M. truncatula* that have different levels of resistance to *A. euteiches* infection: strain DZA45.5 has the highest level of resistance, A17 is an intermediate level, and F83 is the most susceptible. These three strains were analyzed by LC−MS in NI mode and potential biomarkers were highlighted by multivariate data analysis (Supplementary Table S6). The metabolites that were differentially produced in the two most resistant strains (A17 and DZA45.5) when compared

https://dx.doi.org/10.1021/acs.analchem.0c01594
Anal. Chem. 2020, 92, 9971−9981

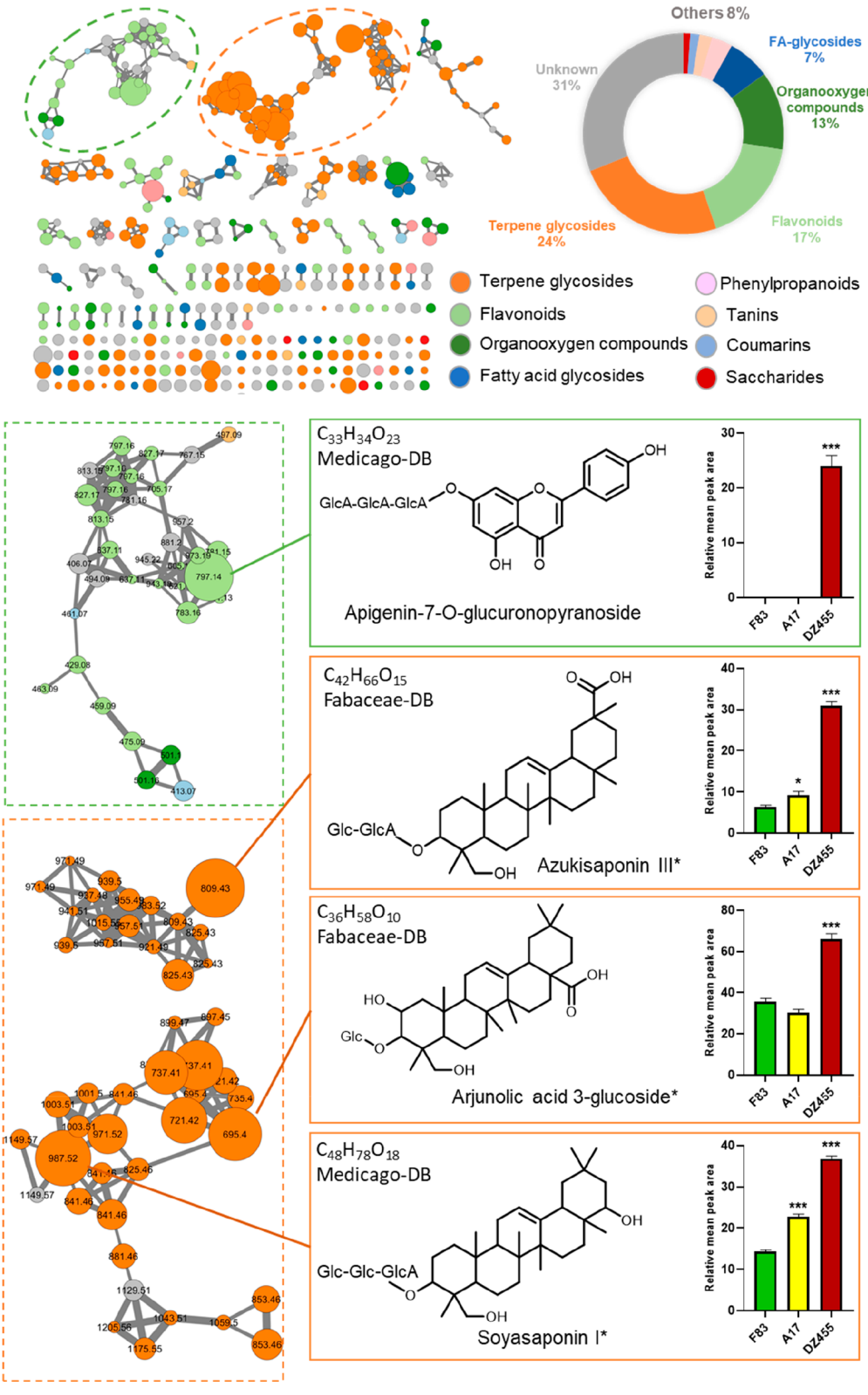


Figure 6. Mass spectral similarity network of *M. truncatula* NI data set (cosine ≥ 0.8). Nodes are colored according to their chemical classes and sized relative to their OPLS regression coefficient score (see text for details). The edge width is proportional to the cosine value. The pie chart displays the annotated chemical class ratio in the LC−MS NI data set (others include the coumarin derivatives, tannins, and saccharides chemical classes). Bar plots display normalized mean peak areas for the four most highly ranked structures by OPLS-regression modeling (Supplementary Table S6). One-way ANOVA and Dunnett's posthoc test ($p \leq 0.05$) were used to assess differences between the sensitive (F83) and resistant (A17 and DZA45.5) *M. truncatula* strains (*$p \leq 0.05$, **$p \leq 0.01$, ***$p \leq 0.001$). Compound names with an asterisk indicate an isobaric annotation match with ref 34 (Glc, glucoside; GlcA, glucuronopyranoside).

https://dx.doi.org/10.1021/acs.analchem.0c01594
Anal. Chem. 2020, 92, 9971−9981

to the more sensitive one (F83) were identified by OPLS regression.

After application of the MS-CleanR workflow, the PCA score plot showed a net clustering of the samples from each strain of *M. truncatula*. QC samples were centered on the PCA plot, demonstrating very good reproducibility (Figure 5). When annotated by reference to DBs from *Medicago* or the legume family Fabaceae, 60% of the data set was annotated (Figure 5) and an additional 9% with MS-FINDER DBs. A molecular spectral similarity network was built to highlight common chemical classes related to resistance traits (Figure 6). Among all annotated features, flavonoids and terpene glycosides compounds were prevalent. This latter class included primarily triterpene sapogenins which appeared to be highly correlated to the resistance traits according to the OPLS regression model. In particular, the four top ranked compounds belonged to two clusters related to sapogenins and one to flavonoids. Our untargeted approach revealed the presence of apigenin-7-*O*-glucuronopyranoside (best MS-FINDER score among several possible match in flavonoid class) only in the resistant DZ45.5 strain. This result corroborated a previous study by our group which demonstrated the implication of flavonoid pathways in resistance.[33] However, other detected flavonoids were not correlated to the resistance contrary to sapogenins class. Among the 151 terpene glycosides annotated in this study, 36 were also identified by a large-scale sapogenin profiling study in various ecotypes of *M. truncatula*[34] (Supplementary Table S6). Interestingly, the three-top ranked sapogenins by the OPLS model (azukisaponin III, arjunolic acid 3-glucoside, and soyasaponin I) displayed an isobaric match with two hederagenin glycosides and a bayogenin derivative, respectively, annotated by Sumner and colleagues. These sapogenins accumulate primarily in the roots rather than the leaves. These organs, however, have distinct profiles of specific saponins, which may be explained by the adaptation of each ecotype to its biotic environment. A previous study showed that saponins derived from hederagenin glycoside in *M. truncatula* have antifungal activity.[35] Our study confirmed a higher level of these compounds in the strains resistant to *A. euteiches* (DZA45.5 and A17) than in the sensitive strain F83. Although the relevance of saponins to resistance of *M. truncatula* to *A. euteiches* has not been confirmed, these findings demonstrate the potential value of applying metabolomics tools to identify biomarkers of plant resistance.

CONCLUSIONS

The main goal of LC−MS-based untargeted metabolomics is to convert chromatographic profiles of complex biological extracts into a comprehensive metabolite list. We demonstrate here that feature degeneracy has a great effect on the final annotated peak list, thus impacting biological knowledge mined from untargeted metabolomic studies. We estimate, based on analysis of standard mixtures, that feature inflation is close to 95%, in agreement with other studies.[11,30] Our package MS-CleanR, with its point-and-click software on a Shiny interface, is a new component in the suite of tools comprising the GUI software MS-DIAL and the annotation capabilities of MS-FINDER. Together, these provide a comprehensive workflow, from raw data to final annotated peak list. MS-CleanR can reduce the number of features by 80−90% and keep most unique metabolite signals without compromising the final data structure. Several parameters are tunable by the user, including the generic filters values and the selection method after feature clustering. The default values displayed in the MS-CleanR interface are a good starting point, but we advise careful inspection of the raw data to set the RSD value as well as the blank filtering threshold. The RMD window can be tuned based on the matrices component knowledge under study. The addition of all generic filters and Pearson correlation clustering extension will provide sharp data set filtering, thus focusing on high quality peaks. Conversely, a wider filtered data set could be obtained by removing some or all of the filters. (Additional tips can be found in the tutorials available at https://github.com/eMetaboHUB/MS-CleanR). The package is also able to combine the PI and NI modes (*A. thaliana* experiment) or to treat only one mode independently (*M. truncatula* study) depending on the study objectives. The opportunity to rank the annotation results with reference to user-defined databases narrows down the final identification possibilities. Each MS-CleanR step takes from a few seconds to a few minutes, even for the clustering and feature selection stage which use a computationally fast heuristic approach.[25] Another advantage of feature filtering prior to MS-FINDER annotation is the drastic reduction of processing time which is divided by several orders of magnitude. We demonstrate the utility of this workflow by analyzing secondary metabolites levels in three *M. truncatula* strains with different susceptibilities to a pathogenic oomycete. We annotated 70% of the data set with 60% at the genus or family level using DBs prioritization. The resulting mass spectral similarity network further supports annotation results as most clusters gathered the same metabolite chemical class. Still, our approach was unable to keep only unique metabolite features regarding the annotation rate comprising between 24 and 45% for standard mixtures. A limitation of our filtering process is its dependence on chromatographic resolution, which can seriously impair the final results by clustering several unique metabolites together. In the present study, we chose a 20 min gradient, like those generally applied in most untargeted metabolomics studies. Extending the elution time might improve the chromatographic resolution but is difficult to apply in day-to-day work, especially for high-throughput experiments. These challenges will be addressed in future developments of MS-CleanR.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.analchem.0c01594.

Supplementary Text S1, detailed workflow of MS-CleanR package and materials used in this study (PDF)

Figure S1, alignment spot screenshot in ESI PI and NI ionization modes showing repeated blank pseudomolecular ions detected in QCs samples with a retention time shift; Figure S2, feature filtering of LC−MS data set from 167 IROA-MS standards library according to generic filters and clustering algorithm (PDF)

Supplementary Table S1, Excel table with cluster annotation, result summary and database for level 2 annotation imported in MS-FINDER for the 51 NPs data set (XLSX)

Supplementary Table S2, Excel table with cluster annotation, result summary and database for level 2

https://dx.doi.org/10.1021/acs.analchem.0c01594
Anal. Chem. 2020, 92, 9971−9981

annotation imported in MS-FINDER for the 167 IROA MS standard library data set (XLSX)

Supplementary Table S3, *Arabidopsis* DB used as input for level 2 annotation in MS-FINDER (ZIP)

Supplementary Table S4, *Brassicaceae* DB used as input for level 2 annotation in MS-FINDER (ZIP)

Supplementary Table S5, *Arabidopsis* data set treated by MS-CleanR and OPLS-DA top ranked features (XLSX)

Supplementary Table S6, *Medicago* data set treated by MS-CleanR and OPLS regression coefficient for feature ranking (XLSX)

Supplementary Table S7, *Medicago* DB used as input for level 2 annotation in MS-FINDER (ZIP)

Supplementary Table S8, *Fabaceae* DB used as input for level 2 annotation in MS-FINDER (ZIP)

AUTHOR INFORMATION

Corresponding Author

Guillaume Marti − *Pharma Dev, Université de Toulouse, IRD, UPS, 31400 Toulouse, France; Laboratoire de Recherche en Sciences Végétales and Metatoul-AgromiX Platform, MetaboHUB, National Infrastructure for Metabolomics and Fluxomics, LRSV, Université de Toulouse, CNRS, UPS, 31400 Toulouse, France; Institut de Recherche en Informatique de Toulouse, Université de Toulouse, UPS, Toulouse 31400, France;* orcid.org/0000-0002-6321-9005; Phone: (+33) 534 32 38 31; Email: guillaume.marti@univ-tlse3.fr

Authors

Ophélie Fraisier-Vannier − *Pharma Dev, Université de Toulouse, IRD, UPS, 31400 Toulouse, France; Institut de Recherche en Informatique de Toulouse, Université de Toulouse, UPS, Toulouse 31400, France*

Justine Chervin − *Laboratoire de Recherche en Sciences Végétales and Metatoul-AgromiX Platform, MetaboHUB, National Infrastructure for Metabolomics and Fluxomics, LRSV, Université de Toulouse, CNRS, UPS, 31400 Toulouse, France*

Guillaume Cabanac − *Institut de Recherche en Informatique de Toulouse, Université de Toulouse, UPS, Toulouse 31400, France;* orcid.org/0000-0003-3060-6241

Virginie Puech − *Laboratoire de Recherche en Sciences Végétales and Metatoul-AgromiX Platform, MetaboHUB, National Infrastructure for Metabolomics and Fluxomics, LRSV, Université de Toulouse, CNRS, UPS, 31400 Toulouse, France*

Sylvie Fournier − *Laboratoire de Recherche en Sciences Végétales and Metatoul-AgromiX Platform, MetaboHUB, National Infrastructure for Metabolomics and Fluxomics, LRSV, Université de Toulouse, CNRS, UPS, 31400 Toulouse, France*

Virginie Durand − *Laboratoire de Recherche en Sciences Végétales, Université de Toulouse, CNRS, UPS, 31400 Toulouse, France*

Aurélien Amiel − *Laboratoire de Recherche en Sciences Végétales, Université de Toulouse, CNRS, UPS, 31400 Toulouse, France; De Sangosse, Bonnel, 47480 Pont-Du-Casse, France*

Olivier André − *Laboratoire de Recherche en Sciences Végétales, Université de Toulouse, CNRS, UPS, 31400 Toulouse, France; De Sangosse, Bonnel, 47480 Pont-Du-Casse, France*

Omar Abdelaziz Benamar − *Laboratoire de Recherche en Sciences Végétales, Université de Toulouse, CNRS, UPS, 31400 Toulouse, France*

Bernard Dumas − *Laboratoire de Recherche en Sciences Végétales, Université de Toulouse, CNRS, UPS, 31400 Toulouse, France*

Hiroshi Tsugawa − *RIKEN Center for Sustainable Resource Science, Yokohama 230-0045, Japan; RIKEN Center for Integrative Medical Science, Yokohama 230-0045, Japan;* orcid.org/0000-0002-2015-3958

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.analchem.0c01594

Notes

The authors declare no competing financial interest.

Raw data from *Arabidopsis* and *Medicago* LC−MS profiling available on Zenodo.org using the DOI 10.5281/zenodo.3744480.

ACKNOWLEDGMENTS

We thank Dr. Stephane Bertani for providing us the standards from the IROA-MS Library. Financial support was received from the French National Infrastructure for Metabolomics and Fluxomics, Grant MetaboHUB-ANR-11-INBS-0010, and the PSPC SOLSTICE Project (SOLutionS pour des Traitements Intégrés dans une Conduite Environnementale) managed by Belchim Crop Protection and partly funded by the French state within the framework of the Programme d'Investissements d'Avenir. We thank E. Amblard, N. Jariais, and C. Jacquet for the *M. truncatula* cultures and A. Haouy for the *A. thalina* cultures and sample preparations. We also acknowledge Carol Featherstone of Plume Scientific Communication Services for the professional scientific editing.

REFERENCES

(1) Zierer, J.; Jackson, M. A.; Kastenmüller, G.; Mangino, M.; Long, T.; Telenti, A.; Mohney, R. P.; Small, K. S.; Bell, J. T.; Steves, C. J.; Valdes, A. M.; Spector, T. D.; Menni, C. *Nat. Genet.* **2018**, *50* (6), 790−795.

(2) Li, H.; Ning, S.; Ghandi, M.; Kryukov, G. V.; Gopal, S.; Deik, A.; Souza, A.; Pierce, K.; Keskula, P.; Hernandez, D.; Ann, J.; Shkoza, D.; Apfel, V.; Zou, Y.; Vazquez, F.; Barretina, J.; Pagliarini, R. A.; Galli, G. G.; Root, D. E.; Hahn, W. C.; Tsherniak, A.; Giannakis, M.; Schreiber, S. L.; Clish, C. B.; Garraway, L. A.; Sellers, W. R. *Nat. Med.* **2019**, *25* (5), 850−860.

(3) Gargallo-Garriga, A.; Sardans, J.; Pérez-Trujillo, M.; Oravec, M.; Urban, O.; Jentsch, A.; Kreyling, J.; Beierkuhnlein, C.; Parella, T.; Peñuelas, J. *New Phytol.* **2015**, *207* (3), 591−603.

(4) Mohimani, H.; Gurevich, A.; Shlemov, A.; Mikheenko, A.; Korobeynikov, A.; Cao, L.; Shcherbin, E.; Nothias, L.-F.; Dorrestein, P. C.; Pevzner, P. A. *Nat. Commun.* **2018**, *9*, 4035 DOI: 10.1038/s41467-018-06082-8.

(5) Wang, S.; Alseekh, S.; Fernie, A. R.; Luo, J. *Mol. Plant* **2019**, *12* (7), 899−919.

(6) Huan, T.; Forsberg, E. M.; Rinehart, D.; Johnson, C. H.; Ivanisevic, J.; Benton, H. P.; Fang, M.; Aisporna, A.; Hilmers, B.; Poole, F. L.; Thorgersen, M. P.; Adams, M. W. W.; Krantz, G.; Fields, M. W.; Robbins, P. D.; Niedernhofer, L. J.; Ideker, T.; Majumder, E. L.; Wall, J. D.; Rattray, N. J. W.; Goodacre, R.; Lairson, L. L.; Siuzdak, G. *Nat. Methods* **2017**, *14* (5), 461−462.

(7) Giacomoni, F.; Le Corguille, G.; Monsoor, M.; Landi, M.; Pericard, P.; Petera, M.; Duperier, C.; Tremblay-Franco, M.; Martin, J.-F.; Jacob, D.; Goulitquer, S.; Thevenot, E. A.; Caron, C. *Bioinformatics* **2015**, *31* (9), 1493−1495.

(8) Pluskal, T.; Castillo, S.; Villar-Briones, A.; Orešič, M. *BMC Bioinf.* **2010**, *11* (1), 395.

(9) Tsugawa, H.; Cajka, T.; Kind, T.; Ma, Y.; Higgins, B.; Ikeda, K.; Kanazawa, M.; VanderGheynst, J.; Fiehn, O.; Arita, M. *Nat. Methods* **2015**, *12* (6), 523−526.
(10) Patti, G. J.; Yanes, O.; Siuzdak, G. *Nat. Rev. Mol. Cell Biol.* **2012**, *13* (4), 263−269.
(11) Mahieu, N. G.; Patti, G. J. *Anal. Chem.* **2017**, *89* (19), 10397−10406.
(12) Dührkop, K.; Fleischauer, M.; Ludwig, M.; Aksenov, A. A.; Melnik, A. V.; Meusel, M.; Dorrestein, P. C.; Rousu, J.; Böcker, S. *Nat. Methods* **2019**, *16* (4), 299−302.
(13) Tsugawa, H.; Kind, T.; Nakabayashi, R.; Yukihira, D.; Tanaka, W.; Cajka, T.; Saito, K.; Fiehn, O.; Arita, M. *Anal. Chem.* **2016**, *88* (16), 7946−7958.
(14) Ruttkies, C.; Schymanski, E. L.; Wolf, S.; Hollender, J.; Neumann, S. *J. Cheminf.* **2016**, *8* (1), 3.
(15) Djoumbou-Feunang, Y.; Pon, A.; Karu, N.; Zheng, J.; Li, C.; Arndt, D.; Gautam, M.; Allen, F.; Wishart, D. S. *Metabolites* **2019**, *9* (4), 72.
(16) Tikunov, Y. M.; Laptenok, S.; Hall, R. D.; Bovy, A.; de Vos, R. C. H. *Metabolomics* **2012**, *8* (4), 714−718.
(17) Broeckling, C. D.; Afsar, F. A.; Neumann, S.; Ben-Hur, A.; Prenni, J. E. *Anal. Chem.* **2014**, *86* (14), 6812−6817.
(18) DeFelice, B. C.; Mehta, S. S.; Samra, S.; Čajka, T.; Wancewicz, B.; Fahrmann, J. F.; Fiehn, O. *Anal. Chem.* **2017**, *89* (6), 3250−3255.
(19) Edmands, W. M. B.; Petrick, L. M.; Barupal, D. K.; Scalbert, A.; Wilson, M.; Wickliffe, J.; Rappaport, S. M. *Anal. Chem.* **2017**, *89*, 3919.
(20) Uppal, K.; Walker, D. I.; Jones, D. P. *Anal. Chem.* **2017**, *89* (2), 1063−1067.
(21) Kuhl, C.; Tautenhahn, R.; Böttcher, C.; Larson, T. R.; Neumann, S. *Anal. Chem.* **2012**, *84* (1), 283−289.
(22) Daly, R.; Rogers, S.; Wandy, J.; Jankevics, A.; Burgess, K. E. V.; Breitling, R. *Bioinformatics* **2014**, *30* (19), 2764−2771.
(23) Senan, O.; Aguilar-Mogas, A.; Navarro, M.; Capellades, J.; Noon, L.; Burks, D.; Yanes, O.; Guimerà, R.; Sales-Pardo, M. *Bioinformatics* **2019**, *35* (20), 4089−4097.
(24) Tsugawa, H.; Nakabayashi, R.; Mori, T.; Yamada, Y.; Takahashi, M.; Rai, A.; Sugiyama, R.; Yamamoto, H.; Nakaya, T.; Yamazaki, M.; Kooke, R.; Bac-Molenaar, J. A.; Oztolan-Erol, N.; Keurentjes, J. J. B.; Arita, M.; Saito, K. *Nat. Methods* **2019**, *16* (4), 295−298.
(25) Blondel, V. D.; Guillaume, J.-L.; Lambiotte, R.; Lefebvre, E. *J. Stat. Mech.: Theory Exp.* **2008**, *2008* (10), P10008.
(26) Salem, M. A.; Jüppner, J.; Bajdzienko, K.; Giavalisco, P. *Plant Methods* **2016**, *12*, 45 DOI: 10.1186/s13007-016-0146-2.
(27) Ekanayaka, E. A. P.; Celiz, M. D.; Jones, A. D. *Plant Physiol.* **2015**, *167* (4), 1221−1232.
(28) Rutz, A.; Dounoue-Kubo, M.; Ollivier, S.; Bisson, J.; Bagheri, M.; Saesong, T.; Ebrahimi, S. N.; Ingkaninan, K.; Wolfender, J.-L.; Allard, P.-M. *Front. Plant Sci.* **2019**, *10*, 1329.
(29) Creek, D. J.; Dunn, W. B.; Fiehn, O.; Griffin, J. L.; Hall, R. D.; Lei, Z.; Mistrik, R.; Neumann, S.; Schymanski, E. L.; Sumner, L. W.; Trengove, R.; Wolfender, J.-L. *Metabolomics* **2014**, *10* (3), 350−353.
(30) Li, Z.; Lu, Y.; Guo, Y.; Cao, H.; Wang, Q.; Shui, W. *Anal. Chim. Acta* **2018**, *1029*, 50−57.
(31) Gaulin, E.; Jacquet, C.; Bottin, A.; Dumas, B. *Mol. Plant Pathol.* **2007**, *8* (5), 539−548.
(32) Bonhomme, M.; André, O.; Badis, Y.; Ronfort, J.; Burgarella, C.; Chantret, N.; Prosperi, J.-M.; Briskine, R.; Mudge, J.; Debéllé, F.; Navier, H.; Miteul, H.; Hajri, A.; Baranger, A.; Tiffin, P.; Dumas, B.; Pilet-Nayel, M.-L.; Young, N. D.; Jacquet, C. *New Phytol.* **2014**, *201* (4), 1328−1342.
(33) Badis, Y.; Bonhomme, M.; Lafitte, C.; Huguet, S.; Balzergue, S.; Dumas, B.; Jacquet, C. *Mol. Plant Pathol.* **2015**, *16* (9), 973−986.
(34) Lei, Z.; Watson, B. S.; Huhman, D.; Yang, D. S.; Sumner, L. W. *Front. Plant Sci.* **2019**, *10*, 850 DOI: 10.3389/fpls.2019.00850.
(35) Abbruscato, P.; Tosi, S.; Crispino, L.; Biazzi, E.; Menin, B.; Picco, A. M.; Pecetti, L.; Avato, P.; Tava, A. *J. Agric. Food Chem.* **2014**, *62* (46), 11030−11036.

https://dx.doi.org/10.1021/acs.analchem.0c01594
Anal. Chem. 2020, 92, 9971−9981

Phytochemistry 163 (2019) 99–110

Contents lists available at ScienceDirect

Phytochemistry

journal homepage: www.elsevier.com/locate/phytochem

Deciphering the phylogeny of violets based on multiplexed genetic and metabolomic approaches

Justine Chervin[a,b,c], Thierry Talou[a], Marjorie Audonnet[b], Bernard Dumas[b], Laurent Camborde[b], Marie-Thérèse Esquerré-Tugayé[b], Christophe Roux[b], Guillaume Cabanac[d], Guillaume Marti[c,*]

[a] Laboratoire de Chimie Agro-industrielle, LCA, Université de Toulouse, INRA, Toulouse, France
[b] Laboratoire de Recherche en Sciences Végétales [LRSV], UMR 5546, UPS/CNRS, Toulouse, France
[c] UMR 152 PharmaDEV, Université de Toulouse, IRD, UPS, Toulouse, France
[d] UMR 5505 IRIT, Université de Toulouse, UPS, Toulouse, France

ARTICLE INFO

Keywords:
Biomarkers
Chemotaxonomy
GC-MS
Metabolomic
Phylogeny
UHPLC-HRMS
Violets

ABSTRACT

Molecular phylogenetics based on nucleotide sequence comparisons has profoundly influenced plant taxonomy. A comprehensive chemotaxonomical approach based on GC-MS and UHPLC-HRMS profiling was evaluated for its ability to characterize a large collection of plants all in the violet family Violaceae (n = 111) and thus decipher the taxonomy. A thorough identification of violets is challenging due to their natural hybridization and phenotypic variability. Phylogenetic inference performed on ribosomal internal transcribed spacer sequences using maximum likelihood and neighbor-joining distance methods allowed the clear identification of 58% of the collection. Metabolomic approaches with multivariate data analysis were performed on SPME/GC-MS chromatograms of volatile compounds emitted by fresh mature flowers and on UHPLC-HRMS/MS leaf extracts for non-volatile compounds. Interestingly, molecular and biochemical approaches provided separate classifications while highlighting several common clusters. The profiling of secondary metabolites was proved most suitable for the classification of hundreds of extracts. The combination of phylogenetic and chemotaxonomic approaches, allowed the classification of 96% of the entire collection. A correlation network revealed specific chemotaxonomic biomarkers, in particular flavonoids, coumarins and cyclotides. Overall, our pioneering approach could be useful to solve misclassification issues within collections of close plant species.

1. Introduction

The Violaceae family of plants contains 23 genera and approximately 900 species distributed all over the world. Its three largest genera represent 98% of its species, including *Viola* with approximately 600 species. This genus is separated into two main groups (sections) widely distributed in the Northern Hemisphere (Ballard et al., 1999): violets in the *Viola* section and pansies in the *Melanium* section (Yockteng et al., 2003). *Viola* are perennial herbs with very short stems, often stolons and typically zygomorphous flowers.

The delimitation of *Viola* species into sections and subsections can be problematic in particular when using morphological characteristics. This is due to hybridization, which is relatively common in vascular plants and has a large impact on speciation events (Erben, 1996). Such hybridization can contribute to rapid diversification of many plant lineages with karyotypic and genomic changes (Arnold, 2006) but can also lead to the demise of rare species (Todesco et al., 2016). Hybridization may reduce a population's growth rate by the production of hybrid seed, which is produced at the expense of conspecific (pure) seed when they competed with them for resources or they possess better vigor and fertility than parents, for instance (Levin et al., 1996). In the past, taxonomists have studied the phylogeny of *Viola* based on anatomical characteristics (Marcussen et al., 2012), chromosome numbers (Clausen, 1927), nuclear ribosomal sequences (Ballard et al., 1999) (Mereda et al., 2011) and chloroplast DNA as complementary tools in particular to detect maternal lineages (Cennamo et al., 2011). Within nuclear ribosomal sequences, internal transcribed spacer (ITS) regions are highly popular in phylogenetic studies (Álvarez, 2003) and have been efficiently used to differentiate plant species. Indeed, they are easily amplified by polymerase chain reaction (PCR) using universal

* Corresponding author.
E-mail addresses: justine.chervin@ensiacet.fr (J. Chervin), thierry.talou@ensiacet.fr (T. Talou), marjorie.audonnet@laposte.net (M. Audonnet), dumas@lrsv.ups-tlse.fr (B. Dumas), camborde@lrsv.ups-tlse.fr (L. Camborde), esquerre@lrsv.ups-tlse.fr (M.-T. Esquerré-Tugayé), roux@lrsv.ups-tlse.fr (C. Roux), guillaume.marti@univ-tlse3.fr (G. Marti).

https://doi.org/10.1016/j.phytochem.2019.04.001
Received 9 January 2019; Received in revised form 27 March 2019; Accepted 10 April 2019

Abbreviations

Diode array detector (DAD)
Electrospray iononisation (ESI)
Headspace solid-phase microextraction (HS-SPME)
Internal transcribed spacer (ITS)
Orthogonal partial least square discriminant analysis (OPLS-DA)
Pair of bases (pb)
Polymerase Chain Reaction (PCR)
Principal component analysis (PCA)
Quality control (QC)

primers (Baldwin et al., 1995) and the nucleotide sequences of ITS1 and ITS2 regions are highly polymorphic between species.

The French National collection of violets held in Toulouse (France) comprises 111 plants individuals from all over the world (Asia, America, and Europe), each characterized by its phenotypic traits like flower shape and colour and to a lesser extent leaf shape. Although they all possess vernacular names, only 21% have been scientifically identified with respect to taxa. Their rigorous classification was therefore necessary to allow better identification and discrimination. A first strategy was to identify species and cultivars according to ITS regions known to be conserved. However, due to hybridization between violets, distinction of individuals based on phenotype and ITS alone is often insufficient and complementary methods such as chloroplast genome comparison are necessary. To improve the resolution of the genotypic characterization of our collection, we conducted chemotaxonomic studies using a metabolomic approach with the aim of facilitating the analyses of large datasets *via* the association of potential biomarkers to specific genotypes (Cox et al., 2014). Our first chemotaxonomic study investigated volatiles of fresh native flowers at maturity using headspace solid-phase microextraction coupled to gas chromatography with mass spectrometry (HS-SPME-GC-MS). The second analyzed the secondary metabolites of ethanolic leaf extracts by ultra-high performance liquid chromatography-high resolution mass spectrometry (UHPLC-HRMS).

2. Results

2.1. Phylogenetic analysis

The whole ITS region (500 bp) was amplified for the 85 plant individuals of the French collection of violets and sequenced. In this way, 80 well-defined sequences were obtained. Other violets were either not sequenced (N = 26) or presented non-exploitable sequences (N = 9), likely due to the presence of heterozygotes and intra-individual polymorphism as suggested by the presence of double peaks for some nucleotide sites. For 31 violets, only a few double peaks along the whole sequence were observed, thus the two sequences were studied separately by manually substituting nucleotides and identification performed for both. In each case, the same species but sometimes different isolates were obtained, revealing the significant hybridization between violets of the same species. Sequence alignment (see Supplementary Fig. S1) highlighted few differences: twenty nucleotide sites were polymorphic in ITS 1 and 12 in ITS 2. Alignment of experimental sequences was made with references of GenBank database from the National Center for Biotechnology Information (NCBI), a curated non-redundant public nucleotide sequence database used internationally as a

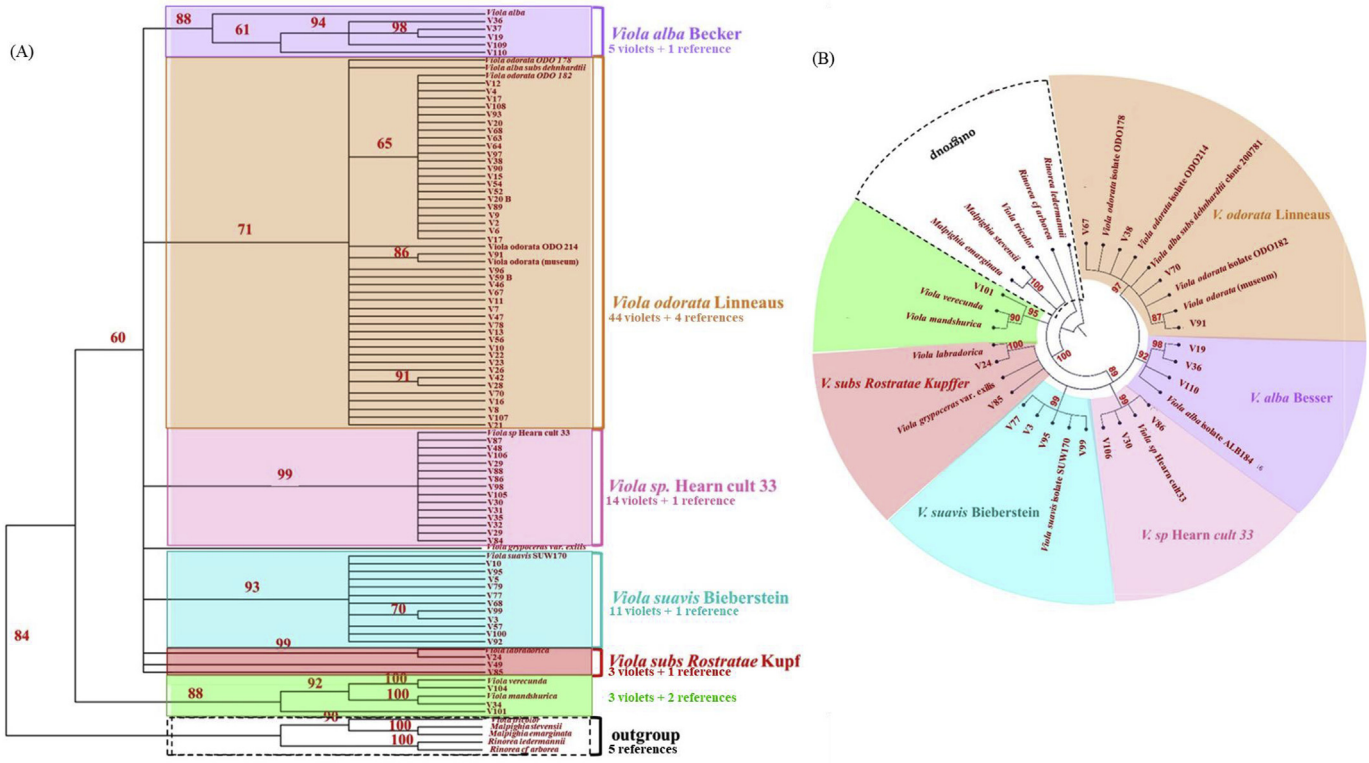


Fig. 1. Cladograms obtained with maximum parsimony (A) and maximum likelihood (B) method showing genetic relationship of selected experimental sequences and references based on ITS sequences. To improve the quality of the figure, only sequences of a few characteristic samples from each clade observed with distance model were selected for B. The outgroup is used to root the tree. Bootstrap values to study robustness are indicated in red (in colour). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

standard for genome annotation (Pruitt, 2004) used in previous pansies (Yockteng et al., 2003) and violet (Conesa at al. 2008) studies. It provided a similarity score and a no-doubt-identification was decided above 99%. This was the case for 64 violets (58% of the whole collection): two were referenced as *V. subs Rostratae* Kupffer (namely *V. labradorica* and *V. grypoceras*) (Marcussen et al., 2010), two as *V. alba* Besser, six as *V. suavis* Bieberstein, twelve as *V.* sp Hearn cult 33, one as *V. verecunda* A Gray, one as *V. mandshurica* Becker and forty as *V. odorata* Linneaus.

Clear defined sequences (N = 80) were compared by a distance based reconstruction method to identify clades with a bootstrap threshold value of 80%. Six clades of various sizes were highlighted and seemed to correlate with specific species (Supplementary Fig. S2). Fifty-five percent of the sequences grouped together in one clade, around 15% in each of two clades and 5% in each of the remaining three. Therefore, while the collection presented some genotype variability, the majority of the violets were found to possess similar DNA sequences.

Comparisons were then made with references found in GenBank nucleotides database to reconstruct maximum parsimony (Fig. 1 A) and the maximum likelihood phylogenetic trees, the latter built with few selected sequences of distinct clades for visual quality (Fig. 1 B). These references were: *Viola alba isolate* ALB184 V6 (GenBank: EU413916), *Viola* sp Hearn cult 33 (GenBank: DQ521291), *Viola suavis isolate* SUW170 V17 (GenBank: EU413930), *Viola mandshurica* isolate TKM201586 (GenBank: KX394615), *Viola alba subsp dehnhardtii* clone 200781 (GenBank: EU430656), *Viola odorata* isolate ODO178 V11 (GenBank: EU413922), isolate ODO214 V8 (GenBank: EU413919), isolate ODO182 V3 (GenBank: EU413918). One *V. odorata* botanically identified obtained for the Museum of Natural History of Toulouse was also included and 99.7% correspondence with *V.* odorata from NCBI was obtained. Cladograms was thus obtained and *Rinorea ledermannii* as species of Violaceae was used to root the tree. As used previously, a bootstrap threshold value of 80% was fixed and same clades were highlighted in both trees with confirmed identifications made as *V. odorata* for the orange clade with 44 violets, *V. alba* for the purple clade with 5 violets, *V. sp* Hearn for the pink clade with 14 violets, *V. suavis* for the turquoise clade with 11 violets, and *V. subs Rostratae* for the red clade with 3 violets; an unidentified cluster in green with 3 violets was well separated from the other clades.

It was quite interesting to also note certain similarities in terms of phenotypic traits of violets belonging to the same clade. A lavender and sometimes white double flower characterized *V. alba* (Supplementary data Fig. S3 a) against mainly big dark purple and sometimes single white flowers for *V.* sp Hearn (Fig. S3 b), small violet single flowers for *V. subs Rostratae* (Fig.S3 c) and single mauve flowers for *V. suavis* (Fig. S3 d). Nevertheless, such phenotypes cannot be described as species-specific since various colors and shapes were observed for different violets all belonging to *V. odorata* (Fig. S3 e), for example either lavender or white double flowers.

Besides, botanical studies based on morphological and anatomical (bracteoles and leaf epidermis) aspects were conducted on three characteristic violets belonging to each clade. For *V. subs Rostratae*, both *V. labradorica* and *V. grypoceras* species were described. All violets possess heart shape leave with either rounded or sharped extremity of various size. Average size of 3–4 cm were observed but bigger and smaller leaves were respectively noted for *V. sp* Hearn and violets belonging to *V. subs Rostratae*. Average flower size is around 1–2 cm and bigger and smaller flowers were as well observed for *V. sp* Hearn and *V. subs Rostratae*.

Main anatomical divergences focused on the presence, shape and localization of glandular and non-glandular trichomes, the size of the anisocytic-type stomata bigger for *V. sp* Hearn and *V. labradorica* as well as the shape of adjoining cells, either rounded or epidermal. Presence of mucilage and hypertrophied epidermal cells were also of varying importance depending on the clade. Tables of resemblances and divergences can be noted in Supplementary Table S2. This study allowed the validation of species identification by comparison with literature (Bonnier and Douin, 1990) (Tutin et al., 1968).

Regarding the identification, data could be more precise than that obtained here, for example violet of Toulouse being identified as *Viola alba* (purple cluster). Indeed, in the literature it is known as *Viola alba sups dehnhardtii* (Malécot et al., 2007). In addition, violets belonging to *Viola odorata* were distributed among the three previously quoted isolates (ODO178, ODO182 and ODO214), however, their sequences were highly analogous with 99% and even 100% similarity. This is indicative of the insufficiency of the ITS sequence for precise identification and the need for further methods such as chloroplast DNA in such cases. Indeed, chloroplast DNA markers showed variability among *Viola* and

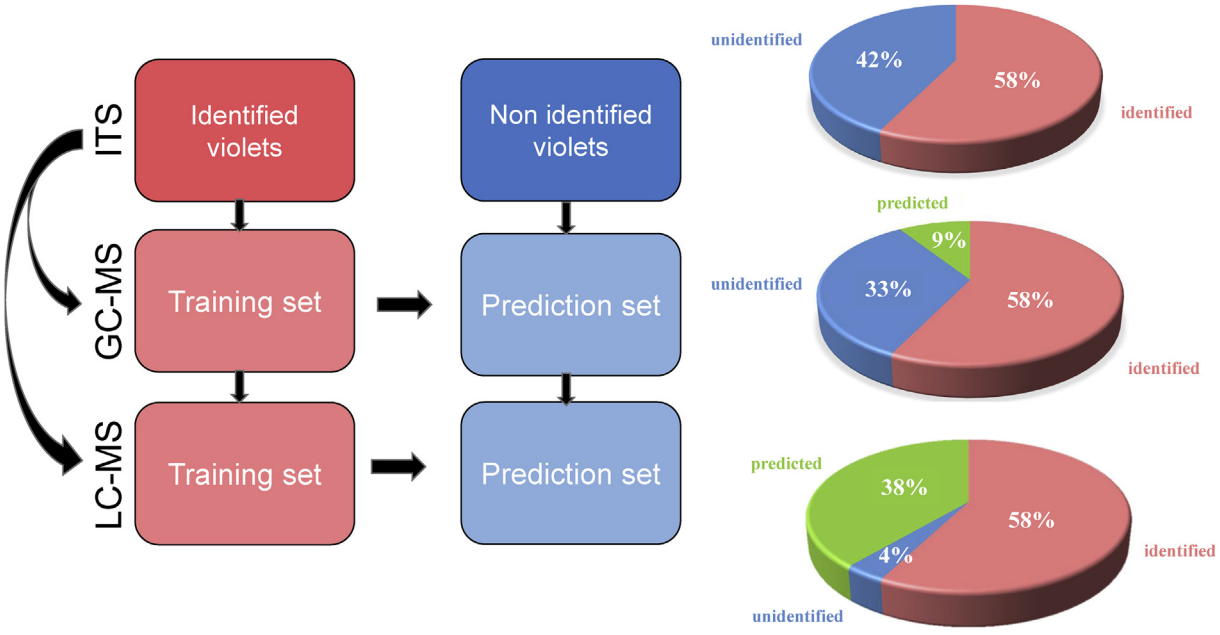


Fig. 2. Strategy of classification based on training and prediction sets. Pie charts show the distribution of classification results (in colour). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

allowed the distinction of sections and subsections as in the analysis of Korean *Viola* (Yoo and Jang, 2010). Nevertheless, since the main goal of our study was to obtain a classification of violets into various clades limiting their identification to the species level was sufficient.

These genetic results were then used as a base model on which to apply a new strategy based on chemotaxonomic studies. From the dataset of violets analyzed by each of two methods (GC-MS and UHPLC-HRMS), genetically well-identified violets (N = 64 violets) were used as training set on which to build statistical models. The unidentified violets were used as a prediction set to try to classify them and attribute a potential identification (Fig. 2).

2.2. HS-SPME-GC-MS

2.2.1. HS-SPME protocol optimization

A second order Doehlert design (Ferreira, 2004) was used to optimize the release and the trapping of volatiles emitted by fresh flowers at maturity. Two time parameters were studied: the time leading up to volatile release and the adsorption time of volatiles trapped onto the SPME fiber. This optimization was undertaken with the violet of Toulouse at room temperature. Plants were kept under a bell-shaped hood to ensure non-degradation of the vegetal matter. Based on one report in the literature, the expertise domain was defined as: a time for volatile release between 20 and 80 min and an adsorption time between 15 and 45 min (Vercammen et al., 2000). The Doehlert matrix indicated nine experiments to perform with two repetitions at the center of the experimental domain (Supplementary Table S3). Two responses were observed, the number of peaks and their global integration (Supplementary Tables S4 and S5). The repeatability of the plan was confirmed by the repetition of the same order of magnitude: number of peaks 13 ± 1 and integration 21 ± 3. Method optimization was sought using the maximum final parameter values: 65 min to release the volatiles and 37 min for the adsorption. This theoretical maximum was experimentally confirmed by the analyses of three different plants on three days (Supplementary Table S6). For routine analyses, a proposed technical maximum of 60 min and 30 min was validated upon similar results to those of the theoretical maximum with a standard deviation of only 2% regarding the peak integration. This practical optimum was therein conserved.

2.2.2. HS-SPME-GC-MS-based metabolomic approach

HS-SPME-GC-MS profiles of all 39 analyzed flowering plants afforded 82 features (*m/z*-RT pairs). As a preliminary step, principal component analysis (PCA) was applied as an exploratory tool to provide an unsupervised overview of the GC-MS volatile fingerprints. Use of binary data based on the presence or absence of peaks in each sample was preferred to avoid intensity variability (Fig. 3). Indeed, analyses of the same violet of Toulouse throughout one season (February-March) provided quantitative but not qualitative variations and the identification of the same profile with only peak intensity variations (Supplementary Fig. S4). This PCA revealed a group composed of only violets belonging to *V. sp* Hearn that were well-separated from the four other clusters. This separation was dependent on PC1 while the other species were separated by PC2. This observation suggests that *V.* sp Hearn features volatile profiles that differ considerably from the other species which among themselves share common features.

We then applied a supervised orthogonal partial least square discriminant analysis (OPLS-DA) in order to try to classify the unidentified violets by projection and thus extend the classification of our collection. A base model was built with only violets identified genetically (N = 19 violets) (Fig. 4A) with distinction of five clusters related to specific species. *V. sp* Hearn cluster is separated by PC1 axis whereas the others by PC2 axis. The quality of model prediction was not satisfactory (R2Y = 0.93, Q2Y = 0.505, CV-ANOVA *p*-value = 0.59) and a permutation test did not assess its validity (Supplementary Fig. S5). This result could be explained by the overlapping of clusters relating to *V. alba*, *V. odorata* and *V. suavis* (Fig. 4B) clearly observed after prediction of unidentified violets (N = 20 violets). It suggests that these three species might share common features which did not allow their clear distinction compared to *V. subs Rostratae* and *V. sp* Hearn which are both well

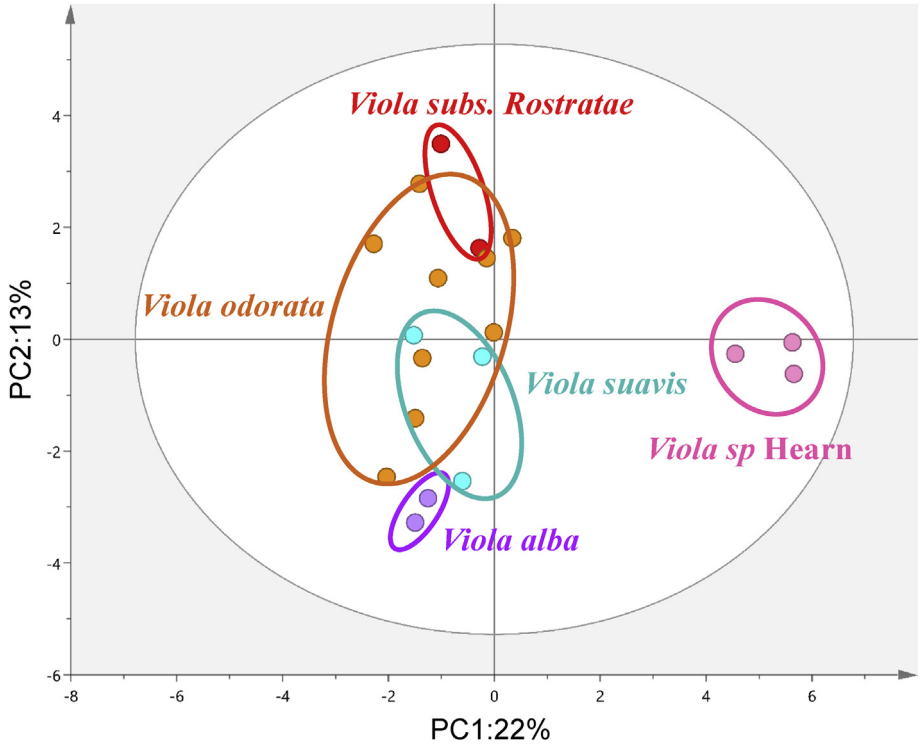


Fig. 3. PCA score plot of GC-EIMS data from SPME flower extracts using binarized data (in colour). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

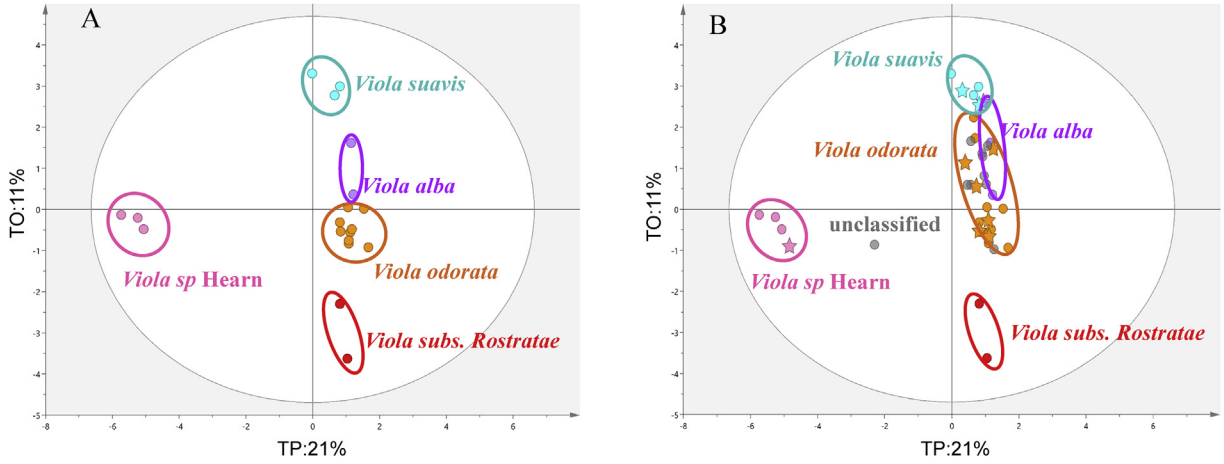


Fig. 4. OPLS-DA base model of GC-EIMS data (A) and prediction set of unidentified violets (B). Well-projected violets are indicated by a star (in colour). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

separated in the variance plot. Prediction results which were quite weak, as only ten violets in twenty (i.e., 50%) were projected and classified into the defined clusters, have to be taken with hindsight as projections in *V. sp* Hearn could be quite certain due to its good separation, but those in the three overlapping clusters remain hypothetical.

2.2.3. Identification of violet species-specific volatile biomarkers

Despite the poor classification of violets with volatile analyses, some species-specific biomarkers were highlighted through database interrogation (Fig. 5). A profile with major amount of limonene (Supplementary Fig. S6.1 a) was revealed for violets belonging to *V.* sp Hearn with the presence of some terpenes such as α-terpinene (S6.1 b). Two ionone profiles enabled a distinction between violets belonging to *V. alba* and *V. suavis*. Regarding *V. alba*, the three ionones classically found for violets were noted, i.e. α-ionone (S6.1 c), β-ionone (S6.1 d) and dihydro-β-ionone (S6 e), as well as one methoxybenzene named methylanisole (S6.1 f). This latter is absent for *V. suavis* but we suspect the presence of other ionone derivatives near the area of well-identified ionones as two new peaks appeared. Finally, violets identified as *V. subs Rostratae* have no quantitative volatiles according to the absence of chromatographic peaks, correlating to nearly no flower fragrance compared to the others. These profiles correlate well with our findings using the OPLS-DA base model, showing high levels of similarity between *V. suavis* and *V. alba* in view of the presence of ionones, in contrast to both *V. sp* Hearn, which while devoid of these compounds

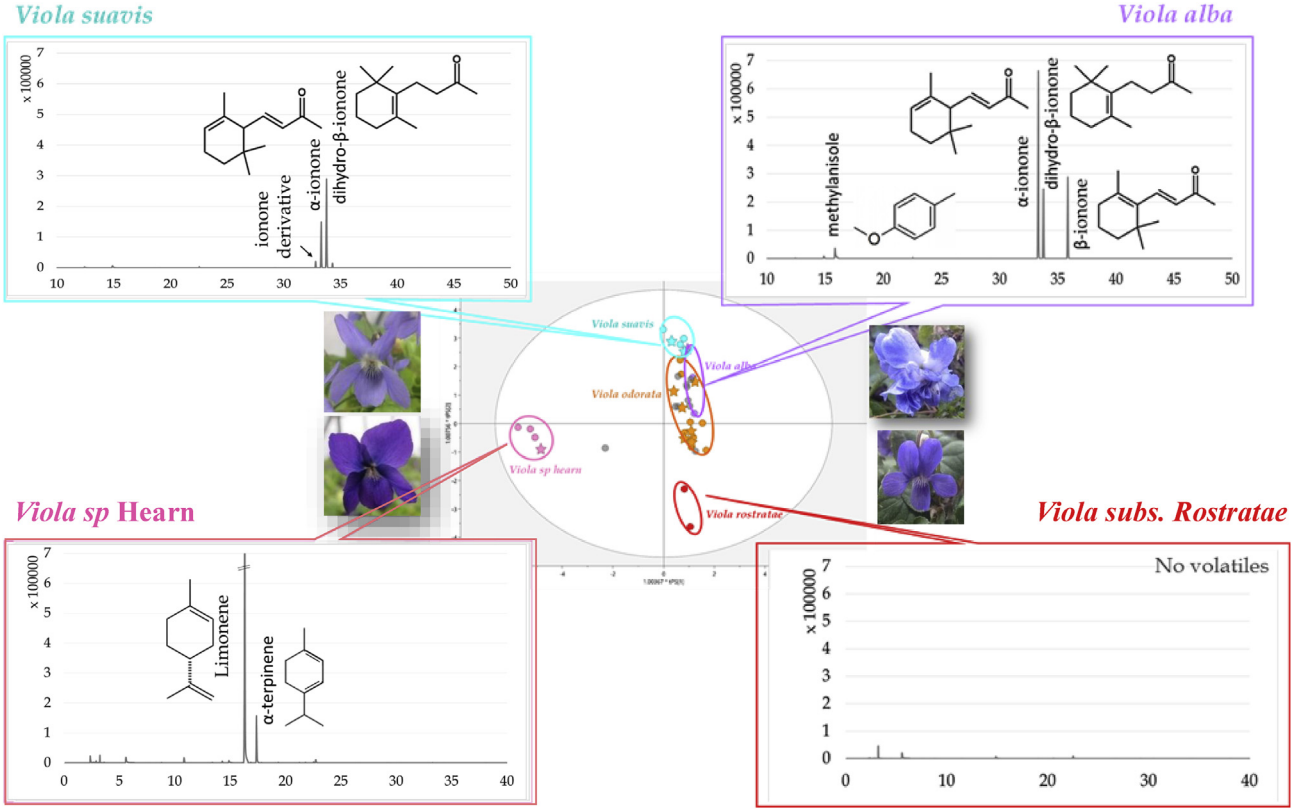


Fig. 5. Specific volatile profiles with identification of characteristic volatile biomarkers for each OPLS-DA cluster of GC-EIMS data. Stars indicate well-projected violets (in colour). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

are rich in others, and *V. subs Rostratae* with no volatiles. These ionone profiles thus explain the well-defined clusters of *V.* sp Hearn and *V. subs Rostratae* and the tendency for the others to merge.

Finally, the standard deviation reaching 7% between retention times of standards and compounds derived from violet extracts, as well as the strong similarity of the experimental mass spectra allowed the validation of all the proposed annotations obtained *in silico* with databases (Supplementary Fig. S6.2).

2.3. UHPLC-HRMS-based metabolomic approach

UHPLC-HRMS profiles of all 119 extracts (110 crude extracts, 1 botanical reference *V. odorata* and 8 QC samples prepared by pooling aliquots of all extracts) afforded 208 and 198 features in NI and PI modes respectively and 120 common features between both ionization modes. The analytical method was optimized so as to obtain as much information as possible in a satisfactory analysis time (about 20 min). The four most intense ions in each scan were fragmented to provide structural information and facilitate annotation of the compounds.

As expected, PCA-X grouped QC near the plot center with violet extracts distributed around this central point (Fig. 6). Clusters correlated with species were highlighted according to genetic results. They matched with the five identified species: *V. subs Rostratae*, *V. suavis*, *V. alba, V. odorata*, and *V.* sp Hearn. These results indicated then a stable chemical composition within aerial parts of violets belonging to the same species. Moreover, it can be noticed the formation of two main groups based on PC1 with on the one hand *V. odorata* and on the other hand *V. suavis*, *V. sp* Hearn, *V. alba* and *V. subs Rostratae*. According to this configuration, *V. alba* and *V. subs Rostratae* seem to present similar profiles compared to the others, especially *V. odorata* and *V. suavis* which are both well-separated from the other clusters suggesting they possess specific metabolites.

After PCA-X, we applied OPLS-DA analysis in order to try to classify the unidentified violets by projection and extend the classification of the collection. A base model was first built with only genetically-identified violets (N = 62 violets) (Fig. 7A). This model was quite well-defined (R2Y = 0.962, Q2Y = 0.822, CV-ANOVA p-value < 0.001) and validated by a permutation test (Supplementary Fig. S7). Unidentified violets (N = 46 violets) were then projected on the same model as prediction set. Forty-two out of the forty-six unknown violets were putatively identified by projection and thereby affiliated to clear-defined clusters characterized by specific chemical profiles and related to specific species (Fig. 7B). Three violets were associated with *V. alba*, two with *V. subs Rostratae*, nine with *V.* sp Hearn, two with *V. suavis* and twenty-six with *V. odorata*. Four violets remained totally separate from these clusters and could therefore be neither classified nor putatively identified.

In our previous genetic results, some violets were not clearly identified due to similarity scores under 99% but were then affiliated to one clade through distance model phylogeny. Upon comparison with the UHPLC-HRMS results, predictions correlated with genetic affiliation thus reinforcing the putative identification of these violets which allowed reaching the classification of 96% of the collection. The four undefined violets were not well correlated to any specific cluster and regarding genetic results, two were not sequenced and the others revealed clear sequence with low identification scores. Thus 4% of our collection of violets remain unidentified.

The stability of leaves metabolome over time was assessed by a kinetic study. Every two months, ten leaves of violet of Toulouse were collected and extracted as depicted in section 2.4. The PCA score plot (Supplementary Fig. S8) containing few selected extracts of each species representative of each cluster was built. All kinetic extracts of violet of Toulouse were distinctly projected within *V. alba* cluster. Thus, the model validity obtained at one precise moment of the season may be more generally interpreted as attested by these kinetic results.

To complete this study, phylogenetic reconstruction was tried by combining ITS sequence and LC-MS data through a parsimony method. Since no standardization method was found in the literature to treat LC-MS data, the binary matrix was built on the average intensity as threshold, independently for each variable. Thus superior values were indexed with 1 against 0 on the contrary. All data were combined and the resulting combined tree (Supplementary Fig. S9) conducted to the observation of the majority of the groups identified in ITS phylogenetic tree. The only difference focused on the scission of *V. odorata* clade into two groups. Complementary hierarchical cluster analysis (HCA) of LCMS data highlighted the same organization as one branch related to *V. odorata* was noted with a separation into two groups (Supplementary Fig. S10). These results could be then explained by the presence of different varieties within *V. odorata* which seem highlighted by combination of data.

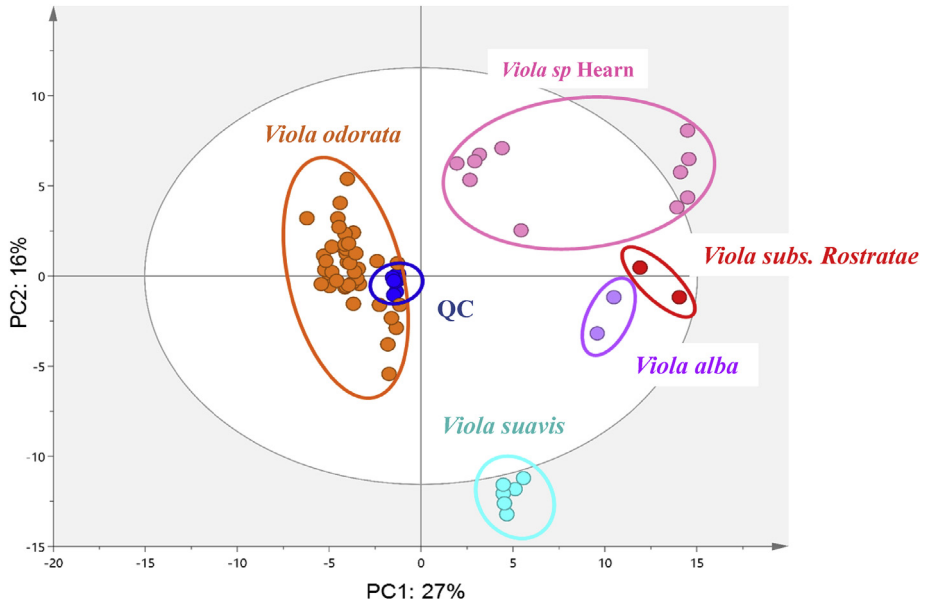


Fig. 6. PCA score plot of UHPLC-HRMS-ESI-NI/PI combined dataset from violets leaf extracts (in colour). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

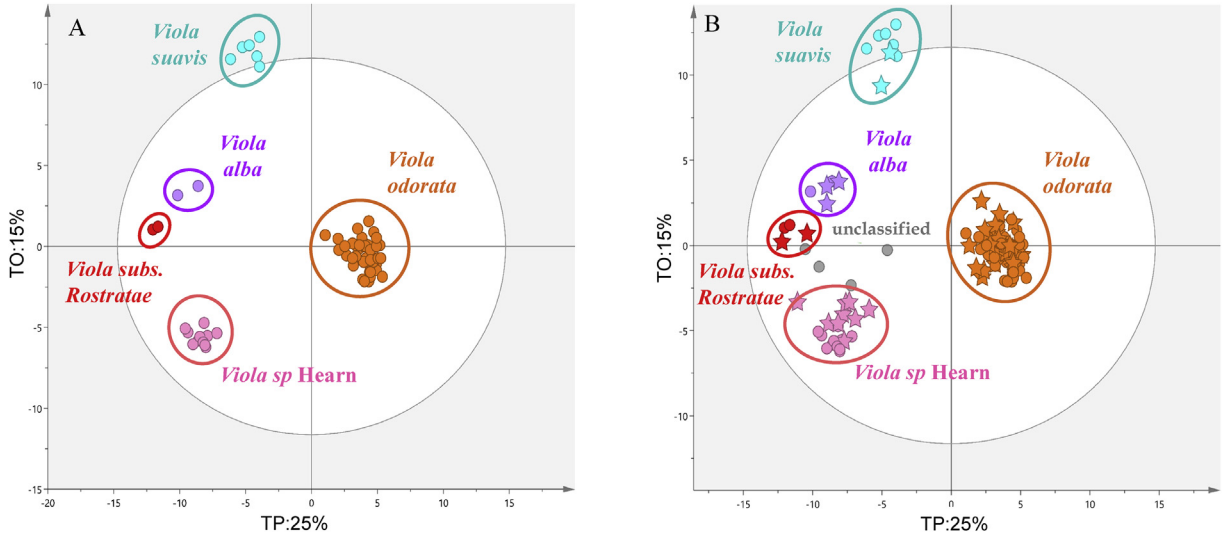


Fig. 7. OPLS-DA base model of UHPLC-HRMS-ESI-NI/PI combined dataset (A) and prediction set of unidentified violets (B). Well projected violets are indicated by a star (in colour). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

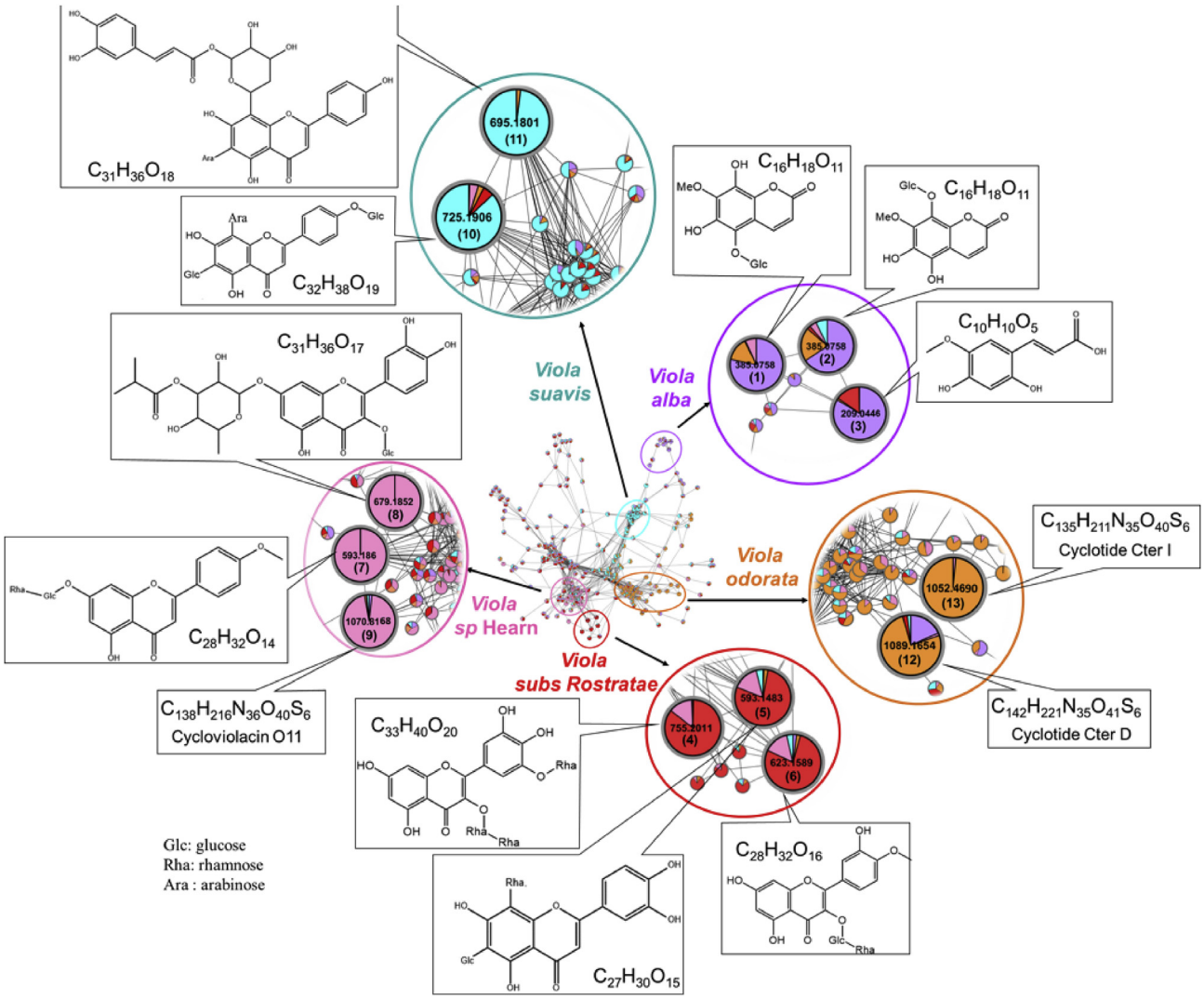


Fig. 8. Correlation network based on ESI-NI/PI combined dataset of UHPLC-HRMS performed on leaf extracts and identification of characteristic biomarkers. Colour tag is based on species identification by genetic analysis. Node size was emphasized based on OPLS-DA coefficient value. Putative structures were based on HRMS and MS/MS spectra and correspond to the first hit in MS-FINDER as annotation illustration (in colour). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 1
Summary of all annotated compounds with MS-FINDER *in silico* matches.

N°	m/z	RT (min)	Detected pseudo-molecular ion	Molecular Formula	Error (ppm)	Chemical class[a]	Putative annotation	Cluster Affiliation	Number of corresponding MS peaks with literature	Biological source	Reference
1	385.0758	2.95	$[M-H]^-$	$C_{16}H_{18}O_{11}$	4.6744	Coumarin glycosides	7-Methoxy-5,6,8-dihydroxycoumarin-5-β-glucopyranoside[b]	V. alba	–	Tetraphis pellucida	Jung et al. (1995)
2	385.0758	3.47	$[M-H]^-$	$C_{16}H_{18}O_{11}$	4.6744	Coumarin glycosides	7-Methoxy-5,6,8-trihydroxycoumarin-8-β-glucoside[b]	V. alba	–	–	–
3	209.0446	3.32	$[M-H]^-$	$C_{10}H_{10}O_5$	4.3053	Hydroxycinnamic acids	Dihydroxy-methoxy-propenoic acid Phenyl[c]	V. alba	ND	Viola betonicifolia	Muhammad et al. (2013)
4	755.2011	3.92	$[M-H]^-$	$C_{33}H_{40}O_{20}$	3.8625	Flavonoid glycosides	Myricetin O-deoxyhexoside-O-deoxyhexoxyl-deoxyhesoside[c]	V. subs Rostratae	2	Ajuga remota	Arot Manguro et al., 2006
5	593.1483	4.46	$[M-H]^-$	$C_{27}H_{30}O_{15}$	4.8891	Flavonoid glycosides	Luteolin-C-deoxyhexoside-C-hexoside[c]	V. subs Rostratae	ND	Plagiomnium elatum	Anhut et al., 2014
6	623.1589	4.63	$[M-H]^-$	$C_{28}H_{32}O_{16}$	4.4932	Flavonoid glycosides	Isorhamnetine-O-hexosyl-deoxyhexoside[c]	V. subs Rostratae	2	Ginkgo biloba	Luo et al. (2013)
7	593.1860	5.42	$[M+H]^+$	$C_{28}H_{32}O_{14}$	0.8429	Flavonoid glycosides	Acateine-O-hexosyl-deoxyhexoside[c]	V sp Hearn	3	Salvia moorcroftiana	Zahid et al. (2003)
8	679.1852	5.71	$[M-H]^-$	$C_{31}H_{36}O_{17}$	3.8281	Flavonoid glycosides	Quercetin O-hexoside-C-acylated-deoxyhexoside[c]	V. sp Hearn	ND	Sinocrassula indica	Xie and Yoshikawa, 2013
9	1070.8168 803.3678	5.94	$[M+3H]^{3+}$ $[M+4H]^{4+}$	$C_{138}H_{216}N_{36}O_{40}S_6$	1.0273	Cyclotide	Cycloviolacin O11[d]	V. sp Hearn	ND	Viola odorata	Craik et al. (1999)
10	725.1906	3.31	$[M-H]^-$	$C_{32}H_{38}O_{19}$	3.8610	Flavonoid glycosides	Apigenin O-hexoside-C-pentoside-C-hexoside[c]	V. suavis	8	Ceratonia siliqua	Batista and Gomes, 1993
11	695.1801	3.52	$[M-H]^-$	$C_{31}H_{36}O_{18}$	4.0277	Flavonoid glycosides	Apigenin C-pentoside-C-hydroxyferuloyl-pentoside[c]	V suavis	5	–	Benayad et al. (2014)
12	1089.1653 817.1260	5.9	$[M+3H]^{3+}$ $[M+4H]^{4+}$	$C_{142}H_{221}N_{35}O_{41}S_6$	3.4889	Cyclotide	Cter D[d]	V. odorata	–	Clitoria ternatea L	Poth et al. (2011)
13	1052.4690	5.42	$[M+3H]^{3+}$	$C_{135}H_{211}N_{35}O_{40}S_6$	1.4252	Cyclotide	Cter I[d]	V. odorata	–	Clitoria ternatea L	Poth et al. (2011)
14	1047.1356 785.6131	5,17	$[M+3H]^{3+}$ $[M+4H]^{4+}$	$C_{133}H_{207}N_{37}O_{39}S_6$	4,8704	Cyclotide	Cycloviolacin[d]	V. odorata	–	Viola odorata	Craik et al. (1999)

[a] Determined with ClassyFire (Djoumbou Feunang et al., 2016).
[b] Compounds confirmed by NMR (Chervin et al., 2017).
[c] Putative annotation based on experimental HRMS, MS/MS and *in silico* fragmentation matches restricting interrogation to Violaceae family and close derivatives (See figure S12). Positional group could not be determined and were removed from proposed names.
[d] Cyclotides were determined based on their characteristic isotopic shape and their triply and/or fourthly charged pseudo-molecular ions.

2.4. Annotation of UHPLC-HRMS chemotaxonomic biomarkers of violet species

In an attempt to take our classification further and identify potential chemotaxonomic biomarkers, we built a correlation network (Fig. 8). This was composed of a set of nodes corresponding to UHPLC-HRMS peaks and correlated by edges using Spearman correlation. This correlation was based on the relative intensity of each peak in each extract. Compounds having the same pattern would thus be correlated together.

The network obtained was composed of around 250 nodes and allowed us to observe, as expected, clear clusters of metabolites specific to identified species. Non-discriminant metabolites shared by different and sometimes all species were also found plotted at the center of the network (Supplementary Fig. S11).

The annotation of characteristic features of each species was undertaken using *in silico* fragmentation to compare against our own in-house *Viola* databases as well as local databases within MS-FINDER (Table 1). Every correspondence between experimental and *in silico* fragments is presented in Supplementary Fig. S12. Moreover, UV spectra were also combined to confirm the chemical class of compounds. The combination of spectral data allows annotation of level 2 (Sumner et al., 2007).

It is interesting to note the abundance of *C*- and *O*-glycosylated coumarin and flavonoid derivatives. It can be noted that *C*- and *O*-glycosylated aglycone can be differentiated by their MS-MS spectra as previously described (Benayad et al., 2014). For *O*-hexose like glucose, a difference of 162 Da is observed; for *O*-deoxyhexose like rhamnose it is a difference of 146 Da and for *O*-pentose like arabinose, this is a difference of 132 Da. On the contrary, for *C*-glycosylated substituents, these losses are not observed but rather 60, 90 or 120 Da (Ferreres et al., 2003). For instance, fragments with mass differences of 162 Da and 15 Da suggest the presence of one methoxy and one *O*-hexose substituents for compounds 1 and 2. NMR characterizations confirmed these substitutions of a coumarin aglycone (Chervin et al., 2017). On the contrary, for compound 5, observation of fragments 503.11 (−90 Da) and 473.21(-120 Da) suggested the presence of a *C*-rhamnose whereas fragments 383.05 and 353.11 highlighted a *C*-glucose. Relative intensities of fragments suggest the rhamnose is linked at the 8 position and the glucose is linked at the 6 position (Benayad et al., 2014).

OPLS-DA coefficients were determined for every feature and for each identified species. Characteristic features were correlated to the highest OPLS-DA coefficient scores. *V. odorata* presents mainly cyclotides as biomarkers with $[M+3H]^{3+}$ showing a characteristic isotopic shape, whereas *V. alba*, *V. subs Rostratae*, *V. sp* Hearn and *V. suavis* seem well characterized by polyphenols, in particular flavonoids and coumarins specific to *V. alba*. As in the above-mentioned UHPLC-HRMS-based metabolomic approach, the identification of two groups was reinforced by the putative distinguished biomarkers.

3. Discussion

We have described the taxonomic delimitation of the French national collection of violets held in the municipal greenhouses of Toulouse performed by combining their genotypic and chemotaxonomic profiles. Untargeted metabolomic approaches coupled to LC-MS (Martucci et al., 2014) as well as SPME GC-MS (Khalil et al., 2017) phytochemical characterizations have already demonstrated their potential in taxonomic classification on different species. As explained previously, regarding the species concept (Smedsgaard and Nielsen, 2005), the classification and understanding of any species rely on three main axes to distinguish their phenotypic characteristics, ecological behavior and genome. However, these aspects are not always sufficient to well-differentiate closely related species or during the speciation process. Since the metabolome represents the final picture of what is extracted resulting from genomic and environmental interaction, metabolite profiling appears to offer the data required to bridge the gap left by genetic-based taxonomical studies performed alone. Implementation of such metabolite profiling should thus provide new clues with which to refine genomic-based classification. Genetic analyses based on the study of ITS sequencesallowed the classification of 72% of the collection of violets into six different clades with distance-based phylogeny reconstruction. Nevertheless, only 58% of the collection clearly matched to one reference found in the NCBI database with a similarity score above 99%. To improve this classification, we undertook chemotaxonomic studies involving native flower volatile profiling obtained by HS-SPME/GC-MS and UHPLC-HRMS analyses of fresh leaf extracts. We found the volatile profiling unsuitable for discrimination of the complete collection, mainly due to the small number of features inducing a less significant variability among the species. In addition, this technique is quite sensitive, in particular to environmental conditions, and is long to perform (around two and half hours for one analysis). Moreover, profiling the volatile emitted by flowers is limited to the flowering period, which is only two winter months for the *Viola* genus. Analysis of a large collection is therefore laborious and was deemed unsuitable in our case.

On the contrary, UHPLC-HRMS analyses worked well for our classification purposes. Indeed, the strategy used here based on genetic results to build a statistical model from which a prediction set was used for unidentified violets, the classification rate rose to 96% of the whole collection composed of 111 violets. Importantly, this technique is easy to implement, in our case requiring only an ethanolic extraction which lasted 30 min using around 1 g of fresh sample (the equivalent of a dozen leaves in the case of *Viola* plantlets). Moreover, separations are of better quality and are achieved in a shorter time (around 30 min per sample) thereby reducing the cost of analysis (Dong and Zhang, 2014) making it better suited to the analysis of a whole collection containing hundreds of samples. The overall good predictive quality of our LCMS-based models could be explained by the larger number of variables involved (around 250 in total). Furthermore, the variability observed over time is higher for volatile compounds compare to non-volatiles, as depicted in Fig. S4 and Fig. S8. As a consequence, the inter-sample variation is lower for leaf extracts. The main limitation remains the false discovery rate (Benjamini and Hochberg, 1995); a common way of circumventing this is to employ the diagnostic tool CV-ANOVA (analysis of variance testing of cross-validated predicted residuals) to assess the reliability of the OPLS model. This tool displays the *p*-value indicating the probability level where a model with one value is the result of chance. A *p*-value lower than 0.05 is correlated to a significant model. Permutation tests are also often applied in validation procedure to diagnostic model overfit (Eriksson et al., 2008). The acquisition of UHPLC–HRMS profiles in data-dependent analysis mode provided accurate mass-to-charge ratios for molecular formula determination along with MS/MS fragments used for peak assignments. Regarding the various databases used, our *Viola* databases provided more pertinent annotations than those based on natural products databases within MS-FINDER, which allowed a more general identification for which mainly chemical class was relevant.

Cross referencing with the genetic classification reinforced the putative identification of unknown violets. In this way, more than half of the collection was identified as *V. odorata*, around 18% as *V.* sp Hearn, 5% as *V. alba*, 7% as *V. suavis* and 4% as *V. subs Rostratae*.

While *V. alba, V. suavis* and *V. odorata* possess similar volatile profiles principally composed of ionones, they could be better differentiated by their non-volatile secondary metabolites from coumarines, flavonoids and various non-phenolic classes of compounds. On the contrary, while *V. sp* Hearn and *V. subs Rostratae* possess similar secondary metabolite markers within the flavonoid class, they could be accurately distinguished by their volatile profiles with the presence of terpenic compounds within the first group and the absence of quantitative volatiles from the second.

Metabolite profiling, as depicted in a previous study of tomato species (Overy, 2004), allowed the examination of biological diversity

with clear discrimination of violet species; this permitted their segregation through multivariate data analysis. The chemical profiling methods used in our study provided complementary information and contributed towards the establishment of a significant relationship with ITS phylogenetic analysis for *Viola* genus.

Consensus tree based on ITS sequences and binary LC-MS matrix validated the majority of the clades correlated to one species (Fig. S9). However, for *V. odorata* a scission was observed conducting to two groups instead of the only one observed with ITS sequence. The first hypothesis was about the potential loss of information due to the binarization of LC-MS data. Indeed, there are continuous data with semi-quantitative information at one variable level. But HCA of LC-MS data highlighted the same phenomenon which could then be explained by the potential presence of several varieties. Comparison of both clades highlighted a highest concentration of cycloviolacin O9 in the major group of *V. odorata*. Independent analysis of ITS data did not permit the clear identification of varieties but combination of information conducted to this revelation thus reinforcing the complementarity of the methods and the relevance of metabolome profiles in classification.

One of the main limitations of our workflow concerns the non-referenced sequences. As depicted for the four unclassified violets (gray dots, Fig. 7b), two of them displayed exploitable ITS sequences, with no satisfactory matches within the NCBI bank (> 99%). According to our combined results, these violets could be new species which warrants further investigation. Another drawback of our procedure is the identification of metabolite biomarkers which is time-consuming and elusive despite recent advances in this area (Perez de Souza et al., 2017). Contrary to other "-omics" based approaches, such as genomics and proteomics, metabolomic is still in its infancy and requires protocol standardization to overcome this identification shortfall. Nevertheless, taxonomy can expect a strong complementary partnership between comprehensive metabolome profiling methods and genetic-based approaches in the near future.

4. Materials and methods

4.1. Plant material

Violet leaves were harvested in spring 2016 from the whole collection of violet plants maintained in the municipal greenhouses of Toulouse, France: two leaves for genetic analysis and ten for UHPLC-HRMS analysis. Directly after collection, they were immerged in liquid nitrogen to stop any enzymatic activity and then stored at −80 °C. Genetic analysis was undertaken on frozen leaves; for UHPLC-HRMS profiling, the vegetable matter was lyophilized and ground into powder using a microtube homogenizer (BeadBug, 40 W). Overall, 108 distinct violet leaves were extracted.

For volatile profiling, the whole native flowering plants were analyzed between February and April 2016 according to their blossoming quality. A total of 39 flowering plants were analyzed.

4.2. DNA amplification, purification and sequencing

Genetic analyses were performed directly on a small portion (approx. 1 mm^2) of a frozen fresh leaf of the various plants without previous DNA extraction. Species discrimination was accomplished through amplification of the whole internal transcribed sequence region ITS1-5.8S-ITS2 by PCR (Mullis et al., 1986) using the KAPA3G Plant PCR kit (Kapabiosystems).

The forward primer ITS (GGAGAAGTCGTAACAAGGTTTCCG) (Aceto et al., 1999) and the reverse primer ITS (GCTCGCCGTTACTAG GGGAATC) were used to amplify the ITS region according to the manufacturer's recommendations. Briefly, a small sample of frozen leaf was mixed with 50 µl of 1× PCR buffer solution, containing 1 mM of $MgCl_2$ and 0.2 mM of dNTPs, 0.3 µM of each primer and 1 Unit of KAPA3G DNA polymerase, in sterile conditions under a fume hood, with filter tips, after a UV cycle of 20 min. Amplification was performed on a Mastercycler (Eppendorf), starting with an initial denaturation step of 10 min at 95 °C followed by 35 cycles composed of a denaturation step at 95 °C for 30 s, followed by an annealing step at 62 °C for 20 s, and an extension step at 72 °C for 30 s. PCR amplicons were resolved on a 1% agarose gel, after 30 min at 100 V. DNA revelation was achieved under UV with ethidium bromide (EtBr) solution. PCR products were then purified with Promega Wizard® SV Gel and PCR Clean-Up System (Promega Corp., Madison, Wisconsin). Finally DNA was collected in 50 µL of nuclease-free water. Sanger sequencing was performed by GATC Biotech (Constance, German). Sequence identification was achieved by matching against references found within the National Center for Biotechnology Information (NCBI) database (Pruitt, 2004).

4.3. Sequence alignment and phylogenetic inference

Phylogenetic analyses were performed using ITS1-5.8S-ITS2 sequences of references available from GenBank (see Supplementary Table S1). Alignment of ITS sequences was performed using the ClustalX program within the CLC Main Workbench software version 7.8.1 (Qiagen Aarhus A/S) and then manually adjusted. Two inferences were then performed: a distance analysis by Neighbor Joining (using Kimura 80 as distance setting) to define sequence similarity, and a phylogenetic analysis using the maximum likelihood method. Maximum parsimony trees of ITS sequences as well as combined ITS and metabolomic data were also built using PAUP* 4.0 software (Swofford, 2001). Searches were heuristic with the parsimony criterion, with MulTrees and ACCTRAN options in effect and TBR branch swapping algorithm. Bootstrap values were derived from 1000 replicates and were calculated for both inferences to define clade robustness. The LC-MS data were binarized according to mean peak intensity with a value of "1" for intensity above mean and "0" on the contrary. Outgroups composed of sequences of *Viola* species from section *Melanium* (pansies), the different genus *Rinorea* from Violaceae and *Malpighia* from Malpighiaceae were also introduced to obtain rooted trees.

4.4. Leaf metabolite extraction

Metabolites were extracted by adding ten volumes of 80% ethanol to the powdered material (80.0 ± 0.2 mg). The solutions were sonicated in a bath (Fisher Scientific, Illkirch, France) at room temperature for 10 min, then centrifuged for 2 min at 5 °C and 14 000 rpm. The supernatant was put aside and this procedure then repeated once on the residue with fresh solvent. In order to discard chlorophyll pigments, the combined liquid extracts were mixed with silica-C_{18} in final weight proportions of 4 silica powder per 1 initial plant material. The mixture was centrifuged and the resulting liquid extract put in an UHPLC-HRMS vial for subsequent UHPLC-HRMS analysis.

4.5. UHPLC-HRMS profiling

UHPLC-HRMS analyses were performed with diode array detector (DAD) on a UHPLC-LTQ Orbitrap XL instrument (Ultimate 3000, Thermo Fisher Scientific, Hemel Hempstead, UK) as previously reported (Chervin et al., 2017). Briefly, the LC–MS system was run using a Acquity UPLC BEH C18 column (100 × 2.1 mm i.d., 1.7 µm, Waters, MA, USA) equipped with a guard column. The mobile phase A (MPA) was water with 0.1% formic acid (FA) and mobile phase B (MPB) was acetonitrile with 0.1% FA. The solvent gradient was: 0 min, 95% MPA; 0.5 min 95% MPA; 12 min, 5% MPA; 15 min, 5% MPA, 15.5 min, 95% MPA; 19 min, 95% MPA. The flow rate was 0.3 mL/min, the column temperature was set to 40 °C and injection volume fixed to 2 µL. The UV detection was performed by a diode array detector (DAD) from 210 to 400 nm. Mass detection was performed using an electrospray source (ESI) in positive ionization (PI) and negative ionization (NI) modes. The mass scanning range was m/z 100–2000 Da. Each full MS scan was

followed by data dependent MS/MS on the three most intense peaks.

4.6. Native plant volatiles emission

Analysis of the volatiles of fresh mature flowers from native plants was performed using HS-SPME coupled to GC-MS. The whole plant was analyzed to ensure that no fragile vegetal matter was destroyed in preparation. Optimization of the protocol was undertaken on the violet of Toulouse with use of a Doelhert experimental design (Ferreira, 2004) and the following parameters were obtained at ambient temperature: time leading up to volatile release of 1 h under a bell followed by 30 min of adsorption on the divinylbenzene/carboxen/polydimethylsiloxane (DVB/CAR/PDMS) SPME fiber (Supelco, Oakville, ON, Canada). This fiber was chosen for its mixed properties. The fiber was retracted and transferred to the GC-MS injection port at 250 °C with desorption time of 5 min.

4.7. GC-MS profiling

GC-MS analyses were performed on an Agilent Technologies GC system 7890B equipped with HP 5 MS column (30 m × 0.25 mm id x 0.25 µm film thickness; Agilent, Santa Clara, USA) and coupled to Agilent Technologies 5977B MSD mass chromatograph. The temperature of the injector was maintained at 250 °C. The following temperature gradient of the oven was used: 60 °C for 10 min, then increase to 130 °C at 4 °C/min, then to 150 °C at 2 °C/min and finally ramped at a rate of 25 °C/min to 250 °C and maintained for 10 min. Helium was the carrier gas with flow rate of 1 mL/min. Splitless injection was used and no solvent delay was fixed. The mass spectrometer detector was operated in EI mode at 70 eV. The scan range was set at 50–550 Da.

4.8. Data processing

The GC-MS and UHPLC-HRMS raw data were converted to abf files (*Reifycs* Abf Converter) and processed with MS-DIAL version 2.74 (Tsugawa et al. 2015) for mass signal extraction respectively between 50 and 600 Da from 4 to 40 min and between 100 and 1500 Da from 0 to 14 min.

For LC-HRMS/MS, respective MS1 and MS2 tolerance were set to 0.01 and 0.2 Da in centroid mode. The optimized detection threshold was set to 4×10^5 for NI and 1.6×10^5 for PI concerning MS1 and 10 for MS2 in both cases. Adducts, correlated peaks among samples and neutral loss fragments from MS/MS found in higher m/z at a RT windows of 0.1 min were identified to exclude them from the final peak list. Finally, the peaks were aligned on a quality control (QC) reference file with a retention time tolerance of 0.1 min and a mass tolerance of 0.025 Da. Identification was performed with a local database built on MS-FINDER model (Tsugawa et al. 2016) and based on *Viola* genus.

Regarding GC-MS, the optimized detection threshold was set to 500. Finally, the peaks were aligned on the reference file of the violet of Toulouse with a retention time tolerance of 0.075 min.

The resulting peak lists were then exported to comma-separated value (CSV) format prior to multivariate data analysis (MVA) using SIMCA-P+ (version 14.0, Umerics, Umea, Sweden).

For cyclotide detection, positive profiles were processed with MZmine 2.29. A peak list was built and deconvoluted. The identification of isotopes was made with a maximum charge fixed at 5, a m/z tolerance of 10 ppm and a retention time tolerance of 0.2 min. The representative isotope selected was the lowest.

4.9. Statistical analysis

Both UHPLC-HRMS CSV files of NI and PI mode were combined using MScombine package (Calderón-Santiago et al., 2016). The resulting CSV and GC-MS files were then separately imported into SIMCA-P+. For MVA, all data were log transformed and pareto scaled. The unsupervised principal component analysis (PCA) allowed the visualization of the global organization of the samples. The supervised orthogonal partial least square discriminant analysis (OPLS-DA) was made with the names of species as qualitative Y input. Clearly genetically identified plant individuals (similarity score > 99%) were the starting point from which to build OPLS-DA models. Unidentified violets were predicted by projection in potential observable clusters. Coefficient scores were used to rank variables according to their chemotaxonomic biomarker potential. For each model, a leave-one-subject-out cross-validation was performed to assess the model fit. The validity of the discriminant model was verified using permutation tests (Y-scrambling).

4.10. Identification of significant features

Using the OPLS-DA regression analysis results, molecular formulae and structural identification of significant features were calculated with MS-FINDER 2.22 (Tsugawa et al. 2016) based on our previously described model (Chervin et al., 2017). An in-house database based on *Viola* matches inside the Dictionary of Natural Products (DNP, CRC press, v25:2) extended to the chemical classes and of local databases within MS-FINDER (KNApSAck, PlantCyc and UNPD) were used. For each compound, the results afforded several candidates and ranked them according to their similarity score, which was based on comparison between experimental MS/MS fragments and *in silico* spectra of candidates.

For cyclotide identification, a homemade database based on cyclotides found in the DNP was built, clarifying m/z of $[M+3H]^{3+}$ and $[M+4H]^{4+}$, and imported within MZmine with a *m/z* tolerance of 10 ppm.

For volatile compounds, identification was based on retention time matching with NIST MS Search 2.2 (National Institute of Standards and Technology, Gaithersburg, MD, USA) mass spectral database and EI-MS databases proposed with MS-DIAL. The injection of standards analyzed with the same GC-MS method was undertaken in order to validate these results by comparing retention times and mass spectra.

4.11. Correlation network

The text file format exported from MS-DIAL was cleaned-up by eliminating adducts and keeping only peaks with MS/MS data before importation into MetamapR (version 1.4.0) (Grapov et al., 2015). A correlation network was created using the Spearman model and a p-value set to 5.10^{-6}. The calculated edge list was then downloaded and processed with Cytoscape 2.8.3 (Shannon et al. 2003). An attribute file containing all processed information was imported to improve network visualization and interpretation. The mean average intensity of each peak was added to specify the relative part of each compound per species with the use of the plugin “Colour Nodes”.

Funding

This work was supported by the Regional Council Occitanie (Project CLE 13053062).

Acknowledgements

The authors would like to thank the Regional Council Occitanie for sponsoring this project (Project CLE 13053062), as well as the municipal greenhouses of Toulouse for making the violet collection available to its collaborators. The authors also would like to thank Professor Isabelle Fourasté for her help in botanical identification of Violets traits.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.phytochem.2019.04.001.

References

Aceto, S., Paolo, C., Salvatore, C., Luciano, G., Aldo, M., 1999. Phylogeny and evolution of orchis and allied genera based on ITS DNA variation: morphological gaps and molecular continuity. Mol. Phylogenet. Evol. 13 (1), 67–76.

Álvarez, I., 2003. Ribosomal ITS sequences and plant phylogenetic inference. Mol. Phylogenet. Evol. 29 (3), 417–434. https://doi.org/10.1016/S1055-7903(03)00208-2.

Anhut, S., Biehl, J., Seeger, T., Mues, R., Zinsmeister, H.D., 2014. Flavone-C-Glycosides from the mosses *Plagiomnium elatum* and *Plagiomnium cuspidatum*. Z. Naturforsch. C Biosci. 47 (9–10), 654–660.

Arnold, M.L., 2006. Evolution through Genetic Exchange. Oxford University Press, Oxford ; New York.

Arot Manguro, L.O., Wagai, S., Lemmen, P., 2006. Flavonol and iridoid glycosides of Ajuga remota aerial parts. Phytochemistry 67 (8), 830–837. https://doi.org/10.1016/j.phytochem.2006.01.005.

Baldwin, B.G., Sanderson, M.J., Porter, J.M., Wojciechowski, M.F., Campbell, C.S., Donoghue, M.J., 1995. The its region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. Ann. Mo. Bot. Gard. 82 (2), 247. https://doi.org/10.2307/2399880.

Ballard, H.E., Sytsma, K.J., Kowal, R.R., 1999. Shrinking the violets: phylogenetic relationships of infrageneric groups in *Viola* (Violaceae) based on internal transcribed spacer DNA sequences. Syst. Bot. 23 (4), 439–458. https://doi.org/10.2307/2419376.

Batista, M.T., Gomes, E.T., 1993. C-glycosylflavones from *Ceratonia siliqua* cotyledons. Phytochemistry 34 (4), 1191–1193. https://doi.org/10.1016/s0031-9422(00)90745-3.

Benayad, Z., Gómez-Cordovés, C., Es-Safi, N., 2014. Characterization of flavonoid glycosides from fenugreek (trigonella foenum-graecum) crude seeds by HPLC–DAD–ESI/MS analysis. Int. J. Mol. Sci. 15 (11), 20668–20685. https://doi.org/10.3390/ijms151120668.

Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society. Series B (Methodological), pp. 289–300.

Bonnier, G., Douin, R., 1990. La Grande Flore En Couleurs, vol. 3 Nouvelles Flores de Bonnier, Belin.

Calderón-Santiago, M., Fernández-Peralbo, M.A., Priego-Capote, F., Luque de Castro, M.D., 2016. MSCombine: a tool for merging untargeted metabolomic data from high-resolution mass spectrometry in the positive and negative ionization modes. Metabolomics 12 (3), 1–12. https://doi.org/10.1007/s11306-016-0970-4.

Cennamo, P., Del Guacchio, E., Jury, S.L., Caputo, P., 2011. Molecular markers in *Viola L. Subsect. Viola* : application and taxonomic implications for the identification of dubious herbarium specimens. Plant Biosyst. -Int. J. Deal. All Aspect. Plant Biol. 145 (2), 306–323. https://doi.org/10.1080/11263504.2011.558681.

Chervin, J., Perio, P., Martins-Froment, N., Pharkeovilay, C., Reybier, K., Nepveu, F., Fabre, N., Talou, T., Bonzon-Ponnet, V., Marti, G., 2017. Dereplication of natural products from complex extracts by regression analysis and molecular networking: case study of redox-active compounds from *Viola alba subsp. dehnhardtii*'. *Metabolomic* 13 (8). https://doi.org/10.1007/s11306-017-1227-6.

Clausen, J., 1927. Chromosome number and the relationship of species in the genus *Viola*. Ann. Bot. 41 (164), 677–714.

Conesa, M., Mus, M., Rosselló, J.A., 2008. Hybridization between insular endemic and widespread species of *Viola* in non-disturbed environments assessed by nuclear ribosomal and CpDNA sequences. Plant Systemat. Evol. 273 (3–4), 169–177. https://doi.org/10.1007/s00606-008-0006-2.

Cox, D.G., Oh, J., Keasling, A., Colson, K.L., Hamann, M.T., 2014. The utility of metabolomic in natural product and biomarker characterization. Biochim. Biophys. Acta 1840 (12), 3460–3474. https://doi.org/10.1016/j.bbagen.2014.08.007.

Craik, D.J., Daly, N.L., Bond, T., Waine, C., 1999. Plant cyclotides: a unique family of cyclic and knotted proteins that defines the cyclic cystine knot structural motif. J. Mol. Biol. 294 (5), 1327–1336. https://doi.org/10.1006/jmbi.1999.3383.

Djoumbou Feunang, Y., Eisner, R., Knox, C., Chepelev, L., Hastings, J., Owen, G., Fahy, E., Steinbeck, C., Subramanian, S., Bolton, E., Greiner, R., Wishart, D.S., 2016. ClassyFire: automated chemical classification with a comprehensive, computable taxonomy. J. Cheminf. 8 (61). https://doi.org/10.1186/s13321-016-0174-y.

Dong, M.W., Zhang, K., 2014. Ultra-high-pressure liquid chromatography (UHPLC) in method development. Trac. Trends Anal. Chem. 63 (December), 21–30. https://doi.org/10.1016/j.trac.2014.06.019.

Erben, M., 1996. The significance of hybridization on the forming of species in the genus *Viola*. Bocconea 5 (1), 113–118.

Eriksson, L., Trygg, J., Wold, S., 2008. CV-ANOVA for significance testing of PLS and OPLS® models. J. Chemom. 22 (11–12), 594–600. https://doi.org/10.1002/cem.1187.

Ferreira, S., 2004. Doehlert matrix: a chemometric tool for analytical chemistry - review. Talanta 63 (4), 1061–1067. https://doi.org/10.1016/j.talanta.2004.01.015.

Ferreres, F., Silva, B.M., Andrade, P.B., Seabra, R.M., Ferreira, M.A., 2003. Approach to the study of C-glycosyl flavones by ion trap HPLC-PAD-ESI/MS/MS: application to seeds of quince (*Cydonia oblonga*). Phytochem. Anal. 14 (6), 352–359. https://doi.org/10.1002/pca.727.

Grapov, D., Wanichthanarak, K., Fiehn, O., 2015. MetaMapR: Pathway independent metabolomic network analysis incorporating unknowns. Bioinformatics, btv194. https://doi.org/10.1093/bioinformatics/btv194.

Jung, M., Geiger, H., Zinsmeister, H.D., 1995. Tri- and tetrahydroxycoumarin derivatives from *Tetraphis pellucida*. Phytochemistry 39 (2), 379–381.

Khalil, M.N.A., Mostafa, I.F., Mohamed, A.F., 2017. Metabolome based volatiles profiling in 13 date palm fruit varieties from Egypt via SPME GC–MS and chemometrics. Food Chem. 217, 171–181. https://doi.org/10.1016/j.foodchem.2016.08.089.

Levin, D.A., Francisco-Ortega, J., Jansen, R.K., 1996. Hybridization and the extinction of rare plant species. Conserv. Biol. 10 (1), 10–16.

Luo, J.-L., Lu, F.-L., Liu, Y.-C., Shih, Y.-C., Lo, C.-F., 2013. Fingerprint analysis of *Ginkgo biloba* extract and *Ginkgo semen* in preparations by LC-Q-TOF/MS. J. Food Drug Anal. 21 (1), 27–39.

Malécot, V., Marcussen, T., Munziger, J., Yockteng, R., Henry, M., 2007. On the origin of the sweet-smelling parma violet cultivars (Violaceae): wide intraspecific hybridization, sterility, and sexual reproduction. Am. J. Bot. 94 (1), 29–41.

Marcussen, T., Karlsson, T., Jonsell, B., Wind, P., 2010. Violaceae. In: Flora Nordica, 6 12–52 The Bergius Foundation, Stockholm.

Marcussen, T., Yousefi, N., Mehrvarz, S.S., 2012. Anatomical studies on selected species of *Viola* (Violaceae). Nord. J. Bot. 30 (4), 461–469. https://doi.org/10.1111/j.1756-1051.2011.01266.x.

Martucci, M.E.P., De Vos, R.C.H., Carollo, C.A., Gobbo-Neto, L., 2014. Metabolomic as a potential chemotaxonomical tool: application in the genus *vernonia schreb*. Edited by jamshidkhan chamani. PLoS One 9 (4), e93149. https://doi.org/10.1371/journal.pone.0093149.

Mereda, P., Hodalova, I., Kucera, J., 2011. Genetic and morphological variation in *Viola suavis* s.l. (Violaceae) in the western balkan Peninsula : two endemic subspecies revealed. Syst. Biodivers. 9, 211–231.

Muhammad, N., Saeed, M., Adhikari, A., Muhammad, K., 2013. Isolation of a new bioactive cinnamic acid derivative from the whole plant of *Viola betonicifolia*. J. Enzym. Inhib. Med. Chem. 28 (5), 997–1001. https://doi.org/10.3109/14756366.2012.702344.

Mullis, K., Faloona, F., Scharf, S., Saiki, R.K., Horn, G.T., Erlich, H., 1986. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. Cold Spring Harbor Symposia on Quantitative Biology, vol. 51. Cold Spring Harbor Laboratory Press, pp. 263–273. http://symposium.cshlp.org/content/51/263.extract.

Overy, S.A., 2004. Application of metabolite profiling to the identification of traits in a population of tomato introgression lines. J. Exp. Bot. 56 (410), 287–296. https://doi.org/10.1093/jxb/eri070.

Perez de Souza, L., Naake, T., Tohge, T., Fernie, A.R., 2017. From chromatogram to analyte to metabolite. How to pick horses for courses from the massive web resources for mass spectral plant metabolomic. GigaScience 6 (7), 1–20. https://doi.org/10.1093/gigascience/gix037.

Poth, A.G., Colgrave, M.L., Philip, R., Kerenga, B., Daly, N.L., Anderson, M.A., Craik, D.J., 2011. Discovery of cyclotides in the fabaceae plant family provides new insights into the cyclization, evolution, and distribution of circular proteins. ACS Chem. Biol. 6 (4), 345–355. https://doi.org/10.1021/cb100388j.

Pruitt, K.D., 2004. NCBI reference sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res. 33 (Database issue): D501–4. https://doi.org/10.1093/nar/gki025.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., et al., 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 13 (11), 2498–2504.

Smedsgaard, J., Nielsen, J., 2005. Metabolite profiling of fungi and yeast: from phenotype to metabolome by MS and informatics. J. Exp. Bot. 56 (410), 273–286. https://doi.org/10.1093/jxb/eri068.

Sumner, L.W., Amberg, A., Barrett, D., Beale, M.H., Beger, R., Daykin, C.A., Fan, T.W.M., 2007. Proposed minimum reporting standards for chemical analysis: chemical analysis working group (CAWG) metabolomics standards initiative (MSI). Metabolomics 3 (3), 211–221. https://doi.org/10.1007/s11306-007-0082-2.

Swofford, D.L., 2001. PAUP*: Phylogenetic Analysis Using Parsimony (And Other Methods) 4.0. B5. Sinauer Associates, Sunderland, Massachusetts.

Todesco, M., Pascual, M.A., Owens, G.L., Ostevik, K.L., Moyers, B.T., Hübner, S., Heredia, S.M., 2016. Hybridization and extinction. Evol. Appl. 9 (7), 892–908. https://doi.org/10.1111/eva.12367.

Tsugawa, H., Cajka, T., Kind, T., Ma, Y., Higgins, B., Ikeda, K., et al., 2015. MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis. Nat. Meth. Adv. https://doi.org/10.1038/nmeth.3393. online publication.

Tsugawa, H., Kind, T., Nakabayashi, R., Yukihira, D., Tanaka, W., Cajka, T., et al., 2016. Hydrogen rearrangement rules: computational MS/MS fragmentation and structure elucidation using MS-FINDER software. Anal. Chem. https://doi.org/10.1021/acs.analchem.6b00770.

Tutin, T.G., Moore, D.M., Valentine, D.H., Walters, S.M., Webb, D.A., 1968. Violaceae. Flora Europaea, vol. 2. Cambridge University Press, pp. 270–275.

Vercammen, J., Sandra, P., Baltussen, E., Sandra, T., David, F., 2000. Considerations on static and dynamic sorptive and adsorptive sampling to monitor volatiles emitted by living plants. J. High Resolut. Chromatogr. 23 (9), 547–553. https://doi.org/10.1002/1521-4168(20000901)23:9%3c547::AID-JHRC547%3e3.0.CO;2-7.

Xie, H.-H., Yoshikawa, M., 2013. Further acylated flavonol bisdesmosides from *Sinocrassula indica*. J. Asian Nat. Prod. Res. 15 (8), 885–890. https://doi.org/10.1080/10286020.2013.800973.

Yockteng, R., Ballard Jr., H.E., Mansion, G., Dajoz, I., Nadot, S., 2003. Relationships among pansies (*Viola* section *Melanium*) investigated using ITS and ISSR markers. Plant Systemat. Evol. 241, 153–170. https://doi.org/10.1007/s00606-003-0045-7.

Yoo, K.-O., Jang, S.-K., 2010. Infrageneric relationships of Korean *Viola* based on eight chloroplast markers. J. Syst. Evol. 48 (6), 474–481. https://doi.org/10.1111/j.1759-6831.2010.00102.x.

Zahid, M., Saeed, M., Asim, M., Ishrud, O., Wu, S., Ahma, V.U., Pan, Y., 2003. New glycosides from salvia moorcroftiana (lamiaceae). Helv. Chim. Acta 86, 2021–2027.

Plant, Cell & Environment

Plant, Cell and Environment (2013) **36**, 621–639 doi: 10.1111/pce.12002

Metabolomics reveals herbivore-induced metabolites of resistance and susceptibility in maize leaves and roots

GUILLAUME MARTI[1], MATTHIAS ERB[2,3], JULIEN BOCCARD[1], GAÉTAN GLAUSER[2], GWLADYS R. DOYEN[2], NEIL VILLARD[2], CHRISTELLE A M. ROBERT[3,4], TED C. J. TURLINGS[2], SERGE RUDAZ[1] & JEAN-LUC WOLFENDER[1]

[1]*School of Pharmaceutical Sciences, EPGL, University of Geneva and University of Lausanne, Geneva Switzerland,* [2]*Laboratory for Fundamental and Applied Research in Chemical Ecology (FARCE), University of Neuchâtel, Neuchâtel, Switzerland,* [3]*Root-Herbivore Interactions Group and* [4]*Department of Biochemistry, Max Planck Institute for Chemical Ecology, Jena, Germany*

ABSTRACT

Plants respond to herbivory by reprogramming their metabolism. Most research in this context has focused on locally induced compounds that function as toxins or feeding deterrents. We developed an ultra-high-pressure liquid chromatography time-of-flight mass spectrometry (UHPLC-TOF-MS)-based metabolomics approach to evaluate local and systemic herbivore-induced changes in maize leaves, sap, roots and root exudates without any prior assumptions about their function. Thirty-two differentially regulated compounds were identified from *Spodoptera littoralis*-infested maize seedlings and isolated for structure assignment by microflow nuclear magnetic resonance (CapNMR). Nine compounds were quantified by a high throughput direct nano-infusion tandem mass spectrometry/mass spectrometry (MS/MS) method. Leaf infestation led to a marked local increase of 1,3-benzoxazin-4-ones, phospholipids, *N*-hydroxycinnamoyltyramines, azealic acid and tryptophan. Only few changes were found in the root metabolome, but 1,3-benzoxazin-4-ones increased in the vascular sap and root exudates. The role of *N*-hydroxycinnamoyltyramines in plant–herbivore interactions is unknown, and we therefore tested the effect of the dominating *p*-coumaroyltyramine on *S. littoralis*. Unexpectedly, *p*-coumaroyltyramine was metabolized by the larvae and increased larval growth, possibly by providing additional nitrogen to the insect. Taken together, this study illustrates that herbivore attack leads to the induction of metabolites that can have contrasting effects on herbivore resistance in the leaves and roots.

Key-words: *Zea mays*; benzoxazinones; induced defence; induced resistance; metabolomics; *p*-coumaroyltyramine; root exudates; *Spodoptera littoralis*; systemic signaling; UHPLC-TOF-MS.

INTRODUCTION

Plants produce more than 200 000 different metabolites, many of which are directly involved in environmental interactions (Hartmann 2007). Some secondary metabolites for example serve as chemical defences that prevent or restrain herbivores (Pickett, Smiley & Woodcock 1999). Since the discovery of the inducible immune system of plants (reviewed in Howe & Jander 2008), many scientists have focused on toxic compounds that accumulate after herbivore attack. It has been argued that inducible secondary metabolites that are synthesized *de novo* or transformed into their active form upon herbivore attack constitute a flexible and cost-saving defence strategy (Heil & Baldwin 2002). A recent example that illustrates this view are 1,3-benzoxazin-4-one derivatives (Bxs), a class of secondary metabolites that are produced by many *Poaceae* (Frey *et al*. 2009). In maize, Bxs are mainly present as glycoside derivatives (Cambier, Hance & de Hoffmann 2000). Herbivore-attack induces the N-*O*-methylation of the major glucosides DIMBOA-Glc and DIM_2BOA-Glc to HDMBOA-Glc and HDM_2BOA-Glc (Oikawa *et al*. 2004; Glauser *et al*. 2011). Upon tissue disruption, both the constitutive and induced glucosides are brought into contact with plastid-derived β-glucosidases, which quickly hydrolyse the Bxs to yield toxic aglycones (Morant *et al*. 2008). The latter react strongly with nucleophilic groups of biomolecules due to their electrophilic nature, and, consequently, confer protection against a wide range of herbivores or pathogens (Sicker *et al*. 2000; Ahmad *et al*. 2011).

However, not all induced secondary metabolites have a direct defensive effect. Volatile organic compounds (VOCs) for example are released by the plant and can be perceived by natural enemies to locate their prey (Heil 2008), but also by herbivores themselves (Carroll *et al*. 2006; Robert *et al*. 2012a). Compounds like indole and isoprene have even been shown to interfere with the attraction of natural enemies (D'Alessandro *et al*. 2006; Loivamaki *et al*. 2008). Some induced secondary metabolites also act as radical scavengers (Vickers *et al*. 2009), antibacterial compounds (Huang *et al*. 2012) or hormonal co-factors (Frebortova *et al*. 2010). Thus, herbivore-inducibility *per se* is not a valid predictor for the function of a specific compound. Focusing explicitly on induced toxins may have skewed our view about the actual role

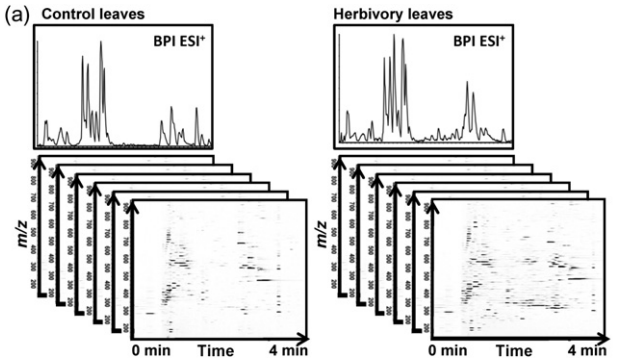
Correspondence: M. Erb. e-mail: merb@ice.mpg.de

and relative importance of herbivore-induced changes in the plant's metabolism for plant resistance (see, for example, Kant & Baldwin 2007; Kroymann 2011).

Another aspect of induced responses that may have been underestimated in the past is that they do not only occur locally in the attacked tissues, but also systemically in non-attacked parts of the plant. For example, alkaloids like nicotine are produced in the roots of tobacco plants, from where they are transported above ground (Dawson & Solt 1959). Furthermore, defensive signals travel to systemic plant parts, where they augment the plant's capacity to respond to subsequent attacks (Heil & Ton 2008). Finally, photoassimilates like sucrose are exported from attacked leaves into non-attacked tissues, possibly to increase the plant's regrowth capacity (Orians, Thorn & Gómez 2011). Systemic changes are particularly important for plant-mediated interactions between herbivores: it has been shown that the metabolic reprogramming of roots following leaf attack has a strong impact on root herbivores, including insect feeders (Soler *et al*. 2007; Erb *et al*. 2011), nematodes (Kaplan *et al*. 2008) and micro-organisms (Yang *et al*. 2011), and *vice versa*, that root herbivory influences above ground resistance against insects (van Dam, Raaijmakers & van der Putten 2005) and pathogens (Erb *et al*. 2009a).

Metabolomics approaches provide the opportunity to assess stress-induced local and systemic changes in plant metabolite patterns without any prior assumptions (Macel, van Dam & Keurentjes 2010). Mass spectrometry (MS)- and nuclear magnetic resonance (NMR)-based methods are increasingly used and have helped to identify novel metabolites of interest (Jansen *et al*. 2009; Leiss *et al*. 2009; Wolfender *et al*. 2009). For example, untargeted short profiling (fingerprints) of leaf extracts based on ultra-high-performance liquid chromatography coupled with high-resolution time-of-flight mass spectrometry (UHPLC-TOF-MS) allowed the identification of biomarkers in wounded *Arabidopsis thaliana* plants (Grata *et al*. 2008; Glauser *et al*. 2008a). Major bottlenecks in MS-based metabolomics are the identification of differentially regulated ions and their absolute quantification *in planta*. This may explain why most metabolomics studies that investigate herbivore-induced changes in plants have only reported general patterns of regulated ions (Sutter & Muller 2011) and their relative changes (Gaquerel *et al*. 2010; Kim *et al*. 2011) or have selected a few compounds of interest for further characterization (Adio *et al*. 2011).

In this study, we employed an unbiased metabolomics approach based on UHPLC-TOF-MS to identify differentially regulated metabolites in maize plants that are attacked by the generalist herbivore *Spodoptera littoralis* (Fig. 1a). Infested maize plants produce a distinct pattern of 1,3-benzoxazin-4-ones (Glauser *et al*. 2011) and are more resistant to *S. littoralis* than uninfested plants (Erb *et al*. 2009a). Furthermore, leaf-attacked plants are more resistant to the root feeder *Diabrotica virgifera*, despite the fact that this specialist is not affected by 1,3-benzoxazin-4-ones (Robert *et al*. 2012b). We therefore hypothesized that apart from 1,3-benzoxazin-4-ones, other locally and systemically induced metabolites may contribute to induced herbivore resistance. In order to obtain a detailed survey of the largest possible number of induced metabolites, we profiled leaves, vascular sap from stems, roots as well as root exudates of *S. littoralis*-infested plants (Fig. 1b). Differentially regulated ions were highlighted in the UHPLC-TOF-MS fingerprints using unsupervised and supervised data mining methods. Taken together, our study provides an unbiased, quantitative view of a wide range of local and systemic herbivore-induced changes of secondary metabolites in maize.



(b)

Organs	Extract	24h after herbivory	48h after herbivory
Leaves	Isopropanol	X	X
Roots	Isopropanol	X	X
Sap	Crude	-	X
Root exudates	Methanol/water	-	X

Figure 1. Schematic representation of the mass spectrometry (MS)-based metabolomics approach used in this study: (a) ultra-high-pressure liquid chromatography time-of-flight mass spectrometry (UHPLC-TOF-MS) fingerprints in positive ionization mode on a short C18 column and the corresponding 2D ion maps of control and *Spodoptera littoralis*-induced leaves (BPI, base peak intensity); (b) summary table of all analysed maize extracts ($n = 12$).

MATERIALS AND METHODS

Plants and insects

Maize (*Zea mays L.*, var. Delprim) was sown in plastic pots (10 cm high, 4 cm diameter) in sand (non-washed, 3–5 mm, JUMBO, Marin, Switzerland) topped with 2 cm of commercial potting soil (Ricoter Aussaaterde, Aarberg, Switzerland) and placed in a climate chamber (23 °C, 60% r.h., 16:8 h L/D, 200 μmol m^{-2}). Plants used for the experiments were 10–12 days old and had 2–3 fully developed leaves. *S. littoralis* eggs were obtained from a laboratory colony from Syngenta (Stein, Switzerland), and reared in plastic boxes ($8 \times 12 \times 3$ cm) under laboratory conditions (24 °C, 40% r.h., 16:8 h L/D, 50 μmol m^{-2}).

Plant treatment, harvest and extraction

To profile the metabolic changes of maize seedlings upon herbivory, plants were transferred to an experimental

chamber equipped with full spectrum light benches (23 °C, 40% r.h., 16:8 h L/D, 90 μmol m^{-2}). The seedlings where then infested with 20 second instar *S. littoralis* larvae or left herbivore free. To stop the herbivores from escaping, polyethylene (PET) tubes (1.5 L, 30 cm height, 7 cm diameter) with their bottom removed were put upside down over the plants. The opening on the top ensured adequate air supply. After 24 and 48 h of induction, the herbivores were removed. The timing was chosen based on previous studies showing a strong induction of defences at these time points (Turlings *et al*. 1998; Erb *et al*. 2011; Glauser *et al*. 2011; Hiltpold *et al*. 2011). The leaf and root tissues were harvested, cleaned from herbivore frass with paper tissue, flash frozen in liquid nitrogen and stored at –80 °C ($n = 12$). Each sample was ground to powder using a mortar previously frozen in liquid nitrogen. The frozen powder was weighed (300 mg ± 2 mg) and 1.5 mL of isopropanol was immediately added for metabolite extraction. Samples were vortexed, sonicated in a bath at room temperature (5200 Bransonic, Danbury, CT, USA) for 20 min, vortexed again and centrifuged at 10 000 rpm for 2 min (Hettich mikrolitter D 7200, Buford, GA, USA). The supernatant was recovered and the extraction procedure was repeated one more time. Each isopropanol extract was dried under vacuum (Genevac HT-4X, Ipswich, UK) and suspended in a mixture of methanol/water 85/15 for a rapid SPE C_{18} enrichment procedure (100 mg C18 cartridge Sep-Pack®, Waters, Milford, MA, USA) to remove highly non-polar compounds. The filtered extracts were dried and dissolved in methanol/water (85/15 v/v) mixture at 1 mg mL^{-1} for UHPLC-TOF-MS analysis. To collect root exudates, the entire root systems of control and *S. littoralis*-infested seedlings (including the potting medium) were placed on a filter funnel of high porosity and set under vacuum 48 h after the beginning of attack in an independent experiment ($n = 12$). The root systems were then washed with 100 mL of a methanol/water mixture (50/50 v/v). This procedure yielded approximately 14 mg of material per plant. Sap from the vascular system was collected in a third experiment by cutting the stems into pieces of 3 cm and placing them in 1.5 mL Eppendorf tubes. The stems where then centrifuged for 5 min at 7000 g. This method yielded approximately 20 μL of xylem and phloem enriched sap per plant. Treatments were identical as in the root exudate collection experiment ($n = 12$).

Rapid metabolite fingerprinting and detailed metabolite profiling

Metabolite analysis was carried out using a UPLC-PDA-TOF-MS instrument (LCT Premier, Waters, MA, USA) equipped with an electrospray (ESI) source. The LC-MS fingerprint of each extract was obtained using a short UPLC BEH C18 Acquity column (50 × 1.0 mm i.d., 1.7 μm, Waters, MA, USA). The mobile phase consisted of 0.1% formic acid (FA) in water (phase A) and 0.1% FA in acetonitrile (phase B). The linear gradient program was as follows: 98% A over 0.2 min, to 100% B over 4.9 min, held at 100% B for a further 1.1 min, then returned to initial conditions (98% A) in 0.1 min for 1.1 min of equilibration before next analysis. The flow rate was 0.3 mL min^{-1}; column temperature was kept at 40 °C. Detection was performed by TOF-MS in W-mode in both electrospray negative (NI) and positive (PI) ion modes in independent runs with the following settings: capillary voltage at 2.8 kV, cone voltage at 40 V, desolvation temperature at 250 °C, source temperature at 120 °C and desolvation gas flow at 600 L h^{-1}. The *m/z* range was 100–1000 Da with a scan time of 0.25 s. The MS was calibrated using sodium formiate, and leucine enkephalin used as an internal reference. The injection volume was 1 μL. The MassLynx software version 4.1 (Waters, MA, USA) was used to control all instruments and determine molecular formulae from accurate *m/z*. In addition, mass spectrometry/mass spectrometry (MS/MS) experiments were carried out using an UPLC-QTOF-MS equipped with an ESI source (QTOF-MS Xevo, Waters, MA, USA). A pool of control and induced plants extracts was separated using an UPLC BEH C18 Acquity column (150 × 2.1 mm i.d., 1.7 μm) with a gradient from 5 to 95% B in 30 minutes at a flow rate of 460 μL min^{-1}. The QTOF-MS was operated in the NI mode at a resolution of approximately 10 000 (full width half maximum). The data were acquired over an *m/z* range of 50–1000 in the MS^E mode using alternating scans of 0.2 s at collision energy of 4 eV and at a collision energy ramp from 15–35 eV. The capillary and cone voltages were set to 2.5 kV and 40 V, respectively. The source temperature was maintained at 120 °C and the desolvation and cone gas flows were set to 900 L h^{-1} at 350 °C and 20 L h^{-1}, respectively.

Data processing and statistical analysis

UHPLC-TOF-MS fingerprints of samples were processed using MarkerLynx 4.1 software (Waters, MA, USA) for mass signal extraction and alignment from 0 to 6 min with *m/z* values from 100 to 1000 Da with the following parameters: noise elimination level was set at 25.00, minimum peak intensity at 200 counts, a mass window of 0.05 and a retention time window of 0.10 min; isotopic peaks were excluded and no internal standard was used. MarkerLynx processing was done for each ionization mode independently. Principal component analysis (PCA) and orthogonal partial least square discriminant analysis (OPLS-DA) models were evaluated with the SIMCA-P software (version 12, Umetrics, Umeå, Sweden). For each model, a leave-one-subject-out cross-validation was performed to assess the model fit. The validity of the discriminant models was verified using permutation tests (Y-scrambling) and CV-ANOVA (P-value < 0.05; Eriksson, Trygg & Wold 2008).

Metabolite identification

Biomarkers were identified by the following approach: (1) the molecular formulae were calculated by MassLynx software based on accurate mass and isotopic pattern recognitions in both PI and NI modes; (2) each suggested molecular

formula was matched with putative structures using the *Dictionary of Natural Products* (Chapman & Hall/CRC), SciFinder Scholar database (SciFinder Scholar™ 2007), KEGG (Kanehisa *et al.* 2010) and KNapSAcK (Shinbo *et al.* 2006); (3) the MS/MS fragmentation of the metabolites was compared with candidate compounds identified in databases or earlier publications, especially when the metabolites were already reported in *Z. mays*; (4) scale-up purification was performed for structural confirmation by ^{1}H-NMR. A discussion on the metabolite identification is provided in the supplementary material and the characteristic MS/MS spectra recorded for representative 1,3-benzoxazin-4-ones are given (Supporting Information Fig. S1).

Metabolite purification

To isolate individual metabolites, about 100 maize seedlings were collected 48 h after infestation by *S. littoralis*. The leaf tissues were harvested, cleaned from herbivore frass, flash frozen in liquid nitrogen and stored at −80 °C. Five hundred and fifty g of leaves were extracted two times in 5 L of isopropanol (Isopropanol purum, VWR, Dietikon, Switzerland) to obtain 8.2 g of crude extract. Extracts were then subjected to medium pressure liquid chromatography (Armen Spot Flash, Saint Ave, France) on a C-18 column (40–63 μm, 37 g, Merck, Rahway, NJ, USA) with a five-step gradient from 100% water to 100% methanol (HPLC grade, VWR) at a flow rate of 10 mL min^{-1} and yielded five fractions. Each fraction was analysed by UHPLC-TOF-MS to localize the compounds of interest. *P*-coumaroyltyramine, DIMBOA-Glc, DIM$_2$BOA-Glc and DIMBOA were found in fraction 3 (640 mg), which corresponded to elution with the 50/50 methanol/water mobile phase composition. The best chromatographic parameters for the final purification step through a semi-preparative C18 column (C18, 250 × 10 mm i.d., 5 μm, XBridge™, Waters, UK) were determined using HPLC modelling software (OSIRIS 4.0, Datalys, Grenoble, France), on the basis of two generic gradients that only differed in slope (Glauser *et al.* 2008b). A gradient of water/acetonitrile from 95/5 to 75/25 in 60 min at a flow rate of 8 mL min^{-1} yielded *p*-coumaroyltyramine (4.5 mg), DIMBOA-Glc (8 mg), DIM$_2$BOA-Glc (8 mg) and DIMBOA (5 mg) in pure form. With the same procedure, DHBOA-Glc (4 mg), HMBOA-Glc (2 mg), HDMBOA-Glc (3 mg) and HDM$_2$BOA-Glc (0.5 mg) were purified from fraction 4. Structures were confirmed by ^{1}H-NMR. NMR analyses were performed on a Varian Unity Inova 500 MHz NMR instrument (Palo Alto, CA, USA) equipped with a 5 μL microflow NMR probe (CapNMR) from Protasis/MRM (Savoy, IL, USA) having an active volume of 2.5 μL. The samples were dissolved in 6.5 μL of DMSO-d$_6$ and parked in the probe with a push volume of 9 μL. The signal of DMSO-d$_6$ at 2.50 ppm was used as reference (temperature 30 °C). L-tryptophan, azelaic acid and rutin were identified by comparison with commercial standards purchased from Sigma-Aldrich (Buchs, Switzerland).

Quantification by MRM

To measure absolute concentrations of elicited compounds in crude plant extracts, a method based on direct nano-infusion MS/MS was developed using the commercially available Advion TriVersa Nanomate chip (Advion BioScience, NY, USA). The Nanomate was equipped with a 96-wheel plate, a rack of 384 conductive pipettes tips and a disposable ESI chip with a 200 × 200 array of nozzles. For each analysis, a new tip and a new nozzle was used to eliminate any possibilities of carry-over. Crude IPA extracts without SPE pretreatment were dissolved at 0.01 mg mL^{-1} in a solution of chloroform/methanol/water 2/7/1 with 5 mM of ammonium acetate and directly infused through the nanochip. Five μL were necessary for each injection. To generate the nano-electrospray, a low delivery of 0.5 psi gas pressure (nitrogen) and a voltage of 1.5 kV were applied. The mass spectrometric response was measured on a triple quadrupole mass spectrometer (TSQ Vantage, Thermo Scientific, Waltham, MA, USA) equipped with an ESI interface using selected reaction monitoring (SRM) in negative ionization mode (Supporting Information Fig. S2). The collision energies and the S-Lens value were optimized for each compound. Absolute quantities were determined using the average value of three standard curves from 0.5 ng to 5000 ng obtained from purified compounds using a 1/x-weighted linear regression model. The matrix effect was estimated by spiking the quantity used to build the calibration curve in a pool of extract at 0.01 mg mL^{-1}. The recovery was above 95% for all quantified compounds. The mass spectrometer parameters were as follows: capillary temperature was maintained at 190 °C, collision gas pressure was set to 1.5 mTorr, Q1 and Q2 peak width were set at 0.05 (FWMH) and a scan width of 0.1 s with a scan time of 0.5 s were applied.

Total nitrogen measurements

To get insight into the role of nitrogen in the induced response of maize, we determined total nitrogen of control and infested maize leaves (48 h, $n = 4$) using the Kjeldahl method (Bremmer and Mulvaney 1982).

P-coumaroyltyramine diet assays

To determine the effect of *p*-coumaroyltyramine on *S. littoralis* performance, the metabolite was extracted and purified from maize seedlings as described. Artificial diet was then prepared as described by (Peñaflor *et al.* 2011) and was spiked with coumaroyltyramine. In a first experiment, the control diet was spiked with 0.6 μg g^{-1}, while the treatment diet was supplemented with 6 μg g^{-1}. These doses correspond to concentrations found *in planta* (see results). Furthermore, taking into account that the diet is substantially more concentrated in nutrients than the plant material, and that an attacking caterpillar will encounter a matrix of induced and uninduced plant tissues, we deemed a concentration of 6 μg g^{-1} suitable to assess the impact of coumaroyltyramine

induction on caterpillar performance. Plastic boxes ($3 \times 3 \times 2$ cm) containing cubes of the different diets were prepared, and four second instar *S. littoralis* larvae were added ($n = 12$). The larvae were left to feed for 10 days. Diet was regularly resupplied to guarantee *ad libitum* feeding. After this time period, the larvae were re-weighed and their weight gain was determined. In a second experiment, we tested whether *S. littoralis* retains or converts after ingestion. Individual third instar larvae were fed with pre-weighed diet cubes (250 mg) that contained 0 or 12 μg g^{-1} coumaroyltyramine ($n = 6$). After 48 h, the remaining diet was weighed. To determine the actual diet consumption by the larvae, additional diet cubes were incubated for 48 h without larvae, and their weight loss due to desiccation was determined for both treatments ($n = 3$). The frass, gut and body of the larvae was collected and weighed as well, and coumaroyltyramine concentrations were determined using the MRM method described above. The stability of coumaroyltyramine was evaluated by measuring coumaroyltyramine concentrations in the diet cubes at the end of the trial. From the obtained tissue weights and concentrations, the total coumaroyltyramine uptake and excretion was calculated. Furthermore, the total amount of coumaroyltyramine in the gut and the body of *S. littoralis* was determined. A full metabolomics analysis of the different tissues was also performed, and the metabolic profiles were analysed with PCA and OPLS-DA as described above.

RESULTS

Short UHPLC-TOF-MS profiles reveal tissue-specific metabolite patterns

Thanks to the rapid UHPLC-TOF-MS gradient conditions that were developed for this metabolomics study, highly reproducible LC-MS fingerprints for more than 280 maize extracts were obtained in both positive (PI) and negative (NI) ESI ion modes (Fig. 1). We detected more than 300 features in leaves and roots of *Z. mays* seedling. 180 features were detected in the vascular sap and 40 features were found in root exudates. Each tissue and biological fluid displayed a specific UHPLC fingerprint (Fig. 2a). The leaf extracts contained predominantly lipophilic compounds, detected at the end of the chromatogram between 3 and 4.5 min, while the root matrices contained high-intensity peaks in the hydrophilic range within the first part of the chromatogram. The vascular saps were equally characterized by a high content of hydrophilic molecules and eluted in the first minute of the gradient. Exudates samples obtained through a soft washing of roots with methanol and water (1/1) showed less complex profiles, with a dominant peak at 1.3 min. Interestingly, this peak was also detected in the roots and leaves.

Unsupervised data mining highlights herbivore-induced changes

Principal component analysis was performed as an exploratory step of data analysis to provide an unsupervised overview of the LC-MS fingerprints. Leaves of herbivore-infested and control plants 24 and 48 h after attack were clearly separated in the PC1 × PC2 score plot (Fig. 2b). The two first principal components explained 65% of the total variance. By contrast, a PCA of vascular sap profiles did not show treatment-specific clustering 48 h after *S. littoralis* infestation. The PCA of root extracts separated the different time points on the first PC axis (67% of the total variance) while no separation was observed between control and leaf-attacked seedlings. Root exudates did not show any treatment-specific clustering.

Supervised data analysis retrieves *S. littoralis*-induced ions in maize leaves

Following the separation of leaf extracts by PCA, a supervised data mining approach (OPLS-DA) was applied to obtain classification models and highlighted putative features involved in the stress response of maize leaves. The OPLS method is designed to separate the predictive part of the data related to the class distinction from the within-class variation that is not related to the response (Trygg & Wold 2002). Separately applied to both time points (Fig. 3), statistically significant models were obtained, providing a clear separation between control and infested leaves. In order to assess the contribution of the detected metabolites to the herbivore attack response, the correlation vectors corr(tp,X), computed from the loadings of the predictive component (first latent variable) of both models (24 and 48 h) were combined to build a shared and unique structure plot (SUS-plot). This representation (Wiklund *et al*. 2008) allowed the detection of features related to the plant defence metabolism according to their position (Fig. 3). Characteristic compounds detected in higher abundance in herbivore-infested leaves 24 and 48 h after infestation are displayed in the upper right square of the SUS plot (features labelled 1–19 in Fig. 3 correspond to the 19 first loadings sorted according to their contribution in OPLS-DA plots). Compounds plotted in the upper left square of the SUS plot are induced only 48 h after attack (20–21), features plotted in the lower right square of SUS plot are induced after 24 h, but not 48 h after attack (31–32, Fig. 3), while compounds in the lower left square are suppressed by herbivore attack (22–30). These biomarkers, highlighted with unbiased statistical methods (Fig. 4, labelled features) where further characterized (see below). OPLS-DA models were also constructed for vascular sap and roots, but did not provide any statistically significant results (P-value > 0.05). However a significant model was obtained for root exudates. Overall, the most discriminant variables highlighted by multivariate data analysis performed independently on PI and NI TOF-MS fingerprints were related to the same compounds. However, the NI mode was generally found to highlight more information mainly for non-polar compounds (e.g. hydroxylated fatty acids) and the main part of this study therefore relies on the NI TOF-MS data.

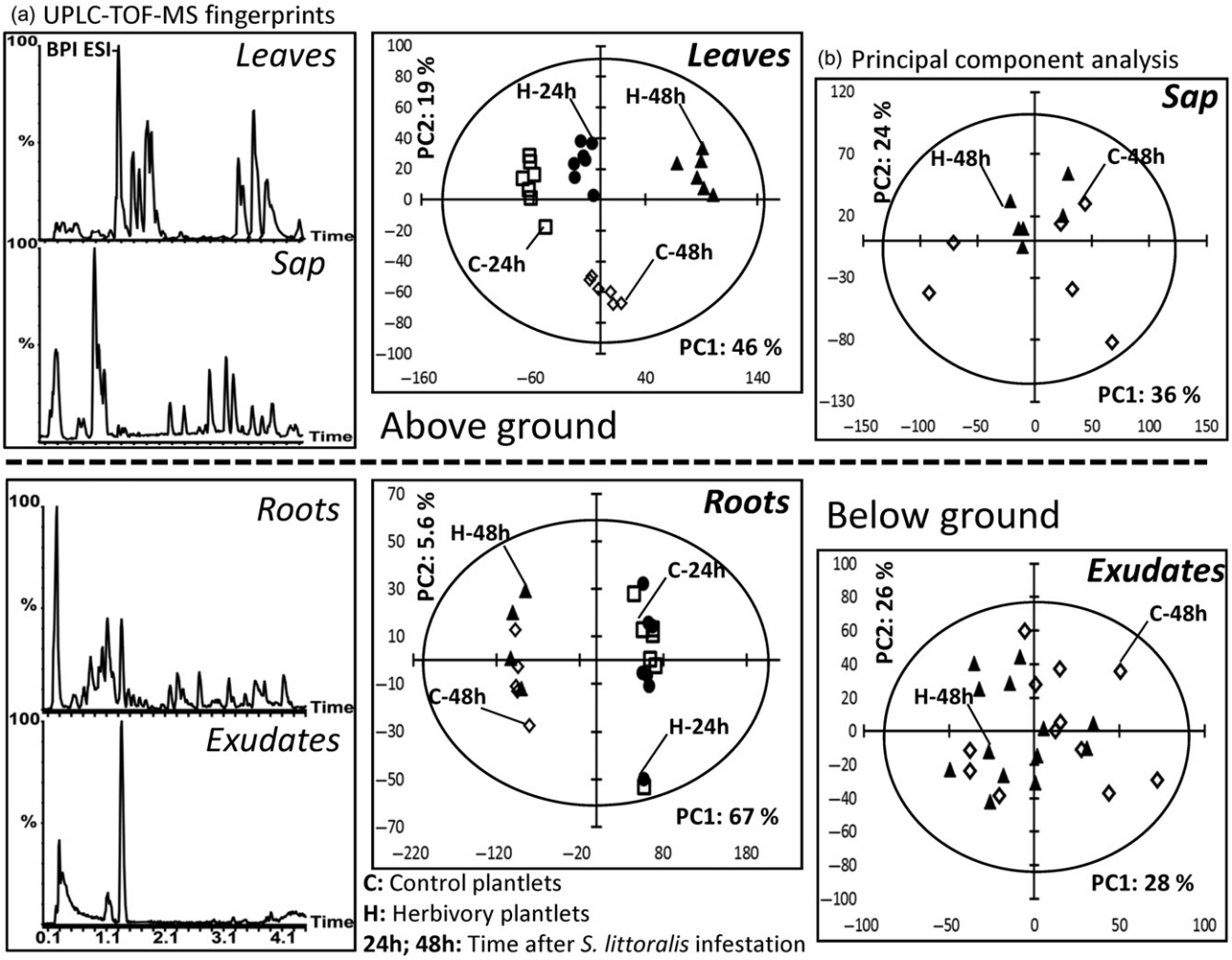


Figure 2. (a) Ultra-high-pressure liquid chromatography time-of-flight mass spectrometry (UHPLC-TOF-MS) fingerprints of different tissues from herbivore-induced maize seedlings. (b) Principal component (PC) analysis of leaves vascular sap, roots and root exudates of control plants (C) and herbivore-attacked plants (H) 24 and 48 h after *Spodoptera littoralis* attack.

Identification of differentially regulated metabolites

The biomarkers related to the 32 features that were most pronouncedly changed by *S. littoralis* infestation were identified by additional detailed high-resolution QTOF-MS/MS profiling (HR-MS/MS) on representative pooled samples (Fig. 4). The LC peak annotation process included molecular formula calculation, heuristic filtering, fragment detection and database matching. When needed, a complete identification was obtained by targeted MS-based isolation of given biomarker and subsequent *de novo* NMR characterization. Taken together, this strategy allowed the full or partial identification of all biomarkers of interest depending on the type of spectroscopic information obtained and previous information on their occurrence in maize (Table 1). A detailed phytochemical account of the applied procedures and fragmentation methods can be found in the supplementary material (Supporting Information Appendix S1).

Evaluation of the relative changes and quantification of the induced compounds

The relative changes in abundance of the 32 biomarkers were measured by manually integrating the peak areas of main ions of their corresponding full MS spectra. Mean values were normalized to the mean of control of each time points (Fig. 5). Unpaired *t*-tests were carried out between control and induced extracts 24 or 48 h after infestation and revealed significant relative changes in intensity. Twenty-four hours after the onset of *S.littoralis* attack, we found a strong increase in tryptophan and the two N-methoxylated benzoxazinone derivatives HDMBOA-Glc and HDM_2BOA-Glc. Moreover, several hydroxycinnamic acid amide derivatives, including coumaroyl- and feruloyl-tyramine conjugates along with coumaroyl-tryptamine, increased in abundance. Two glycerogalactolipids (galactopyranosyl-linaloylglycerol and digalactopyranosyl-linaloylglycerol) and several hydroxylated fatty acids

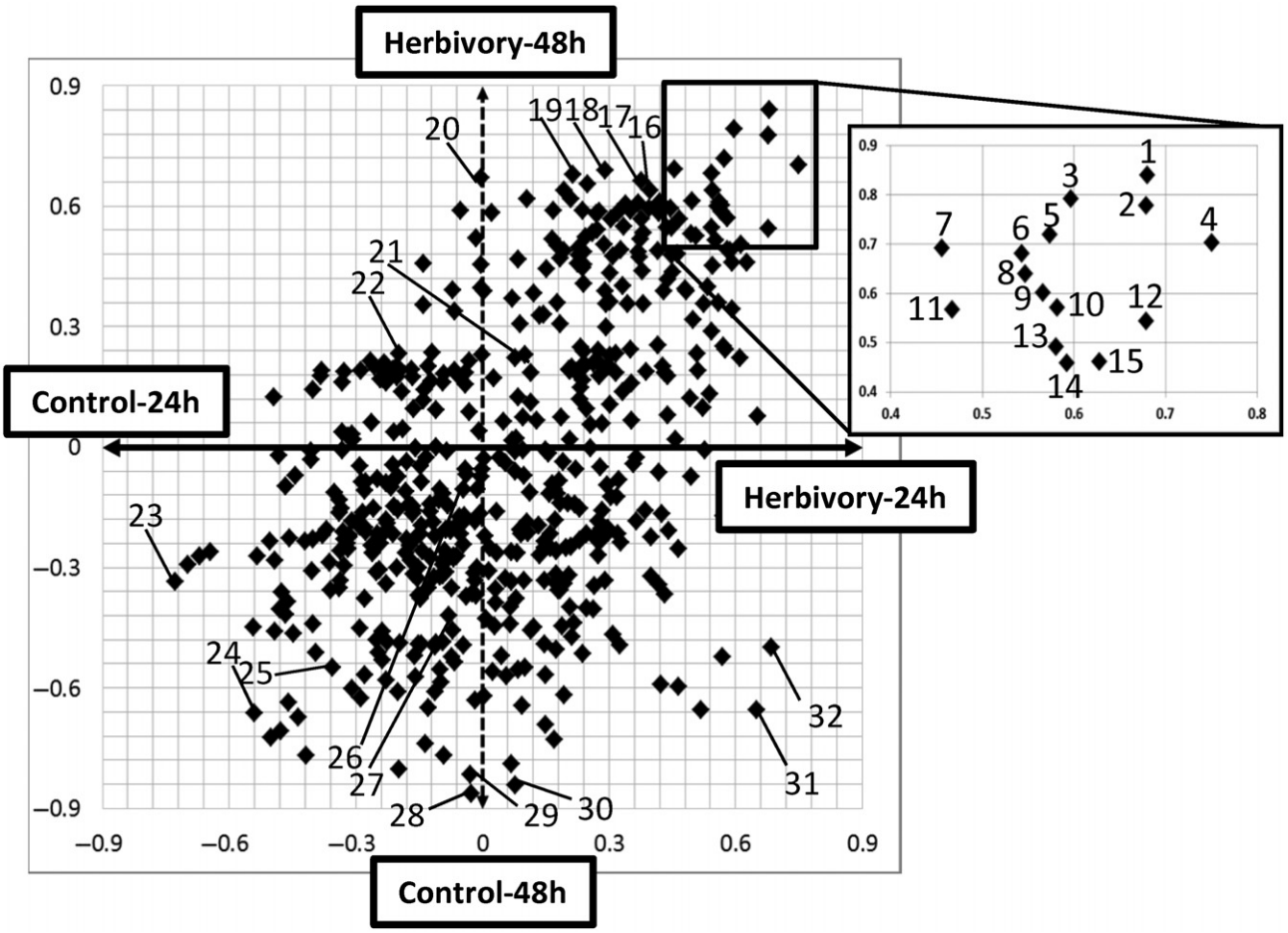


Figure 3. Shared and unique (SUS) plot analysis highlighting the most significant features that change after *Spodoptera littoralis* attack in the leaves of maize plants. *X*-axis: compounds that are of higher abundance in control samples (left) or herbivore-induced samples (right) after 24 h of induction. *Y*-axis: compounds that are of higher abundance in control samples (down) or herbivore-induced samples (up) after 48 h of induction. Features in the upper right quadrant are induced by herbivory after 24 and 48 h. Inset: top-ranked, herbivore-induced features. See Table 1 for compound names.

(18:1-3OH, 18:2-1OH, 18:3-2OH, 18:3-3OH) also showed a slight increase 24 h after infestation compared to controls. Two days after the beginning of herbivore attack, we found an even stronger accumulation of the same compounds. The highest increase was observed for 18:3-3OH, which was induced over 100-fold in infested plants. Moreover, several hydroxylated octadecanoic fatty acids were found to increase. Higher intensity values of the dicarboxylic azelaic acid along with several lysophospholipids were also detected. Contrarily, the relative abundance of the flavonoids rutin, kaempferol-3-*O*-rutinoside and maysin along with the benzoxazinoids DIMBOA and DIMBOA-Glc were reduced in herbivore-attacked leaves. Purification of several Bxs (DIMBOA-Glc, DIM_2BOA-Glc, DHBOA-Glc, HMBOA-Glc, HDMBOA-Glc and HDM_2BOA-Glc), along with the hydroxycinnamic acid amide coumaroyltyramine, allowed setting up an accurate high throughput tandem MS quantification method based on direct nano-infusion of the crude extracts. This simple, original method that does not require chromatographic separation of the analytes complemented the metabolite profiling results by providing accurate quantification of the isolated biomarkers (Fig. 6).

S. littoralis converts and benefits from coumaroyltyramine

Interestingly, caterpillars that fed on a diet with 6 μg g^{-1} coumaroyltyramine gained 30% more mass than caterpillars feeding on 0.6 μg g^{-1} 5 d after the start of the experiment. At the end of the feeding period, caterpillars on the high coumaroyltyramine diet were two times heavier than the controls (Fig. 7a). Over a period of 48 h, *S. littoralis* larvae consumed similar amounts of diet irrespective of the presence or absence of coumaroyltyramine, indicating that the compound does not act as a feeding stimulant (Fig. 7a, inset). On average, each larva feeding on diet containing 12 μg g^{-1} coumaroyltyramine ingested 1 μg of coumaroyltyramine over 2 d of feeding. The amounts of coumaroyltyramine in the gut and body were only about 4 ng, and the frass contained an average of 66 ng per larva (Fig. 7b),

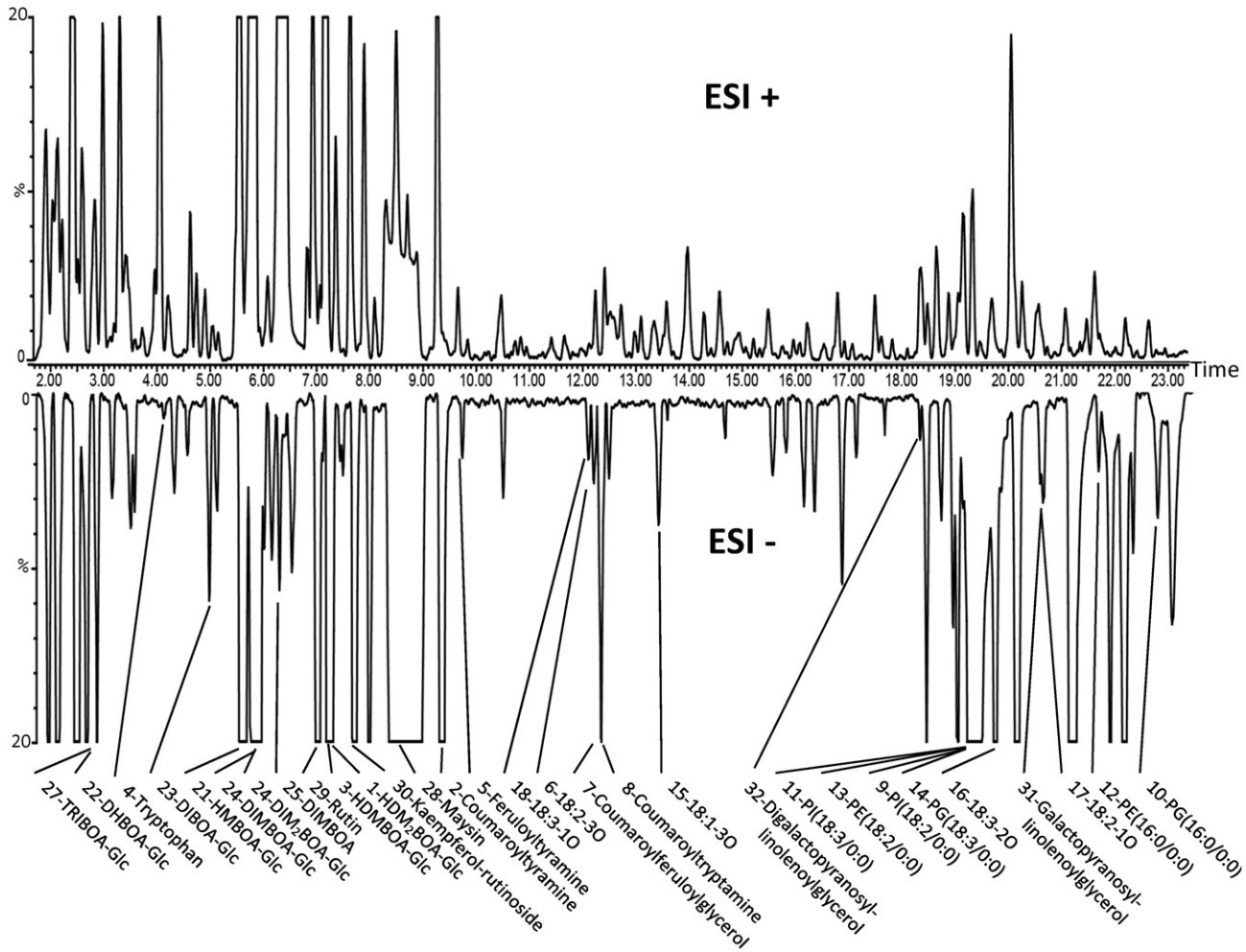


Figure 4. Detailed ultra-high-pressure liquid chromatography time-of-flight mass spectrometry (UHPLC-TOF-MS) profile of pooled leaf extracts from maize seedlings that were attacked by *Spodoptera littoralis* for 48 h. All identified peaks are annotated. Total ion chromatograms were recorded in positive (upper side) and negative ion mode (lower side) using a C18 column with a 30 min water/acetonitrile + 0.1% formic acid gradient. Note that peak intensities do not always correspond to absolute compound abundance due to differential ionization.

showing that *S. littoralis* converted 93% of the coumaroyltyramine to other metabolic products during the digestion process (Fig. 7b, inset). Metabolomics analysis revealed no clear differences between larval tissues from control and coumaroyltyramine containing diet (Fig. 7d), suggesting that the coumaroyltyramine was integrated into the metabolic matrix of the larvae rather than being stored or retained as a detectable derivative. The induction of nitrogen-containing compounds by *S. littoralis* increased the total N content of the leaves marginally (*t*-test: $P = 0.08$, Fig. 7c).

DISCUSSION

This study was motivated by the fact that maize strongly reacts to herbivore attack. Upon the perception of specific elicitors in the saliva of *Spodoptera* larvae (Alborn *et al.* 1997), the plant starts producing large amounts of volatile organic compounds, including oxylipin breakdown products, aromatic compounds and sesquiterpenes (Turlings & Tumlinson 1992). Recently, we found that plants attacked by *S. littoralis* also become more resistant to subsequent infestation by the same species in the leaves (Erb *et al.* 2009a) and against *D. virgifera* in the roots (Erb *et al.* 2011). Induced immunity in the leaves is positively correlated with volatile emissions (Erb *et al.* 2011), but the volatiles themselves do not seem to have any direct toxic or repellent effect on the herbivores (Turlings & Tumlinson 1991), suggesting that non-volatile changes in the metabolome may account for the increased defensive capacity of the plant. Although several toxic secondary metabolites are known to be produced by maize plants (Byrne *et al.* 1996; Frey *et al.* 1997; Nuessly *et al.* 2007; Huffaker *et al.* 2011b), little is known about induced metabolites, and we hypothesized that untargeted metabolomics may provide a

Table 1. Identified compounds in maize leaves

Compound number	Identification[a,b,c]	Retention time (min)	Molecular formula	HR-MS ES (+)[d]	HR-MS ES (–) HR-MS/MS	Calculated molecular formula	Neutral loss (intensity)	Adducts putative loss from $[M-H]^-$	Error (mDa)[e]
1	$HDM_2BOA\text{-}Glc$[a,b]	1.442	$C_{17}H_{23}NO_{11}$	440.1176 $[M+Na]^+$	462.1229	$C_{18}H_{24}NO_{13}$		$[M+FA-H]^-$	1.9
					416.1183	$C_{17}H_{22}NO_{11}$		$[M-H]^-$	1.0
					254.0553	$C_{11}H_{12}NO_6$	162 (40)	-Glc	0.7
					224.0561	$C_{10}H_{10}NO_5$	192 (100)	-Glc-OMe	0.2
					194.0450	$C_9H_8NO_4$	222 (60)	-Glc-2OMe	0.3
2	CT[a,b]	1.730	$C_{17}H_{17}NO_3$	284.1281 $[M+H]^+$	282.1130	$C_{17}H_{16}NO_3$		$[M-H]^-$	0.0
					119.0494	C_8H_7O	163 (100)		0.3
					162.0561	$C_9H_8NO_2$	120 (15)		0.6
3	HDMBOA-Glc[a,b]	1.441	$C_{16}H_{21}NO_{10}$	410.1067 $[M+Na]^+$	432.1142	$C_{17}H_{22}NO_{12}$		$[M+FA-H]^-$	1.6
					386.1087	$C_{16}H_{20}NO_{10}$		$[M-H]^-$	1.0
					224.0559	$C_{10}H_{10}NO_5$	162 (40)	-Glc	0.5
					194.0453	$C_9H_8NO_4$	192 (100)	-Glc-OMe	0.7
					164.0348	$C_8H_6NO_3$	222 (60)	-Glc-2OMe	0.4
4	Tryptophan[a,c]	1.208	$C_{11}H_{12}N_2O_2$	205.0953 $[M+H]^+$	203.0825	$C_{11}H_{11}N_2O_2$		$[M-H]^-$	0.4
					116.0511	C_8H_6N	87 (100)		1.1
5	FT[a]	1.777	$C_{18}H_{19}NO_4$	314.1391 $[M+H]^+$	312.1230	$C_{18}H_{18}NO_4$		$[M-H]^-$	0.6
					148.0529	$C_9H_8O_2$	164 (100)		
6	18:2–3O[a,b]	2.149	$C_{18}H_{32}O_5$	N.D.	327.2175	$C_{18}H_{31}O_5$		$[M-H]^-$	0.4
					309.2091	$C_{18}H_{29}O_4$	18 (20)	$-H_2O$	2.5
					171.1003	$C_9H_{15}O_3$	156 (100)		1.8
					211.1305	$C_{12}H_{19}O_3$	116 (50)		2.9
7	CFg[a]	2.145	$C_{22}H_{22}O_8$	437.1230 $[M+Na]^+$	413.1223	$C_{22}H_{21}O_8$		$[M-H]^-$	1.3
					193.0501	$C_{10}H_9O_4$	220 (60)		0.2
					163.0391	$C_9H_7O_3$	250 (40)		0.4
8	CTr[a]	2.156	$C_{19}H_{18}N_2O_2$	329.1257 $[M+Na]^+$	305.1280	$C_{19}H_{17}N_2O_2$		$[M-H]^-$	1.0
					119.0481	C_8H_7O	186 (100)		1.6
9	PI(18:2/0:0)[a]	3.249	$C_{27}H_{49}O_{12}P$	619.2849 $[M+Na]^+$	595.2889	$C_{27}H_{48}O_{12}P$		$[M-H]^-$	0.6
					279.2328	$C_{18}H_{31}O_2$	316 (100)		0.4
10	PG(16:0/0:0)[a]	3.645	$C_{22}H_{45}O_9P$	507.2769 $[M+Na]^+$	483.2718	$C_{22}H_{44}O_9P$		$[M-H]^-$	0.5
					255.2311	$C_{16}H_{31}O_2$	228 (100)		1.3
11	PI(18:3/0:0)[a]	3.041	$C_{27}H_{47}O_{12}P$	617.2722 $[M+Na]^+$	593.2728	$C_{27}H_{46}O_{12}P$		$[M-H]^-$	0.1
					277.2197	$C_{18}H_{29}O_2$	316 (100)		1.9
12	PE(16:0/0:0)[a]	3.483	$C_{21}H_{44}NO_7P$	454.2937 $[M+H]^+$	452.2780	$C_{21}H_{43}NO_7P$		$[M-H]^-$	0.3
					255.2306	$C_{16}H_{31}O_2$	197 (100)		1.8
13	PE(18:2/0:0)[a]	3.255	$C_{23}H_{44}NO_7P$	478.2913 $[M+H]^+$	476.2780	$C_{23}H_{43}NO_7P$		$[M-H]^-$	0.3
					279.2343	$C_{18}H_{31}O_2$	197 (100)		1.9
14	PG(18:3/0:0)[a]	3.214	$C_{24}H_{43}O_9P$	529.2528 $[M+Na]^+$	505.2562	$C_{24}H_{42}O_9P$		$[M-H]^-$	0.4
					277.2179	$C_{18}H_{29}O_2$	228 (100)		1.1

Table 1. *Continued*

Compound number	Identification[a,b,c]	Retention time (min)	Molecular formula	HR-MS ES (+)[d]	HR-MS ES (–) HR-MS/MS	Calculated molecular formula	Neutral loss (intensity)	Adducts putative loss from $[M-H]^-$	Error (mDa)[e]
15	18:1–3OH[a]	2.256	$C_{18}H_{34}O_5$	N.D.	329.2323	$C_{18}H_{33}O_5$		$[M-H]^-$	0.5
					171.1021	$C_9H_{15}O_3$	158 (100)		0.9
					211.1308	$C_{12}H_{19}O_3$	118 (50)		1.4
16	18:3-2OH[a]	3.029	$C_{18}H_{30}O_4$	N.D.	309.2058	$C_{18}H_{29}O_4$		$[M-H]^-$	0.8
					291.1961	$C_{18}H_{27}O_3$	18 (20)	$-H_2O$	0.1
					197.1167	$C_{11}H_{17}O_3$	112 (100)		
17	18:2-1OH[a]	3.317	$C_{18}H_{32}O_3$	N.D.	295.2289	$C_{18}H_{31}O_3$		$[M-H]^-$	1.6
					279.2310	$C_{18}H_{31}O_2$	18 (100)	$-H_2O$	1.4
					195.1336	$C_{12}H_{19}O_2$	100(20)		4.5
18	18:3-1OH[a]	3.133	$C_{18}H_{30}O_3$	N.D.	293.2108	$C_{18}H_{29}O_3$		$[M-H]^-$	0.9
					275.2012	$C_{18}H_{27}O_2$	18 (100)	$-H_2O$	0.1
					171.1021	$C_9H_{15}O_3$	122 (70)		0
19	18:3-3OH[a]	2.413	$C_{18}H_{30}O_5$	N.D.	325.2021	$C_{18}H_{29}O_5$		$[M-H]^-$	0.6
					307.1898	$C_{18}H_{27}O_4$	18 (20)	$-H_2O$	1.1
					201.1154	$C_{10}H_{17}O_4$	114 (100)		2.7
20	Azelaic acid[a,c]	1.644	$C_9H_{16}O_4$	N.D.	187.0958	$C_9H_{15}O_4$		$[M-H]^-$	1.2
					141.0889	$C_8H_{13}O_2$			2.7
					125.0961	$C_8H_{13}O$			0.5
21	HMBOA-Glc[a,b]	1.288	$C_{15}H_{19}NO_9$	380.0949 $[M + Na]^+$	356.0970	$C_{15}H_{18}NO_9$		$[M-H]^-$	1.2
					194.0449	$C_9H_8NO_4$	162 (100)	-Glc	0.4
					166.0502	$C_8H_8NO_3$	190 (20)	-Glc-CO	0.2
					138.0552	$C_7H_8NO_2$	218 (40)	-Glc-2CO	0.3
22	DHBOA-Glc[a,b]	1.204	$C_{14}H_{17}NO_9$	366.0804 $[M + Na]^+$	342.0809	$C_{14}H_{16}NO_9$		$[M-H]^-$	1.6
					180.0293	$C_8H_6NO_4$	162 (100)	-Glc	0.4
					152.0358	$C_7H_6NO_3$	190 (30)	-Glc-CO	1.5
					124.0405	$C_6H_6NO_2$	218 (90)	-Glc-2CO	0.6
23	DIBOA-Glc[a,b]	1.261	$C_{14}H_{17}NO_9$	366.0818 $[M + Na]^+$	388.0894	$C_{15}H_{18}NO_{11}$		$[M + FA-H]^-$	1.4
					342.0805	$C_{14}H_{16}NO_9$		$[M-H]^-$	2.0
					180.0296	$C_8H_6NO_4$	162 (5)	-Glc	0.1
					162.0204	$C_8H_4NO_3$	180 (10)	-Glc-H_2O	1.3
					134.0251	$C_7H_4NO_2$	208 (100)	-Glc-CO_2	0.9
24	DIMBOA-Glc[a,b]	1.300	$C_{15}H_{19}NO_{10}$	396.0909 $[M + Na]^+$	418.0965	$C_{16}H_{20}NO_{12}$		$[M + FA-H]^-$	2.1
					372.0931	$C_{15}H_{18}NO_{10}$		$[M-H]^-$	0.0
					210.0408	$C_9H_8NO_5$	162 (10)	-Glc	0.6
					192.0296	$C_9H_6NO_4$	180 (10)	-Glc-H_2O	0.1
					164.0337	$C_8H_6NO_3$	208 (90)	-Glc-CO_2	1.1
					149.0115	$C_7H_3NO_3$	223 (100)	-Glc-CO_2-CH_3	0.2

25	DIMBOA[a]	1.375	$C_9H_9NO_5$	212.0559 $[M + H]^+$	210.0407	$C_9H_8NO_5$		$[M-H]^-$	0.5
					192.300	$C_9H_6NO_4$	18 (5)	$M-H_2O$	0.3
					164.0344	$C_8H_6NO_3$	46 (30)	$M-H_2O-CO_2$	0.4
					149.0116	$C_7H_3NO_3$	61 (100)	$M-H_2O-CO_2-CH_3$	0.3
26	DIM_2BOA-Glc[a,b]	1.315	$C_{16}H_{21}NO_{11}$	426.0994 $[M + Na]^+$	448.1101	$C_{17}H_{22}NO_{13}$		$[M + FA-H]^-$	1.0
					402.1043	$C_{16}H_{20}NO_{11}$		$[M-H]^-$	0.7
					240.0528	$C_{10}H_{10}NO_6$	162 (10)	M-Glc	2.0
					222.0421	$C_{10}H_8NO_5$	180 (10)	$-Glc-H_2O$	1.9
					194.0448	$C_9H_8NO_4$	208 (90)	$-Glc-CO_2$	0.5
					179.0216	$C_8H_5NO_4$	223 (100)	$-Glc-CO_2-CH_3$	0.3
					163.9996	$C_7H_2NO_4$	239 (40)	$-Glc-CO_2-2CH_3$	1.2
27	TRIBOA-Glc	1.192	$C_{14}H_{17}NO_{10}$	382.0754 $[M + Na]^+$	404.0824	$C_{15}H_{18}NO_{12}$		$[M + FA-H]^-$	0.5
					358.0796	$C_{14}H_{16}NO_{10}$		$[M-H]^-$	2.2
28	Maysin[a]	1.690	$C_{27}H_{28}O_{14}$	599.1389 $[M + Na]^+$	575.1390	$C_{27}H_{27}O_{14}$		$[M-H]^-$	1.1
					411.0700	$C_{21}H_{15}O_9$	164 (100)		1.6
29	Rutin[a,c]	1.403	$C_{27}H_{30}O_{16}$	633.1428 $[M + Na]^+$	609.1459	$C_{27}H_{29}O_{16}$		$[M-H]^-$	0.3
					301.0328	$C_{15}H_9O_7$	301 (100)		2.0
30	Kaemp-Rut[a]	1.615	$C_{27}H_{30}O_{15}$	595.1669 $[M + H]^+$	593.1495	$C_{27}H_{29}O_{15}$		$[M-H]^-$	1.1
					285.0413	$C_{15}H_9O_6$	308 (100)		1.4
31	GLg[a]	3.279	$C_{27}H_{46}O_9$	537.3044 $[M + Na]^+$	559.3118	$C_{28}H_{47}O_{11}$		$[M + FA-H]^-$	1.3
					513.3049	$C_{27}H_{45}O_9$		$[M-H]^-$	1.5
					277.2167	$C_{18}H_{29}O_2$	236 (100)		0.1
					253.0923	$C_9H_{17}O_8$	260 (10)		0.9
32	DiGLg[a]	3.0456	$C_{33}H_{56}O_{14}$	699.3584 $[M + Na]^+$	721.3659	$C_{34}H_{57}O_{16}$		$[M + FA-H]^-$	1.2
					675.3780	$C_{33}H_{55}O_{14}$		$[M-H]^-$	1.2
					415.1461	$C_{15}H_{27}O_{13}$	260 (40)		0.9
					397.1338	$C_{15}H_{25}O_{12}$	278 (100)		0.8
					277.2174	$C_{18}H_{29}O_2$	236 (60)		0.6

[a]Putative identification by HR-MS and HR-MS/MS; [b]identification by NMR analysis after purification; [c]comparison with commercial standard; [d]major adducts detected; [e]error based on ESI NI mode. Compounds have been numbered according to Fig. 3. N.D., not detected; CT, coumaroyltyramine; FT, feruloyltyramine; CFg, coumaroylferuloylglycerol; CTr, coumaroyltryptamine; Kaemp-Rut, Kaempferol-rutinoside; GLg, Galactopyranosyl-linolenoylglycerol; DiGLg, Digalactopyranosyl-linolenoylglycerol.

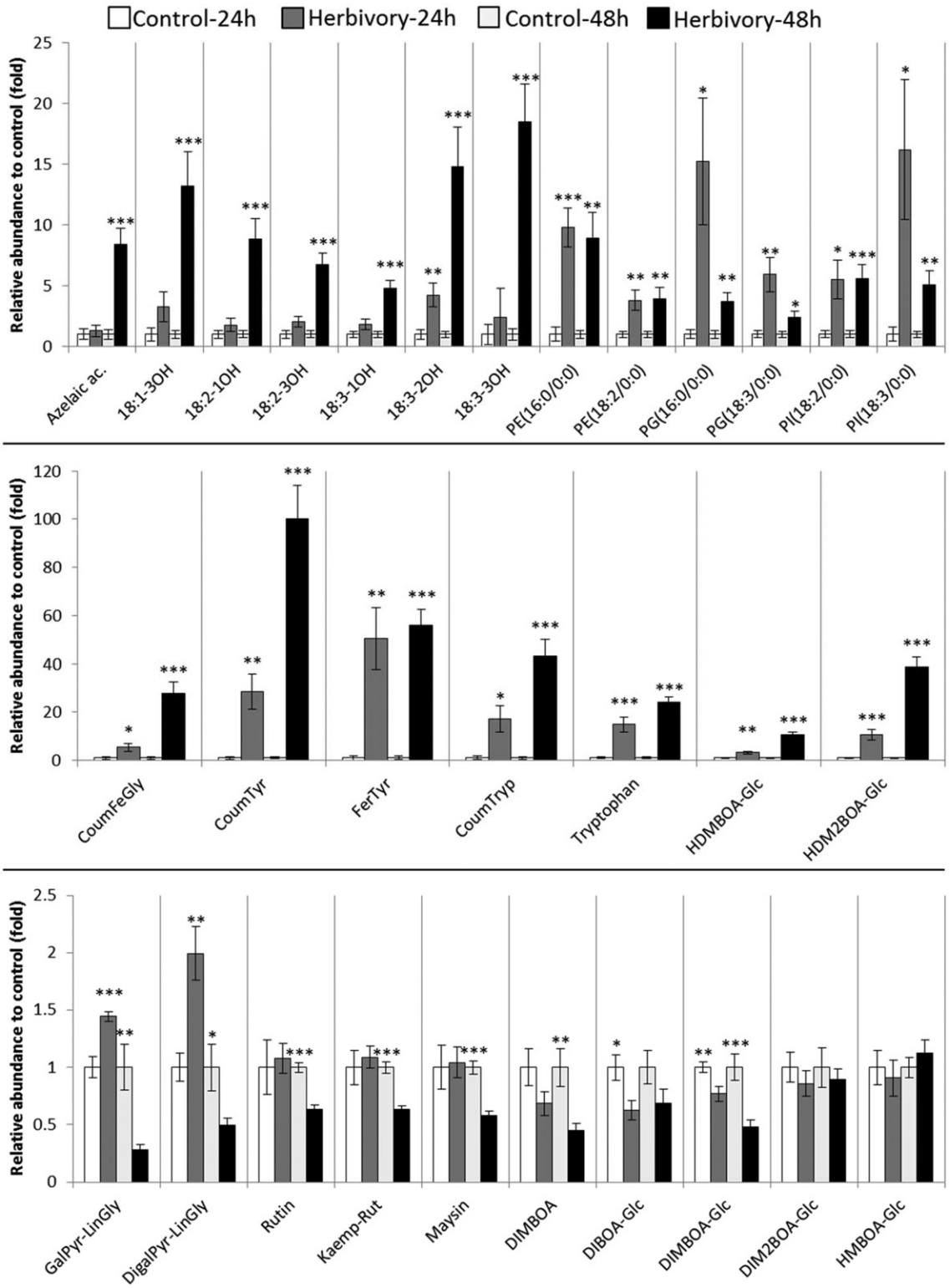


Figure 5. Relative fold changes of 31 biomarkers in herbivore-attacked plants. Values have been calculated from the relative mean peak area of each compound and normalized to control levels ($n = 12$). The two first bars represent control and herbivore-attacked leaves after 24 h and the two last ones control and herbivore-attacked leaves after 48 h (±SEM). Stars indicate significant differences calculated by *t*-tests for 24 h and 48 h time points within time points: $^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$. CoumFeGly, coumaroylferuloylglycerol; CoumTyr, coumaroyltyramine; CoumTryp, coumaroyltryptamine; DigalPyr-LinGly, digalactopyranosyl-linoylglycerol; FerTyr, feruloyltyramine; GalPyr-LinGly, galactopyranosyl-linoylglycerol; Kaemp-Rut, kaempferol-rutinoside; PE, lysophosphatidylethanolamines; PG, lysophosphatidylglycerols; PI, lysophosphatidylinositols.

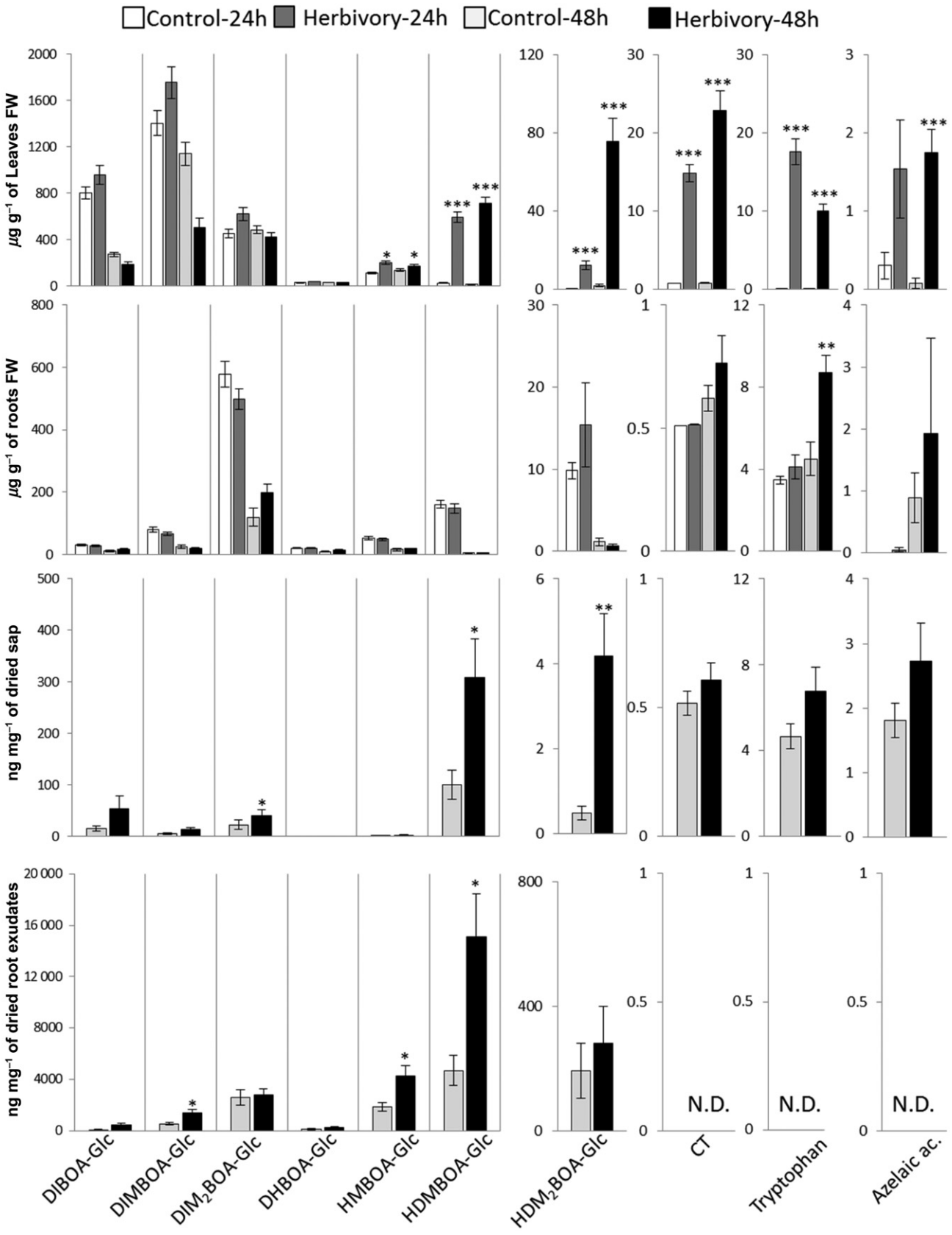


Figure 6. Absolute quantification of 10 herbivore-induced metabolites by nano-infusion-MS/MS. Stars indicate significant differences between treatments within time points: $*P < 0.05$; $**P < 0.01$; $***P < 0.001$. CT, *p*-coumaroyltyramine; N.D., not detected.

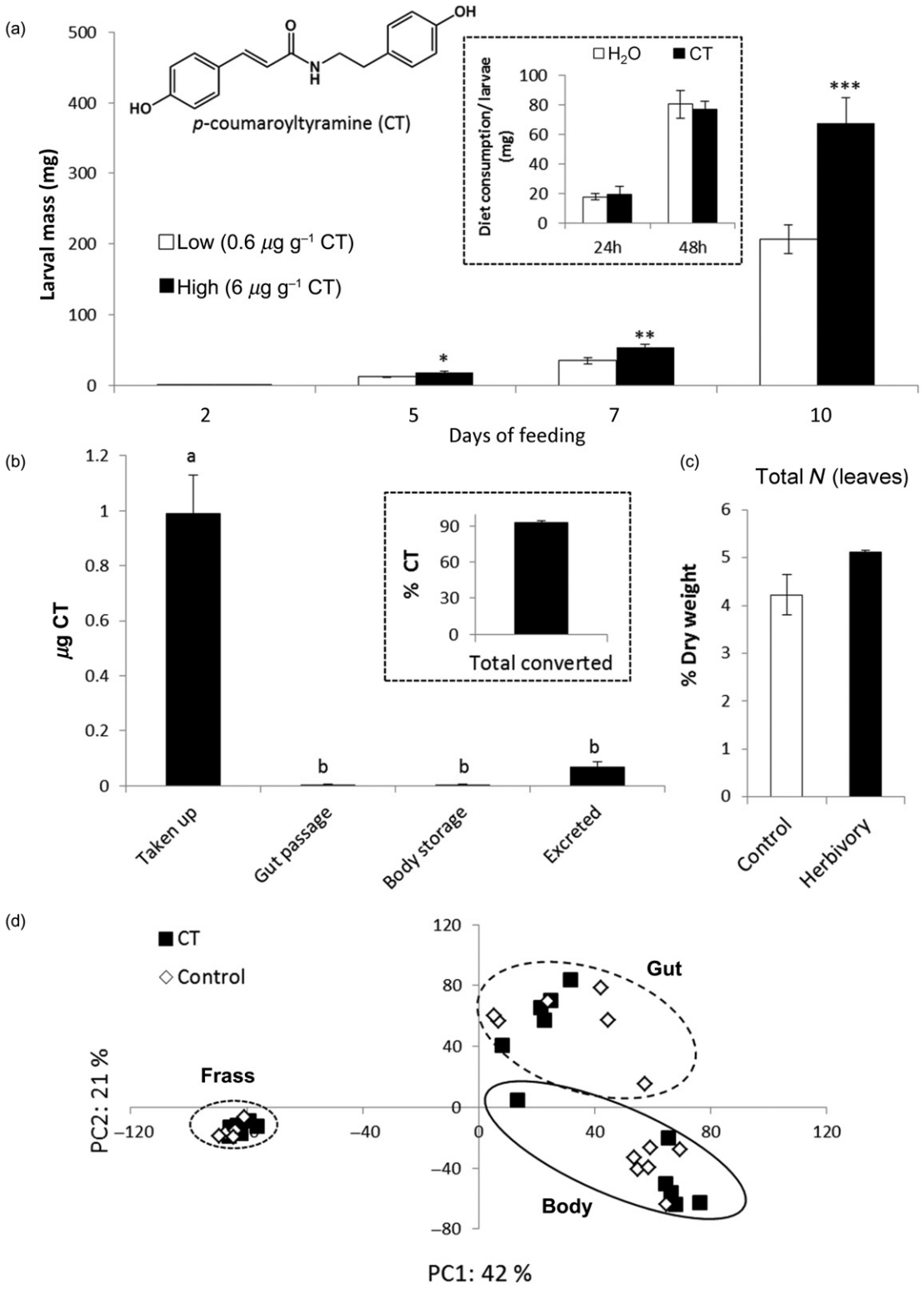


Figure 7. Effect of *p*-coumaroyltyramine (CT) on *Spodoptera littoralis*. (a) Average larval weight gain (mg ± SE) of *S. littoralis* larvae feeding on diet containing 6 μg g^{-1} of CT relative to the weight gain of larvae feeding on diet with 0.6 μg g^{-1}. Inset: diet consumption of *S. littoralis* on control diet and diet containing 12 μg g^{-1} of *p*-coumaroyltyramine. (b) Calculated total amounts of CT that were taken up and retained in the midgut, body and frass of *S. littoralis*. (c) Total nitrogen content of attacked (48 h) and unattacked maize leaves. (d) Principal component analysis of the metabolomic profiles from tissues of *S. littoralis* caterpillars fed on control or CT-containing diet. Stars indicate significant differences between treatments within time points (*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$). Letters indicate significant differences between tissues ($P < 0.05$).

comprehensive overview of the biochemical changes involved in the stress response of leaves and roots.

Our UHPLC-TOF-MS-based metabolomics approach revealed 32 features that are induced upon *S. littoralis* feeding 24 and 48 h after the onset of attack (Fig. 3). The main compounds (1–20) corresponding to these features were identified (Table 1), either by peak annotation based on HR-MS/MS or full characterization by CapNMR (see Supporting Information Fig. S1). The chosen strategy also allowed a clear discrimination between the herbivore-induced response and growth effects (Fig. 2b). For example, the main differences between control leaves 24 and 48 h could be attributed to a slight increase in lysophospholipid content (Fig. 5). The main biomarkers can be categorized into three major classes: fatty acid derivatives, hydroxycinnamic acid amides and 1,3-benzoxazin-4-ones. The strong induction of lysophospholipids in attacked maize leaves (20- to 100-fold after 48 h) is particularly noteworthy (Fig. 5). To our knowledge, this phenomenon has not been observed in herbivore-attacked plants before, but it has been previously shown that wounding induces important changes in the lipid content of plant cells. For instance, a rapid and systemic elevation of phosphatidic acid and lysophospholipids was found in tomato leaves in response to wounding (Lee *et al*. 1997). Following wounding or application of systemin, a phospholipase A is activated, with concomitant release of lysophosphatidylcholine (Narvaez-Vasquez, Florin-Christensen & Ryan 1999). An increase in lysophospholipids is consistent with the fact that wounding induces the release of polyunsaturated fatty acids (Ryu & Wang 1998), which are precursors for the formation of oxylipins such as jasmonic acid. The possible connection of phospholipids to jasmonates is illustrated by the fact that silencing a phospholipase D in rice reduces induced jasmonic acid levels (Qi *et al*. 2011). The accumulation of lysophospholipids in infested maize leaves may thus be involved in the activation of defence signalling cascades via the production of oxylipins (Schmelz, Alborn & Tumlinson 2003; Erb *et al*. 2009a). However, our untargeted approach did not reveal accumulation of jasmonates at the time points considered. These signalling molecules are known to occur at very low concentration upon wounding or herbivory and might have been present below the detection limits of our generic fingerprinting method. Unsaturated, hydroxylated fatty acids were also induced 48 h after *S. littoralis* attack (Fig. 5). Oxidation products of linolenic and linoleic acids have antifungal properties (Masui, Kondo & Kojima 1989) and are structural components of the plant cuticle (Kolattukudy 1974). The reason for the induced changes in these compounds in maize remains to be elucidated. Interestingly, we also found a pronounced induction of azelaic acid 48 h after *S. littoralis* attack. Azelaic acid has been implicated in systemic priming of *Arabidopsis thaliana* following pathogen attack (Jung *et al*. 2009), and it remains to be determined whether this compound is involved in similar processes in maize.

One of the striking changes that are likely to be mediated by the herbivore-induced signals is the observed induction of benzoxazinoid derivatives (Bxs). The two most strongly elicited 1,3-benzoxazin-4-ones, HDMBOA-Glc and HDM_2BOA-Glc (Fig. 5), have previously been found to accumulate in the leaves upon jasmonic acid treatment (Oikawa *et al*. 2001), elicitation with the peptide PEP1 (Huffaker, Dafoe & Schmelz 2011a) wounding (Oikawa *et al*. 2004) and *S. frugiperda* attack (Glauser *et al*. 2011). Upon tissue disruption, the glucosides are hydrolysed by β-glucosidase to release the toxic aglycones. The aglycone of HDMBOA-Glc in particular is very reactive and can deter both generalist and specialist herbivores (Glauser *et al*. 2011). HDMBOA-Glc is thought to be formed from DIMBOA-Glc (Oikawa, Ishihara & Iwamura 2002), and the reduction of DIBOA-Glc, DIMBOA-Glc and DIMBOA in attacked leaves (Fig. 5) may reflect this induced methylation process. In addition, we observed a strong accumulation of tryptophan in attacked leaves, possibly as a result of an increase in the activity of the shikimate pathway that supplies the benzoxazinoid branch.

Another interesting class of metabolites that were strongly induced in the leaves of *S. littoralis*-attacked maize plants are *N*-hydroxycinnamoyltyramines. (*E*)-feruloyltyramine and *(E)-p*-coumaroyltyramine have been found to accumulate in tomato in a JA-independent manner following wounding and chitosan treatment (Pearce *et al*. 1998), in pepper leaves inoculated with the pathogen *Xanthomonas campestris*, and in wounded *Nicotiana attenuata* and maize tissues (Ishihara *et al*. 2000). Coumaroyltryptamine has been reported to occur in maize kernels (Ehmann 1974), and 1-*p*-coumaroyl-3-feruloylglycerol was found in *Populus tremula* buds (Isidorov *et al*. 2008). Our study shows that these two compounds are induced by herbivory as well (Fig. 5). Despite the fact that coumaroyl- and feruloyl-conjugates seem to be stress inducible across many different taxa, little is known about their actual biological role. Two hypotheses have been proposed in this context. Firstly, the conjugates may have antimicrobial effects (Zacares *et al*. 2007). Secondly, they may refortify cell walls as phenolic barriers (Negrel & Jeandet 1987; Ishihara *et al*. 2000). As a third possibility, we wanted to find out whether *p*-coumaroyltyramine also has a direct effect on attacking insect herbivores. Surprisingly, *S. littoralis* larvae grew significantly better on diet containing 6 $\mu g\ g^{-1}$ of *p*-coumaroyltyramine than on diet containing 10 times less of this metabolite. The fact that *p*-coumaroyltyramine did not have any stimulatory effect on the larvae and that it was almost completely converted to unknown products during digestion (Fig. 7) suggests that this compound may be used as a nitrogen source by the caterpillars. However, as the additional nitrogen derived from the *p*-coumaroyltyramine in the artificial diet represents only a fraction of the additional nitrogen that the larvae would need to grow as big as in Fig. 7a, it is likely that the compound may also have other positive effects on the insect. Total nitrogen marginally increased in attacked plants (Fig. 7c), and our study shows that herbivore attack in maize induces nitrogen-containing secondary metabolites that actually increase the susceptibility of the plant. From the perspective of the plant, it remains

to be determined whether *N*-hydroxycinnamoyltyramines are induced to reduce the colonization of wounding sites by microbes, or whether they increase plant resistance indirectly by fortifying cell walls.

Interestingly, most of the herbivore-induced changes in the maize secondary metabolism remained localized in the leaves, as we did not detect any significant changes in metabolite abundance in the roots. The main differences in the root PCAs were related to growth effects (Fig. 2b and Supporting Information Fig. S3). The only significant change in the roots following leaf attack was an increase of tryptophan (Fig. 6). Other studies have found changes in free amino acids in the roots of leaf-attacked tomato (Steinbrenner *et al*. 2011) and wild tobacco plants (Kim *et al*. 2011). In general, roots seem to reconfigure their primary metabolism to support plant defences and tolerance (Schwachtje & Baldwin 2008), and the increase of free tryptophan that we detected may help maize seedlings to satisfy the increased demand for this compound in the leaves. Our experiments confirm the results of an earlier study that found no overlap in transcriptional changes in the leaves and roots of *S. littoralis*-attacked plants (Erb *et al*. 2009b). Clearly, the metabolic changes that occur in the roots are distinct from the leaves, and the main metabolites that are induced locally in the leaves cannot explain the systemic increase in resistance in the roots against the root feeder *D. virgifera* (Erb *et al*. 2011). In contrast to the roots, we did detect changes in the vascular sap and root exudate patterns following leaf herbivory: *S. littoralis*-infested plants exuded significantly more DIMBOA-Glc, HMBOA-Glc and HDMBOA-Glc than control plants (Fig. 6). As we extracted the complete rhizosphere, we may in theory have included small amounts of *S. littoralis* frass. However, the higher quantities of Bxs in the collected exudates are unlikely to stem from the frass, as *S. littoralis* does not excrete any HDMBOA-Glc (Glauser *et al*. 2011). As benzoxazinoid synthesis does not seem to be induced in the roots, it is possible that Bxs may be transported from the leaves to the roots via the vascular bundles and released into the rhizosphere. Alternatively, they may be synthesized in higher amounts in the roots and directly exported via the exudates. A recent study shows that Bxs recruit the plant growth promoting bacterium *Pseudomonas putida* to maize roots (Neal *et al*. 2012), and it is tempting to speculate about the possibility that attacked maize plants could recruit beneficial microbes to help reduce the negative effects of leaf herbivory.

Overall, our study demonstrates that leaf herbivory induces a variety of local and systemic changes in the plants' metabolome. These changes can have contrasting effects on herbivore resistance and may affect other organisms that are directly or indirectly connected to the plant. Untargeted metabolomics can therefore not only lead to the discovery of novel dynamically regulated metabolites, but can also help to create novel hypotheses about how insect herbivores influence plant–environment interactions via induced responses. The approach we present here may serve as a template to study plant stress responses in an unbiased manner, from metabolic profiling to activity mapping.

ACKNOWLEDGMENTS

The work was supported by the National Centre of Competence in Research (NCCR) Plant Survival, a research programme of the Swiss National Science Foundation. This work is also supported by a Swiss National Science Foundation Fellowship to M.E. (PBNEP3-134930). We are grateful to Howard and Isabelle Riezman (UNIGE) for their technical help for the nano-infusion MS/MS experiments and to Laurence Marcourt (UNIGE) for assistance in the NMR measurements.

REFERENCES

Adio A., Casteel C., Vos M.D., Kim J., Joshi V., Li B., Juéry C., Daron J., Kliebenstein D. & Jander G. (2011) Biosynthesis and defensive function of n-acetylornithine, a jasmonate-induced Arabidopsis metabolite. *The Plant Cell* **23,** 3303–3318.

Ahmad S., Veyrat N., Gordon-Weeks R., *et al*. (2011) Benzoxazinoid metabolites regulate innate immunity against aphids and fungi in maize. *Plant Physiology* **157,** 317–327.

Alborn H.T., Turlings T.C.J., Jones T.H., Stenhagen G., Loughrin J.H. & Tumlinson J.H. (1997) An elicitor of plant volatiles from beet armyworm oral secretion. *Science* **276,** 945–949.

Bremmer J.M. & Mulvaney C.S. (1982) Nitrogen-total. In *Methods of Soil Analysis, Part 2. Chemical and Microbiological Properties* (eds A.L. Page, R.H. Miller & D.R. Keeney), pp. 595–624. American Society of Agronomy, Madison, WI, USA.

Byrne P.F., McMullen M.D., Snook M.E., Musket T.A., Theuri J.M., Widstrom N.W., Wiseman B.R. & Coe E.H. (1996) Quantitative trait loci and metabolic pathways: genetic control of the concentration of maysin, a corn earworm resistance factor, in maize silks. *Proceedings of the National Academy of Sciences of the United States of America* **93,** 8820–8825.

Cambier V., Hance T. & de Hoffmann E. (2000) Variation of DIMBOA and related compounds content in relation to the age and plant organ in maize. *Phytochemistry* **53,** 223–229.

Carroll M.J., Schmelz E.A., Meagher R.L. & Teal P.E.A. (2006) Attraction of *Spodoptera frugiperda* larvae to volatiles from herbivore-damaged maize seedlings. *Journal of Chemical Ecology* **32,** 1911–1924.

D'Alessandro M., Held M., Triponez Y. & Turlings T. (2006) The role of indole and other shikimic acid derived maize volatiles in the attraction of two parasitic wasps. *Journal of Chemical Ecology* **32,** 2733–2748.

van Dam N.M., Raaijmakers C.E. & van der Putten W.H. (2005) Root herbivory reduces growth and survival of the shoot feeding specialist *Pieris rapae* on *Brassica nigra*. *Entomologia Experimentalis et Applicata* **115,** 161–170.

Dawson R.F. & Solt M.L. (1959) Estimated contributions of root and shoot to the nicotine content of the tobacco plant. *Plant Physiology* **34,** 656–661.

Ehmann A. (1974) N-(ρ-coumaryl)-tryptamine and N-ferulyl tryptamine in kernels of *Zea mays*. *Phytochemistry* **13,** 1979–1983.

Erb M., Flors V., Karlen D., de Lange E., Planchamp C., D'Alessandro M., Turlings T.C.J. & Ton J. (2009a) Signal signature of aboveground-induced resistance upon belowground herbivory in maize. *The Plant Journal* **59,** 292–302.

Erb M., Lenk C., Degenhardt J. & Turlings T.C.J. (2009b) The underestimated role of roots in defense against leaf attackers. *Trends in Plant Science* **14,** 653–659.

Erb M., Robert C.A.M., Hibbard B.E. & Turlings T.C.J. (2011) Sequence of arrival determines plant-mediated interactions between herbivores. *Journal of Ecology* **99,** 7–15.

Eriksson L., Trygg J. & Wold S. (2008) CV-ANOVA for significance testing of PLS and OPLS (R) models. *Journal of Chemometrics* **22,** 594–600.

Frebortova J., Novak O., Frebort I. & Jorda R. (2010) Degradation of cytokinins by maize cytokinin dehydrogenase is mediated by free radicals generated by enzymatic oxidation of natural benzoxazinones. *The Plant Journal* **61,** 467–481.

Frey M., Chomet P., Glawischnig E., Stettner C., Grun S., Winklmair A., Eisenreich W., Bacher A., Meeley R. & Briggs S. (1997) Analysis of a chemical plant defense mechanism in grasses. *Science* **277,** 696–699.

Frey M., Schullehner K., Dick R., Fiesselmann A. & Gierl A. (2009) Benzoxazinoid biosynthesis, a model for evolution of secondary metabolic pathways in plants. *Phytochemistry* **70,** 1645–1651.

Gaquerel E., Heiling S., Schoettner M., Zurek G. & Baldwin I.T. (2010) Development and validation of a liquid chromatography-electrospray ionization-time-of-flight mass spectrometry method for induced changes in *Nicotiana attenuata* leaves during simulated herbivory. *Journal of Agricultural and Food Chemistry* **58,** 9418–9427.

Glauser G., Grata E., Rudaz S. & Wolfender J.L. (2008a) High-resolution profiling of oxylipin-containing galactolipids in *Arabidopsis* extracts by ultra-performance liquid chromatography/time-of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry* **22,** 3154–3160.

Glauser G., Guillarme D., Grata E., Boccard J., Thiocone A., Carrupt P.A., Veuthey J.L., Rudaz S. & Wolfender J.L. (2008b) Optimized liquid chromatography-mass spectrometry approach for the isolation of minor stress biomarkers in plant extracts and their identification by capillary nuclear magnetic resonance. *Journal of Chromatography A* **1180,** 90–98.

Glauser G., Marti G., Villard N., Doyen G.A., Wolfender J.-L., Turlings T.C.J. & Erb M. (2011) Induction and detoxification of maize 1,4-benzoxazin-3-ones by insect herbivores. *The Plant Journal* **68,** 901–911.

Grata E., Boccard J., Guillarme D., Glauser G., Carrupt P.A., Farmer E.E., Wolfender J.L. & Rudaz S. (2008) UPLC-TOF-MS for plant metabolomics: a sequential approach for wound marker analysis in *Arabidopsis thaliana*. *Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences* **871,** 261–270.

Hartmann T. (2007) From waste products to ecochemicals: fifty years research of plant secondary metabolism. *Phytochemistry* **68,** 2831–2846.

Heil M. (2008) Indirect defence via tritrophic interactions. *New Phytologist* **178,** 41–61.

Heil M. & Baldwin I.T. (2002) Fitness costs of induced resistance: emerging experimental support for a slippery concept. *Trends in Plant Science* **7,** 61–67.

Heil M. & Ton J. (2008) Long-distance signalling in plant defence. *Trends in Plant Science* **13,** 264–272.

Hiltpold I., Erb M., Robert C.A.M. & Turlings T.C.J. (2011) Systemic root signaling in a belowground, volatile-mediated tritrophic interaction. *Plant, Cell & Environment* **34,** 1267–1275.

Howe G. & Jander G. (2008) Plant immunity to insect herbivores. *Annual Review of Plant Biology* **59,** 41–66.

Huang M., Sanchez-Moreiras A.M., Abel C., Sohrabi R., Lee S., Gershenzon J. & Tholl D. (2012) The major volatile organic compound emitted from *Arabidopsis thaliana* flowers, the sesquiterpene *(E)*-β-caryophyllene, is a defense against a bacterial pathogen. *New Phytologist* **193,** 997–1008.

Huffaker A., Dafoe N.J. & Schmelz E.A. (2011a) ZmPep1, an ortholog of Arabidopsis elicitor peptide 1, regulates maize innate immunity and enhances disease resistance. *Plant Physiology* **155,** 1325–1338.

Huffaker A., Kaplan F., Vaughan M.M., Dafoe N.J., Ni X., Rocca J.R., Alborn H.T., Teal P.E.A. & Schmelz E.A. (2011b) Novel acidic sesquiterpenoids constitute a dominant class of pathogen-induced phytoalexins in maize. *Plant Physiology* **156,** 2082–2097.

Ishihara A., Kawata N., Matsukawa T. & Iwamura H. (2000) Induction of N-hydroxycinnamoyltyramine synthesis and tyramine N-hydroxycinnamoyltransferase (THT) activity by wounding in maize leaves. *Bioscience Biotechnology and Biochemistry* **64,** 1025–1031.

Isidorov V.A., Brzozowska M., Czyzewska U. & Glinka L. (2008) Gas chromatographic investigation of phenylpropenoid glycerides from aspen (Populus *tremula* L.) buds. *Journal of Chromatography A* **1198,** 196–201.

Jansen J.J., Allwood J.W., Marsden-Edwards E., van der Putten W.H., Goodacre R. & van Dam N.M. (2009) Metabolomic analysis of the interaction between plants and herbivores. *Metabolomics* **5,** 150–161.

Jung H.W., Tschaplinski T.J., Wang L., Glazebrook J. & Greenberg J.T. (2009) Priming in systemic plant immunity. *Science* **324,** 89–91.

Kanehisa M., Goto S., Furumichi M., Tanabe M. & Hirakawa M. (2010) KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Research* **38,** D355–D360.

Kant M.R. & Baldwin I.T. (2007) The ecogenetics and ecogenomics of plant-herbivore interactions: rapid progress on a slippery road. *Current Opinion in Genetics & Development* **17,** 519–524.

Kaplan I., Halitschke R., Kessler A., Rehill B.J., Sardanelli S. & Denno R.F. (2008) Physiological integration of roots and shoots in plant defense strategies links above- and belowground herbivory. *Ecology Letters* **11,** 841–851.

Kim S.G., Yon F., Gaquerel E., Gulati J. & Baldwin I.T. (2011) Tissue specific diurnal rhythms of metabolites and their regulation during herbivore attack in a native tobacco, *Nicotiana attenuata*. *PLoS ONE* **6,** e26214.

Kolattukudy P.E. (1974) Biosynthesis of a hydroxy fatty acid polymer, cutin. Identification and biosynthesis of 16-oxo-9- or 10-hydroxypalmitic acid, a novel compound in *Vicia faba*. *Biochemistry* **13,** 1354–1363.

Kroymann J. (2011) Natural diversity and adaptation in plant secondary metabolism. *Current Opinion in Plant Biology* **14,** 246–251.

Lee S., Suh S., Kim S., Crain R., Kwak J.M., Nam H.G. & Lee Y. (1997) Systemic elevation of phosphatidic acid and lysophospholipid levels in wounded plants. *The Plant Journal: For Cell and Molecular Biology* **12,** 547–556.

Leiss K.A., Choi Y.H., Abdel-Farid I.B., Verpoorte R. & Klinkhamer P.G.L. (2009) NMR Metabolomics of thrips (*Frankliniella occidentalis*) resistance in senecio hybrids. *Journal of Chemical Ecology* **35,** 219–229.

Loivamaki M., Mumm R., Dicke M. & Schnitzler J.P. (2008) Isoprene interferes with the attraction of bodyguards by herbaceous plants. *Proceedings of the National Academy of Sciences of the United States of America* **105,** 17430–17435.

Macel M., van Dam N.M. & Keurentjes J.J.B. (2010) Metabolomics: the chemistry between ecology and genetics. *Molecular Ecology Resources* **10,** 583–593.

Masui H., Kondo T. & Kojima M. (1989) An antifungal compound, 9,12,13-trihydroxy-(E)-10-octadecenoic acid, from *Colocasia antiquorum* inoculated with *Ceratocystis fimbriata*. *Phytochemistry* **28,** 2613–2615.

Morant A.V., Jorgensen K., Jorgensen C., Paquette S.M., Sanchez-Perez R., Moller B.L. & Bak S. (2008) Beta-glucosidases as detonators of plant chemical defense. *Phytochemistry* **69,** 1795–1813.

Narvaez-Vasquez J., Florin-Christensen J. & Ryan C.A. (1999) Positional specificity of a phospholipase A activity induced by wounding, systemin, and oligosaccharide elicitors in tomato leaves. *The Plant Cell* **11,** 2249–2260.

Neal A.L., Shakoor A., Gordon-Weeks R. & Ton J. (2012) Benzoxazinoids in root exudates of maize attract *Pseudomonas putida* to the rhizosphere. *PLoS ONE* **7,** e35498.

Negrel J. & Jeandet P. (1987) Metabolism of tyramine and feruloyltyramine in TMV inoculated leaves of *Nicotiana tabacum*. *Phytochemistry* **26,** 2185–2190.

Nuessly G.S., Scully B.T., Hentz M.G., Beiriger R., Snook M.E. & Widstrom N.W. (2007) Resistance to *Spodoptera frugiperda* (Lepidoptera : noctuidae) and *Euxesta stigmatias* (Diptera : ulidiidae) in sweet corn derived from exogenous and endogenous genetic systems. *Journal of Economic Entomology* **100,** 1887–1895.

Oikawa A., Ishihara A., Hasegawa M., Kodama O. & Iwamura H. (2001) Induced accumulation of 2-hydroxy-4,7-dimethoxy-1,4-benzoxazin-3-one glucoside (HDMBOA-Glc) in maize leaves. *Phytochemistry* **56,** 669–675.

Oikawa A., Ishihara A. & Iwamura H. (2002) Induction of HDMBOA-Glc accumulation and DIMBOA-Glc 4-O-methyltransferase by jasmonic acid in poaceous plants. *Phytochemistry* **61,** 331–337.

Oikawa A., Ishihara A., Tanaka C., Mori N., Tsuda M. & Iwamura H. (2004) Accumulation of HDMBOA-Glc is induced by biotic stresses prior to the release of MBOA in maize leaves. *Phytochemistry* **65,** 2995–3001.

Orians C., Thorn A. & Gómez S. (2011) Herbivore-induced resource sequestration in plants: why bother? *Oecologia* **167,** 1–9.

Pearce G., Marchand P.A., Griswold J., Lewis N.G. & Ryan C.A. (1998) Accumulation of feruloyltyramine and p-coumaroyltyramine in tomato leaves in response to wounding. *Phytochemistry* **47,** 659–664.

Peñaflor M.F.G., Erb M., Robert C.A.M., Miranda L.A., Werneburg A.G., Dossi F.C.A. & Turlings T.C.J. & Bento J.M.S. (2011) Oviposition by a moth suppresses constitutive and herbivore-induced plant volatiles in maize. *Planta* **234,** 207–215.

Pickett J.A., Smiley D.W.M. & Woodcock C.M. (1999) Secondary metabolites in plant-insect interactions: dynamic systems of induced and adaptive responses. In *Advances in Botanical Research Incorporating Advances in Plant Pathology* (ed. J.A. Callow) Vol. **30**, pp. 91–115. Elsevier, Amsterdam, the Netherlands.

Qi J., Zhou G., Yang L., Erb M., Lu Y., Sun X., Cheng J. & Lou Y. (2011) The chloroplast-localized phospholipases D α4 and α5 regulate herbivore-induced direct and indirect defenses in rice. *Plant Physiology* **157,** 1987–1999.

Robert C.A.M., Erb M., Duployer M., Zwahlen C., Doyen G.R. & Turlings T.C.J. (2012a) Herbivore-induced plant volatiles mediate host selection by a root herbivore. *New Phytologist* **194,** 1061–1069.

Robert C.A.M., Veyrat N., Glauser G., *et al.* (2012b) A specialist root herbivore exploits defensive metabolites to locate nutritious tissues. *Ecology Letters* **15,** 55–64.

Ryu S.B. & Wang X.M. (1998) Increase in free linolenic and linoleic acids associated with phospholipase D-mediated hydrolysis of phospholipids in wounded castor bean leaves. *Biochimica Et Biophysica Acta-Lipids and Lipid Metabolism* **1393,** 193–202.

Schmelz E., Alborn H. & Tumlinson J. (2003) Synergistic interactions between volicitin, jasmonic acid and ethylene mediate insect-induced volatile emission in *Zea mays*. *Physiologia Plantarum* **117,** 403–412.

Schwachtje J. & Baldwin I.T. (2008) Why does herbivore attack reconfigure primary metabolism? *Plant Physiology* **146,** 845–851.

Shinbo Y., Nakamura Y., Altaf-Ul-Amin M., Asahi H., Kurokawa K., Arita M., Saito K., Ohta D., Shibata D. & Kanaya S. (2006) KNApSAcK: a comprehensive species-metabolite relationship database. In *Plant Metabolomics* (eds K. Saito, R.A. Dixon & L. Willmitzer), pp. 165–181. Springer, Berlin, Heidelberg, Germany.

Sicker D., Frey M., Schulz M. & Gierl A. (2000) Role of natural benzoxazinones in the survival strategy of plants. In *International Review of Cytology – A Survey of Cell Biology* (ed. K.W. Jeon) Vol. 198, pp. 319–346. Academic Press Inc, San Diego, CA, USA.

Soler R., Bezemer T.M., Cortesero A.M., Van der Putten W.H., Vet L.E.M. & Harvey J.A. (2007) Impact of foliar herbivory on the development of a root-feeding insect and its parasitoid. *Oecologia* **152,** 257–264.

Steinbrenner A.D., Gomez S., Osorio S., Fernie A.R. & Orians C.M. (2011) Herbivore-induced changes in tomato (*Solanum lycopersicum*) primary metabolism: a whole plant perspective. *Journal of Chemical Ecology* **37,** 1294–1303.

Sutter R. & Muller C. (2011) Mining for treatment-specific and general changes in target compounds and metabolic fingerprints in response to herbivory and phytohormones in *Plantago lanceolata*. *New Phytologist* **191,** 1069–1082.

Trygg J. & Wold S. (2002) Orthogonal projections to latent structures (O-PLS). *Journal of Chemometrics* **16,** 119–128.

Turlings T.C.J. & Tumlinson J.H. (1991) Do parasitoids use herbivore-induced plant chemical defenses to locate hosts? *Florida Entomologist* **74,** 42–50.

Turlings T.C.J. & Tumlinson J.H. (1992) Systemic release of chemical signals by herbivore-injured corn. *Proceedings of the National Academy of Sciences of the United States of America* **89,** 8399–8402.

Turlings T.C.J., Lengwiler U.B., Bernasconi M.L. & Wechsler D. (1998) Timing of induced volatile emissions in maize seedlings. *Planta* **207,** 146–152.

Vickers C.E., Possell M., Cojocariu C.I., Velikova V.B., Laothawornkitkul J., Ryan A., Mullineaux P.M. & Hewitt C.N. (2009) Isoprene synthesis protects transgenic tobacco plants from oxidative stress. *Plant, Cell & Environment* **32,** 520–531.

Wiklund S., Johansson E., Sjostrom L., Mellerowicz E.J., Edlund U., Shockcor J.P., Gottfries J., Moritz T. & Trygg J. (2008) Visualization of GC/TOF-MS-based metabolomics data for identification of biochemically interesting compounds using OPLS class models. *Analytical Chemistry* **80,** 115–122.

Wolfender J.L., Glauser G., Boccard J. & Rudaz S. (2009) MS-based plant metabolomic approaches for biomarker discovery. *Natural Product Communications* **4,** 1417–1430.

Yang J.W., Yi H.-S., Kim H., Lee B., Lee S., Ghim S.-Y. & Ryu C.-M. (2011) Whitefly infestation of pepper plants elicits defence responses against bacterial pathogens in leaves and roots and changes the below-ground microflora. *Journal of Ecology* **99,** 46–56.

Zacares L., Lopez-Gresa M.P., Fayos J., Primo J., Belles J.M. & Conejero V. (2007) Induction of p-coumaroyidopamine and feruloyidopamine, two novel metabolites, in tomato by the bacterial pathogen *Pseudomonas syringae*. *Molecular Plant-Microbe Interactions* **20,** 1439–1448.

Received 30 March 2012; accepted for publication 13 July 2012

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Figure S1. UHPLC-QTOF-MS/MS fragmentation pattern in ESI negative ionization mode for three benzoxazinones.
Figure S2. Multiple reaction monitoring transition table used for nano-infusion-MS/MS quantification and structures of monitored standards. Abbreviation: CT: *p*- coumaroyltyramine.

Figure S3. Relative fold changes of 11 leaf-biomarkers detected in the roots of leaf-herbivore attacked plants (±SEM). Values have been calculated from the relative mean peak area of each compound ($n = 12$) and normalized to controls at 24 h.
Appendix S1. Identification of biomarkers by high resolution UHPLC-QTOF-MS/MS and 1HNMR.