



**HAL**  
open science

## Estimations et déterminants de la diversité génétique

Thibault Leroy

► **To cite this version:**

Thibault Leroy. Estimations et déterminants de la diversité génétique. Génétique des populations [q-bio.PE]. Université d'Angers, 2022. tel-04065311

**HAL Id: tel-04065311**

**<https://hal.science/tel-04065311>**

Submitted on 11 Apr 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

- Université d'Angers -

**HABILITATION A DIRIGER DES RECHERCHES  
(SPÉCIALITÉ SCIENCES DE LA VIE)**

*Estimations et déterminants  
de la diversité génétique*



**Par :**

**Thibault LEROY**

Institut de Recherche en Horticulture et Semences  
(INRAE, Institut Agro, Université d'Angers)

*Soutenance publique : 13 Juin 2022*

**Jury :**

Rapporteurs :

**M. Sylvain GLEMIN**, Directeur de recherche, CNRS Rennes  
**Mme Maud TENAILLON**, Directrice de recherche, CNRS Paris-Saclay  
**M. Xavier VEKEMANS**, Professeur des universités, Université de Lille

Examinatrices :

**Mme Anne ROULIN**, Professeur assistant, Université de Zurich  
**Mme Carole SMADJA**, Directrice de recherche, CNRS Montpellier

# TABLE DES MATIERES

RESUME	3
REMERCIEMENTS	4
CURRICULUM VITAE	6
SYNTHESE DES ACTIVITES SCIENTIFIQUES	13
<i>I – ESTIMATIONS ET EVOLUTION DE LA DIVERSITE GENETIQUE</i>	13
A) <i>ORDRE DE GRANDEUR DE LA VARIATION GENETIQUE ENTRE ESPECES</i>	13
B) <i>ISOLEMENT GEOGRAPHIQUE VIS-A-VIS DE SES CONGENERES</i>	16
C) <i>ÉVOLUTION DE LA DIVERSITE GENETIQUE CHEZ LES ESPECES DOMESTIQUEES</i>	20
D) <i>PAYSAGES GENOMIQUES DE LA DIVERSITE ET EVOLUTION AU COURS DE LA SPECIATION</i>	25
<i>II – DETERMINANTS DE LA DIVERSITE GENETIQUE</i>	29
A) <i>TAILLES EFFICACES</i>	29
B) <i>TAUX DE MUTATION</i>	33
PROJET DE RECHERCHE ET PERSPECTIVE DE CARRIERE	38
<i>I – EVOLUTION DE LA DIVERSITE GENETIQUE DANS LE COMPARTIMENT CULTIVE</i>	38
<i>II – DIVERSITE GENETIQUE ET BIOLOGIE DE LA CONSERVATION</i>	39
<i>III – TAUX DE MUTATIONS ET EFFICACITE DE LA SELECTION</i>	41
BIBLIOGRAPHIE	46

## RESUME

### *Estimations et déterminants de la diversité génétique*

La diversité génétique au sein des espèces est le substrat autorisant les changements évolutifs. Il s'agit de toute une série de différences génétiques (appelés polymorphismes) au sein des génomes d'une même espèce, de différentes natures, très simples à très complexes. Bien que la majorité de ces mutations n'ait aucun effet sur la survie et la fécondité d'un individu, certaines mutations peuvent être défavorables ou favorables. Une partie de cette diversité est donc essentielle à l'adaptation des espèces aux changements environnementaux, notamment aux changements climatiques actuels et futurs.

Bien que cette prise de conscience de l'importance de la diversité génétique soit plus récente par rapport à d'autres dimensions de la biodiversité, la Convention sur la diversité biologique, un traité international placé sous l'égide du programme des Nations Unies pour l'environnement, établit désormais des ambitions fortes pour la protection de la diversité génétique des espèces. Cela inclut notamment une sauvegarde d'au moins 90% de leur diversité génétique à un horizon 2050, à la fois chez les espèces sauvages et domestiquées.

Cette diversité génétique est en effet non seulement essentielle au maintien d'une grande variété d'espèces sauvages, mais elle est aussi cruciale pour la préservation et l'amélioration des espèces domestiquées. Les programmes d'amélioration visent à répondre à des enjeux variés de société et d'environnement, tels que la réduction de l'usage de pesticides en agriculture, une attente majeure en occident, ou une meilleure tolérance à la sécheresse, un enjeu de plus en plus crucial dans différentes régions du monde pour contribuer à assurer le droit fondamental de l'humain à être à l'abri de la faim.

Après avoir fait une rétrospective de mes activités de recherche autour de l'étude de la diversité génétique d'espèces variées, sauvages et domestiquées, je me concentrerai sur la présentation des résultats de mes études les plus récentes autour des déterminants de cette diversité génétique : les tailles efficaces et le taux de mutations. Je conclurai en évoquant des pistes de recherche que je souhaite continuer à explorer en lien avec ces déterminants.

## REMERCIEMENTS

A la fin de mes remerciements de thèse, je concluais par un « *dès que le vent soufflera, je repartira* ». A l'heure d'écrire ce mémoire, mes remerciements vont donc à toutes celles et tous ceux qui m'ont permis de construire ce parcours, rafale après rafale, lors des 10 dernières années. J'ai en particulier eu la chance d'intégrer des beaux projets (ANRs Genoak & BirdIslandGenomic, ERC Treepeace, entre autres) et une grande partie des lauriers doivent revenir à leurs porteurs, puisqu'en l'absence de contrat pérenne, je n'ai eu qu'une marge de manœuvre limitée sur les choix scientifiques.

La première rafale a probablement été la plus déterminante. La présentation de ce mémoire implique que je sois resté dans l'académique. C'était pourtant loin d'être écrit et cela, je le dois en particulier à Christophe Plomion et à Antoine Kremer qui ont cru en moi, m'ont fait (re)découvrir ce métier et, au fil des années, m'ont laissé toujours plus d'autonomie dans mes choix scientifiques. Au-delà de Christophe et Antoine, ce passage à Pierroton m'a permis de rencontrer de nombreuses autres personnalités, toutes aussi brillantes et attachantes les unes que les autres (je ne me risquerai pas à toutes les lister). Mais j'ai une pensée particulière pour Stefanie Wagner et Isabelle Lesur, mes chères « co-bureaux » pour ces belles années de discussions plus ou moins scientifiques, ainsi que Catherine Bodénès, Sophie Gerber, Myriam Heuertz, Céline Lalanne, les membres de la plateforme PGTB, parmi bien d'autres personnalités pierrotonnaises. Je n'oublie pas la très regrettée Marjorie Vidal. J'y ajoute tous les (post)doctorants que j'ai pu rencontrer au cours de ces années-là, Christelle Fraïsse, Camille Roux, Quentin Rougemont, et bien d'autres.

La seconde rafale a été un bon coup de tramontane qui m'a soufflé jusqu'à l'ISEM. Ces deux années ont été pour moi un vrai régal. Tout d'abord, pour les dimensions purement scientifiques associées à un projet génial. Un grand merci à Benoit Nabholz de m'avoir confié la fin de cette ANR et de m'avoir fait confiance sur un projet d'évolution moléculaire. Cela m'a ouvert des horizons scientifiques nouveaux, qui j'espère seront perceptibles à la lecture de ce mémoire. Les moments incroyables de terrain, à la Réunion notamment, ne le seront pas par contre : la beauté pure de la nature (éruption et coulées de lave, couchers de soleil au-dessus des nuages), ainsi que les nuitées inoubliables à grelotter en bivouac à 2300m d'altitude. Au-delà de Benoit, c'est toute la « Zobo team » que je remercie : Borja Milá, Christophe Thébaud, Maya Mould & Maëva Gabrielli. En outre, ces deux années à l'ISEM ont été pour moi riches en réflexion sur la situation de la recherche, sur ses dérives aussi, en particulier celles d'un système de publication productiviste et économiquement ruineux, un système en somme complètement désuet.

Ensuite, une bourrasque m'a emporté jusqu'en Autriche (telle à faire décoller mon quintal au dessus des Alpes). J'ai été plus que fier que Christian Lexer me fasse confiance et m'invite à le rejoindre comme professeur assistant, plus que dévasté quelques mois plus tard par son départ soudain à seulement quelques jours de Noël 2019. Je n'étais ni préparé à cette déchirure ni aux responsabilités nouvelles que cela allait engendrer. Sans l'aide d'Ovidiu Paun, une particulièrement belle personne, je n'aurais raisonnablement pas pu prendre la suite de Christian pour l'encadrement des 3 thèses en cours, dont les thèses déjà bien avancées de Huiying Shang et Gil Yardeni, et celle à peine initiée de Clara Groot Crego, pas plus que l'intégralité de la responsabilité du stage de « M2 » de Francesca Beclin. Pour ne rien arranger, une pandémie mondiale s'en est ensuite rapidement mêlée. Dans un contexte aussi compliqué humainement, socialement et scientifiquement, je ne peux être qu'admiratif du travail et de la ténacité de mes doctorantes et étudiantes. Sur un tout autre registre, je suis aussi admiratif de mes étudiants pour avoir suivi et positivement évalué mes enseignements universitaires en... français !

Enfin, un nouveau coup de vent, ou peut-être plus vraisemblablement un oiseau migrateur - je l'ignore -, m'a ramené au nid. Après plus d'une année isolé au-delà des frontières, la décision de quitter l'équipe de Vienne pour revenir chez moi sur Angers a été très difficile à prendre. Je remercie Laurence Hibrand, Fabrice Foucher et Jérémy Clotault pour avoir rendu cela possible. Cela m'a permis de découvrir le modèle rosier, son histoire et sa génétique si complexe, ainsi que les belles personnes qui y consacrent leurs vies, Thérèse Loubert en particulier. En outre, revenir dans son université m'a permis de faire une rétrospective de mon parcours, ce qui explique en partie l'existence de cette HDR.

L'heure est venue de me faire souffler dessus (pas trop j'espère !) par les membres du jury. Un grand merci à Anne, Carole, Maud, Sylvain et Xavier d'avoir immédiatement accepté ma proposition. J'en suis extrêmement honoré. Faute de temps, le document final n'est pas vraiment à la hauteur de ce que j'avais rêvé qu'il soit, je m'en excuse. Au détour, en parlant de souffle, je ne peux qu'espérer que le prochain coup de vent sera le dernier, un peu las de tous ces déracinements, années après années.

Une partie de mes recherches ayant lieu en Ukraine, je ne peux avoir à l'heure de l'écriture de ce document que des pensées émues en évoquant l'invasion de l'Ukraine et le sort de mes collaborateurs sur place, leurs familles, leurs amis et au final tout un peuple. Au-delà de l'Ukraine, je tiens à noter aussi que l'ensemble de mes recherches se déroule sur la Terre. Comme de nombreux biologistes, je garde avant tout l'amour pour la nature et le goût pour son observation. Qu'en restera-t-il demain ? Puisse le monde ouvrir les yeux sur l'ensemble de notre œuvre collective de destruction.

*Dès que le vent tournera, on s'assagira.*

# CURRICULUM VITAE

**Thibault LEROY,**  
Né le 19 mai 1986

IRHS, Centre INRAE Angers-Nantes  
BP 60057, 49071 Beaucouzé  
thibault.leroy@inrae.fr

## Parcours universitaire

2009 - 2012 : Thèse de doctorat, Université d'Angers, France

2007 - 2009 : Master Biologie Végétale Intégrative: Gène, Plante, Agrosystème (BioVIGPA), avec spécialisation en génétique et pathologie des plantes, Université d'Angers et Agro-Campus Rennes, mention AB

2004 - 2007 : Licence Sciences de la vie et géosciences, parcours biologie cellulaire, moléculaire et physiologie, avec spécialisation en microbiologie et physiologie des plantes, Université d'Angers, France, mention B (obtention de la bourse du mérite)

2004 : Baccalauréat scientifique, mention AB.

## Expériences professionnelles

### **2009- 2012:**

Doctorant, équipe écologie évolutive des champignons (Ecofun), IRHS, INRAE/Université d'Angers, France – Financement Conseil Général 49

*Au cours de ma thèse, je me suis intéressé à l'évolution de *Venturia inaequalis*, un champignon, agent responsable de la tavelure, une des principales maladies du pommier cultivé et de ses espèces apparentées. Mes travaux se sont intéressés aux conséquences évolutives et épidémiologiques d'un contact secondaire, c'est-à-dire des échanges génétiques (flux de gènes) de populations initialement isolées géographiquement qui se retrouvent à nouveau en contact de manière fortuite, avec les programmes de sélection du pommier. J'ai ainsi pu contribuer à mettre en évidence une augmentation de l'agressivité du pathogène en lien avec le programme de sélection pour la résistance.*

### **2013:**

Postdoctorant, équipe écologie évol. des champignons (Ecofun), IRHS – Financement ANR

*Suite à ma thèse, je suis resté quelques mois dans le même laboratoire pour travailler sur la génomique de ce pathogène et l'identification des gènes associés au pouvoir pathogène, gènes codant des protéines du champignon et qui sont sécrétées hors de la cellule du champignon pour faciliter l'infection.*

### **2014 - 2018:**

Postdoctorant, Biogeco, INRA Bordeaux, France – Financement ANR (1 an) et ERC (3 ans)

*J'ai ensuite eu la chance de travailler avec Christophe Plomion et Antoine Kremer, ainsi que de nombreux autres collègues bordelais, pendant 4 ans. Sur un site à taille humaine, j'ai profité d'une grande liberté thématique. Si mes*

recherches étaient au départ associées à la seule étude des gènes de la spéciation dans le complexe d'espèce des chênes blancs, j'ai rapidement pu élargir le nombre de questions étudiées, en m'intéressant à l'histoire évolutive de ces espèces, aux transferts d'adaptations entre elles et au processus de mutation.

### **2018 - 2019:**

Postdoctorant, ISEM (Institut des sciences de l'évolution de Montpellier, ISEM), Université de Montpellier, France – Financement ANR

*J'ai aussi eu la chance de rejoindre l'ISEM, et de profiter du travail quotidien stimulant avec des généticiens intéressés par une grande diversité de questions des sciences de l'évolution. Dans le cadre de l'ANR BirdIslandGenomic porté par Benoit Nabholz, j'ai pu m'ouvrir à une nouvelle thématique (évolution moléculaire) et à de nouveaux modèles (passereaux), en analysant un jeu de données composé de centaines de génomes complets d'espèces d'oiseaux insulaires et continentaux.*

### **2019 - 2021 :**

Assistant universitaire (équivalent d'Assistant Professor), département de botanique et de biodiversité, - équipe de génomique évolutive des plantes et systématique – Université de Vienne, Autriche

*A ce stade de ma carrière, j'ai ensuite voulu acquérir à la fois une expérience internationale et une expérience d'enseignement. J'ai rejoint le laboratoire de Christian Lexer à l'Université de Vienne sur un contrat offrant d'excellentes conditions de travail, incluant notamment une liberté totale d'action pour 50% de mon temps de recherche. Peu après mon arrivée, le décès brutal de Christian a mis un frein aux projets initiaux. J'ai alors mis de côté mes propres axes de recherche pour me concentrer sur ceux de l'équipe, avec comme priorité l'aboutissement des thèses de trois doctorants du laboratoire. Cela a été une expérience extrêmement forte. Après plus d'un an de restrictions associées au COVID-19, j'ai décidé de rentrer en France et de continuer à conseiller et épauler les étudiants depuis la France.*

### **2021 - 2022:**

Postdoctorant, Équipe Génétique et diversité des ornementales – Institut de recherche en horticulture et semences (IRHS) – INRAE - Université d'Angers - Institut Agro – Financement RFI Objectif Végétal

*Depuis mai 2021, je suis post-doctorant à l'IRHS dans l'équipe de génétique et diversité des ornementales et je m'intéresse particulièrement à l'évolution des rosiers de jardin au cours du XIX<sup>e</sup> siècle. Mon contrat arrivera à son terme en juin 2022. Le projet me permet d'explorer des dimensions non seulement populationnelles mais aussi quantitatives (GWAS pour de nombreux caractères).*

## **Intérêts scientifiques**

- Evolution de la diversité génétique au sein et entre les espèces
- Adaptations locales aux environnements changeants
- Diversification, spéciation, zones hybrides
- Flux de gènes, introgression (neutre, adaptative ou maladaptative)
- Taux et spectres de mutation
- Évolution moléculaire, théorie neutre et quasi-neutre de l'évolution



- Génétique des pathogènes et épidémiologie évolutive
- Apports interdisciplinaires des connaissances en génétique, en particulier à la philosophie des sciences
- Bioinformatique et développements méthodologiques en génétique des populations et quantitatives (inférences démographiques, scans génomiques, génétique d'association)

### **Activités d'enseignement**

#### Université de Vienne

Activités dans le cadre de mon contrat d'assistant universitaire (2019-2021). Tous mes enseignements sont en anglais.

- Cours magistraux du module d'enseignement 300127 : “*Approches génomique pour la variation, adaptation et spéciation : une feuille de route*” en génétique des populations (étudiants en master et doctorat)
- Travaux pratiques en génétique des populations organisés du module d'enseignement 300369 : “*Approches génomiques pour la variation, adaptation et spéciation*”, organisé à distance (étudiants en master et doctorat)
- Journal Club – Écologie et génomique évolutive (semestre d'été) et Génomique de la spéciation (semestre d'hiver)
- Cours magistral et travaux dirigés en biostatistiques : Introduction aux statistiques en biologie évolutive (niveau licence 2)

#### Autres enseignements ponctuels :

- Depuis 2016 : enseignements ponctuels et individuels d'introduction aux langages informatiques, à la bioinformatique ou à la génétique (enseignement à distance, plateforme “Superprof”)

### **Activités d'encadrement**

*Doctorantes* : depuis fin 2019, je co-encadre avec Ovidiu Paun 3 doctorantes à l'université de Vienne : Huiying Shang (thèse soutenue le 4 juin 2021), Gil Yardeni (soutenance prévue au second semestre 2022) et Clara Groot Crego (soutenance prévue courant 2023). Cette situation n'a pas été anticipée puisqu'elle est liée au décès soudain de Christian Lexer. Dans le cadre de la thèse de Huiying sur l'histoire évolutive des peupliers (*Populus* section *Populus*) et des paysages génomiques de la divergence, j'ai pris la responsabilité scientifique du second chapitre (Shang *et al.* en révision). Les thèses de Gil et Clara font le pont entre la phylogénomique et la génomique des populations et se focalisent sur les *Tillandsias*, des plantes épiphytes ayant connu une radiation évolutive récente. En plus de ces trois thèses, j'ai conseillé plusieurs autres doctorants, dont deux de manière plus longue et active (articles en second auteur), Nastasia Merceron (INRAE de Bordeaux, 2013-2016) et Ahmad Muhammad (Université de Vienne, 2018-2021).

*Etudiants en master* : J'ai régulièrement accueilli des étudiants en master, particulièrement au cours de ma thèse (2 étudiantes en L3, 4 étudiants de M1), dont 2 qui sont désormais docteurs : Déborah Merda, qui a soutenu sa thèse en 2016 et qui est actuellement chercheuse titulaire à l'ANSES et Guillaume Baudouin qui a soutenu sa thèse en 2017. Plus récemment, entre mars 2020 et juin 2021,

j'ai encadré Francesca Beclin, une étudiante en M2 à l'université de Vienne. Francesca est actuellement en thèse avec Magnus Nordborg (Gregor Mendel Institute, Vienne). Son sujet de stage portait sur l'inférence démographique et les scans génomiques chez deux espèces de peupliers. Ces travaux sont encore en cours et visent à être valorisés par une publication à l'avenir. Je co-encadre actuellement le stage de M1 d'Amandine Touzé (université de Bordeaux) avec Sophie Gerber (avril-juin 2022).

*Postdoctorant* : Depuis février 2021, je co-encadre Sylvain Schmitt (INRAE Antilles-Guyane), avec Myriam Heuertz et Niklas Tysklind. Sylvain est recruté sur le projet TreeMutation (CEBA, 200k€), notamment pour assurer la partie détection des mutations somatiques (WP4, dont je suis le coordinateur). Deux articles sont prévus, le premier faisant appel à un gros travail de simulations et de réanalyses de données existantes (Schmitt *et al.* soumis) et le second renseignant des effectifs et des spectres de mutations entre branches fortement ou faiblement exposées aux UV.

*Comités de suivi de thèse* : Maëva Gabrielli sur l'histoire de la divergence des *Zosterops* des Mascareignes (Université de Toulouse, 2018 – 2020, soutenance: le 3 juillet 2020), encadrée par Christophe Thébaud et Boja Milá. ; Susanne Reier (muséum d'histoire naturelle de Vienne, 2021 – 2024) qui étudie les échanges génétiques entre populations de poissons via les karsts, encadrée par Luise Kruckenhauser.

### **Autres activités de recherche**

- Activité de relecture dans des journaux scientifiques (*reviewer*) : Annals of Botany, BMC Plant Biology, Biology Letters, Ecology and Evolution, European Journal of Forest Research, Evolutionary Applications, Forestry, Frontiers in Plant Science, Molecular Ecology, Molecular Ecology Resources, New Phytologist, Peer J, Plant People Planet, Plos Genetics, Tree Genetics & Genomes, Tree Physiology

- Missions de terrain principales :

2014-2017 : participant à diverses missions de terrain dans les forêts de l'ONF ou en Espagne

2018 : participant à 3 semaines d'échantillonnage des oiseaux à La Réunion à 2700m d'altitude, avec campement

2018 coordinateur de la mission de datation par dendrochronologie d'arbres à Tchernobyl (3 jours)

- Coordinateur de "Work Package"

2020-2023 : projet CEBA Treemutation, projet porté par Myriam Heuertz et Niklas Tysklind (200k€) - WP4 : "Caractérisation des processus de mutations chez les arbres néotropicaux", encadrement d'un post-doctorant (Sylvain Schmitt, UMR EcoFoG, INRAE).

### **Engagements éthiques**

- J'apporte mon soutien aux sociétés savantes et autres alternatives aux grands éditeurs mondiaux de la littérature scientifique, notamment via la communauté PCI (*Peer Community In...*). Je suis favorable au partage de l'ensemble des manuscrits en particulier les préprints (BioRxiv et

similaires), données, pipeline informatiques et autres codes, ainsi qu'à des processus de relecture plus transparents, incluant notamment la correspondance avec les auteurs.

- Je soutiens le programme SAFE EVOLUTION qui promeut un environnement de travail sécurisé, inclusif et professionnel pour toutes et tous en biologie évolutive.

- Je refuse les récompenses et prix scientifiques individuels, et plus largement le narcissisme et la mégalomanie en recherche. Je soutiens une science collective et ouverte sur le monde.

- Je suis signataire de l'appel des scientifiques du monde concernant l'urgence climatique (Ripple *et al.* 2020) et de la lettre de soutien au plan stratégique post-2020 pour protéger et restaurer les forêts de l'Union Européenne ("European Green Deal", Commission Européenne, 2021).

### **Articles scientifiques**

*Articles publiés (revues à comité de lecture) - \* Co-1er auteur , • Co-dernier auteur*

Thibault Leroy & Benoit Nabholz. Response to Kratochvíl and Rovatsos. *Current Biology* 2022, 32(1):PR30-31

Dounia Saleh, Jun Chen, Jean-Charles Leplé, Thibault Leroy, Laura Truffaut, Benjamin Dencausse, Céline Lalanne, Karine Labadie, Isabelle Lesur, Didier Bert, Frédéric Lagane, François Morneau, Jean-Marc Aury, Christophe Plomion, Martin Lascoux & Antoine Kremer. Genome-wide evolutionary response of European oaks since the Little Ice Age, *Evolution Letters*, 2022, 6:40-20 (available on BioRxiv, 2021.05.25.445558)

Thibault Leroy, Yoann Anselmetti, Marie-Ka Tilak, Sèverine Bérard, Laura Csukonyi, Maëva Gabrielli, Céline Scornavacca, Borja Milá, Christophe Thébaud, Benoit Nabholz. A bird's white-eye view on avian sex chromosome evolution. *Peer Community Journal*, Volume 1 (2021), article no. e63 (available on BioRxiv: 505610)

Gil Yardeni, Juan Viruel, Margot Paris, Jaqueline Hess, Clara Groot Crego, Marylaure de La Harpe, Norma Rivera, Michael H. J. Barfuss, Walter Till, Valeria Guzmán-Jacob, Thorsten Krömer, Christian Lexer•, Ovidiu Paun• & Thibault Leroy•. Taxon-specific or universal? Using target capture to study the evolutionary history of a rapid radiation. *Molecular Ecology Resources* (available on bioRxiv, 2021.05.20.444989)

Muhammad Ahmad, Thibault Leroy, Nikos Krigas, Eva M. Temsch, Hanna Weiss-Schneeweiss, Christian Lexer, Eva Maria Sehr & Ovidiu Paun. Spatial and ecological drivers of genetic structure in Greek Populations of *Alkanna tinctoria* (Boraginaceae), a polyploid medicinal herb. *Frontiers in Plant Science* 2021, 12:706574 (available on BioRxiv, 2021.05.06.442843v1).

Thibault Leroy, Marjolaine Rousselle, Marie-Ka Tilak, Aude Caizergues, Céline Scornavacca, Maria Recuerda Carrasco, Jérôme Fuchs, Juan Carlos Illera, Dawie H. De Swardt, Christophe Thébaud, Borja Milá & Benoit Nabholz. Endemic island songbirds as windows into evolution in small effective population sizes, *Current Biology* 2021, 31(6):1303-1310 (available on BioRxiv, 2020.04.07.030155).

Antoine Kremer, Christophe Plomion, Thibault Leroy, Hermine Alexandre, Laure Truffaut, Sylvain Delzon, Alexis Ducouso, Brigitte Musch, Jose Manuel Torres Ruiz, Stefanie Wagner. Evolution passée et contemporaine des chênes. *Revue forestière française* 2020, 72(4):297-318 (available on hal, hal-03206333, in French).

Quentin Rougemont, Jean-Sébastien Moore, [Thibault Leroy](#), Eric Normandeau, Eric B. Rondeau, Ruth E. Withler, Donald M. Van Doornik, Penelope A. Crane, Kerry A. Naish, John Carlos Garza, Terry D. Beacham, Ben F. Koop & Louis Bernatchez. Demographic history shaped geographical patterns of deleterious mutation load in a broadly distributed Pacific Salmon. *Plos Genetics*. 2020. Available on bioRxiv (732750).

Maëva Gabrielli, Benoit Nabholz, [Thibault Leroy](#), Borja Milá, Christophe Thébaud. Within-island diversification in a passerine bird. *Proceedings of the Royal Society B*. 2020, 287:20192999.

[Thibault Leroy](#), Jean-Marc Louvet, Céline Lalanne, Grégoire Le Provost, Karine Labadie, Jean-Marc Aury, Sylvain Delzon, Christophe Plomion & Antoine Kremer. Adaptive introgression as a driver of local adaptation to climate in European white oaks. *New Phytologist* 2020; 226: 1171-1182 (special issue: oak genomics).

[Thibault Leroy](#), Quentin Rougemont, Jean-Luc Dupouey, Catherine Bodénès, Céline Lalanne, Caroline Belser, Karine Labadie, Grégoire Le Provost, Jean-Marc Aury, Antoine Kremer, Christophe Plomion: Massive postglacial gene flow between European white oaks uncovered genes underlying species barriers. *New Phytologist* 2020, 226: 1183-1197 (special issue: oak genomics).

[Thibault Leroy](#), Antoine Kremer & Christophe Plomion. Oak symbolism in the light of genomics. *New Phytologist*, 2020; 226:1012–1017 (special issue: oak genomics).

Caroline Pont\*, [Thibault Leroy](#)\*, Michael Seidel\*, Alessandro Tondelli\*, Wandrille Duchemin\*; David Armisen\*; Daniel Lang\*, Daniela Bustos-Korts\*, Nadia Goué, François Balfournier, Marta-Molnar-Lang, Jacob Lage [+20 co-authors including the WHEALBI consortium]: Tracing the ancestry of modern bread wheats. *Nature Genetics* 05/2019; 51 (5): 905

Gerald. A. Tuskan; Andrew T. Groover, Jeremy Schmutz, Stephen Paul DiFazio, Alexander Myburg, Dario Grattapaglia, Lawrence B. Smart, Tongming Yin, Jean-Marc Aury, Antoine Kremer, [Thibault Leroy](#), [+9 co-authors] Hardwood Tree Genomics: Unlocking Woody Plant Biology. *Frontiers in Plant Science*. 12/2018, 9(1799)

Monika Michalecka, Sylwester Masny, [Thibault Leroy](#), Joanna Puławska: Population structure of *Venturia inaequalis*, a causal agent of apple scab, in response to heterogeneous apple tree cultivation. *BMC Evolutionary Biology* 12/2018; 18(1):5

Christophe Plomion\*, Jean-Marc Aury\*, Joelle Amsellem\*, [Thibault Leroy](#)‡, Florent Murat‡, Sébastien Duplessis, Sébastien Faye, Nicolas Francillonne, Karine Labadie, Grégoire Le Provost [+56 other co-authors]: Oak genome reveals facets of long lifespan. 05/2018 *Nature Plants*, 4(7)

Stefanie Wagner, Frédéric Lagane, Andaine Seguin-Orlando, Mikkel Schubert, [Thibault Leroy](#), Erwan Guichoux, Emilie Chancerel, Inger Bech-Hebelstrup, Vincent Bernard, Cyrille Billard [+25 other coauthors]: High-Throughput DNA sequencing of ancient wood. *Molecular Ecology* 02/2018; 27(5)

Nastasia R. Merceron, [Thibault Leroy](#), Emilie Chancerel, Jeanne Romero-Severson, Daniel S. Borkowski, Alexis Ducouso, Arnaud Monty, Annabel J. Porté, Antoine Kremer: Back to America: Tracking the origin of European introduced populations of *Quercus rubra* L. *Genome* 07/2017; 60(9)

[Thibault Leroy](#), Camille Roux, Laure Villate, Catherine Bodénès, Jonathan Romiguier, Jorge A P Paiva, Carole Dossat, Jean-Marc Aury, Christophe Plomion, Antoine Kremer: Extensive recent secondary contacts between four European white oak species. *New Phytologist* 01/2017; 214(2)

[Thibault Leroy](#), Valérie Caffier, Jean-Marc Celton, Nicolas Anger, Charles-Eric Durel, Christophe Lemaire, Bruno Le Cam: When virulence originates from nonagricultural hosts: Evolutionary and epidemiological consequences of introgressions following secondary contacts in *Venturia inaequalis*. *New Phytologist* 02/2016

Christophe Lemaire, Marie De Gracia, Thibault Leroy, Monika Michalecka, Hanne Lindhard-Pedersen, Fabien Guerin, Pierre Gladieux, Bruno Le Cam: Emergence of new virulent populations of apple scab from nonagricultural disease reservoirs. *New Phytologist* 09/2015; 209(3)

Christophe Plomion, Jean-Marc Aury, Joëlle Amselem, Tina Alacitabar, Valérie Barbe, Caroline Belser, Hélène Bergès, Catherine Bodénès, Nathalie Boudet, Christophe Boury, [30 other authors including Thibault Leroy]: Decoding the oak genome: Public release of sequence data, assembly, annotation and publication strategies. *Molecular Ecology Resources* 05/2015; 16(1)

Thibault Leroy, Bruno Le Cam, Christophe Lemaire: When virulence originates from non-agricultural hosts: New insights into plant breeding. *Infection, genetics and evolution* 01/2014; 27., DOI:10.1016/j.meegid.2013.12.022

Thibault Leroy, Christophe Lemaire, Frank Dunemann, Bruno Le Cam: The genetic structure of a *Venturia inaequalis* population in a heterogeneous host population composed of different *Malus* species. *BMC Evolutionary Biology* 03/2013; 13(1):64

Aymeric Joubert, Nelly Bataille-Simoneau, Claire Champion, Thomas Guillemette, piétrick Hudhomme, Béatrice Iacomini-Vasilescu, Thibault Leroy, Stéphanie Pochon, Pascal Poupard, Philippe Simoneau: Cell wall integrity and high osmolarity glycerol pathways are required for adaptation of *Alternaria brassicicola* to cell wall stress caused by brassicaceous indolic phytoalexins. *Cellular Microbiology* 01/2011; 13(1):62-80

Pierre Gladieux, Xiu-Guo Zhang, Isabel Róldan-Ruiz, Valérie Caffier, Thibault Leroy, Martine Devaux, Sabine Van Glabeke, Els Coart, Bruno Le Cam: Evolution of the population structure of *Venturia inaequalis*, the apple scab fungus, associated with the domestication of its host. *Molecular Ecology* 02/2010; 19(4):658-74

#### *Articles soumis*

Huiying Shang, Martha Rendon-Anaya, Ovidiu Paun, David Field, Jaqueline Hess, Claus Vogl, Jianquan Liu, Pär K. Ingvarsson\*, Christian Lexer\* & Thibault Leroy\*. Conserved genomic landscapes of differentiation across *Populus* speciation continuum, in revision for *Peer Community in Evolutionary Biology* (available on bioRxiv, 2021.08.26.457771v2)

Sylvain Schmitt, Thibault Leroy, Myriam Heuertz, Niklas Tysklind. Somatic mutation detection in plants: a critical evaluation through simulations and reanalyses (available on bioRxiv, 2021.10.11.462798)

#### *Autres communications*

Thibault Leroy & Sophie Gerber. 2021 « Arbres et forêts, entre corps et cœurs » The Conversation.

<https://theconversation.com/arbres-et-forets-entre-corps-et-coeurs-170331>

#### *Ouvrages, chapitres d'ouvrages, rapports diplômants*

Thibault Leroy & Quentin Rougemont. Book chapter: Introduction to population genomics methods. In: *Molecular Plant Taxonomy*. 2020. Springer Protocols, pp. 287-324.

Thibault Leroy. Thèse: Conséquences évolutives d'un contact secondaire chez les pathogènes: une alternative aux hypothèses écologiques? 2012. Université d'Angers, 175 pages

Thibault Leroy. Rapport de Master: Hétérogénéité inter-locus de structure génétique entre populations de *Venturia inaequalis* pathogènes de différentes espèces de *Malus*. 2009. Université d'Angers, 38 pages

## SYNTHESE DES ACTIVITES SCIENTIFIQUES

Plutôt que de faire un déroulé chronologique de mes travaux scientifiques depuis ma thèse (tel que retracé plus haut dans mon CV), j'ai choisi d'organiser la synthèse de mes travaux scientifiques de manière thématique. Bien qu'ayant travaillé sur plusieurs espèces différentes dans le cadre de mes projets de recherche et de mes collaborations, le dénominateur commun de mon parcours est le goût pour l'étude de la diversité au sens large. En particulier, je me suis intéressé aux patrons de diversité génétique et ceux associés à l'étude de la spéciation, le processus de divergence permettant *in fine* la formation de nouvelles espèces reproductivement isolées. J'ai étudié des espèces appartenant à des taxons très variés tels que des champignons pathogènes, des céréales, des arbres, des plantes épiphytes, des ornementales, des passereaux ou encore des salmonidés. La première partie de cette synthèse sera axée sur la caractérisation de la diversité génétique au sein et entre espèces, et l'évolution de cette diversité, à la fois chez des espèces sauvages et domestiquées. La seconde partie sera axée sur l'étude des processus évolutifs à la base de cette diversité génétique, les tailles efficaces et des taux de mutations.

### *I – Estimations et évolution de la diversité génétique*

Je vais ici m'intéresser à la diversité génétique associée à des polymorphismes simples de l'ADN (variation d'une base ou de quelques bases, c'est-à-dire les SNPs et éventuellement les indels), en écartant les sources de variations de nature plus complexes (variants structuraux, épimutations, etc), qui restent encore difficiles à cataloguer à l'échelle des populations chez la plupart des espèces. Une mesure classique et assez simple de la diversité génétique, pour une population donnée, est la diversité nucléotidique. Elle correspond au nombre moyen de nucléotides différents entre toutes les paires de séquences d'ADN disponibles ( $\pi$ ) (Nei & Li, 1979).

#### a) Ordre de grandeur de la variation génétique entre espèces

Les espèces sont loin d'être sur un pied d'égalité en ce qui concerne les niveaux de diversité nucléotidique, puisque plusieurs ordres de grandeur peuvent les séparer (Leffler *et al.* 2012). Notre propre espèce en est un bon exemple puisqu'elle présente une diversité génétique relativement

faible comparativement à la majorité des espèces animales, avec environ 1 polymorphisme de type SNP observé toutes les 1000 bases d'ADN en moyenne entre deux séquences prises au hasard (Fig. 1A, Li et Sadler 1991). Dans le cadre d'un travail postdoctoral avec Benoit Nabholz (ISEM, 2018-2019), j'ai entrepris de mesurer la variation de diversité génétique chez 25 espèces de passereaux continentales et insulaires sur la base de l'ensemble de leurs génomes (295 génomes au total). Cette étude a révélé qu'entre l'espèce présentant le moins de diversité génétique, *Fringilla teydae*, un pinson bleu endémique des îles Canaries et celle qui en présentait le plus, *Phylloscopus trochilus*, le Pouillot fittis, une espèce commune en Eurasie, la diversité génétique varie d'un ordre de grandeur de 1 à 17 (Leroy *et al.* 2021, voir aussi plus bas). À une échelle plus globale, notre étude a démontré une forte différence de diversité génétique entre les espèces continentales et les espèces insulaires, au détriment des insulaires. Cette forte variation de la diversité génétique est encore plus visible sur des échelles phylogénétiques plus grandes, elle est de l'ordre de 1 à 200 par exemple, entre *Lynx lynx*, le lynx eurasiatique et *Drosophila melanogaster*, la très célèbre mouche du vinaigre (Fig. 1A, Leffler *et al.* 2012).

Chez les plantes, les premières estimations à grande échelle ont révélé des ordres de grandeur de variation bien moindres, de l'ordre de 1 à 10 entre *Amborella trichopoda*, une espèce endémique des forêts humides de Nouvelle-Calédonie appartenant à la lignée la plus primitive des plantes à fleurs actuelles, et *Capsella grandiflora*, une plante sauvage de la famille des Brassicacées (Chen *et al.* 2017). Bien qu'ayant observé une plus grande homogénéité de cette diversité génétique chez les plantes, l'étude de Chen *et al.* avait néanmoins mis en évidence une diversité génétique moyenne plus élevée chez les plantes de l'étude que chez les espèces animales. Chez ces plantes, cette diversité génétique varie en fonction de quelques traits d'histoire de vie des espèces, notamment la longévité et les systèmes de reproduction, avec une plus faible diversité chez les espèces longévives et/ou pratiquant l'autofécondation (Chen *et al.* 2017). Il est intéressant de noter que dans le cadre de mon projet postdoctoral à l'INRAE de Bordeaux (2014-2018), où j'ai travaillé sur la diversité génétique du chêne pédonculé (Fig. 1B, Plomion *et al.* 2018), une espèce particulièrement longévive, les analyses que nous avons menés à grande échelle (génomes complets) ont permis de révéler le caractère singulier des chênes au regard des autres espèces décrites par Jun Chen et ses collaborateurs (2017 ; Fig. 1C). Le chêne présente une diversité et un fardeau parmi les plus élevés décrits chez les plantes, et n'ayant pas d'équivalent actuellement identifiés chez les arbres. Des analyses plus larges de la diversité génétique des chênes blancs européens se poursuivent et ont par exemple fait l'objet d'une étude récente menée avec de

nombreux collaborateurs sur l'évolution récente des chênes sessiles en lien avec l'évolution du climat depuis le petit âge glaciaire (Saleh *et al.* 2022).

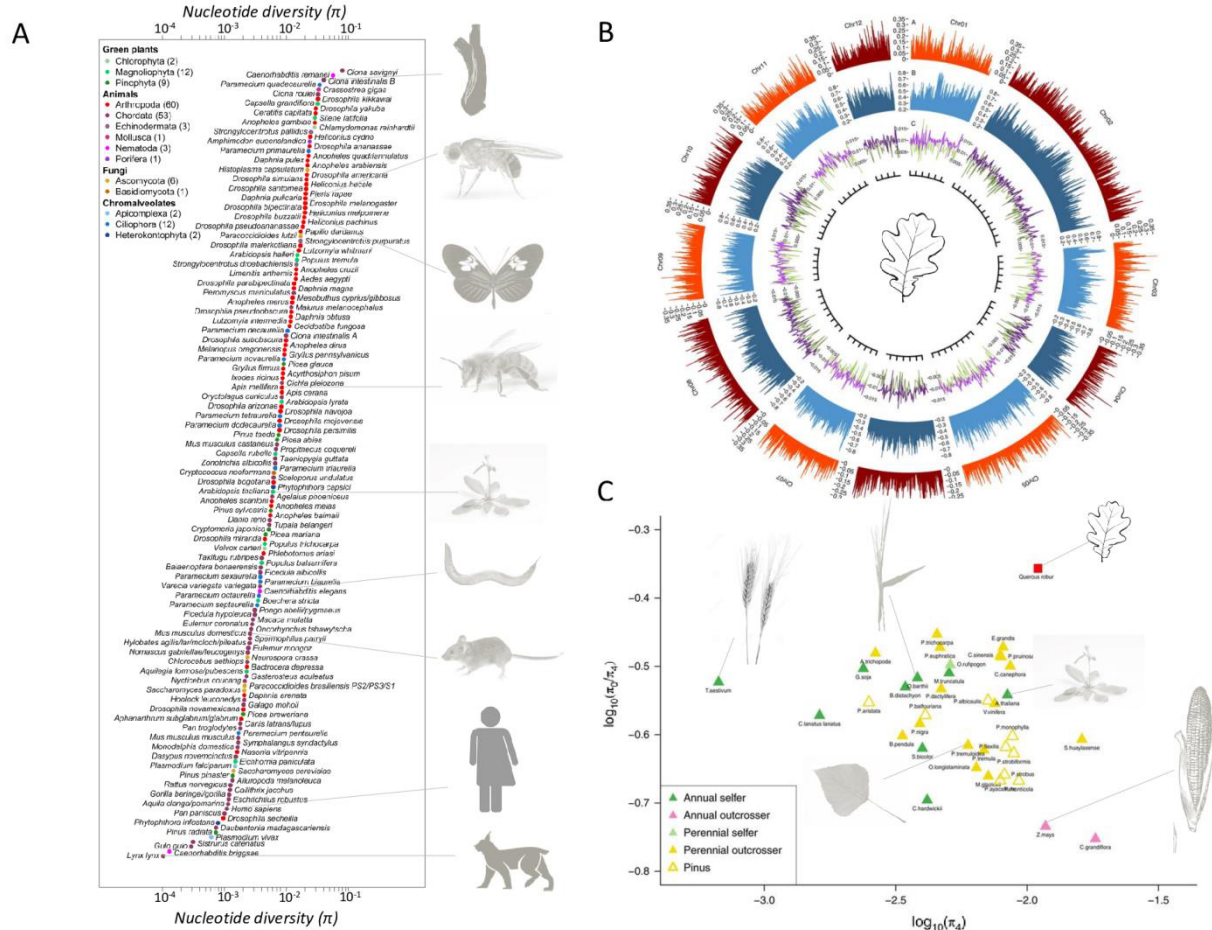


Figure 1 : Quelques repères au sujet de la diversité génétique des espèces. A) Diversité nucléotidique ( $\pi$ ) de plus de 150 espèces appartenant à différents taxons (modifié de Leffler *et al.* 2012). L'échelle est  $\log_{10}$ -transformée. Quelques silhouettes correspondant à des espèces modèles sont indiquées (de la moins diverse à la plus diverse) : Lynx lynx, Homo sapiens, Mus musculus domesticus, Caenorhabditis elegans, Arabidopsis thaliana, Apis mellifera, Heliconius melpomene, Drosophila melanogaster & Ciona intestinalis B. B) Hétérogénéité génomique du contenu en gène, éléments transposables et de la diversité génétique au sein des chênes pédonculé. Cette variation de la diversité génétique (qu'on appellera « paysage » par la suite) est suivie par deux estimateurs le pourcentage de sites hétérozygotes appelés après positionnement des lectures de l'individu 3P sur le génome de référence du chêne (3P, violet). En vert, l'estimation de la diversité génétique est effectuée sur la base des données d'un séquençage en mélange d'une population de chêne pédonculé publié dans Leroy *et al.* 2020a. C) Variation des diversités (estimée sur la base des sites 4 fois dégénérés ( $\pi_4$ ), abscisse) et fardeaux génétiques (en utilisant le proxy des diversités aux sites 0 et 4 fois dégénérés ( $\pi_0/\pi_4$ ), ordonnée) chez un grand nombre d'espèces de plantes (voir Chen *et al.* 2017 & Plomion *et al.* 2018). Certaines espèces végétales importantes, dont certaines particulièrement discutées dans ce mémoire sont indiquées par un logo, Triticum aestivum (blé tendre, épis), Oryza barthii (progéniteur sauvage du riz africain, épi), Populus tremuloides (feuille), Arabidopsis thaliana (rosette), Quercus robur (chêne pédonculé, feuille) et Zea mays (maïs, épi). Les deux échelles sont  $\log_{10}$ -transformées.



## b) Isolement géographique vis-à-vis de ses congénères

L'intensité des échanges génétiques (les flux de gènes) entre populations et espèces proches est un autre facteur déterminant de la diversité génétique. Lorsqu'elles ne sont pas complètement reproductivement isolées de leurs espèces proches, c'est-à-dire lorsque le processus de spéciation est encore en cours, certaines espèces ont la capacité de maintenir des flux de gènes avec leurs espèces apparentées. Les chênes blancs européens, dont le chêne pédonculé que je citais ci-dessus, en sont un exemple emblématique pour le règne végétal. Au cours de ce même contrat postdoctoral à l'INRAE de Bordeaux (2014-2018), j'ai caractérisé en détail l'histoire de ces chênes blancs européens en utilisant des méthodes d'inférences démographiques dont du calcul bayésien approché (ABC) (Leroy *et al.* 2017, 2020a). L'idée était de comparer les niveaux de variations, et autres statistiques de génétique des populations observées, sur un jeu de données réelles à des millions de jeux de données simulées sous différents scénarios évolutifs. Dans le cas présent, la comparaison était basée sur un jeu de données réelles de quelques milliers ou dizaines de milliers de polymorphismes SNPs. Le principe de base est que le scénario évolutif le plus probable est celui dont les statistiques résumées reproduisent le plus fréquemment les statistiques résumées observées. Ce travail a permis de révéler que, très vraisemblablement, les 4 principales espèces de chênes européens ont évolué de manière isolée les unes des autres pendant l'essentiel de leur divergence (Fig. 1A). Par contre, depuis le dernier maximum glaciaire (~20 000 ans), la trajectoire de ces espèces a été profondément modifiée et elles se sont mises à échanger massivement des gènes, et cela sur une grande partie de leurs génomes (Fig. 1A ; Leroy *et al.* 2017, 2020a). Un avantage de ces analyses est qu'elles prenaient en compte les hétérogénéités génomiques de flux génique et de tailles efficaces le long des génomes, qui sont connus biaiser les inférences démographiques. Un inconvénient majeur de ces analyses toutefois est que le scénario général (Fig. 2A) n'est construit qu'en considérant les résultats indépendants obtenus par l'inférence par paires, ce qui reste une limite majeure de l'exercice. Si l'hypothèse récente du contact secondaire massif semble bien soutenue, à la fois de par la cohérence des inférences en ABC (Leroy *et al.* 2017 ; Leroy *et al.* 2020a), en approximation de diffusion ( $\partial a \partial i$ , Leroy *et al.* 2017) et au regard des connaissances par ailleurs sur la structure génétique et les lignées de recolonisation postglaciaires (Petit *et al.* 1997 ; 2002), la séquence des événements de spéciation dans ce complexe reste très incertaine. Les progrès continus sur ces approches d'inférence (p.ex. Fraïsse *et al.* 2021) et l'extension de ces approches à plus que deux espèces ouvrira la voie à l'avenir à une réanalyse de ces données, en procédant à une inférence conjointe des 4 espèces, ce qui permettra notamment d'inférer

simultanément tous les temps de divergence et de mieux tester le scénario évolutif proposé (Fig. 2A). En tout cas, l'hypothèse générale est que ce contact secondaire massif a contribué à réhomogénéiser massivement ces génomes, qui étaient en cours de différenciation depuis des millions d'années. Ce résultat a alors ouvert deux axes de recherche.

Le premier axe est associé à l'étude de la spéciation. En effet, dans certaines zones du génome, cette réhomogénéisation n'est pas possible de par l'existence de barrières au flux génique, c'est-à-dire de combinaisons d'allèles empêchant la formation des hybrides ou défavorables, intrinsèquement ou au sein de l'environnement, quand elles sont présentes chez les hybrides. Les échanges massifs ont donc vraisemblablement contribué à augmenter le contraste entre régions réhomogénéisées et non-réhomogénéisées. Il a ainsi été possible d'identifier des zones du génome qui contribuent malgré tout à maintenir la différenciation de ces espèces, et ensuite d'identifier les gènes candidats au sein de ces régions, notamment les gènes responsables d'adaptations différentielles à leurs environnements. Il est en effet bien établi qu'aujourd'hui encore, ces espèces de chênes ont des préférences écologiques particulières, notamment par rapport à la disponibilité en eau et au pH du sol (Leroy *et al.* 2020a). Plusieurs cadres conceptuels sont possibles pour identifier les régions qui présentent un excès de différenciation comparé à l'attendu neutre. Dans le cadre de ces analyses, j'avais choisi de faire des simulations neutres sous les postérieurs des valeurs de paramètres obtenus avec les meilleurs scénarios par paires. La stratégie était donc de faire des simulations sous le meilleur scénario ABC et donc de fournir un attendu neutre pour la variance de la différenciation interspécifique, et ce, au plus près de l'histoire démographique reconstruite. En raisonnant en densité locale en marqueurs présentant un excès de différenciation (par la suite appelés « outliers », Fig. 2B), ce travail avait permis d'identifier des régions particulièrement riches en outliers et ce, sur une ou plusieurs paires d'espèces. Un travail d'annotation manuelle des gènes a ensuite été réalisé sur toutes ces régions. Ce travail a permis d'identifier des gènes candidats pour l'adaptation des chênes aux stress hydriques, à la tolérance au froid et à la résistance aux pathogènes, ainsi que de nombreux gènes associés à des processus plus internes, mais tout aussi importants du point de vue du processus de spéciation, tels que la floraison, le développement du grain de pollen, la germination ou le développement de l'embryon (Leroy *et al.* 2020a).

Le second axe est associé aux conséquences adaptatives de ces échanges, c'est-à-dire au transfert de mutations avantageuses entre les espèces, on parle alors d'introgessions adaptatives. Cette étude a été menée dans le cadre du séquençage de 18 populations de chênes sessiles, dont 10

populations de plaines sur un gradient latitudinal couvrant des populations françaises, allemandes et irlandaises, ainsi que 8 populations de deux gradients altitudinaux pyrénéens proches (vallée d'Ossau et vallée de Luz). Pour mieux renseigner l'histoire de la divergence de ces populations, j'avais généré un simple arbre de dérive avec TreeMix (Leroy *et al.* 2020b). Cela avait permis d'observer à quel point les populations latitudinales, pourtant isolées par des centaines voire milliers de kilomètres, sont génétiquement plus proches entre elles que les populations de gradients altitudinaux, bien que séparées seulement de quelques kilomètres ou dizaines de kilomètres. Cela conforte des résultats qui avaient été par ailleurs décrits sur ces populations, notamment via l'étude des  $F_{ST}$  par paires sur la base de la variation observée à une centaine de gènes ciblés (Alberto *et al.* 2013). En autorisant quelques événements d'introgession de façon à maximiser la variance totale expliquée par l'arbre reconstruit, la topologie change peu mais suggère des échanges génétiques entre les chênes pédonculés et sessiles, particulièrement les chênes sessiles présents en altitude (Fig. 2C, outre la population irlandaise 124 qui présente un fond génétique hybride sessile-pédonculé). Au-delà de cette description globale, l'approche que j'ai utilisée ensuite est basée sur l'identification de marqueurs présentant un excès de différenciation et une corrélation aux variations climatiques des sites d'étude (moyenne annuelle des températures, des précipitations, dates de débourrement en jardin commun) en utilisant BayPass, une méthode de scan génomique développée par Mathieu Gauthier (2015). La détection des marqueurs présentant un excès de différenciation et une covariation aux gradients climatiques et/ou phénotypiques a été possible après obtention d'un attendu neutre sur la base des jeux de données pseudo-observés. Remarquablement, les lots de marqueurs détectés présentant un excès de différenciation au sein des populations de chênes sessiles ou une covariation des fréquences alléliques au sein du gradient climatique ou phénotypique (niveau intraspécifique) présentaient aussi des niveaux de différenciation bien plus élevés à l'échelle interspécifique, notamment pour les  $F_{ST}$  entre chênes sessiles et pédonculés (Leroy *et al.* 2020b). Cela suggérerait donc que les adaptations détectées sont la résultante des flux géniques interspécifiques, conformément à un attendu d'introgession adaptative. Ce transfert entre espèces aurait ainsi permis le maintien des populations de chêne sessile en altitude dans les Pyrénées, par transferts d'adaptations provenant vraisemblablement du chêne pédonculé, une espèce ayant une distribution plus septentrionale et orientale que le chêne sessile en Europe, et donc vraisemblablement plus adaptée aux conditions froides (Leroy *et al.* 2020b, Kremer *et al.* 2020).

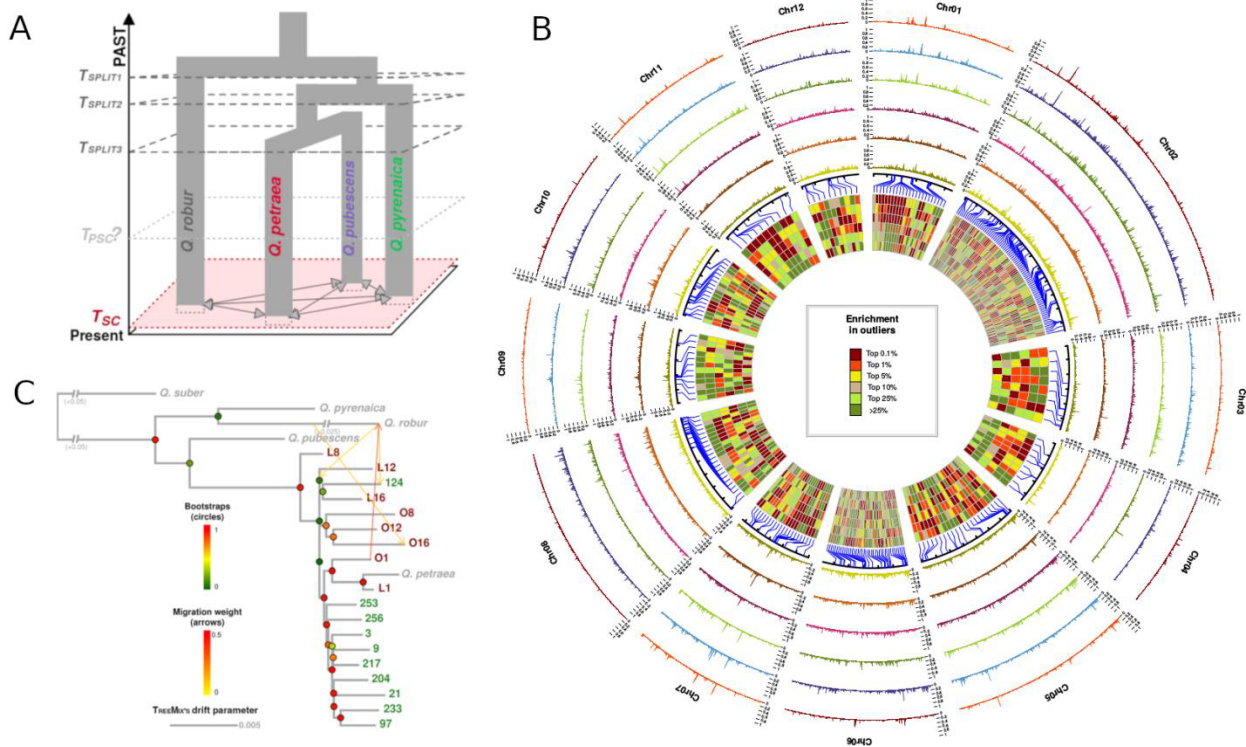


Figure 2 : Histoire évolutive résumée des chênes blancs européens. A) Représentation schématique de l'histoire évolutive probable sur la base d'un arbre des espèces et d'inférences ABC par paires d'espèces (Leroy et al. 2017 ; Leroy et al. 2020a). Chez toutes les paires d'espèces, le signal inféré est celui d'un contact secondaire massif récent ( $T_{SC}$ ), compatible avec un évènement d'échange massif de gènes postérieur au dernier épisode glaciaire. B) Densité locale en « outliers », ici définis comme marqueurs présentant un excès de différenciation comparativement aux attendus sur la base de simulations neutres sous les postérieurs des meilleurs scénarios ABC (incluant notamment les contacts secondaires récents) pour toutes les paires d'espèces, de l'extérieur vers le centre : Q. robur/Q. petraea, Q. pyrenaica/Q. petraea, Q. pubescens/Q. petraea, Q. robur/Q. pubescens, Q. robur/Q. pyrenaica and Q. pubescens/Q. pyrenaica. Au centre, chaque rectangle indique s'il y a un enrichissement local en marqueurs outliers pour chaque paire dans le même ordre (281 régions enrichies en outliers pour une paire au moins, Leroy et al. 2020a). C) Topologie d'arbre inférée sous TreeMix pour décrire l'histoire de la divergence de populations de chênes sessiles (Q. petraea) de plaines d'un gradient latitudinal (France, Allemagne et Irlande, numéros indiqués en vert) ou de populations de gradients altitudinaux dans les Pyrénées (X1 à X16 indiqués en rouge, correspondant de 100 à 1600 mètres d'altitude). On notera que les phases d'introgession inférées sont entre le chêne pédonculé et les chênes sessiles, dont la population irlandaise (124) et les populations d'altitudes (modifié de Leroy et al. 2020b).

Sur un tout autre exemple, le cas d'un pathogène majeur du pommier, *Venturia inaequalis*, j'avais dans le cadre de mon travail de thèse mis en évidence les conséquences évolutives que peuvent avoir cette reprise du flux de gènes. Ce pathogène a des capacités de migration intrinsèques très limitées (Leroy *et al.* 2013). Sa dispersion à plus grande distance, et même plus globalement son succès invasif mondial, est donc essentiellement liée à sa dissémination par l'homme (Gladieux *et al.* 2008). Cette caractéristique isole de nombreuses populations du pathogène dans la nature, sur des espèces de *Malus* différentes du pommier domestique. Au cours du programme de sélection du pommier, les sélectionneurs ont utilisé des résistances génétiques provenant d'espèces sauvages

contre cette maladie, comme cela est souvent l'usage en sélection, dans le cas présent une autre espèce de *Malus*. Cela a fortuitement conduit à mettre en contact deux populations de ce pathogène initialement isolées géographiquement et ainsi permis des échanges de diversité génétique de ces populations, incluant des adaptations différentes. Ces échanges ont eu pour conséquence une augmentation du pouvoir pathogène de cette espèce, puisque les hybrides entre ces deux populations étaient plus agressifs en serre que les souches représentatives des deux populations initiales sur des cultivars de pommier commerciaux tels que Gala (Leroy *et al.* 2016). L'hypothèse était que ces hybrides ont combiné les facteurs d'agressivité provenant des deux populations préalablement isolées. Si les chênes constituaient un exemple où le transfert d'adaptations s'avérait bénéfique pour la biodiversité, ce second exemple montrait une conséquence défavorable de ces échanges dans un contexte épidémiologique.

### c) Évolution de la diversité génétique chez les espèces domestiquées

Si jusqu'ici je m'intéressais à l'évolution de populations naturelles, l'étude de la diversité génétique est tout aussi importante chez les espèces domestiquées. D'abord, la domestication de ces espèces est présumée avoir conduit à une perte de diversité génétique, car les premiers agriculteurs ont sélectionné des caractères particuliers d'intérêt contribuant à une perte de certains caractères phénotypiques non-désirés, et de ce fait, d'une diversité génétique. C'est ce qu'on appelle le goulot d'étranglement associé à la domestication (*domestication bottleneck*). A titre d'exemple, dans le cadre d'un chapitre de livre d'introduction à la génomique des populations, co-écrit avec Quentin Rougemont, un de mes collaborateurs actuels principaux, je me suis lancé dans la réanalyse d'une cinquantaine de génomes de riz africains obtenus dans le cadre d'une étude plus large (Cubry *et al.* 2018), en comparant des population du progéniteur sauvage *Oryza barthii* au niveau du centre de domestication et de l'espèce domestiquée *O. glaberrima* (Leroy & Rougemont, 2021). Cette analyse suggère qu'environ un tiers de la diversité génétique totale initiale a été perdue au cours de la domestication du riz africain. Il est à noter qu'il faut rester prudent sur ces estimations, tant la tâche reste ardue, sur la seule base de génomes contemporains (en l'absence de données de séquençage d'ADN provenant d'échantillons fossiles).

Dans un second temps, les programmes de sélection conduisent à créer une deuxième vague d'érosion de diversité. Dans le cadre d'une large étude sur le blé tendre, j'ai non seulement analysé

des génomes de blé sauvage et de blé tendre afin d'estimer la part de diversité génétique qui a été perdue au cours de la domestication (voir aussi Haudry *et al.* 2007), mais surtout au cours du programme de sélection au XXe siècle. Chez cette espèce à l'architecture génomique particulièrement complexe (allohexaploïdie, donc 3 sous-génomes, 17 Gb), la diversité génétique du blé tendre a diminué de 33 % et 28 % pour les sous-génomes A et B, et de 6% pour le sous-génome D en un siècle tout au plus. Seul le sous-génome D présente donc une réduction de diversité plus modérée au cours du XXe siècle, mais il est à noter que ce sous-génome présente une diversité déjà extrêmement réduite au regard des deux autres (Fig. 3A). Cette érosion massive et rapide contraste avec les ambitions actuelles « post-2020 » de la convention de la biodiversité de conserver 90% de la diversité génétique des espèces sauvages et domestiquées. Evidemment ces repères sont complexes et sources d'interrogation : quels référentiels considérer dans ce cas, l'ère pré-industrielle (les « landraces », c'est-à-dire les variétés locales) ou la situation actuelle déjà détériorée ? Quels estimateurs faire primer, la diversité génétique ou les tailles efficaces (Hoban *et al.* 2020) ? Une telle érosion doit en tout cas alerter. Face aux enjeux climatiques et sanitaires actuels et à venir, la réintroduction de la diversité génétique dans les programmes de sélection apparaît aujourd'hui fondamentale pour la sécurité alimentaire mondiale.

En analysant la variation par chromosome, il est possible d'identifier des larges régions où la diversité actuelle a été réduite de plus de 80% par rapport aux variétés anciennes locales (Fig. 3C, correspondant quand même à 9% du génome total). Certaines de ces régions pouvant être associées à des gènes importants de domestication, d'autres à des zones qui restent encore à caractériser à ce stade. Dans les prochains mois, je vais continuer à travailler sur la détection des patrons de balayages sélectifs chez différentes espèces de céréales, notamment le blé dur et le blé tendre pour essayer de mieux caractériser les empreintes génomique de la sélection. Mon objectif sera en particulier d'étudier les cibles communes de la sélection artificielle chez différentes espèces domestiquées, via l'étude des balayages sélectifs communs aux gènes orthologues (cf Chen *et al.* 2022 pour une illustration récente).

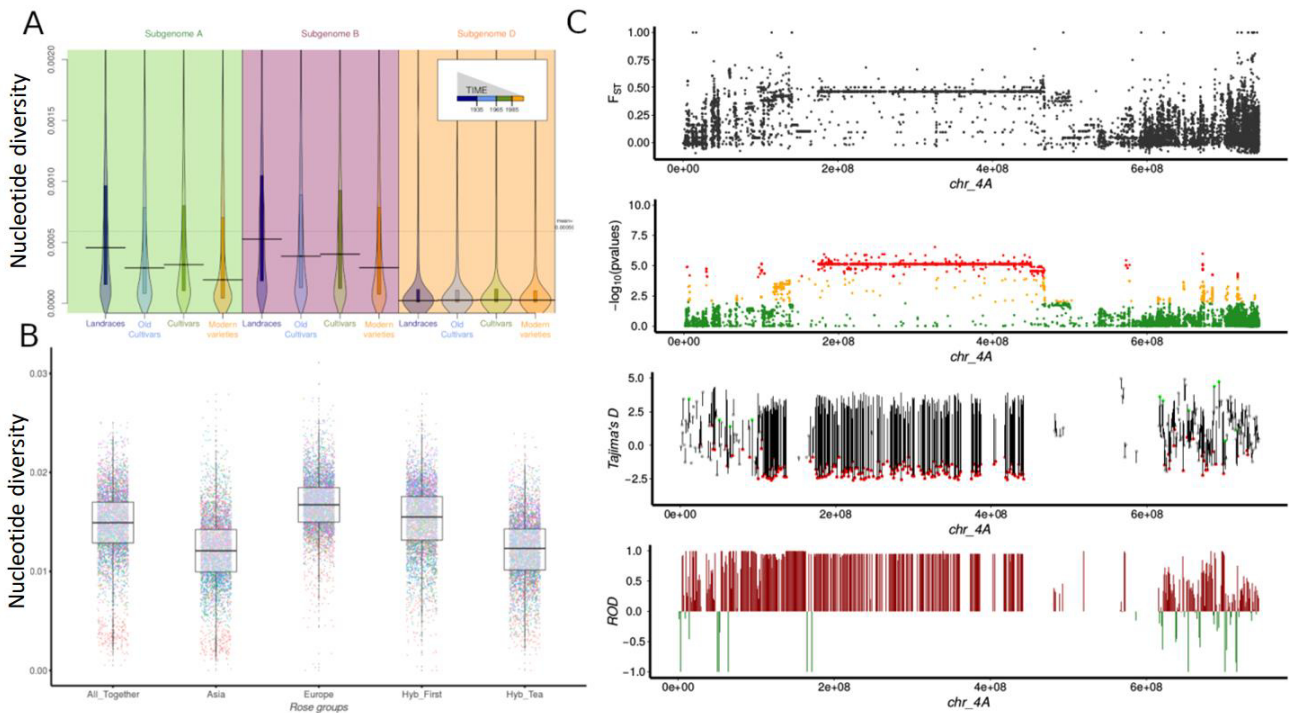


Figure 3 : Evolution de la diversité nucléotidique au cours des programmes de sélection du blé (A, C, figures modifiées de Pont et al. 2019) et du rosier de jardin (B, Leroy et al. en préparation). A) Distribution de la diversité observée en fonction des sous-génomes de blé tendre par période de sélection (bleu foncé : variétés « landraces » (avant 1935), blé clair : cultivars anciens (1935-1965), vert : cultivars (1965-1985) et jaune : cultivars modernes (postérieurs à 1985). Ces années correspondent aux dates d'enregistrement dans les catalogues de variétés. B) Variation de la diversité nucléotidique des rosiers de jardins entre les cultivars asiatiques, européens, premiers hybrides et hybrides de Thé. La catégorie « All\_Together » regroupe l'ensemble des cultivars. Chaque point correspond à la diversité sur une fenêtre de 100 kb, avec une couleur différente par chromosome. C) Signature de sélection sur le chromosome 4 du sous-génome A du blé.  $F_{st}$  entre landraces et variétés modernes,  $-\log_{10}$  des p-values d'un scan génomique sous PCAdapt, évolution des D de Tajima entre les groupes (le segment est proportionnel à la variation entre landraces et variétés modernes). Le point rouge indique le D estimé chez les variétés modernes (point en rouge si à la baisse, en vert si à la hausse par rapport aux landraces). Réduction de diversité (ROD) calculée par  $1 - (\pi_{modernes} / \pi_{landraces})$ , les valeurs en rouge indiquent une perte nette de diversité génétique ( $I =$  toute la diversité perdue). On notera que l'on suit ici la variation le long d'un chromosome d'une longueur  $\sim 750$  Mb (l'équivalent du génome entier du chêne ou encore de 5 fois celui d'*Arabidopsis thaliana*)!

Dans le cadre de mon projet post-doctoral actuel (IRHS Angers, 2021-2022), je m'intéresse à la caractérisation de l'évolution de la diversité génétique des rosiers de jardin. Les rosiers font partie des exemples d'espèces ayant eu la domestication la plus tardive avec une intense période de sélection au cours du XIX<sup>e</sup> siècle, un siècle crucial pour la diversification de certains caractères chez le rosier, notamment via l'utilisation des rosiers en provenance d'Asie. C'est au cours de ce siècle qu'ont été, par exemple, introduits les allèles de remontée de floraison dans des fonds génétiques européens, permettant d'allonger considérablement la période de floraison des rosiers de nos jardins (Iwata *et al.* 2012 ; Soufflet-Freslon *et al.* 2021). L'analyse de la structure génétique sur la base de séquençages complets de quelques dizaines de variétés de rosiers permet de retracer

assez fidèlement cette histoire. En combinant les analyses de la structure génétique et de l'apparentement génomique, j'ai pu fournir des informations assez précises. Les pools génétiques asiatiques et européens ont été croisés entre eux permettant l'obtention des premiers hybrides asiatiques-européens, qui ont ensuite été rétrocroisés avec le fond génétique asiatique pour l'obtention des hybrides de Thé. Comme le fond génétique asiatique était initialement moins diversifié que le fond génétique européen ( $\pi$  moyens :  $1,20 \times 10^{-2}$   $1,67 \times 10^{-2}$ , respectivement, Fig. 3B), l'histoire de la sélection est donc celle d'une érosion de la diversité au fur et à mesure de l'introduction du fond génétique asiatique ( $\pi$  moyens :  $1,52 \times 10^{-2}$  chez les premiers hybrides,  $1,21 \times 10^{-2}$  chez les hybrides de Thé). On pourrait donc alors considérer que cette évolution de la diversité résulte uniquement des rétrocroisements avec le fond génétique le moins polymorphe (le fond asiatique). Mais il est à noter qu'en utilisant les marqueurs « diagnostiques » sur mes individus séquencés (allèles avec  $f=0$  dans un pool génétique et  $f=1$  dans l'autre pool), j'ai calculé un indice hybride moyen pour les premiers hybrides et pour les hybrides de Thé qui m'a permis d'obtenir des estimations précises de la part des génomes des deux fonds génétiques. Bien que cette sélection se soit réalisée vraisemblablement sur plusieurs générations, les premiers hybrides présentent un fond génétique remarquablement proche du 50/50 (50,9% asiatique), alors que les hybrides de Thé présentent un fond génétique aux trois quarts asiatique (74,8%). Considérant les écarts de diversité et la part du fond génétique européen restant dans les hybrides de Thé (25,2% donc), la diversité moyenne attendue se situait autour de  $1,32 \times 10^{-2}$ , là où la diversité moyenne observée n'est que de  $1,21 \times 10^{-2}$ . Ce résultat suggère donc un effet particulièrement net de la sélection artificielle chez les hybrides de Thé. Les rosiers de jardin étant une ressource biologique que l'on peut maintenir par reproduction végétative, ces travaux montrent l'importance de veiller au maintien des collections de variétés anciennes. Au-delà des effets globaux, mes travaux visent à présent à identifier des zones génomiques présentant une signature claire de sélection, en combinant des approches populationnelles et de génétique d'association. L'idée est de détecter des signatures compatibles avec des balayages sélectifs et de renseigner des caractères potentiellement sélectionnés, et éventuellement ensuite les gènes contrôlant ces caractères (Fig. 4 pour un exemple).



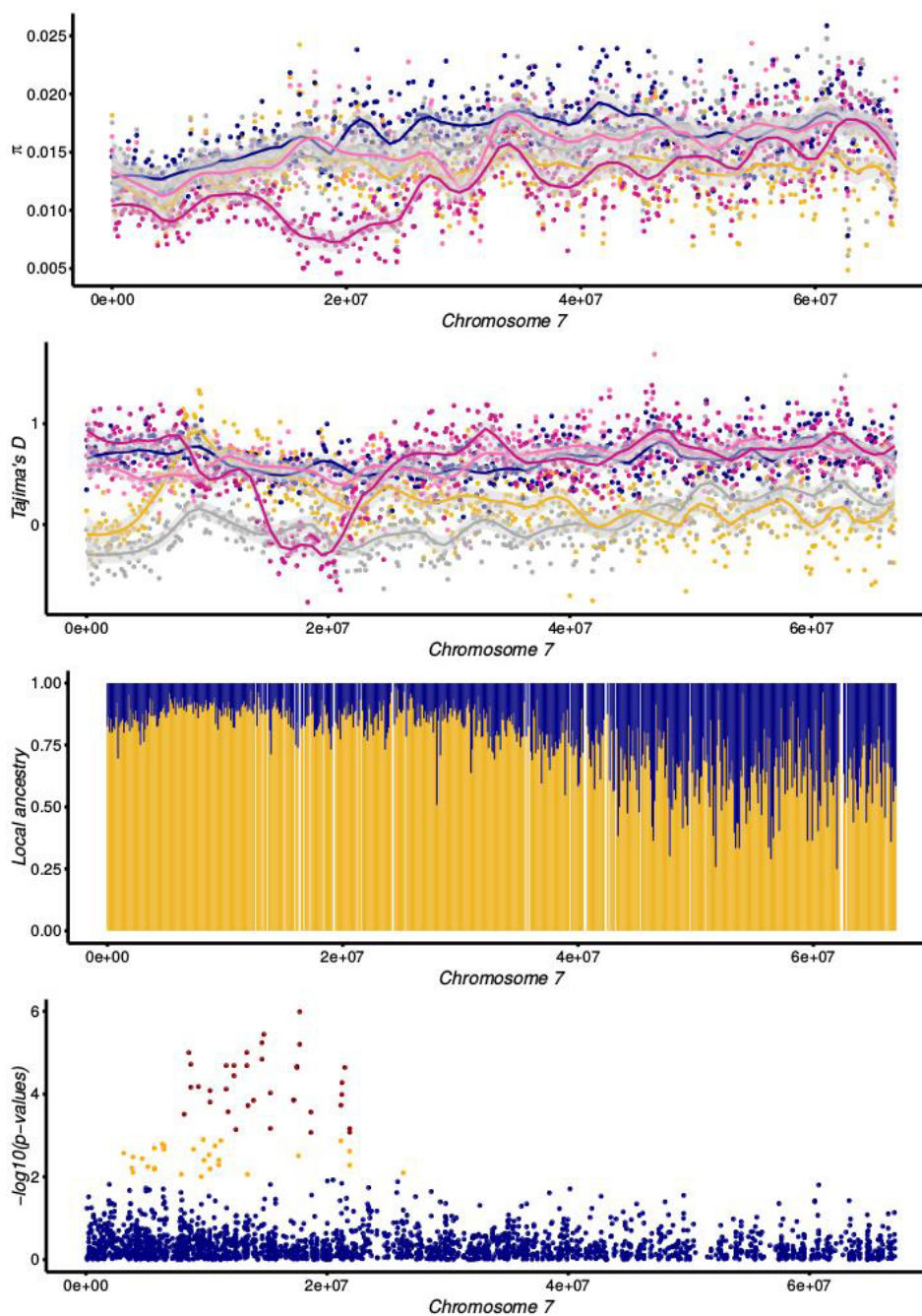


Figure 4 : Signature de sélection sur le chromosome 7 du rosier potentiellement en lien avec la sélection pour le parfum. Du haut vers le bas : a) Variation de la diversité nucléotidique estimée sur des fenêtres de 100 kb (points) pour les accessions européennes (bleu), asiatiques (jaune), premiers hybrides (rose clair) and hybrides de Thé (rose foncé). Les régressions locales (loess) par groupe sont indiquées. b) Variation du D de Tajima le long du chromosome 7, c) variation des valeurs d'indices hybrides estimées par la moyenne des fréquences alléliques aux allèles diagnostics. d) analyse d'association pour la production de géraniol sur le chromosome 7, en considérant le caractère qualitatif (génotypes produisant pas ou peu vs. beaucoup de géraniol). Les marqueurs significativement associés à une q-valeur de 1% et de 0.1% sont indiqués en orange et rouge, respectivement. On notera ici la difficulté associée au modèle rosier, avec un très large déséquilibre de liaison en raison du nombre très limité de générations de recombinaison. Le ou les gènes candidats restent à identifier.

#### d) Paysages génomiques de la diversité et de la divergence et évolution au cours de la spéciation

Outre l'étude de la variation des diversités génétiques locales ou de différenciation dans le but de détecter des empreintes de balayages sélectifs ou des barrières aux flux de gènes, des réflexions plus globales peuvent être menées sur les paysages de la diversité génétique et de la divergence et leurs évolutions au cours du processus de spéciation. L'essor des méthodes de séquençage à haut débit a en effet permis un accès à la description des paysages génomiques chez de nombreuses espèces au cours de la dernière décennie. Ces patrons se sont révélés particulièrement hétérogènes (p.ex. Ellegren, *et al.* 2012; Martin, *et al.* 2013; Renaut *et al.* 2013 Lamichhaney, *et al.* 2015). De nombreux processus sont supposés contribuer à cette hétérogénéité, la variation d'intensité du flux génique en raison des barrières d'isolement reproductif, la variation génomique des tailles efficaces à cause de la sélection liée (autostop génétique (Smith and Haigh 1974) et sélection d'arrière-plan (Charlesworth, *et al.* 1993)), la variation des taux de recombinaison, des taux de mutations, etc. Ces paysages sont depuis longtemps reconnus comme complexes à interpréter (Cruickshank & Hahn 2014). Certains attendus (prudents) peuvent toutefois être dérivés sur la base des corrélations génomiques attendues entre les statistiques résumées sous différents scénarios de divergence (Fig. 5, Han *et al.* 2017 ; Irwin *et al.* 2018 ; Shang *et al.* en révision).

Dans le cadre de la thèse de Huiying Shang, nous avons essayé de caractériser ces paysages et d'étudier leurs évolutions au cours d'un continuum de divergence, en utilisant le genre *Populus* comme modèle (Fig. 6). Les analyses ayant été réalisées à partir de séquençages complets de 7 espèces du genre (1,3 à 4,8 millions d'années de divergence estimées sur la base d'une précédente étude, Shang *et al.* 2020).

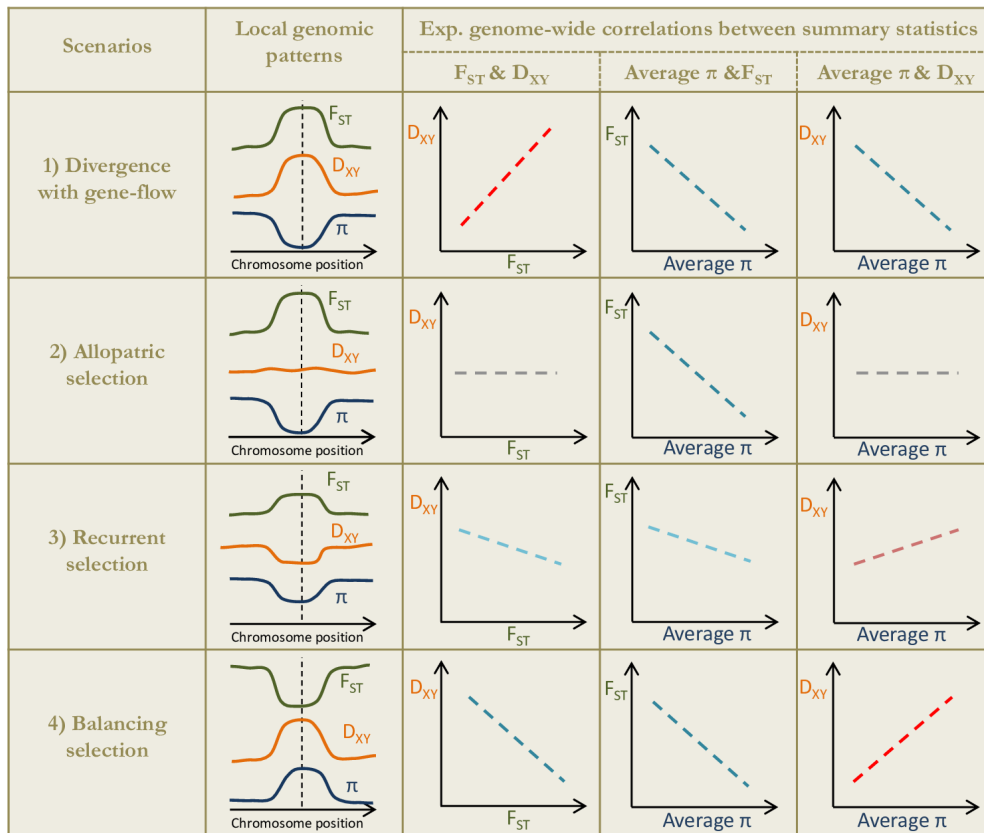


Figure 5 : Corrélations attendues entre paires de statistiques résumées décrivant les paysages génomiques de la diversité génétique et de la différenciation associé aux 4 scénarios proposé par Han et al. (2017) et Irwin et al. (2018). Les régressions positives et négatives sont montrées en rouge et en bleu en fonction du sens de la corrélation attendue. Les patrons locaux représentent 1) une barrière d'isolement reproductif, 2) un balayage sélectif dans une des deux populations, 3) une signature de sélection ancestrale qui reste en cours au sein des espèces filles et 4) une empreinte associée à de la sélection équilibrante. Le  $\pi$  moyen correspond à la moyenne du  $\pi$  entre les deux espèces impliquées dans la paire. Pour plus de détails, voir Shang et al. (en révision).

Huiying a d'abord procédé à une annotation de l'ensemble des régions présentant des signatures compatibles avec ces modèles locaux sur la base des distributions observées des différentes métriques (en utilisant des quantiles (ici les 5% de valeurs les plus élevées) comme seuils pour  $F_{ST}$ ,  $D_{XY}$  et  $\pi$ ) estimés sur des fenêtres coulissantes non-chevauchantes de 10kb (Fig. 6 B & C). Le modèle qui a de loin été le plus conforté est le modèle de sélection allopatric (74,3 à 78,7% de l'ensemble des fenêtres identifiées selon les paires). Ce modèle est compatible avec les signatures d'empreintes récentes de sélection positive ou de la sélection d'arrière-plan. Une détection de balayage a été effectuée avec SweepFinder2 pour essayer d'identifier le nombre et l'étendue des régions de balayages sélectifs et cette détection ne semble pas compatible avec les effectifs de fenêtres détectées globalement, suggérant que c'est surtout la sélection d'arrière-plan qui a un rôle prédominant dans l'établissement de ces paysages génomiques. Les autres modèles les

plus fréquents arrivant ensuite sont le scénario de « sélection équilibrante » (11.6% à 13.9% suivant les paires), puis de divergence avec flux de gènes (5.5% à 8.1%, « seulement »).

Plus globalement, on peut étudier à quel point ces patrons sont corrélés entre espèces ou paires d'espèces et comment évoluent ces corrélations le long du gradient de spéciation. Dans ce complexe d'espèce, les paysages restent relativement stables à travers toutes ces espèces avec des corrélations significatives de même signe quel que soit le stade de la spéciation des paires étudiées. Cela suggère une bonne conservation de la synténie et du paysage de la recombinaison, conduisant ainsi à un maintien relatif des corrélations même en contrastant des espèces au début et en fin de processus de spéciation. Dans le détail, les paysages de  $F_{ST}$  génomiques observées chez ces espèces deviennent de plus en plus corrélés entre eux au fur et à mesure que la divergence moyenne augmente (Fig. 6D). Aucun patron aussi net ne semble toutefois se dégager pour le  $D_{XY}$ . Comme le  $F_{ST}$  est une mesure relative de la divergence, elle est affectée par les niveaux de diversité locaux, contrairement au  $D_{XY}$  (Cruickshank & Hahn 2014 ; Wolf & Ellegren 2016). L'hypothèse serait donc que les effets de la sélection d'arrière-plan contribuent à amplifier l'hétérogénéité du paysage de  $F_{ST}$  au fur et à mesure que la divergence augmente, renforçant donc le niveau des corrélations. La comparaison des paysages moyens de la diversité entre paires d'espèces (sur la base des  $\pi$  moyens des deux espèces composant la paire) semble soutenir l'hypothèse d'une réduction des niveaux de corrélations au fur et à mesure de la divergence. Pour étudier plus finement l'évolution de ces corrélations, une possibilité est d'utiliser la moyenne génomique du  $d_A$  comme proxy du temps de divergence et de reporter les valeurs de corrélations pour toutes les paires possibles (Fig. 6E). Cela permet par exemple d'observer une évolution des niveaux de corrélations entre les paysages du  $D_{XY}$  et du  $\pi$  moyen qui deviennent de moins en moins corrélés à mesure que le temps de divergence augmente, un patron qui s'expliquerait alors par la réduction des polymorphismes partagés dans les niveaux de diversité, puisque le  $D_{XY}$  est sensible à la diversité ancestrale. Des patrons assez similaires ont aussi été retrouvés empiriquement par l'étude de plusieurs espèces de la famille *Mimulus* par exemple (Stankowski *et al.* 2019). Bien que délicate à étudier et à interpréter, l'étude fine de l'évolution de ces paysages génomiques à travers l'étude de complexes d'espèces ouvrira la voie à une meilleure compréhension des processus gouvernant leurs évolutions.

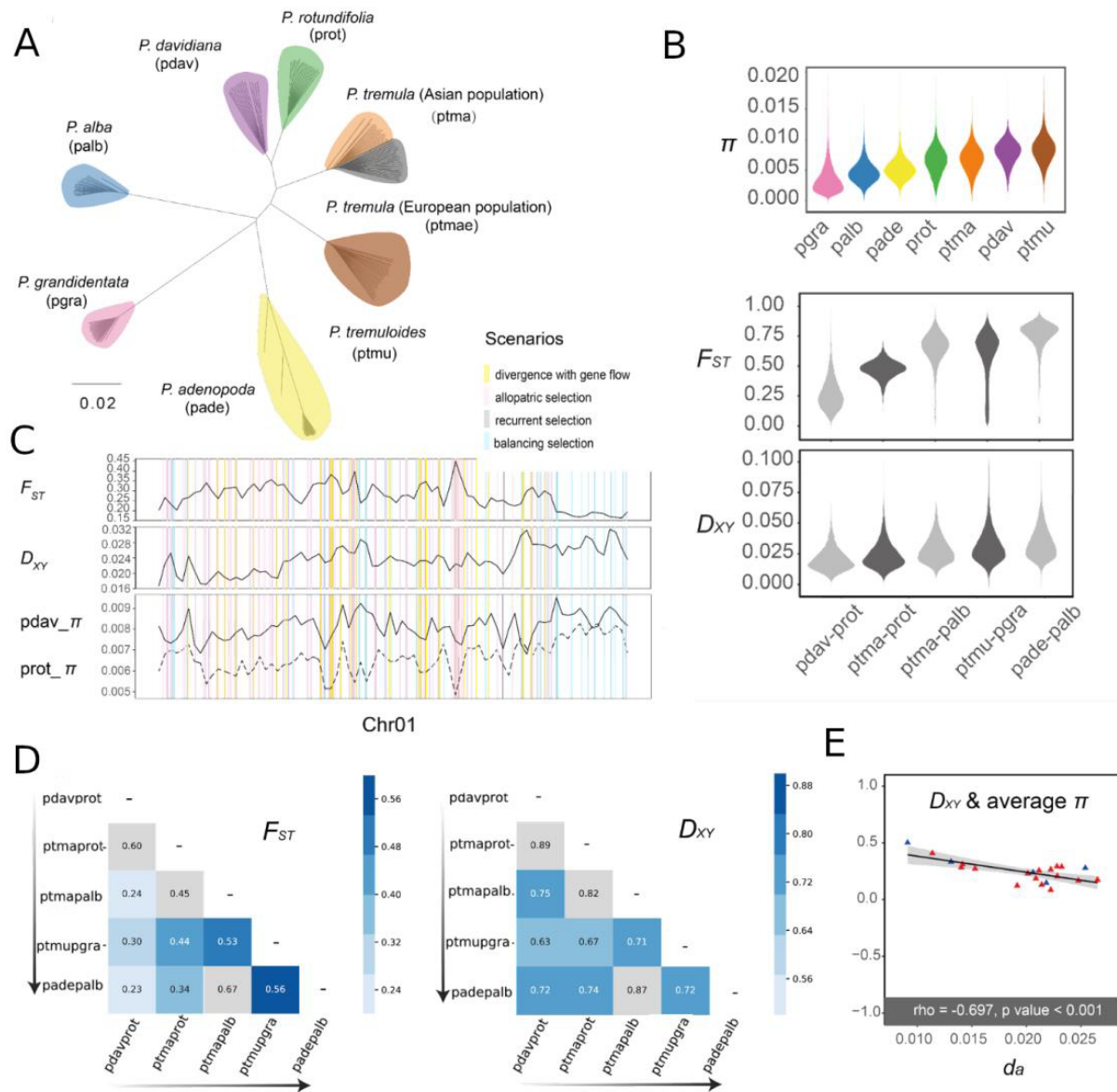


Figure 6 : Evolution des paysages génomiques au sein de 7 espèces du complexe d'espèce des *Populus* section *Populus*. A) Arbre phylogénétique simple reconstruit par la méthode du Neighbor Joining. Pour plus d'information sur la structure génétique au sein du complexe, voir Shang et al. (en révision). B) Variance observée du  $\pi$  au sein des espèces et du  $F_{ST}$  &  $D_{XY}$  pour 5 paires choisies pour être représentatives des différents stades le long du continuum de divergence. C) Exemple de paysages observés pour la paire proche *P. davidiana* and *P. rotundifolia*. A noter que pour la lisibilité de la figure, la variation des métriques n'est pas la même échelle que celles utilisées pour la détection des fenêtres se conformant aux scénarios évolutifs présentés en Fig. 5. D) Variation des niveaux de corrélations des paysages de  $F_{ST}$  et de  $D_{XY}$  le long du gradient de divergence (flèche). E) Variation de la corrélation entre le  $D_{xy}$  et la moyenne du  $\pi$  pour les deux espèces impliquées dans chaque paire en fonction du  $d_a$  estimé sur la moyenne des fenêtres et utilisé comme proxy du temps de divergence. Les 5 paires précédemment discutées sont indiquées par des points bleus, les 16 autres par des points rouges.

## ***II – Déterminants de la diversité génétique***

En l'absence de sélection et à l'équilibre, la diversité génétique est supposée dépendre uniquement de trois facteurs, le niveau de ploïdie ( $c$ ) pour l'espèce donnée, les tailles efficaces ( $N_e$ ) et le taux de mutation ( $\mu$ ) via la relation  $2cN_e\mu$  ( $4N_e\mu$  pour une espèce diploïde, Nei & Tajima, 1981). Ainsi ramené en équation, on pourrait alors se dire que les mystères de la diversité génétique sont derrière nous, pourtant nous n'avons encore que peu de connaissances sur l'évolution de ces deux déterminants : les tailles efficaces et les taux de mutations.

### **a) Tailles efficaces**

Malgré le fait que j'ai déjà fait ponctuellement appel à la notion de taille efficace dans les précédentes sections, il semble ici essentiel de bien définir ce que recouvre cette notion. Bien que ce paramètre soit essentiel en génétique des populations et en biologie de la conservation, sa définition n'est pas si triviale. Il s'agit de la taille (en nombre d'individus) d'une population idéale qui aurait le même taux de perte de diversité génétique (c'est-à-dire la même dérive génétique) que la population réelle. Les tailles efficaces sont donc une métrique qui permet de capturer l'intensité de la dérive génétique. Plus les tailles efficaces sont grandes, moins la dérive est importante. A noter que pour la grande majorité des espèces, ces tailles efficaces sont beaucoup plus faibles que le nombre réel d'individus dans les populations (Palstra & Fraser 2012). L'émergence des technologies de séquençage à coût réduit a permis un essor des méthodes permettant l'estimation des tailles efficaces et l'inférence de son évolution dans le passé relativement proche (un ou quelques millions d'années tout au plus), mais ces approches restent relativement récentes et simplistes. A mes yeux, un des premiers grands succès a été l'estimation, sur la base de l'ADN mitochondrial, de l'évolution des tailles efficaces humaines depuis la sortie de l'homme du continent africain (Atkinson *et al.* 2008). Des développements méthodologiques considérables ont été entrepris depuis 10 ans, permettant d'inférer l'évolution des tailles efficaces et *in fine* de mieux comprendre les diversités génétiques observées actuellement, comme par exemple la très faible diversité du lynx eurasiatique, pour reprendre un exemple préalablement discuté (Lucena-Perez *et al.* 2020). Il n'en demeure pas moins que malgré ces progrès, les estimations directes de  $N_e$  restent

déliçates et doivent être manipulées avec précaution. En pratique, de nombreux estimateurs indirects de  $N_e$  restent utilisés comme alternatives (voir ci-après).

Au cours de mon projet postdoctoral à Montpellier (ISEM, 2018-2019), j'ai étudié un nombre importants d'espèces de passereaux (25 espèces ou sous-espèces) sur la base d'un nombre total de 295 génomes complets. Au début de ce mémoire, j'ai décrit la variation des niveaux de diversité génétique observée au sein des espèces de ce jeu de données, en particulier les niveaux plus élevés observés chez les espèces continentales comparativement aux espèces endémiques d'îles isolées (Fig. 7A). Au-delà des diversités nucléotidiques, ici utilisées comme proxy du  $N_e$ , le même patron a pu être retrouvé en considérant les tailles efficaces inférées par PSMC ( $N_e$  moyen sur le dernier million d'année, Fig. 7C). Il s'agissait là en premier lieu de vérifier l'hypothèse cadre de l'ANR BirdIslandGenomic de Benoit Nabholz selon laquelle les espèces insulaires sont des bons modèles pour étudier les conséquences d'une évolution en petite taille de populations. Notons ici que la contrainte sur le  $N_e$  est associée à l'aire de distribution et non à un effet de l'insularité *per se* (Leroy *et al.* 2021 ; Leroy & Nabholz 2022). L'idée est que la petite surface disponible sur une île induit une limite haute possible des tailles efficaces. Les espèces endémiques insulaires sont donc des systèmes simples permettant d'étudier les conséquences d'une vie en petite taille de population. Il est toutefois à noter que des espèces continentales qui présenteraient de manière stable (sur des dizaines ou des centaines de milliers de générations) des petites tailles de populations seraient tout aussi intéressantes à étudier dans ce contexte d'une évolution en petite taille de populations. Mais comparativement aux espèces insulaires, l'hypothèse que ces espèces aient pu avoir des grands  $N_e$  dans le passé est plus difficile à exclure. Il est important de préciser cet élément car dans le détail, la relation que nous avons identifiée est continue en fonction des surfaces des aires de répartitions. Nous avons ainsi pu reporter sur ce jeu de données des corrélations significatives entre des estimateurs plus ou moins directs de  $N_e$  et les tailles réelles des populations (Fig. 7E), ou encore les tailles efficaces et les aires de distributions des espèces (Fig. 7F ; Leroy *et al.* 2021). Si de telles corrélations peuvent sembler intuitives, rares sont les jeux de données à avoir pu l'observer empiriquement.

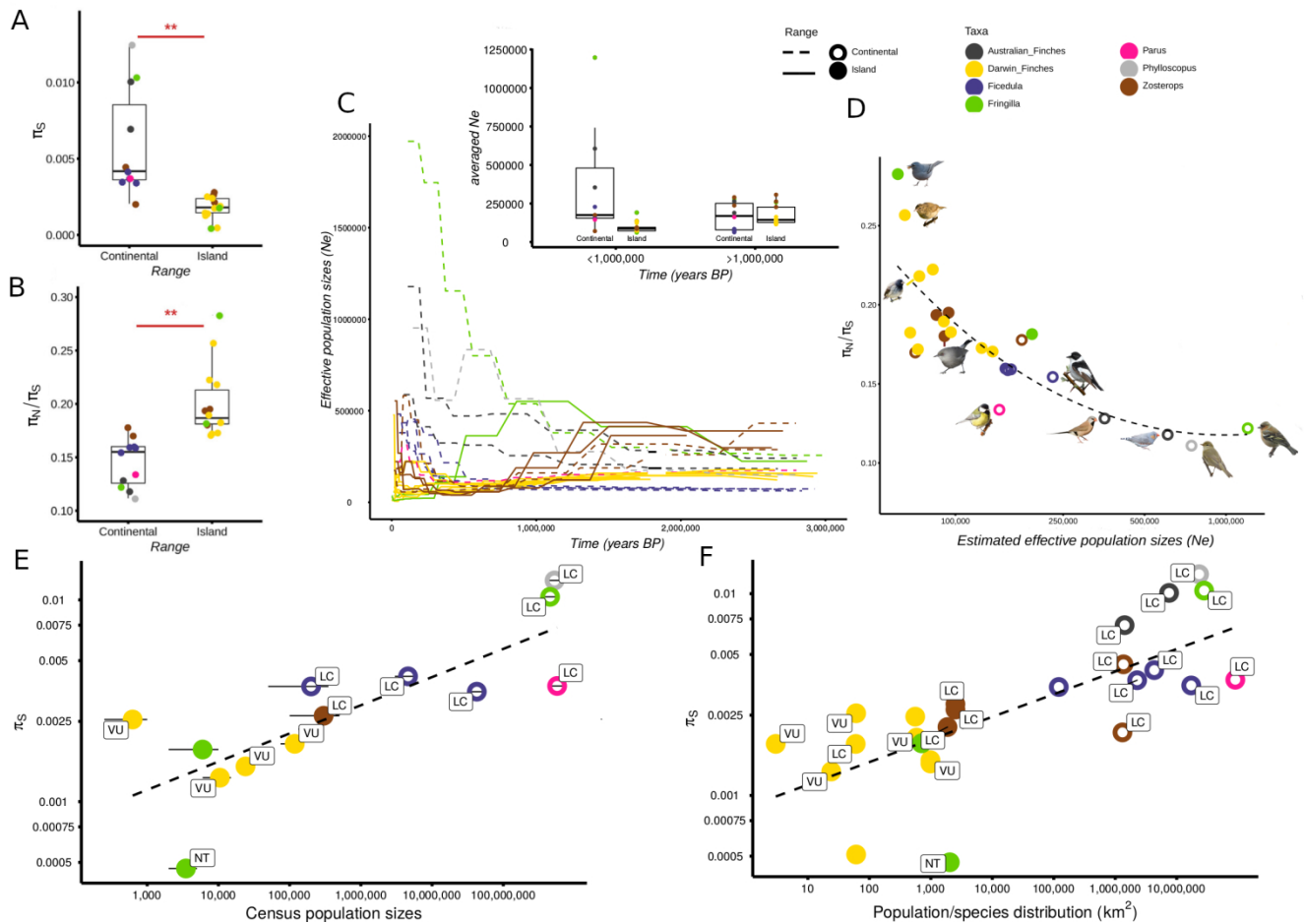


Figure 7 : Les espèces insulaires comme modèles pour l'évolution en petite taille de population. A & B) Variation des niveaux de diversité nucléotidiques (A) et des  $\pi_N/\pi_S$  (B) entre espèces endémiques insulaires et continentales. C) Variation des  $N_e$  inférés par PSMC pour chaque espèce, sur la base d'un individu choisi aléatoirement pour chaque espèce. La variation des valeurs moyenne lors du dernier million d'année (après exclusion des 4 pas de temps les plus récents pour éliminer le bruit de fond associé aux inférences du  $N_e$  récent) a ensuite été calculée et est comparée entre insulaires et continentales. D) Régression locale entre le  $\pi_N/\pi_S$  et la moyenne des  $N_e$  précédemment indiquée. E & F) Régression linéaire entre la diversité nucléotidique et les effectifs réels d'individus des populations (E) et pour les surfaces des aires de répartition estimées des espèces (F). Lorsque connus, les statuts de conservation de l'union internationale pour la conservation de la nature (IUCN) sont indiqués (LC : préoccupation mineure, NT : quasi-menacée, VU : vulnérable).

En plus de fournir tous ces enseignements sur les tailles efficaces, ces analyses ont permis d'étudier les conséquences de la variation de  $N_e$ , en particulier au regard de l'adaptabilité et de l'efficacité de la sélection à purger les mutations défavorables de leurs génomes. Nous avons ainsi pu observer que les ratios de diversités nucléotidiques non-synonymes et synonymes ( $\pi_N/\pi_S$ ), un proxy de la teneur en mutations faiblement délétères dans les génomes, est de 40% plus élevé en moyenne sur le groupe des insulaires que chez les continentales (Fig. 7B). Là encore cette variation est en réalité continue en fonction du  $N_e$  et ne représente pas un effet d'insularité *per se* (Fig. 7D).



Au-delà de l'accumulation de mutations faiblement délétères, une autre conséquence attendue de l'évolution en petite taille de populations est une moindre adaptation, de par le plus faible nombre de nouvelles mutations qui entre la population (Rousselle *et al.* 2020) ou de par le fait que les mutations faiblement avantageuses peuvent se comporter de manière effectivement neutres en petite taille de population ( $s < \frac{1}{2Ne}$ , Lanfear *et al.* 2014). Nous avons en effet pu démontrer que les espèces de notre jeu de données présentant les plus faibles  $Ne$  ont des taux de substitutions adaptatives plus réduites ainsi qu'une moindre efficacité de la sélection, tel qu'attendu suivant le modèle quasi-neutre de l'évolution (Ohta, 1992). Conjointement, ces analyses suggèrent que les espèces à faible  $Ne$ , ici représentées par les espèces insulaires endémiques, sont doublement affectées du fait d'une adaptabilité réduite et d'une augmentation du fardeau génétique (Fig. 8, Leroy *et al.* 2021 pour plus de détails). A ce stade des connaissances, les conséquences en termes de risque accru d'extinction restent non-renseignées, mais certains auteurs suggèrent que cela pourrait contribuer aux ultimes étapes du processus d'extinction des espèces, en tout cas, une fois un point de bascule atteint (Rogers and Slatin 2017).

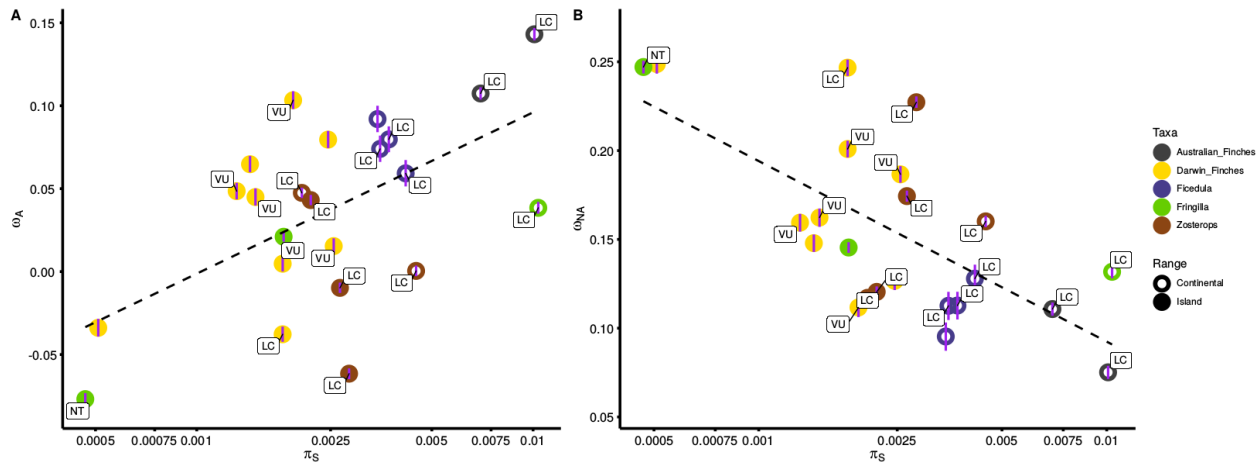


Figure 8 : Relations entre les taux de substitutions adaptatives (A) et non-adaptatives (B) et la diversité nucléotidique utilisée comme proxy des  $Ne$ . Brièvement, ces proportions sont estimées par comparaison des  $d_N/d_S$  observés et attendus sous quasi-neutralité sur la base des polymorphismes (i.e.  $pN/pS$ ) via la méthode développée par Galtier (2016) basée sur l'ajustement du modèle sur les SFS aux sites synonymes et non-synonymes avec  $\omega_A = \alpha(d_N/d_S)$  &  $\omega_{NA} = (1 - \alpha)(d_N/d_S)$ . Les estimations ont été obtenues en considérant tous les sites et le modèle Gamma-Expo. Les intervalles de confiance à 95% des estimations sont indiqués par le segment violet. Les meilleures régressions linéaires sont indiquées par la ligne pointillée ( $p=0.014$  et  $p=0.002$  pour  $\omega_A$  et  $\omega_{NA}$ , respectivement).

## b) Taux de mutation

La mutation est le moteur de l'évolution, la source de toute variation génétique. Son étude est donc par définition cruciale en sciences de l'évolution. Le nombre de mutations produites sur un temps donné (une année, une génération), qu'on appellera taux de mutation, est un caractère supposé très fortement sélectionné au sein des espèces. L'absence complète de mutations est considérée comme une impasse évolutive, ne permettant pas aux espèces l'adaptation aux conditions environnementales changeantes et donc leur pérennité dans le temps. À l'opposé, un trop grand nombre de mutations augmenterait le fardeau et représenterait une autre source de menace pour la pérennité de l'espèce. L'attendu est donc que les espèces évoluent vers un taux de mutations relativement faible, mais strictement non-nul (Kondrashov *et al.* 1988). Encore plus que pour les tailles efficaces, le manque de connaissances sur les taux de mutations est patent. Chez les plantes, très rares sont les estimations directes des taux de mutations disponibles, à ce jour essentiellement *Arabidopsis thaliana*, *Oryza sativa* et *Prunus mira* (Hanlon *et al.* 2019). Plus globalement, malgré l'importance que cela revêt du point de vue des maladies génétiques des enfants par exemple, on connaît que de manière relativement récente, le nombre moyen de mutations qu'un père et une mère transmettent à leur enfant, et la relation avec l'âge des parents à la conception (Jónsson *et al.* 2017).

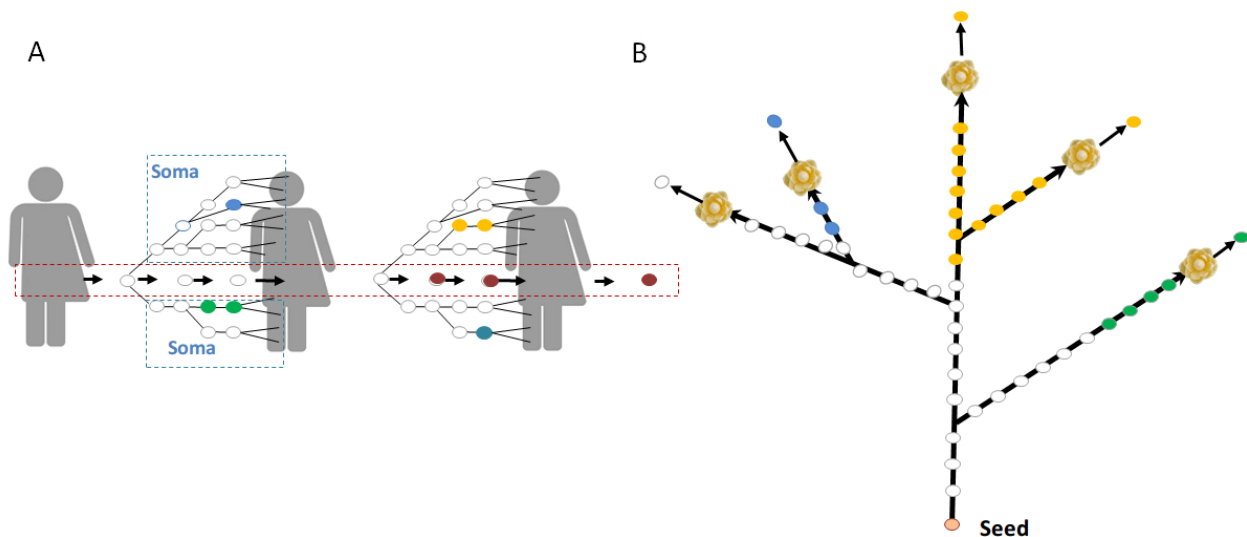


Figure 9 : Héritabilité des mutations somatiques dans le modèle animal (A) et le modèle végétal (B). A) Chez la majorité des espèces animales, seules les mutations accumulées dans la lignée germinale (compartiment rouge) sont susceptibles d'être transmises à la génération suivante. Les mutations somatiques (compartiment bleu) n'ont pas de devenir évolutif au-delà de la vie de l'individu (sauf cas rares des cancers transmissibles). B) La croissance des plantes étant assurée par des millions de générations de divisions cellulaires avant floraison, les mutations somatiques accumulées au sein des méristèmes primaires assurant cette croissance sont supposées présentes au sein des gamètes et donc héréditaires, en tout cas, sous l'hypothèse d'une absence de séparation des lignées germinales et somatiques chez les plantes (voir Lanfear 2018 pour des scénarios alternatifs).

Chez les humains, comme chez la plupart des animaux, seules les mutations des lignées germinales sont supposées héritables, du fait de l'isolement physique des cellules germinales du reste des lignées cellulaires (Fig. 9A). Chez les plantes, l'hypothèse générale est qu'il n'y a pas de séparation des lignées somatiques et germinales et donc que les plantes accumulent des mutations tout au cours de leur croissance et les transmettent à la génération suivante (Fig. 9B).

J'ai commencé à explorer ces questions au cours de mon stage postdoctoral à l'INRAE de Bordeaux (2014-2018). Nous avons conçu un schéma expérimental assez simple consistant à échantillonner des bourgeons sur quelques branches différentes d'un chêne et à séquencer ces échantillons avec une forte couverture (Fig. 10C). L'idée était de découvrir la diversité génétique au sein d'un même individu, c'est-à-dire des mutations nouvelles qui sont apparues au cours de la croissance du tronc ou des branches. En utilisant une méthode développée en cancérologie (MuTect, Cibulskis *et al.* 2013), cette étude m'a permis d'identifier quelques dizaines de mutations accumulées au cours de la croissance de cet arbre centenaire (Plomion *et al.* 2018, Leroy *et al.* 2020c). Il est à noter que de manière parallèle, une étude similaire avait été effectuée sur un chêne de Lausanne, mais en utilisant une approche basée sur des outils génériques (Schmid-Siegert *et al.* 2017), conduisant à une identification d'un nombre plus réduit de mutations (~10 fois moins de mutations, après pondération pour les âges différents des chênes étudiés). L'intérêt de l'utilisation d'une méthode développée pour la cancérologie réside dans le fait que pour les cancers comme pour les tissus végétaux, les nouvelles mutations apparaissent en basses fréquences et les détecter spécifiquement est supposé plus efficace en contrastant tissus sains et tumoraux (Fig. 10A). En utilisant une approche par simulation (Schmitt *et al.* en révision), Sylvain Schmitt, dans le cadre de son postdoctorat, a ainsi pu tester la performance des méthodes de détection des variants somatiques Fig. 10B). Au détour, ce travail fournit des enseignements sur les SNPs callers qui dépassent la seule détection des mutations somatiques. Pour illustrer ce dernier point, nous pouvons noter que le logiciel freebayes (Garrison & Marth 2012) génère une proportion élevée de faux positifs dans la détection des variants (Précision, Fig. 10B), or ce résultat je l'avais déjà observé empiriquement sur les jeux de données obtenus par Gil Yardeni dans le cadre de sa thèse. Il s'agissait là pourtant d'un contexte totalement différent avec une détection de variants en phylogénomique sur un jeu de données de plusieurs espèces couvrant la radiation des *Tillandsia* (Yardeni *et al.* 2021).

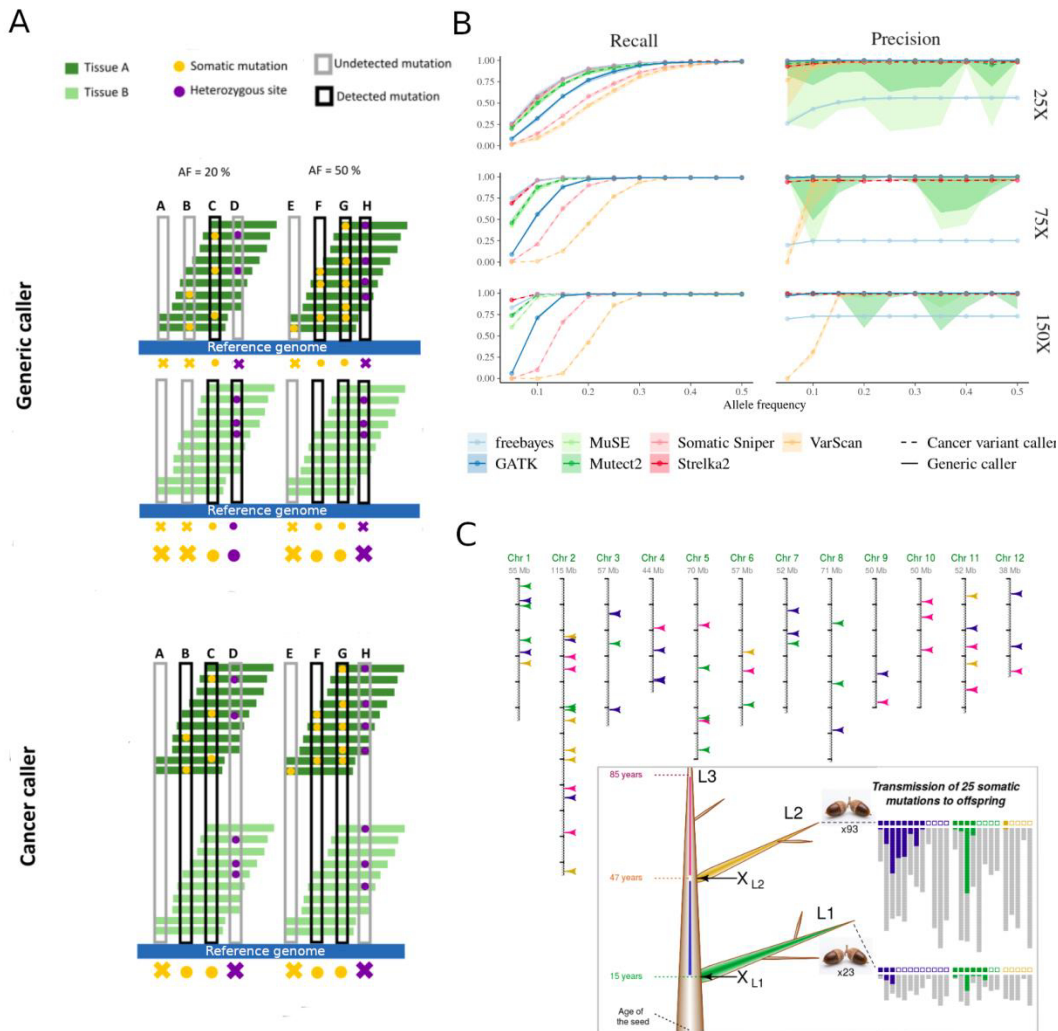


Figure 10 : A) Attendus généraux de détection de mutations somatiques avec des logiciels d'appel de variants génériques (GATK, freebayes) ou les méthodes spécifiquement développées pour la cancérologie (Mutect2, Strelka2, entre autres). Les méthodes utilisant des variants génériques nécessitent de faire spécifique sur chaque tissu et de raisonner ensuite de manière soustractive, là où les approches initialement développées pour les cancers ont des détections qui sont directement basées sur la comparaison des deux tissus (de type tissu sain vs. tumeur). En faible couverture (mutations A et éventuellement E), certaines mutations peuvent être ratées par les deux approches. Avec des couvertures intermédiaires (B & F), certaines mutations basses fréquences peuvent être ratées par les callers génériques. En forte couverture, les mutations somatiques doivent être facilement détectées par les deux approches (C, & G). Mais considérant la rareté des mutations somatiques au regard des sites hétérozygotes (D & H), certains effets d'échantillonnage des allèles peuvent conduire à la non-détection d'un site hétérozygote chez l'un ou chez l'autre des tissus créant la détection d'une mutation potentielle par les approches basées sur des scans génomiques (cercle violet) alors qu'il s'agit là d'un faux positif, ce qui handicape la détection des mutations somatiques. B) Performance des différents logiciels sur la base de simulations. Le « recall » correspond à la capacité à redétecter les mutations simulées. La précision correspond à la proportion de mutations simulées sur le nombre de mutations totales détectées (faux positifs inclus). Chaque point correspond à la moyenne estimée sur 10 simulations considérant des seuils de fréquences alléliques de la mutation et de profondeur de séquençage (les surfaces indiquent la variation observée à travers les 10 répliques d'analyse), pour plus de détail voir Schmitt et al. (en révision). C) Localisation des mutations somatiques détectées par Plomion et al. (2018) sur les 12 chromosomes du chêne pédonculé. Le cadre indique si les mutations somatiques ont été retrouvées au sein des embryons des descendants issus des branches L1 & L2. Le nombre total de copies analysées est indiqué par la somme des surfaces colorées (allèle muté observé) et et grises (allèle de référence non-muté observé).

Pour revenir aux mutations somatiques, Sylvain a non seulement procédé à ce test de performance des méthodes, mais a aussi procédé à une nouvelle détection des variants somatiques sur les jeux de données de chêne précédents (Schmid-Siegert *et al.* 2017 et Plomion *et al.* 2018). Sylvain a utilisé la méthode la plus performante sur la base des simulations (Strelka2 pour les deux jeux de données) et a ensuite procédé à la réanalyse des données. Ainsi, Sylvain a détecté 2 fois plus de mutations somatiques que dans le cas de ma précédente analyse (Plomion *et al.* 2018) et 7 fois plus que dans l'analyse suisse (Schmid-Siegert *et al.* 2017), ce qui contribue à rendre les chiffres entre ces études plus comparables et tend aussi à les rapprocher un peu plus des attendus théoriques en terme d'effectifs de mutations somatiques (Schoen & Schultz 2019). Bien que nous ayons détecté un nombre de mutations somatiques supérieur à celui des précédentes études, les effectifs demeurent très faibles, de l'ordre de quelques dizaines ou centaines de mutations somatiques sur des génomes qui font plusieurs centaines de millions de bases (750 Mb dans le cas du chêne). L'objectif de ces études est d'acquérir des connaissances fondamentales permettant d'étudier le processus de mutation. Dans ce contexte, il me semble important de communiquer à chaque fois sur ces résultats afin de veiller à ce que les interprétations qui en découlent ne soient pas ensuite malheureuses (p.ex. « *chaque branche développe son propre génome* », dossier spécial chêne, Science & Vie, Octobre 2018).

Un autre résultat majeur de mon postdoctorat à l'UMR Biogeco est la validation de la transmission des mutations somatiques aux descendants (Fig. 10C). Pour cela, nous avons utilisé une nacelle pour ensacher les branches basses précédemment utilisées pour la détection (L1 et L2) et ainsi récolter des lots de glands. Après extraction ADN sur les embryons de ces glands et génotypage des mutations somatiques, j'ai ainsi pu vérifier s'ils étaient porteurs des mutations. Environ la moitié des mutations somatiques testées ont été retrouvées dans certains embryons (Fig. 10C), démontrant une hérédité des mutations acquises au cours de la croissance chez les chênes, conformément au modèle végétal (Fig. 9B). Des résultats similaires ont été obtenus en détectant des mutations somatiques chez *Prunus mura* et en les recherchant chez des descendants séquencés en basse couverture (Wang *et al.* 2019). Des dispositifs expérimentaux relativement similaires avaient été mis en place en parallèle par les mêmes auteurs chez *Arabidopsis thaliana*, *Brachypodium distachyon* et *Oryza sativa*. Ils n'ont permis d'identifier que 1,7 à 6,3% des mutations somatiques au sein des lots de graines de ces espèces. Sur la base de ces quelques résultats, qui nécessiteront d'être très largement testés, une différence entre plantes pérennes et annuelles pourrait être suggérée au regard de la conformation au modèle végétal. En tout cas, cette illustration de résultats

autour de l'hérédité des mutations montre bien l'étendue des connaissances fondamentales qui restent à acquérir au sujet des mutations, en particulier chez les plantes.

Dans le cadre du projet TreeMutation (2020-2023) porté par Niklas Tysklind & Myriam Heuertz, dont je suis le responsable pour la partie détection des mutations (WP4), nous sommes en train de questionner le rôle des UV dans la production des mutations. L'idée est de comparer des échantillons prélevés sur des branches de pleine lumière de la canopée ou des branches plus basses, en se focalisant sur des espèces tropicales de la forêt guyanaise pour disposer d'un maximum de contraste. Je reviendrais sur ces points dans le projet de recherche, mais l'idée est de renseigner les taux et spectres de ces mutations, c'est-à-dire non seulement de les dénombrer, mais aussi d'identifier les contextes génomiques de ces mutations. Là encore, des parallèles pourront être effectués avec la cancérologie (p. ex. les mélanomes). Mais d'ores et déjà, il est bon d'introduire le fait que les contextes génomiques des mutations ont été, jusqu'à très récemment, très peu décrits chez les plantes, en comparaison au travail effectué sur l'humain, que ça soit pour les mutations des cellules germinales (p. ex. Milholland *et al.* 2017) et somatiques (p. ex. Martincorena & Campbell, 2015). La réanalyse des données des deux projets d'identification des mutations somatiques chez le chêne effectuée par Sylvain Schmitt ayant permis de fournir des effectifs de mutations plus importants, cela a rendu ces investigations possibles. Sylvain a ainsi pu renseigner le contexte trinuécléotidique des mutations chez les chênes et le caractère répétable de ces spectres à travers les deux jeux de données (Fig. 11).

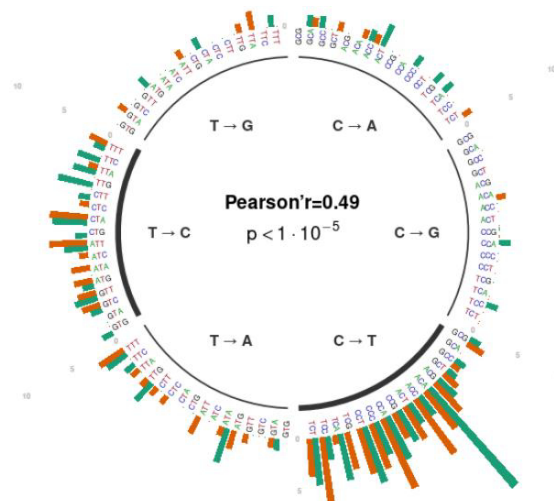


Figure 11 : Pourcentage des mutations somatiques détectées en fonction du contexte génétique local de la mutation (triplet : base en amont, mutation, base en aval) pour Schmid-Siegert et al. 2017 (vert) et Plomion et al. 2018 (orange). Les spectres de mutations sont fortement corrélés (Pearson  $r=0.49$ ,  $p < 1 \cdot 10^{-5}$ ). Les transitions sont indiquées par un trait plein plus large (T -> C et C -> T) comparativement aux transversions (les autres). Toute ressemblance avec le logo de l'émission de télévision Quotidien est fortuite !

## PROJET DE RECHERCHE ET PERSPECTIVES DE CARRIERE

Bien que n'ayant pas encore de poste permanent, les grands axes de recherches sur lesquels je souhaite continuer à travailler sur le court et moyen terme sont clairs pour moi. Ceci bien sûr n'exclut pas pour moi d'explorer d'autres dimensions en fonction des laboratoires de recherche où j'irai à l'avenir. J'espère en tout cas pouvoir continuer à étudier une large diversité de modèles biologiques au cours des prochaines années. J'ai la profonde conviction que la diversité des thématiques abordées, tout autant que des modèles biologiques, est essentielle en recherche académique, et qu'elle est probablement encore plus nécessaire pour l'étude de questions associées à la biodiversité. La notion de référentiel est cruciale afin de pouvoir dessiner des trajectoires. Par exemple, la diversité génétique des plantes domestiquées nécessite aussi d'étudier celle des progéniteurs sauvages, de la même façon que la diversité des espèces menacées nécessite d'avoir une bonne description de celles de ses proches apparentées non-menacées. A terme, on peut imaginer que la diversité génétique s'étudiera aussi sur un grand nombre d'espèces constitutives d'un écosystème particulier, afin d'en qualifier sa trajectoire et donc sa vulnérabilité.

### *I - Evolution de la diversité génétique dans le compartiment cultivé*

Un des grands axes que je souhaite approfondir est l'étude de l'évolution de la diversité d'espèces domestiquées. Bien que récente et trop timide (Hoban *et al.* 2021), une lente prise de conscience émerge de l'importance du maintien d'une diversité génétique suffisamment large pour répondre aux enjeux actuels et futurs (cf CBD, ONU 2021). La prise de conscience concerne en premier lieu les espèces les plus essentielles à l'alimentation humaine, le blé, le riz et le maïs, qui représentent à elles seules la moitié des calories consommées par les humains. Mais au-delà, ce sont de nombreuses autres espèces, des légumes comme des fruits, des espèces ornementales ou encore animales, et ultérieurement, la diversité sauvage, qu'il faudrait intégrer dans ce panorama. Et je tiens à faire remarquer au passage que, quel que soit le futur de l'amélioration génétique, que ça soit par l'introggression classique par générations de croisements successifs ou par les biotechnologies de transformation végétales (transgenèse, CRISPR-CAS9, etc), la disponibilité et la caractérisation de la diversité génétique existante resteront cruciales.

Même à l'heure de la génomique haut-débit, l'étude de la diversité des espèces domestiquées reste une tâche difficile. De nombreuses domestiquées sont par exemple polyploïdes (blé, caféier, colza, fraisier, pomme de terre, ...), alors que le ou les progéniteurs sauvages de ces espèces sont souvent diploïdes (dont certains sont d'ailleurs présumés disparus). L'exemple du rosier de jardin avec ses niveaux de ploïdie variables de 2x à 10x et son grand nombre de progéniteurs sauvages en est un bon exemple. Dans le cadre de mon expérience postdoctorale, je me suis surtout concentré sur son évolution au cours du XIX<sup>e</sup> siècle. Aucun gain net de la diversité génétique, que ça soit comparativement à la diversité ancestrale initialement présente en Europe ou même à la diversité réduite du pool génétique asiatique, n'a pu être observé. Le croisement de deux fonds génétiques bien distincts n'a donc pas contribué à augmenter sa diversité génétique. A court terme, je souhaite continuer à travailler sur l'évolution de la diversité génétique d'espèces domestiquées, comme celle du rosier, sur la base de marqueurs génétiques relativement simples. Si jusqu'alors, j'ai réalisé des analyses indépendantes sur plusieurs espèces domestiquées, le riz africain, le blé tendre ou encore le rosier, une évaluation plus globale reste à effectuer. Je souhaite travailler à l'avenir dans ce sens, en commençant par plusieurs autres espèces de céréales (orge, blé dur, etc). Il s'agirait là d'établir un diagnostic et de mettre la lumière sur cet enjeu de sécurité sanitaire mondial qui reste trop souvent ignoré. A plus long terme, j'aimerais profiter de l'essor des technologies de séquençage de longs fragments pour étudier les bases génétiques d'une diversité plus large. Cela devrait permettre d'étudier la diversité associée à des mutations complexes (variants structuraux) et ainsi mieux expliquer notamment le maintien ou l'arrêt des échanges génétiques entre les espèces domestiquées et leurs progéniteurs sauvages.

## ***II - Diversité génétique et biologie de la conservation***

La liste rouge des espèces menacées de l'Union Internationale pour la Conservation de la Nature (IUCN en anglais) établit les priorités de conservation. Cette liste rouge a pour objectif de guider les efforts de conservation sur des échelles de temps assez courtes (un siècle ou moins) et se base surtout sur des informations écologiques (nombre d'individus, etc.). Une question importante, et qui reste paradoxalement encore assez ouverte, est de savoir si les espèces incluses sur ces listes présentent une diversité génétique moindre que les espèces considérées comme moins préoccupantes. Sur la base de quelques espèces de passereaux, menacées ou non d'après l'IUCN, nous n'avons pas observé de différences majeures en ce qui concerne la proportion de substitutions



adaptatives et légèrement délétères (Leroy *et al.* 2021). Mais une telle comparaison implique de raisonner à des échelles de temps différentes, le “temps humain” de destruction des habitats naturels et autres menaces anthropiques induites, et le temps évolutif d’autre part. Il n’en demeure pas moins que l’accumulation de mutations légèrement délétères et surtout la faible efficacité de la sélection dans les espèces à petite taille efficace pourrait contribuer, sur les temps longs, à leur extinction (Lynch *et al.* 1995). Pour ces espèces fragilisées, une fois un point critique atteint, la voie de l’extinction serait toute tracée (Rogers & Slatkin, 2017). Ces conclusions semblent être appuyées par le séquençage de l’ADN d’un des derniers échantillons de mammoth laineux (~4,300 years) retrouvé sur l’île Wrangel, un territoire au nord de la Russie (Palkopoulou *et al.* 2015). La paléogénomique, c’est-à-dire l’étude de l’ADN d’échantillons fossiles, représente un tournant majeur de notre discipline.

Lors de la dernière décennie, nombre de mes travaux ont fait appel à de l’inférence démographique, ce qui revient à trouver des scénarios et paramètres démographiques capables d’expliquer les niveaux de diversité génétique observés au sein des populations actuelles. J’ai moi-même fait appel régulièrement à ces approches (Leroy *et al.* 2017, 2020a, Merceron *et al.* 2017, Rougemont *et al.* 2020 etc). Ce sont sans nul doute des outils puissants, mais ils nécessitent aussi de faire des hypothèses fortes. La paléogénomique est de nature à permettre des comparaisons plus directes, en utilisant des échantillons allochroniques. Une limite à la paléogénomique reste bien sûr l’acquisition d’échantillons fossiles. Par exemple, l’acquisition d’échantillons fossiles de chênes d’une qualité suffisamment correcte pour envisager un séquençage complet du génome a été un travail de longue haleine. Pour trouver les meilleurs échantillons de chêne, il a fallu plonger au fond de lacs pour trouver des bois qui avaient été utilisés pour les constructions lacustres (Wagner *et al.* 2018). Au côté de Stefanie Wagner (INRAE Toulouse) ou d’autres collaborateurs, je souhaite contribuer à l’analyse d’échantillons d’ADN ancien dans les prochaines années. Dans le cadre du chêne, un des premiers objectifs serait de valider des résultats précédemment obtenus sur la base d’inférence, en particulier au regard du contact secondaire massif entre les 4 espèces majeures présentes en Europe et d’événements antérieurs. Mais cet objectif ne se limite pas seulement aux chênes. A l’avenir, l’apport de la paléogénomique à l’étude de l’évolution de la diversité génétique, notamment des espèces domestiquées, devrait être massif. J’espère avoir l’occasion de contribuer à l’analyse de telles données, dans les contextes des compartiments cultivés comme sauvages.

### ***III - Taux de mutations et efficacité de la sélection***

Si les sujets précédents sont des sujets d'intérêt dans la pleine lignée de mes travaux passés, il y en a un autre qui me passionne depuis des années et que je n'arrive pas pour le moment à explorer pleinement. Il s'agit de l'étude des processus de mutations. Comme étudiant, j'ai appris que les mutations étaient causées par des erreurs de l'ADN polymérase à une fréquence d'environ  $1 \times 10^{-9}$  par paire de bases et par génération et que ce processus était complètement aléatoire. Des années m'auront été nécessaires pour déconstruire l'idée que ce processus était bien connu et que les mutations n'étaient pas un processus strictement aléatoire. Aujourd'hui, j'en suis arrivé à la conviction que la mutation est la force évolutive qui reste la moins bien décrite en science de l'évolution.

Pour bien comprendre cette partie de mon projet, il faut bien saisir que chez la plupart des animaux, il y a une séparation physique qui intervient très tôt au cours du développement entre les lignées somatiques et germinales, ce qui conduit à un nombre faible de divisions cellulaires dans les tissus qui deviendront les lignées germinales et par conséquent peu de mutations hérissables (voir aussi Fig. 9). Chez les plantes, il est couramment admis que cette différenciation se produit au contraire très tard, en bout de branches, de telle façon que les mutations accumulées au cours de la croissance des tissus somatiques du méristème primaire des parties aériennes, typiquement les branches et les troncs, peuvent aussi être transmises à la génération suivante. Le modèle végétal serait donc en accord avec une partie des mutations qui est acquise au cours du développement et une autre partie au cours de la méiose, c'est-à-dire une contribution somatique et gamétique (notamment méiotique) aux taux de mutation des plantes. A ce jour, nous n'avons aucune idée de la part relative de ces deux types de mutations. Grâce à l'étude des patrons d'évolution moléculaire, en faisant quelques hypothèses, notamment celle de neutralité sélective, on sait que les espèces pérennes présentent un taux de mutation par génération plus fort que les espèces annuelles (Lanfear *et al.* 2013), ce qui serait donc explicable par le plus grand nombre de générations de mitoses, mais pour autant, ce taux n'est pas plus élevé, au contraire même, lorsqu'il est ramené à une même unité de temps (p. ex. une année; Lanfear *et al.* 2013, Hanlon *et al.* 2019). Il est donc probable que les mutations somatiques hérissables aient une contribution massive chez les arbres, mais que les effectifs restent plus limités que ceux des mutations générées chez les annuelles sur la même période de temps, de par l'impact plus grand des mutations gamétiques (plusieurs générations de reproduction sexuée). Je tiens à préciser ici que seul le processus de mutation et le fait qu'il soit

moteur de diversité est au cœur de mes considérations. Bien que très à la mode, les hypothèses plus ou moins étayées concernant l'éventuel avantage procuré à l'intérieur de l'arbre par les mutations somatiques pour l'adaptation en temps réel aux changements de conditions environnementales ne seront pas considérées.

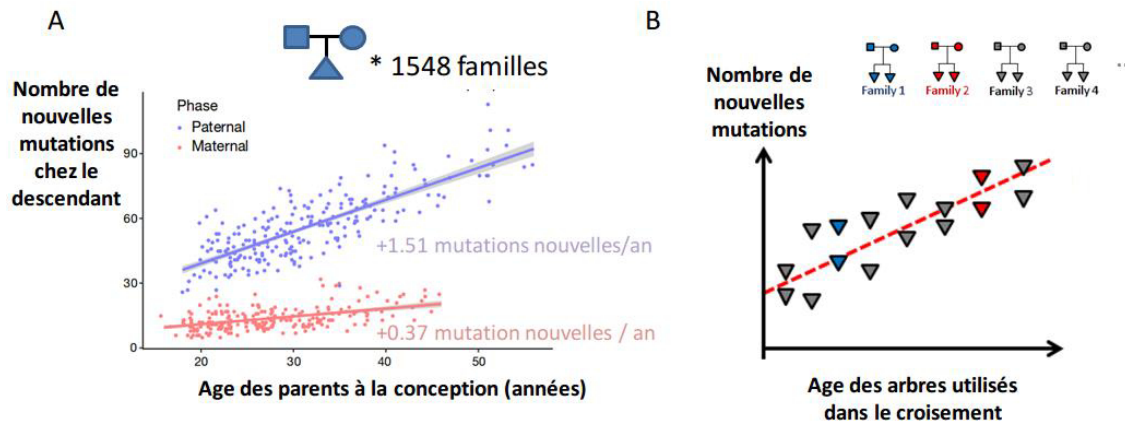


Figure 12 : Nombre de mutations nouvelles détectées transmises à l'enfant en fonction de l'âge des parents à la conception de l'enfant sur la base de plus de 1500 familles d'Islandais (adapté de Jónsson et al. 2017). Les mutations transmises par le père sont indiquées en bleu et les mutations en rouge sont celles de la mère. Les coefficients directeurs des meilleures régressions linéaires sont indiqués. B) Modèle général pour l'estimation des contributions gamétiques (ordonnée à l'origine) et somatiques (pente) aux taux de mutations des chênes.

Un de mes objectifs principaux est d'estimer le taux de mutation et de décomposer la contribution somatique et gamétique de ce taux de mutation, dans un premier temps en utilisant le chêne comme modèle. Concrètement, mon objectif est de transposer la stratégie qui a été développée chez l'humain pour le séquençage de plus d'un millier de familles islandaises (Fig. 12A, Jónsson *et al.* 2017). Pour cela, mes collègues bordelais procèdent actuellement à des croisements entre deux parents de même âge et ce pour toute une gamme d'âges différents, des croisements entre des parents jeunes (p.ex. 20 ans) et des croisements avec des parents plus âgés (p.ex. 250 ans), ainsi que de nombreux intermédiaires. Des glands correspondant à ces fleurs fécondées sont ensuite récupérés et plantés, de telle façon qu'un séquençage des parents et des descendants soit possible pour identifier les nouvelles mutations. Plusieurs croisements contrôlés ont déjà été obtenus et d'autres sont en cours pour obtenir l'ensemble du matériel nécessaire à l'étude. Un financement Défi Scientifique INRAE a déjà été acquis pour ce projet, couvrant l'acquisition de ce matériel biologique, mais les frais de séquençage ne sont actuellement pas

couverts, ce qui devra donc faire l'objet d'un appel à projet spécifique. On notera qu'ici, je fais fi du fait que j'ai soumis ce projet à deux reprises au programme Marie Skłodowska-Curie sans succès (sans que les raisons des rejets aient été motivées). Par la suite, je pars du principe qu'un jour ces recherches seront reconnues dignes d'être financées, ce qui est loin d'être certain ! Le projet initial visait à être non-seulement effectué sur chêne, mais aussi être répliqué sur bouleau, avec comme partenaire principal Martin Lascoux de l'Université d'Uppsala (Suède).

Pour estimer ces taux de mutation somatiques et gamétiques, l'idée que je souhaite développer est la suivante. En déterminant la meilleure des régressions dans le nuage de points du nombre de mutations en fonction de l'âge des parents, il me sera possible d'identifier ainsi la pente, qui serait donc associée aux mutations somatiques héritables produites au cours de la croissance de l'arbre, et l'ordonnée à l'origine qui serait la fraction invariable, liée alors aux mutations gamétiques (Fig. 12B). Par simplicité, je considère ici cette relation comme strictement linéaire, mais il est possible qu'elle ne le soit pas, ce qui reste à étudier. Si le nombre de mutations est suffisamment important pour une telle étude, on pourrait aussi analyser les spectres de ces mutations, c'est-à-dire les types et les contextes génomiques où elles se produisent. Des travaux similaires chez l'homme et la souris ont par exemple révélé que les signatures des mutations somatiques varient énormément entre ces deux espèces, là où les spectres de mutations gamétiques sont beaucoup plus semblables, bien que non rigoureusement identiques (Milholland *et al.* 2017).

Nous ne savons, peu ou prou, rien sur les agents mutagènes de l'ADN chez les plantes. Les UV sont supposés être des sources importantes de mutations car de nombreux organes sont fortement exposés aux rayonnements. Dans le cadre du projet TreeMutation porté par Myriam Heuertz et Niklas Tysklind (CEBA, 2020-2023), nous étudions ce processus en détail. Le schéma expérimental est assez simple, bien que d'ores et déjà complexe à mettre en place, en pratique. L'idée est de contraster le nombre de mutations observées sur des branches de pleine lumière de la canopée de la forêt guyanaise, à des branches d'ombre. L'échantillonnage a déjà été effectué. Outre le nombre de mutations détectées, il sera possible d'étudier le type de mutations induites par ces rayonnements, en comparant les spectres observés pour les branches à l'ombre et en pleine lumière. Une question très générale qui peut être posée est de savoir si les mutations induites par les UV sur l'ADN humain sont les mêmes que les mutations induites chez des arbres tropicaux. Ce type de questions très générales, pourra être vérifié, car le spectre des mutations humaines associées aux UV est bien décrit. Elles sont quasi-exclusivement des mutations C→T (ou G→A) et pour 50%

d'entre-elles environ, elles ont lieu dans un contexte précis, avec un T en 5' et un G en 3' de la base mutée, c'est-à-dire des mutations TCG → TTG (ou CGA → CAA).

Les radiations sont un autre type d'agents mutagènes dont les spectres sont mal décrits. Nous ne connaissons que très mal les spectres associés à la radioactivité chez l'humain, et même au-delà chez les animaux (Saclier *et al.* 2020). Pour étudier ce processus, j'ai réuni plusieurs collaborateurs internationaux en Europe, Amérique du Nord et Australie, afin d'utiliser les arbres des forêts de Tchernobyl comme modèles. Deux projets sont envisagés. Une mission de prospection que j'ai coordonnée en 2018 a permis de valider la faisabilité de ces deux parties. D'abord, j'aimerais utiliser le même dispositif que précédemment présenté dans le cadre des chênes en zone non-contaminée (estimation des taux de mutations par génération), mais cette fois-ci dans les forêts à proximité du réacteur #4 (< 5 kilomètres). Pour cela, il faudra retrouver les relations de parentés au sein des forêts par génotypage d'un grand nombre d'individus et ensuite séquencer certaines familles. Cela pourrait renseigner sur l'évolution du taux de mutations dans des zones fortement irradiées (Fig. 13A), des conditions environnementales que les espèces n'ont pas rencontrées auparavant, au cours de leur évolution. De la même façon que précédemment, ce projet vise à être dupliqué sur le bouleau, dans le cadre d'une collaboration avec Martin Lascoux, puisque les chênes et les bouleaux sont relativement fréquemment retrouvés dans la zone d'exclusion de Tchernobyl. Ensuite, j'aimerais utiliser les arbres qui ont survécu à la catastrophe pour retracer la dynamique temporelle d'accumulation des mutations le long de l'arbre (Fig. 13C). En effet, si la radioactivité a eu des effets ponctuels, alors on devrait observer une accumulation de mutations au niveau des tissus produits en 1986 et une décroissance forte depuis, à mesure de la baisse des niveaux de radiation, avec des niveaux actuels proches de ceux observés dans les tissus mis en place avant la catastrophe. Au contraire, si les radiations ont des effets chroniques, alors on devrait observer des effectifs de mutation maintenus à un niveau élevé. Cela revient à considérer l'arbre comme un biomarqueur. L'étude fine des spectres de mutations pourrait en outre permettre de mieux décrire les types de mutations induites par les rayonnements, via une comparaison avec les mutations observées en zones non-contaminées (Fig. 13B). A noter que la comparaison des spectres obtenus par l'approche intergénérationnelle et intra-individuelle pourrait fournir des informations sur l'évolution des spectres de mutations somatiques et germinales, a fortiori si le dispositif est lui aussi répliqué sur bouleau (à l'image de ce qui a pu être effectué sur les spectres somatiques et germinaux de la souris et de l'humain, Milholland *et al.* 2017). A très long terme, lorsque les données génétiques individuelles ne seront en rien limitantes (type médecine de précision, voire

plus encore) et que les signatures de mutations associées aux différents mutagènes seront bien décrites, il sera a priori possible de retracer les conditions environnementales auxquelles des populations ont été exposées, via l'étude du nombre et du spectre des mutations, ce qui serait susceptible de fournir des connaissances extrêmement variées sur l'histoire de certains individus et/ou populations.

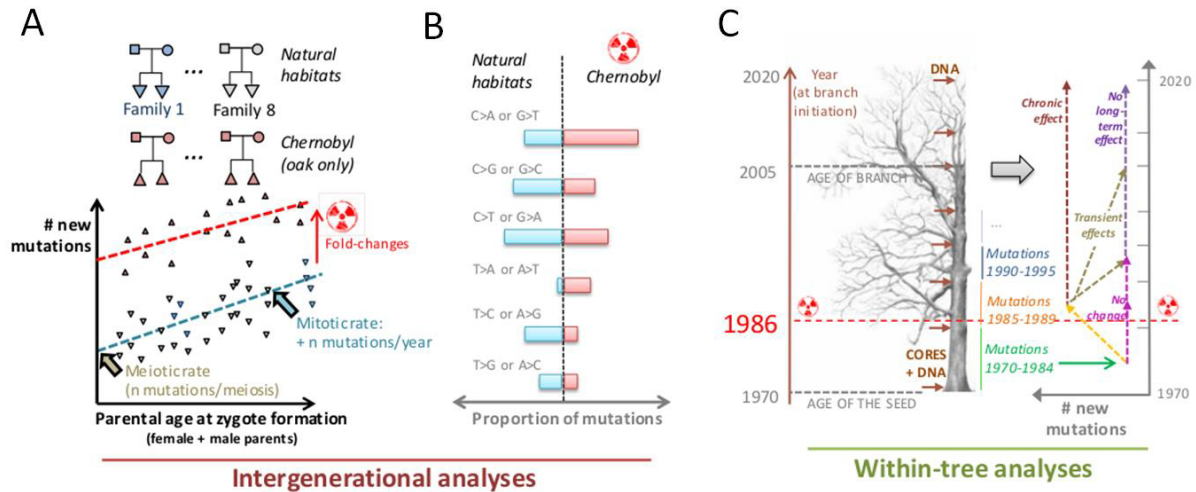


Figure 13 : Résumé du projet d'étude des processus de mutation à Tchernobyl, tel que présenté dans la dernière version du feu projet de Marie Curie. A) Étude des conséquences des taux de mutations intergénérationnels et plus spécifiquement sur les composantes méiotiques et mitotiques en faisant appel à des séquençages en trios ou quartets (idéalement quartets). B) Comparaison des spectres de mutations en conditions naturelles et à Tchernobyl. Ici par simplicité de représentation, seuls les grands types de mutations ont été représentés, mais le projet vise aussi ici à étudier précisément les contextes génomiques. C) Représentation schématique du projet visant à étudier le nombre de mutations à différents niveaux d'un même arbre ayant survécu à la catastrophe. A noter que plusieurs de ces dates ont déjà été identifiées. Les point d'échantillonnage seront datés par dendrochronologie, ce qui donnera des repères sur les phases d'accumulation des mutations, et ainsi permettra de renseigner des effets à courts ou à longs termes des radiations.

## REFERENCES

- Alberto, F.J., Derory, J., Boury, C., Frigerio, J.-M., Zimmermann, N.E., and Kremer, A. (2013). Imprints of Natural Selection Along Environmental Gradients in Phenology-Related Genes of *Quercus petraea*. *Genetics* 195, 495–512.
- Atkinson, Q.D., Gray, R.D., and Drummond, A.J. (2008). mtDNA Variation Predicts Population Size in Humans and Reveals a Major Southern Asian Chapter in Human Prehistory. *Molecular Biology and Evolution* 25, 468–474.
- Charlesworth, B., Morgan, M.T., and Charlesworth, D. (1993). The effect of deleterious mutations on neutral molecular variation. *Genetics* 134, 1289–1303.
- Chen, J., Glémin, S., and Lascoux, M. (2017). Genetic Diversity and the Efficacy of Purifying Selection across Plant and Animal Species. *Molecular Biology and Evolution* 34, 1417–1428.
- Chen W., Chen L., Zhang X., Yang N., Guo J., Wang M., Ji S., Zhao X., Yin P., Cai L., *et al.* Convergent selection of a WD40 protein that enhances grain yield in maize and rice. *Science* 375, eabg7985.
- Cibulskis, K., Lawrence, M.S., Carter, S.L., Sivachenko, A., Jaffe, D., Sougnez, C., Gabriel, S., Meyerson, M., Lander, E.S., and Getz, G. (2013). Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nature Biotechnology* 31, 213–219.
- Cruickshank, T.E., and Hahn, M.W. (2014). Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology* 23, 3133–3157.
- Ellegren, H., Smeds, L., Burri, R., Olason, P.I., Backström, N., Kawakami, T., Künstner, A., Mäkinen, H., Nadachowska-Brzyska, K., Qvarnström, A., *et al.* (2012). The genomic landscape of species divergence in *Ficedula* flycatchers. *Nature* 491, 756. .
- Fraïsse, C., Popovic, I., Mazoyer, C., Spataro, B., Delmotte, S., Romiguier, J., Loire, É., Simon, A., Galtier, N., Duret, L., *et al.* (2021). DILS: Demographic inferences with linked selection by using ABC. *Molecular Ecology Resources* 21, 2629–2644.
- Galtier, N. (2016). Adaptive Protein Evolution in Animals and the Effective Population Size Hypothesis. *PLOS Genetics* 12, e1005774.
- Garrison, E., and Marth, G. (2012). Haplotype-based variant detection from short-read sequencing.
- Gautier, M. (2015). Genome-Wide Scan for Adaptive Divergence and Association with Population-Specific Covariates. *Genetics* 201, 1555–1579.
- Gladieux, P., Zhang, X.-G., Afoufa-Bastien, D., Valdebenito Sanhueza, R.-M., Sbaghi, M., and Le Cam, B. (2008). On the Origin and Spread of the Scab Disease of Apple: Out of Central Asia. *PLOS ONE* 3, e1455.
- Han, F., Lamichaney, S., Grant, B.R., Grant, P.R., Andersson, L., and Webster, M.T. (2017). Gene flow, ancient polymorphism, and ecological adaptation shape the genomic landscape of divergence among Darwin’s finches. *Genome Research* 27, 1004–1015.
- Hanlon, V.C.T., Otto, S.P., and Aitken, S.N. (2019). Somatic mutations substantially increase the per-generation mutation rate in the conifer *Picea sitchensis*. *Evolution Letters* 3, 348–358.
- Haudry, A., Cenci, A., Ravel, C., Bataillon, T., Brunel, D., Poncet, C., Hochu, I., Poirier, S., Santoni, S., Glémin, S., *et al.* (2007). Grinding up wheat: a massive loss of nucleotide diversity since domestication. *Mol Biol Evol* 24, 1506–1517.
- Hoban, S., Campbell, C.D., da Silva, J.M., Ekblom, R., Funk, W.C., Garner, B.A., Godoy, J.A., Kershaw, F., MacDonald, A.J., Mergeay, J., *et al.* (2021). Genetic diversity is considered important but interpreted narrowly in country reports to the Convention on Biological Diversity: Current actions and indicators are insufficient. *Biological Conservation* 261, 109233.

- Irwin, D.E., Milá, B., Toews, D.P.L., Brelsford, A., Kenyon, H.L., Porter, A.N., Grossen, C., Delmore, K.E., Alcaide, M., and Irwin, J.H. (2018). A comparison of genomic islands of differentiation across three young avian species pairs. *Molecular Ecology* 27, 4839–4855.
- Iwata, H., Gaston, A., Remay, A., Thouroude, T., Jeauffre, J., Kawamura, K., Oyant, L.H.-S., Araki, T., Denoyes, B., and Foucher, F. (2012). The TFL1 homologue KSN is a regulator of continuous flowering in rose and strawberry. *The Plant Journal* 69, 116–125.
- Jónsson, H., Sulem, P., Kehr, B., Kristmundsdóttir, S., Zink, F., Hjartarson, E., Hardarson, M.T., Hjorleifsson, K.E., Eggertsson, H.P., Gudjonsson, S.A., *et al.* (2017). Parental influence on human germline de novo mutations in 1,548 trios from Iceland. *Nature* 549, 519. .
- Kondrashov, A.S. (1988). Deleterious mutations and the evolution of sexual reproduction. *Nature* 336, 435.
- Kremer, A., Plomion, C., Leroy, T., Alexandre, H., Truffaut, L., Delzon, S., Ducouso, A., Musch, B., Torres-Ruiz, J.M., and Wagner, S. (2020). Évolution passée et contemporaine des chênes. *Revue Forestière Française* 72, 297–318. .
- Lamichhaney, S., Berglund, J., Almén, M.S., Maqbool, K., Grabherr, M., Martinez-Barrio, A., Promerová, M., Rubin, C.-J., Wang, C., Zamani, N., *et al.* (2015). Evolution of Darwin’s finches and their beaks revealed by genome sequencing. *Nature* 518, 371–375.
- Lanfear, R. (2018). Do plants have a segregated germline? *PLOS Biology* 16, e2005439.
- Lanfear, R., Ho, S.Y.W., Jonathan Davies, T., Moles, A.T., Aarssen, L., Swenson, N.G., Warman, L., Zanne, A.E., and Allen, A.P. (2013). Taller plants have lower rates of molecular evolution. *Nature Communications* 4, 1879.
- Lanfear, R., Kokko, H., and Eyre-Walker, A. (2014). Population size and the rate of evolution. *Trends in Ecology & Evolution* 29, 33–41.
- Leffler, E.M., Bullaughey, K., Matute, D.R., Meyer, W.K., Ségurel, L., Venkat, A., Andolfatto, P., and Przeworski, M. (2012). Revisiting an Old Riddle: What Determines Genetic Diversity Levels within Species? *PLOS Biology* 10, e1001388.
- Leroy, T., and Nabholz, B. (2022). Response to Kratochvíl and Rovatsos. *Current Biology* 32, R30–R31.
- Leroy, T., and Rougemont, Q. (2021). Introduction to Population Genomics Methods. In *Molecular Plant Taxonomy* (Springer Nature), p.287-324.
- Leroy, T., Lemaire, C., Dunemann, F., and Le Cam, B. (2013). The genetic structure of a *Venturia inaequalis* population in a heterogeneous host population composed of different *Malus* species. *BMC Evolutionary Biology* 13, 64.
- Leroy, T., Caffier, V., Celton, J.-M., Anger, N., Durel, C.-E., Lemaire, C., and Le Cam, B. (2016). When virulence originates from nonagricultural hosts: evolutionary and epidemiological consequences of introgressions following secondary contacts in *Venturia inaequalis*. *New Phytologist* 210, 1443–1452.
- Leroy, T., Roux, C., Villate, L., Bodénès, C., Romiguier, J., Paiva, J.A.P., Dossat, C., Aury, J.-M., Plomion, C., and Kremer, A. (2017). Extensive recent secondary contacts between four European white oak species. *New Phytologist* 214, 865–878.
- Leroy, T., Rougemont, Q., Dupouey, J.-L., Bodenes, C., Lalanne, C., Belser, C., Labadie, K., Le Provost, G., Aury, J.-M., Kremer, A., *et al.* (2020a). Massive postglacial gene flow between European white oaks uncovered genes underlying species barriers. *New Phytologist*, 226: 1183-1197
- Leroy, T., Louvet, J.-M., Lalanne, C., Le Provost, G., Labadie, K., Aury, J.-M., Delzon, S., Plomion, C., and Kremer, A. (2020b). Adaptive introgression as a driver of local adaptation to climate in European white oaks. *New Phytologist*, 226: 1171-1182.
- Leroy, T., Plomion, C., and Kremer, A. (2020c). Oak symbolism in the light of genomics. *New Phytologist* 226, 1012–1017.
- Leroy, T., Rousselle, M., Tilak, M.-K., Caizergues, A.E., Scornavacca, C., Recuerda, M., Fuchs, J., Illera, J.C., De Swardt, D.H., Blanco, G., *et al.* (2021). Island songbirds as windows into evolution in small populations. *Current Biology* 31, 1303-1310.e4.
- Li, W.H., and Sadler, L.A. (1991). Low nucleotide diversity in man. *Genetics* 129, 513–523.



Lucena-Perez, M., Marmesat, E., Kleinman-Ruiz, D., Martínez-Cruz, B., Węcek, K., Saveljev, A.P., Seryodkin, I.V., Okhlopkov, I., Dvornikov, M.G., Ozolins, J., *et al.* (2020). Genomic patterns in the widespread Eurasian lynx shaped by Late Quaternary climatic fluctuations and anthropogenic impacts. *Molecular Ecology* 29, 812–828.

Lynch, M., Conery, J., and Bürger, R. (1995). Mutational Meltdowns in Sexual Populations. *Evolution* 49, 1067–1080.

Martin, S.H., Dasmahapatra, K.K., Nadeau, N.J., Salazar, C., Walters, J.R., Simpson, F., Blaxter, M., Manica, A., Mallet, J., and Jiggins, C.D. (2013). Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome Research*

Martincorena, I., and Campbell, P.J. (2015). Somatic mutation in cancer and normal cells. *Science* 349, 1483–1489.

Merceron, N.R., Leroy, T., Chancerel, E., Romero-Severson, J., Borkowski, D.S., Ducouso, A., Monty, A., Porté, A.J., and Kremer, A. (2017). Back to America: tracking the origin of European introduced populations of *Quercus rubra* L. *Genome* 60, 778–790.

Milholland, B., Dong, X., Zhang, L., Hao, X., Suh, Y., and Vijg, J. (2017). Differences between germline and somatic mutation rates in humans and mice. *Nature Communications* 8, 15183.

Nei, M., and Li, W.H. (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc Natl Acad Sci USA* 76, 5269.

Nei, M., and Tajima, F. (1981). DNA polymorphism detectable by restriction endonucleases. *Genetics* 97, 145–163. .

Ohta, T. (1992). The Nearly Neutral Theory of Molecular Evolution. *Annu. Rev. Ecol. Syst.* 23, 263–286.

Palkopoulou, E., Mallick, S., Skoglund, P., Enk, J., Rohland, N., Li, H., Omrak, A., Vartanyan, S., Poinar, H., Götherström, A., *et al.* (2015). Complete genomes reveal signatures of demographic and genetic declines in the woolly mammoth. *Curr Biol* 25, 1395–1400.

Palstra, F.P., and Fraser, D.J. (2012). Effective/census population size ratio estimation: a compendium and appraisal. *Ecology and Evolution* 2, 2357–2365.

Petit, R.J., Pineau, E., Demesure, B., Bacilieri, R., Ducouso, A., and Kremer, A. (1997). Chloroplast DNA footprints of postglacial recolonization by oaks. *Proc Natl Acad Sci USA*. 94, 9996–10001. .

Petit, R.J., Csaikl, U.M., Bordács, S., Burg, K., Coart, E., Cottrell, J., van Dam, B., Deans, J.D., Dumolin-Lapègue, S., Fineschi, S., *et al.* (2002). Chloroplast DNA variation in European white oaks: Phylogeography and patterns of diversity based on data from over 2600 populations. *Forest Ecol Manag.* 156, 5–26.

Plomion, C., Aury, J.-M., Amselem, J., Leroy, T., Murat, F., Duplessis, S., Faye, S., Francillon, N., Labadie, K., Le Provost, G., *et al.* (2018). Oak genome reveals facets of long lifespan. *Nature Plants* 4, 440–452.

Pont, C., Leroy, T., Seidel, M., Tondelli, A., Duchemin, W., Armisen, D., Lang, D., Bustos-Korts, D., Goué, N., Balfourier, F., *et al.* (2019). Tracing the ancestry of modern bread wheats. *Nature Genetics* 51, 905–911.

Renaut, S., Grassa, C.J., Yeaman, S., Moyers, B.T., Lai, Z., Kane, N.C., Bowers, J.E., Burke, J.M., and Rieseberg, L.H. (2013). Genomic islands of divergence are not affected by geography of speciation in sunflowers. *Nature Communications* 4, 1827.

Rogers, R.L., and Slatkin, M. (2017). Excess of genomic defects in a woolly mammoth on Wrangel island. *PLOS Genetics* 13, e1006601.

Rougemont, Q., Moore, J.-S., Leroy, T., Normandeau, E., Rondeau, E.B., Withler, R.E., Van Doornik, D.M., Crane, P.A., Naish, K.A., Garza, J.C., *et al.* (2020). Demographic history shaped geographical patterns of deleterious mutation load in a broadly distributed Pacific Salmon. *PLOS Genetics* 16, e1008348.

Rousselle, M., Simion, P., Tilak, M.-K., Figuet, E., Nabholz, B., and Galtier, N. (2020). Is adaptation limited by mutation? A timescale-dependent effect of genetic diversity on the adaptive substitution rate in animals. *PLOS Genetics* 16, e1008668.

Saclier, N., Chardon, P., Malard, F., Konecny-Dupré, L., Eme, D., Bellec, A., Breton, V., Duret, L., Lefebure, T., and Douady, C.J. (2020). Bedrock radioactivity influences the rate and spectrum of mutation. *ELife* 9, e56830.

Schmid-Siegert, E., Sarkar, N., Iseli, C., Calderon, S., Gouhier-Darimont, C., Chrast, J., Cattaneo, P., Schütz, F., Farinelli, L., Pagni, M., *et al.* (2017). Low number of fixed somatic mutations in a long-lived oak tree. *Nature Plants* 3, 926–929.

Schmitt, S., Leroy, T., Heuertz, M., and Tysklind, N. (soumis). Somatic mutation detection in plants: a critical evaluation through simulations and reanalyses. Disponible sur BioRxiv (2021.10.11.462798)

Schoen, D.J., and Schultz, S.T. (2019). Somatic Mutation and Evolution in Plants. *Annu. Rev. Ecol. Evol. Syst.* 50, 49–73.

Shang, H., Hess, J., Pickup, M., Field, D.L., Ingvarsson, P.K., Liu, J., and Lexer, C. (2020). Evolution of strong reproductive isolation in plants: broad-scale patterns and lessons from a perennial model group. *Philosophical Transactions of the Royal Society B: Biological Sciences* 375, 20190544.

Shang, H., Rendón-Anaya, M., Paun, O., Field, D.L., Hess, J., Vogl, C., Liu, J., Ingvarsson, P.K., Lexer, C., and Leroy, T. (en révision pour *PCI Evolutionary Biology*). Conserved genomic landscapes of differentiation across *Populus* speciation continuum. (disponible sur BioRxiv 2021.08.26.457771).

Smith, J.M., and Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genetical Research* 23, 23–35.

Soufflet-Freslon, V., Araou, E., Jeauffre, J., Thouroude, T., Chastellier, A., Michel, G., Mikanagi, Y., Kawamura, K., Banfield, M., Oghina-Pavie, C., *et al.* (2021). Diversity and selection of the continuous-flowering gene, RoKSN, in rose. *Horticulture Research* 8, 76.

Stankowski, S., Chase, M.A., Fuiten, A.M., Rodrigues, M.F., Ralph, P.L., and Streisfeld, M.A. (2019). Widespread selection and gene flow shape the genomic landscape during a radiation of monkeyflowers. *PLOS Biology* 17, e3000391.

Wagner, S., Lagane, F., Seguin-Orlando, A., Schubert, M., Leroy, T., Guichoux, E., Chancerel, E., Bech-Hebelstrup, I., Bernard, V., Billard, C., *et al.* (2018). High-Throughput DNA sequencing of ancient wood. *Molecular Ecology* 27, 1138–1154.

Wang, L., Ji, Y., Hu, Y., Hu, H., Jia, X., Jiang, M., Zhang, X., Zhao, L., Zhang, Y., Jia, Y., *et al.* (2019). The architecture of intra-organism mutation rate variation in plants. *PLOS Biology* 17, e3000191.

Wolf, J.B.W., and Ellegren, H. (2017). Making sense of genomic islands of differentiation in light of speciation. *Nature Reviews Genetics* 18, 87–100.

Yardeni, G., Viruel, J., Paris, M., Hess, J., Groot Crego, C., de La Harpe, M., Rivera, N., Barfuss, M.H.J., Till, W., Guzmán-Jacob, V., *et al.* (2022). Taxon-specific or universal? Using target capture to study the evolutionary history of rapid radiations. *Molecular Ecology Resources* 22, 927–945.