

Disentangling of the ecotoxicological signal using "omics" analyses, a lesson from the survey of the impact of cyanobacterial proliferations on fishes

Benjamin Marie

▶ To cite this version:

Benjamin Marie. Disentangling of the ecotoxicological signal using "omics" analyses, a lesson from the survey of the impact of cyanobacterial proliferations on fishes. Science of the Total Environment, 2020, 736, pp.139701. 10.1016/j.scitotenv.2020.139701. hal-03065995

HAL Id: hal-03065995 https://hal.science/hal-03065995

Submitted on 15 Dec 2020

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.

Disentangling of the ecotoxicological signal using "omics" analyses, a lesson from the survey of the impacts of cyanobacterial proliferations on fishes

Benjamin Marie¹

¹ Muséum National d'Histoire Naturelle, UMR 7245, CNRS, MNHN Molécules de Communication et Adaptation des Micro-organismes (MCAM), équipe "Cyanobactéries, Cyanotoxines et Environnement", 12 rue Buffon - CP 39, 75231 Paris Cedex 05, France.

E-mail: benjamin.marie@mnhn.fr

Abstract

Omics technologies offer unprecedented perspectives for the rational investigation of complex systems. Indeed, omics present the capability of providing a profound insight into the biochemistry and physiology of the cell and any perturbing effects of xenobiotics through the joint investigation of thousands of molecular responses simultaneously; then it has led to an enthusiastic adoption by research ecotoxicologists. Beyond the presentation of recent advances, we have recently performed on the omics investigation of cyanobacterial deleterious effects on various fishes from both microcosm, mesocosm and field sampling experiments using transcriptomics, proteomics, metabolomics, multivariable statistic, and system biology tools and pipelines, the present prospective paper re-explores the promising perspectives and also the pitfalls of such holistic investigations of the ecotoxicological response of organisms for environmental assessment.

Keywords: metabolomics; proteomics; transcriptomics; data integration; fish ecotoxicology; cyanobacteria

1. Introduction

1.1 Context of the omics area for ecotoxicology

Different end-point techniques, ranging from toxicology to molecular tools (antioxidant enzymes, stress proteins, DNA physiological impairs damages, or neuroendocrine parameters), have been classically used attempting to provide early warnings of phenotypic alterations of organisms exposed to environmental stress or contaminants (Hook et al., 2014). However, during the last 20 years, the field of aquatic ecotoxicology has progressively integrated various emerging omics techniques including transcriptomics, proteomics, and metabolomics, in order to enlarge the investigation range of potential molecular biomarkers (Mezhoud & Edery, 2005). To this end, the use of model organisms appears first as a good strategy that can address some of the challenges associated with biological variation because their genomes are well characterized, supporting proteomics and transcriptomics investigations, and these organisms can be easily manipulated and tested within a controlled laboratory environment (Malécot et al., 2008).

However, despite the high potential of omics for the understanding of the molecular mechanisms implicated in the ecotoxicological responses of organisms, the investigation of new qualitative and quantitative biomarkers characterizing the interactions between natural populations and their biotopes in various ecological context remains still rare and difficult to achieve (Cappello et al., 2016). After an initial descriptive period of their technical skills and capabilities, and of proof-of-concept announcements (Ankley et al., 2010), ecotoxicological omics now gain at rationally integrate tool in a systemic approach addressing appropriately all conceptual and statistical constraints in order to hopefully reach their promising potential (Sauer et al., 2017; Martynuik 2018; Kim et al., 2017).

1.2 The ecotoxicological problematic of cyanobacterial proliferations

The joint effects of eutrophication and climate change promote cyanobacterial blooms in continental aquatic ecosystems, which pose potential risks to ecosystem sustainability (O'Niel et al., 2012). Besides, bloom-forming freshwater cyanobacteria produce a wide range of secondary metabolites potentially toxic, so-called cyanotoxins. These metabolic by-products are mainly stored intracellularly and can be released into the water notably after bloom collapses, leading to potential toxic effects to exposed Human populations as well as on aquatic organisms (Janssen 2019). Microcystis, Aphanizomenon, Anabaena. Cylindrospermopsis, and Planktothrix are the main bloomforming freshwater cyanobacterial genera found during summers in lentic water bodies, and all have been often reported to produce potent toxic compounds (Codd et al., 2015). Among them, the microcystins (MCs), a family of hepatotoxins consisting of more than 230 variants, are the most studied due to their high biological activity and their wide occurrence during freshwater cyanobacterial blooms (Catherine et al., 2017). The effects of freshwater cyanobacteria and their respective cyanotoxins, notably the MCs, have been widely studied on fish maintained in microcosm or mesocosm conditions, and fish have been proposed as valuable indicators of environmental disturbances associated to cyanobacteria proliferation (Malbrouck and Kestemont 2006; Le Manach et al., 2018).

However, the actual knowledge on the genuine cyanobacteria bloom impairs for natural populations of fishes is mainly deduced from short-time experimentation generally performed in micro- or mesocosms with high concentrations of purified toxins (Sotton et al., 2017; Malbrouck & Kestemont 2006; Pavagadhi & Balasubramanian 2013). Furthermore, most of these studies were focused on the mechanisms involved in the dynamic of the MC accumulation-detoxification. It appears that there is still a lack in our understanding of the real ecotoxicological effects of cyanobacterial biomass, particularly on the natural ichtyofauna population. Indeed, cyanobacterial blooms are producing at the same time a "cocktail" of potentially bioactive compounds, among them some cyanotoxins (Le Manach et al., 2018), and are also potentially modifying other important ecological parameters of the water bodies (Sotton et al., 2019). The main challenge that now face the ecotoxicology community that tries to investigate and to propose management strategies of such complex natural proliferation of cyanobacteria and various metabolite production is to be able to distinguish from individual variability of its various biological traits the subtle contribution of low stressor deleterious effects on its phenotypic expression.

For this purpose, omics technologies offer promising perspectives for rational investigation of complex systems. Indeed, omics technologies present the capability of providing a profound insight into the biochemistry and physiology of the cell and any perturbing effects of xenobiotics. They have recently led to an enthusiastic adoption by research ecotoxicologists (Gonzalez & Pierron, 2015).

2. The benefits of the global picture for ecotoxicological investigation

By assimilating concepts from toxicology and ecology, ecotoxicology aims at investigating the environmental effects of toxic substances, such as microbial toxins, on the health of ecosystems and their respective populations, species, and individuals. Many toxic substances can generate adverse effects at all levels of biological organization ranging from the molecular level to the communities and the ecosystems. To support the operational chemical toxicological assessment for contaminated environments, analytical supports such as those provide by approaches using either direct (source-to-outcome indirect (adverse outcome pathway) pathway) or methodologies have been developed; They allow describing cascade chains of causal events occurring at the different levels of biological organization inducing objective ecotoxicological effect. Perhaps the greatest challenge facing nowadays ecotoxicology is to determine how to reliably assess community and ecosystem impacts of chemical pollutants taking supported of the detailed individual investigation that can now be performed thank to the various omics and system biology tools that are currently been developed (Sauer et al., 2017).

Omics technologies have the potential to enable qualitative and quantitative measurement of early changes from the molecular levels that precede following changes at latter cellular, tissular, individual to community levels. It is because omics analyses present the ability to provide a global view of the cellular processes of an individual in response to an environmental change, that they are believed to represent promising opportunity to challenge the complexity of the various molecular variation occurring in an organism in response to a specific but sub-acute if not faint toxicantassociated stressor, in order to extract specific signature of the deleterious processes (Martynuik 2017).

Classical toxicology or ecotoxicology approaches usually focus on the alteration of a restricted number of pertinent biomarkers by histology, protein quantification or gene expression studies (Hook et al., 2014). However, these analytical strategies do not allow to either investigate or to globally characterize the toxicity of emergent pollutants such as a complex mixture of compounds produced by cyanobacteria. Nowadays, with the advances in highthroughput analytical methods, the complexity of such toxicant interactive effects could be further described without a priori hypothesis. The omics tools allow a rapid identification of potential effects through modification on gene or protein expression and regulation. The major advantage of using omics methods in a systemic approach is that one can assess hundreds to thousands of molecular responses simultaneously within an organism, facilitating a more holistic

understanding of the organism's physiological and toxicological status (Biales et al., 2016).

2.1 New hypothesis generation

In principle, omics are ideal tools for discovering key events within adverse outcome pathways (Ankley et al., 2010). The high dimensionality of omics data sets suggests the possibility of developing omics-based fingerprints for a chemical or activated biological pathway. These highthroughput high-output fingerprints have the potential to be applied to both exposure and hazard assessment in an unsupervised or non-targeted manner to simultaneously screen de novo all activated biological pathways, requiring no a priori information regarding the mode of action (Biales et al., 2016). Such a holistic approach is believed to be specifically appropriate in order to investigate the subtle biological signal induced by low/environmental conditions that could statistically be supported by the replication of the characterized variables. Then, it may represent a valuable tool for the proposition of new hypotheses and the generation of either an environmental diagnosis of stressing environmental conditions or the deconvolution of the organism responses to identified stressors for mechanistic toxicological purposes.

To this end, we investigate the global proteomic effect induced by environmental/low dose exposure to cyanobacterial extracts by 28-days balneation on the liver of adult medaka fish, in association with anatomopathological observation (Le Manach et al., 2016). This analysis clearly revealed a distinguishable sex-dependent response of medaka fish to the various hepatotoxic Microcystis treatments according to the group clustering, which was based on the pattern of relative abundance of up-regulated (blue) and downregulated (yellow) proteins (Figure 13A). Metabolism and homeostasis appeared to be the two most modified protein categories in both male and female medaka livers. Primarily, these observations support the idea that chronic exposure to MC-containing extracts induced a very different molecular response in medaka than the response induced by acute and short term treatments (Marie et al., 2012), and now need to be judiciously decorticated and interpreted.



Figure 13. The quantitative proteomic analysis of liver of the medaka exposed to different chronic exposure to hepatotoxic microcystins. Protein quantifications in male and female livers exposed to different cyanobacterial hepatotoxic treatments (CHT1-3) represented in a heatmap with hierarchical clustering (adapted from Qiao et al., 2016a) (A). Comparison of the lists of dysregulated proteins in male and female livers from two similar analysis performed using identical protocols and pipelines (Le Manach et al., 2016; Qiao et al., 2016a) (B).

More in detail, the cyanobacterial hepatotoxic treatments (comprising microcystins) dysregulate various proteins of various molecular function categories comprising lipid, amino acid, carbohydrate and TCA metabolisms and detoxification processes. The sex-specificity of the liver molecular response suggests that identical hepatic stress could impact these various molecular processes differently, potentially inducing dissimilar biological repercussions in the organisms of two sexes. In our example, male-enriched protein dysregulations concern rather TCA, steroid, fatty acid, amino acid, and vitamin B6-7 metabolism pathway categories, whereas female-enriched protein dysregulations are rather related to tRNA biosynthesis, amino acid, glutathione, xenobiotic and drug metabolism pathways. This example demonstrates the importance of sub-classifying and distinguishing results based on sex prior to interpretation of omics data on ecotoxicological issues induced by sub-acute and chronical contaminant exposures. Furthermore, recent reviews also confirm that sexspecific molecular responses are the norm with chemical exposures (Liang et al., 2017). It was reported that there was approximately less than 20% congruence between male and female proteomes following a chemical challenge; however, with increased sensitivity in omics technologies, there could likely be higher congruency in omics responses between the two sexes.

Also, measuring variability in molecular responses is necessary for identifying strong endpoints that one can measure confidently and consistently, notably across replicated experimentations and analyses performed within the same and also within other laboratories. Considering omics analyses, focussed investigations are still needed to address intra- and inter-laboratory reproducibility within the context of ecotoxicology before any robust implementation can occur. To this end, we undergo to compare proteomic results replicated from identical exposure protocol and analytical procedure generated in the same lab, with fishes for the respective parent and descendant generations (Le Manach et al., 2016; Qiao et al., 2016b). The comparison between the two studies of the dysregulated proteins (Figure 13B) reveals a similar hepatic alteration pattern under the same cyanotoxin treatments (namely, MC5 and Ext5). The proteins associated with amino acid metabolism, detoxification, nuclear receptor signaling, translation and transport being globally upregulated in both genders. Particularly, these data suggest that the cellular defense against oxidative stress and the immune response are activated by the MC containing treatments. On the other hand, the proteins involved in lipids and nucleotides metabolisms, protein modification and oviparous specific proteins appear similarly down-regulated within the two analysis.

However, in order to be able to generate a valuable interpretation of mechanistic toxicology from omics dataset,

specific attention should now be attempted on the representativeness of the differential picture of molecular dysregulation. Additionally, one should keep in mind that for technical and financial reasons of experimental design sustainability, such proteomic investigations are often performed on a limited set of samples that can also result of the pooling of different individuals from the same categories, that could distort and homogenate the accurate toxicological response experienced by the different individuals.

2.2 Mode of action and general ecotoxicological mechanisms

One of the main objectives for omics applied to environmental science is to identify biologically meaningful molecular features that describe the toxicological mode of action leading to negative impacts on individual fitness, and that can be used for the description of specific molecular signatures. Moreover, omics data can be specially used to refine potential adverse outcome pathways, and to unveil novel molecular events that potential can initiate undetermined outcomes linkable to ecologically relevant adverse phenotypes (*e.g.* reproduction, growth, energy, development, behavior). It is important to recognize that these molecular pathways are not linear, but are rather integrated and complex - perturbations at one point being potentially inducing collateral consequences for other integrated pathways (Conolly et al., 2017).

However, one should consider when observing the biological response of an organism to a stressor that it results from the combined action of adaptive processes, the organism modulating its physiology to better cope with environmental variation, and also the potential molecular dysregulation that have been induced by the direct toxicologic action of the toxicant itself. In order to be able to disentangle the joint effects of these two different processes, it is crucial to basically take into account the temporal dimension of the biological response.

One other key question in environmental monitoring concerns the recovery following mitigation of a chemical stressor. Recovery is defined as a return to a normal state of health. This underscores the necessity of first quantifying what is normal (often defined from an initial state, observed just preceding a specific stress occurrence) in the context of an omics profile before we can discuss how to use molecular profiles to measure a return to an initial state. In a recent study, we have investigated by LC-MS based untargeted metabolomics approach the molecular effects occurring on the liver of medaka fish exposed to sub-acute dose of anatoxin-a (Colas & Marie, *in prep.*). This analysis reveals that this molecule was rapidly excreted by the organisms and that the molecular dysregulation of the liver metabolite content also rapidly rescues to its initial state, as the level of all dysregulated metabolites recover their respective initial levels, whatever they were up- or down-regulated (Figure 14). The question remains to know to which extent it can be considered as a full "recovery" and returned to the initial un-stressed states (Marjan et al., 2017b).



Figure 14. Monitoring of global effects induced by a single sub-lethal dose of anatoxin-a on the medaka fish liver investigated by non-targeted metabolomics. Principal component analysis individual plot representation according to components 1-2 (A) and 1-3 (B). Heatmap and hierarchical clustering of significantly dysregulated metabolites (ANOVA P<0.01) and representative blox-plots of metabolite variation and recovery after the first post-exposure hours (adapted Colas & Marie *in prep.*).

A take-home lesson of these considerations could be that toxicological baseline (using classical cellular markers of toxic evidence) and time-series dataset (to discriminate direct, indirect or collateral effects) should be required prior to determining whether a molecular response is adaptive, compensatory, homeostatic or toxicologic. Although there are no specific guidelines concerning the appropriate duration of exposure to measure a specific omics response, it is also critical to match the exposure time to the question or hypothesis that is most relevant in the environment. For example, some long-term exposures to low dose of contaminant cocktails, expected to be most relevant for environmental and realistic exposure that occurs in natural environments, have shown that it can be then very difficult to predict effects of chronic exposure from mechanistic knowledge on contaminant toxicological mechanisms because of the dynamic and the complexity of the organism toxicological response (Hamilton et al., 2016). On the contrary, chronic exposures at low concentrations experienced by wild fish to some chemicals, such as the hepatotoxin microcystins, have been found to adversely impact secondary toxicological targets such as reproductive output, suggesting harmful effects on wild fish populations may then occurs (Qiao et al., 2016b), when higher and shorter-term exposure could only describe more direct effects on liver (Marie et al., 2012).

3. Integration of system biology into the ecotoxicological view

3.1 Multi-omics integration using system biology concepts

On one hand, transcriptomics analyses using RNASeq in ecotoxicology have raised nearly as many questions as they have answered, not only due to the complexity of the aquatic animals analyzed but also to the complexity of the chemicals and mixtures to which they are exposed (Bertucci et al., 2018). On the other hand, due to their high technical complexity, proteomics-based studies remain still restrained in the ecotoxicological field, even if some studies have demonstrated the utility of proteomics in ecological species (Degli-Esposti et al., 2019). Proteins can also undergo post-translational modifications, complexifying, even more, the functional interpretation of the proteomic results. In theory, the joint use of transcriptomics and proteomics might provide information that supported and strengthened each approach, in addition to identifying effects that each technique used in isolation may not detect.

However, the comparison of the molecular picture that they both provide, respectively, at the mRNA and protein level might also highlight the difference that can be observed between these two biological compartments (the second resulting from the translation of the first one), especially when observed from a single instantaneous picture. Indeed, in a previous analysis, we had investigated the hepatic effects induced by microcystin balneation of adult medaka fish (Qiao et al., 2016b). When analyzing the global pictures obtained from transcriptomic and proteomic investigations, we used the Ingenuity Pathway Analysis tools in order to globally integrate and compare these two large data sources (Figure 15). Surprisingly, the comparison of the dysregulated gene and protein entries display a limited overlap between these two analyses (only 13 out of a total of 225 proteins or 654 genes are in common). However, noticeable compatibility has been observed between the transcriptome analysis and the proteome investigation at this molecular and cellular function level. Furthermore, transcriptome data even shows a clearer and more notable cellular response than that depicted by proteomes, which is highlighted by a clear up and downregulation between some opposing function aspects, such as cell death and cell survival, illustrating the difference depicted with these two molecular levels of analysis.



Figure 15. Heatmap representation of the overall alteration of the hepatic transcriptomes with gene ontology classification (A), of the hepatic proteomes with gene ontology classification (B) and of the significantly affected cellular and molecular functions determined with transcriptome and proteome data through IPA (adapted from Qiao et al., 2016b).

One of the current limitations in transcriptomics studies is the extrapolation of the results to the higher levels ranging from proteins to the individual. Some studies reported a positive correlation between gene expression and cellular modifications (Ortega et al., 2015), but no correspondence on gene versus individual levels could have been found. In addition, gene expression may not always reflect a physiological or morphological effect (Grifftt et al., 2012).

Nowadays, metabolomics studies have become a relevant approach to describe and analyze the integrated response of the organisms under specific environmental context and scenario (Capello et al., 2016; Gil-Solsona et al., 2017; Tang et al., 2017; Reverter et al., 2017). Also, metabolomics directly addresses the terminal molecular level of the central concept of cellular biology, the substrate. Then, the changes in the primary metabolite concentrations help to provide valuable and direct information concerning the physiological processes involved in the homeostatic or dysregulated responses of the organisms encountering environmental stresses from potential multiple origins. For those reasons, various ecotoxicological studies now integrate metabolomics analyses in their standard omics pipelines, as it appears remarkably relevant to address molecular and functional impairs induced by stressors (Wei et al. 2018).

One way to address biological response to ecotoxicological pressures is to focus on molecular functions related to the observed system changes. Functional datasets tend to be based on qualitative observation and usually contain discrete values and count of a significantly discriminated dataset. The fluctuations in the expression and the presence of non-detects become less relevant and do not drastically affect the interpretation of data using a function dependent approach of system biology. One of the powerful principles of systems biology is that the presence of a disease or a chemical exposure is considered in an organism as a perturbation that affects the molecular network of the functional interaction protein/genes and metabolites within relevant pathways. The comparison of the condition of a normal and a perturbed network provides better molecular knowledge on the physiological condition of an organism and on the toxicological and homeostatic processes induced by the stressor. While systems biology is being broadly used in the human health field, ecotoxicologists are just beginning to use such an approach.



Figure 16. Example of Ingenuity Pathway Analysis performed on female medaka fish exposed to microcystin-producing strain conditions (Mcy) using both proteomics and metabolomics data. Heatmap (A) and top dysregulated molecular pathways, bar chart (B) and significantly dysregulated molecular networks (C). Molecular functions that are specifically down- and up-regulated are indicated in blue and orange, respectively (adapted from Le Manach et al., 2018).

Indeed, different bioinformatic tools, such as Ingenuity Pathway Analysis (Qiagen®), have been developed in order

to over-pass the difficulty of integrating and be able to visualize and to analyse the complex dataset generated from

(multi-)omics approaches. These solutions based on the interrogation of a large, if not exhaustive, corpus of genetic, protein and molecular pathways generated from all available scientific/biomedical knowledge from Human, rat and mouse models, can be attemptedly used for other non-model organisms, such as fishes, considering that most of the homologous proteins may have similar function and relationship between those organisms. This approach proposes a functional integration of genes/proteins and molecules that have been determined in parallel to be significantly dysregulated between tested conditions but does not allow statistical investigation of big data set from a large cohort of organisms (together with its microbiote) considered individually. However, once these characteristics are well defined computationally, they still require experimental validation.

Recently, we performed an integrative ecotoxicological approach on chronic effect induced by various cyanobacteria strain exposures to adult medaka fish using combined anatomopathology, proteomics and metabolomics analyses (Le Manach et al., 2018). To investigate and visualize the biological connectivity of the cyanobacteria stress-enriched metabolites and proteins, the network-generating algorithm of ingenuity pathway analysis (IPA) was used to maximize the interconnectedness of molecules based on all known connectivity in the database developed from Human molecular knowledge in the liver. The results of the IPA biological function analysis for female medaka exposed to Mcy conditions (microcystin-producing strain), represented as a heatmap and a bar chart (Figure 16 A-B, respectively). The IPA network search shows within the top networks predominantly regulated the respective interaction of related proteins and metabolites (Figure 16C).

Ultimately, a system toxicology approach that has functional perspective gains in biological relevance and meaning. Although it is based on functional analogy deduced from molecular similarities with model organism from which derived molecular pathway corpus, most omics datasets obtained from non-model organism aim at being investigated using system biology tools for putative functional information and explicative molecular scenario, that will remain to be further explored.

However, one should also keep in mind that the demonstration of genuine toxicologic effects ultimately can not rely on the sole observation of molecular dysregulation of molecules or genes, even if they are numerous and known to be potentially involved in classical toxicologic response, and that obvious evidence of cellular or characteristic endpoints may support the ultimate toxicological diagnostic. Thus, phenotypic anchoring demonstrates that changes in molecular response highlighted by omics aim at being correlated with physiological outcomes characterized by orthogonal analyses.

3.2 Experiment, field sampling and batch integration

One shortfall of various omics studies, in general, has been the lack of consideration of statistical power in the analysis design in order to ensure appropriate statistical power to test hypotheses. Although proteomics and transciptomics can generate quantitative observations on a very large set of variables (protein quantity or gene expression levels), their application remains relatively complicated and expensive, that the number of replicates concretely limited of ecotoxicological analyses. Oberg and Vitek (2009) address the various statistical issues encountered with the principal designs of quantitative omics experiments. They discuss the benefits and the drawback of each scenario, including effective, randomization, replication, or pooling, and the appropriate statistical test that can be performed in each case to identify differences between samples. For example, a statistical power analysis performed on data obtained from label-free proteomics determined that sample size should be of a minimum of 9-12 individuals in order to ensure to detect 25% changes in protein quantity (Simmons & Sherry 2013). This observation demonstrates that a trade-off must occur if the number of experimental groups is increased, such as multiple doses and time points, and selecting replicate numbers so that the experiment remains cost-effective and maintains adequate statistical power.

The large quantity of data produced from omics approaches has rapidly necessitated the use of adapted data analysis tools for the classification of sample groups. Among the most common are multivariate statistical models, such as principle component, partial least squares discriminant, redundancy or canonical correlation analyses (PCA, PLS-DA, RDA, CCA, ...) (Hervé et al., 2018). Recent effort has now been performed in order to develop new multi-omics tools aiming at quantitatively integrating multi-block omics dataset from a homogeneous or heterogeneous origin (Liquet et al, 2012; Gonzalez-Ruiz et al., 2019; Rohart et al., 2017). The application of multivariate analysis should be performed after that data have been appropriately processed in order to match identifiable biological targets as the variable loadings onto the components reflect true biological variation and not technical artefacts that could, for example, arise from batch-to-batch variation.

Indeed, among the different situation of multi-block analyses, the batch comparison is among the most currently investigated one (Mehl et al., 2015; Boccard et al., 2019). Indeed, because it is difficult to control exposures in the field, laboratory experiments under more controlled conditions are often developed in ecotoxicology in order to make inferences regarding what might happen under similar conditions in the field. Most often, laboratory exposures are used to ascribe specific chemical exposures with health outcomes to control for confounding factors and provide levels of replication that are difficult to obtain with wild fish. This concept relies on the postulate that lab experiment performed under controlled conditions with well-known organisms might present a very low replication variance that ensures the repeatability of the observation and its transferability to less constrained systems (from mesocosms to natural ecosystems). However, true dataset often shows that even very similar analyses present remarkably high batch effects and that the use of dedicated approach and data treatment is pre-requisite in order to be able to distinguish the portion of the variable variance that is relative to batch effects or to the specific stressor treatment. In pilot analyses on chronic effects on medaka induced by *Microcystis* bloom balneation exposure (Le Manach et al., 2018; Sotton et al., 2017a), we had explored the liver metabolomics variations within these two similar but distinct experiments. Now, we attempt to integrate and to compare the ¹H-NMR metabolomics results obtained from each one of these two analyses using either the MINT function or regularized canonical correlation analysis of the MixOmics *R* package (Lê Cao et al., 2018).



Figure 17. Illustration of MINT analysis performed in mixOmics for the integration of homologous dataset from independent batch experiments from Sotton *et al.* (2017a) (BS) and Le Manach *et al.* (2018) (SM). A) MINT sPLS-DA graphical outputs using *plotIndiv* function, from SB and SM metabolomics analysis of livers from medaka exposed to similar *Microcystis* bloom constraints. B) Global sPLS-DA sample integrating metabolomics data from both analyses. C) Coefficient weight of the features selected on component 1 in each study using *plotLoadings* function. D) Global coefficient weight of the features selected on component 1 integrating metabolomics data from both analyses. Colours indicate the class with a maximal mean expression value for each metabolite (adapted from Sotton et al., 2017a; Le Manach et al., 2018).

MINT is an extension of the multi-group PLS framework (Rohart et al., 2016a), where 'groups' represent the two independent studies in a supervised framework with variable/feature selection. MINT seeks to identify a common projection space for all studies that are defined on a small subset of discriminative variables and that display analogous discrimination of the samples across the two studies. Alike to sPLS-DA, MINT selects a combination of features on each PLS-component. The function *plotIndiv* and its representation graphics can act as a quality control step to detect studies that cluster outcome classes differently to other studies and the *plotLoadings* graphic display the coefficient weights of the features that are co-selected by the model. Figure 17 displays some outputs easily obtained by calls to those functions, including MINT sPLS-DA sample and variable contribution plots for both individual batches and combined dataset, highlighting the most discriminant samples and variables, respectively, according to the whole dataset series. Such data analysis tool offers integrative approach for disentangling the informative variance from the various parameters that may influence omics responses of tested organisms.

Similarly, an alternative approach consists in analyzing such multi-batch dataset with regularized canonical correlation analysis (rCCA), also available from the MixOmics R package (Gonzalez et al., 2008). This multivariate exploratory approach aims to highlight correlations between two data sets acquired using the same experimental pipeline. In the same vein as PCA, CCA seeks for linear combinations of the variables (called canonical variables) to reduce the dimensions of the data sets, but this time while trying to maximize the correlation between the two variables (the canonical correlation). One objective of such analysis is in a set of experiments to highlight the variables that are correlated with specific parameters without interfering with experiment reproducibility or batch variation.



Figure 18. ¹H-NMR liver metabolomes and discriminant metabolites of fish exposed to MC-producing and non-producing cyanobacterial strains from BS and SM experimentations. The individual plot of regularized canonical correlations analysis (rCCA) for dimensions 1-2 (A), dimensions 3-4 (B) and dimensions 5-6 (C) are illustrating the individual discrimination according to the experiment (axe 1) and the sex (axe 2), the exposure to N-mcy and Mcy strains in BS experiment (axe 3 and 4, respectively), and the exposure to N-mcy and Mcy strains in SM experiment (axe 5 and 6, respectively). Relevance networks providing from rCCA analysis on dimension 3 and 5 (D, right and left, respectively) show the variables/metabolites that are the most correlated with the Mcy condition (exposure to microcystin producing *Microcystis* strains) in SB and SM experiments, respectively (adapted Sotton et al., 2017a; Le Manach et al., 2018).

Using the rCCA approach, samples can be neatly separated according to the *cyanobacteria strains exposure* factor, avoiding the interference due to the variability between different experiments. This approach can serve to extract the variables that are the most correlated with the different parameters tested and project their respective variances on the different components that are the most correlated with latent parameters. In our comparison of the metabolomic dataset generated from the ecotoxicological investigation of *Microcystis* effects on medaka fish livers from BS and SM experiments, we have also experienced rCCA analyses in order to extract variables that are the most correlated with experimental parameters (cyanobacterial exposure with MC producing and non-producing *Microcystis* strains) within the

two batches (Figure 18). This analysis allows the specific extraction of the variables of the X matrix (comprising the metabolite semi-quantification) that present the best correlation with interrogated parameters integrated into the Y matrix (composed by experimental descriptors and explicative metadatasets).

These two types of multivariable integrative data analyses appear to be efficient to simultaneously extract from two joint datasets, the variables that are the most correlated with toxinogenous cyanobacterial strains, and then allows stress signature determination, constituting promising tools for the investigation of complex data from environmental omic evaluations.

4. Multi-omics stress signature investigation from multi-species analyses

As ecotoxicology aims at characterizing the toxicological condition and processes of organisms within their natural environments, most of the ecotoxicological study concerns various species of non-model organisms according to the authenticity of their natural exposure to environmental contaminants. Then, concerning the search for toxicological mode of action and for relevant biomarkers, one great challenge of ecotoxicology consists to find a relevant molecular signature of specific stressors for these non-model, but relevant, organisms. To this end, metabolomics appears to present all the advantages to reach such objective. Indeed, it is a powerful genome independent approach that can generate the same quality and quantity of data for all organisms (presenting sufficient amount of material to be analysed), that can be equally considered as *good* models for environmental metabolomics investigations.



Figure 19. ¹H-NMR liver metabolomes and relevance network offish exposed to MC-producing and non-producing cyanobacterial strains. The individual plots of regularized canonical correlations analysis (rCCA) for dimensions 1–2 (A), dimensions 2–3 (B) and dimensions 3–4 (C). Perch-control (green diamond), perch-N-mcy (blue diamond), perch-Mcy (red diamond), crucian carp-control (green triangle), crucian carp-N-mcy (blue triangle), crucian carp-Mcy (red triangle), roach-control (green circle), roach-N-mcy (blue circle), roach Mcy (red circle). Relevance network providing from rCCA analysis on dimension 3 and 4 (D). Circles indicate metabolites are significantly dysregulated in exposed fish according to additional two-ways ANOVAs (adapted from Sotton et al., 2017b).

In order to investigate the global response of fish to punctual cyanobacterial event, three representative fish species (the roach, the crucian carp and the common perch) of freshwater ponds from the European temperate regions (frequently encountering cyanobacterial blooms), were 96-h environmental high concentrations exposed to of cyanobacteria in the context of an experimental approach in mesocosms designed to mimic natural conditions. Metabolomic analyses were performed on the fish liver in order to investigate the global molecular responses of the different fish species, and potentially identify a general metabolomic signature of cyanobacterial exposure in fishes (Sotton et al., 2017b). All metabolomics data from the three species were integrated and investigated together with rCCA (Figure 19). When canonical components 1 and 2 clearly discriminate the metabolome of the different species whatever the treatment considered (the species being the main driver of the metabolome difference between the individuals), the dimension 3 and 4 firmly discriminate the 3 experimental conditions (control, microcystin and non-microcystin producing strain exposure). The relevance network based on the third and the fourth rCCA dimensions identifies metabolites discriminating the different treatment groups (Figure 19D). Strong positive correlations of Mcy-treated fish group with the relative concentrations of the metabolites annotated as hypotaurine, 1-methylhistidine and gamma-

glutamylcysteine are then highlighted by the relevant network. Although this study shows that rapid and significant metabolic changes occur in 3 representative fish species exposed senescent cyanobacterial strains, it also demonstrates the great capability of metabolomics to investigate non-model organism metabolome and, when combined with rCCA, extracts the metabolite common changes induced by the exposure conditions, beyond the large difference of the respective metabo-type of each three species. Thus, it appears that both cyanobacteria induce molecular changes in a straindependent way. According to the respective difference of metabolite production between the two strains (Mcy and N-Mcy), it is not surprising to observe distinct fish metabo-types resulting from the different cyanobacterial treatments. These metabolome changes are very likely linked to the stress response of fish in order to overcome the negative consequences of the different cyanobacterial secondary metabolites absorbed by the fish cells, and notably, the potent hepatotoxins MCs as the producing strain induces even more molecular changes.

Considering these first observations, one can now suppose that in a natural ecosystem, after a bloom exposure, fishes may present some characteristic metabolic signatures, which could be reversely used as potential bio-indicators of the ecological constraints induced by the presence of specific cyanobacterial genera and/or secondary metabolites they produce. This hypothesis has been recently investigated using the metabolomes of fish sentinel species collected from a field sampling campaign within a gradient of cyanobacteria proliferation (Sotton et al., 2019). In this way, during the summer 2015, young fish of two representative and common species of freshwater lakes from the European temperate regions, the common perch (Perca fluviatilis) and the pumpkin-seed sunfish (Lepomis gibbosus), have been sampled in eight peri-urban lakes of the Île-de-France region contrasted by their phyto-planktonic community composition ("presence" or "absence" of cyanobacterial blooms). ¹H-NMR metabolomics analyses were performed on the fish liver in order to investigate the global metabolome local specificities of the two fish species collected from a gradient of distinct ecological contexts (comprising the cyanobacteria dominance) and to further identify the metabolic signatures related to these potential specific phenotypic responses using joint rCCA, MANOVAs and two-ways ANOVAs, as previously described (Figure 20).



Figure 20. ¹H-NMR liver metabolomes and relevance network of fish sampled in the different lakes. The individual plots of regularized canonical correlations analysis (rCCA) for dimensions 1 and 2 (A). Perch individuals are on the left side and pumpkinseed individuals on the right side of the graphic. Lakes are represented by their respective letters that in blue correspond to control lakes and in green to perturbed lakes. Relevance network providing from rCCA analysis on the dimension 2 (B). Green edges correspond to negative correlations with the discriminant ecological factors (adapted from Sotton et al., 2019).

A net species effect is observable through the dimension 1 with *Perca* and *Lepomis* clearly separated by the first dimension whatever the lake considered. However, on the dimension 2, a large effect correlated with cyanobacterial biomass is observable, as the metabolomes of fishes coming from cyanobacterial dominated lakes (in green) are clearly separated by this dimension whatever the fish species considered. The relevance network based on the second rCCA dimension specifically highlights metabolites discriminating the fish from dominated and non-dominated cyanobacteria lakes and linking to the correlated environmental factors (mainly the cyanobacteria concentration/BBEcya concentrations but also the pH) (Figure 20B). Overall, it appears that all these metabolites shown by this network

exhibit negative correlation with both cyanobacterial concentrations and pH values. Globally, it appears that the metabolome of livers of fishes sampled in Triel, Fontenay and Verneuil lakes, and in Varenne-sur-Seine lake in a lesser extent, exhibit the most significant differences in the relative concentrations of the highlighted metabolites compared to fish coming from other investigated lakes, where cyanobacteria are notably in low or not dominant proportions. Further studies need now to be performed in order to disentangle the specific and/or synergic effects of the various bioactive metabolite production together with those of other physicochemical parameters associated with cyanobacterial blooms such as the pH increased.

Interestingly, the fish collected from the T lake exhibits one of the most divergent global metabo-types, suggesting that it could potentially be influenced by the occurrence of multistress conditions, comprising cyanobacteria proliferation together with the other heavy metal, hydrocarbon or other organic contaminants, that have been observed in this specific environments (Catherine et al., 2016). Indeed, T lake was not characterized by the highest concentration of cyanobacteria observed during this study. However, due to the presence of other pollutants already monitored in past studies in this pound and not in the other one, this observation suggests that additive and/or synergistic effects of multi-pollutants together with cyanobacterial bloom seem to be involved in similar metabolic variations than those of fish from pounds which are the most stressed by cyanobacterial blooms.

Another question addressed by Sotton et al. (2019) consists to determine which descriptive factors are also relevant for environmental monitoring and which of these variables matter more for the fish metabolome determinism - is genetic selection, developmental state, season or connectivity other explicative factors also affecting the metabolome local signature specificity and patterns observed between the different lakes?

5. The (metabol)omics signatures as bio-indicator of environmental conditions?

Although data obtained from transcriptomics and proteomics approaches are still highly comparable, <u>metabolomics methods currently appear to be the most</u> <u>consistent between laboratories and instruments, and maybe</u> <u>gain at being the new standard for environmental omics</u>, especially for non-model organisms. Indeed, of all the molecular entities (genes, transcripts, proteins, metabolites), metabolites have the closest relationship to expressed phenotype as they are the final end-points of upstream biochemical processing. Although metabolite production is the results of genetically controlled processes, the metabolomics pipeline is completely independent of a background genomic dataset, and for this reason, metabolomics investigations produce the same quality and quantity of data for all organisms, whatever it is a model organism, or not. Quantitative analysis of metabolite abundance reflects both cellular processes and xenobiotic occurrence, this latter being physicochemically distinct from the molecular entities that originate in the host. During the last years, various innovative metabolomics investigations have been developed in order to explore original bio-indicative compartment using both invasive and non-invasive analysis of endogenous and exogenous chemicals (Bouslimani et al., 2016; Davis et al., 2016).

The selection of sites, species, organisms, development stages, biofluids or tissue types is a crucial aspect of experimental design in ecotoxicology. Further effort is still needed to obtain baseline datasets in ecologically relevant species and integrative biological compartments:

- Fishes being widely considered as relevant bioindicators of human impact on aquatic ecosystems (Clavel et al. 2013), various studies have demonstrated relationships between abiotic characteristics and the diversity and abundance of fishes assemblages (Zhao et al. 2016). A step further, long-life organisms such as fishes exhibit numerous relevant molecular characters that could be investigated by omic molecular approaches in order to perform more accurate longitudinal follow-up. These aspects make that fishes are among the most currently monitored organisms for environmental assessment and ecotoxicology.

- Although other tissues or species may respond differently, metabolomic-based analyses of livers may provide considerable insight into a variety of contaminants and their associated impacts on biological processes. During our previous analyses on hepatotoxicity of cyanobacterial blooms using fish metabolomics, because we focused on endogenous metabolites in livers, assessments of the biological importance of explicative factors (including cyanobacteria dominance and cyanotoxin occurrence) were restricted to those contaminants affecting liver processes. Because it is an important site for biotransformation of exogenous contaminants, the liver can be susceptible to various cyanotoxin/contaminant highly exposure; many of those contaminants being hepatotoxins (Codd et al., 2005). It also performs essential roles in many biological processes (e.g., immune response, energy metabolism, and hormone production); thus, perturbations affecting other organs can also alter liver structure and function, making liver many the most integrative organ of ecotoxicological processes. However, application of similar approaches to other tissues (e.g., mucus) or biological measurements (e.g., transcriptomics) could further highlight additional biologically active metabolites and provide a more complete assessment of respective potential toxicity mechanisms.

- Gut microbial communities constitute a compartment of crucial importance in the regulation of homeostasis of multiple host physiological functions as well as in resistance towards environmental pollutants. Many chemical contaminants were shown to constitute a major threat to gut bacteria. Changes in the gut microbiome could lead to alteration of host health, but can also represent indirect evidence of environmental dysregulation due to a toxicological pressure increase (Evariste et al., 2019). To this end, the democratisation of high-throughput sequencing, that permitted a recent great expansion of microbiome importance in human health research. may now support the development of ecotoxicological studies integrating the characterization of the free or associated microbial community, that may significantly contribute to the better understanding of relevant toxicological mechanisms for environmental assessment (Astudillo-Garcia et al., 2019). One should estimate that such microbiome analysis may rapidly represent the next key reference end-point of near-future ecotoxicological studies.

- On the other side, the relevance and appropriateness of the stressor gradient within a larger set of reference sites are of critical importance in field-sampling-based omics studies, as it directly influence the interpretation of omics data specificities observed locally. Indeed, the difficulties in selecting a unique set of reference sites stem from the reality that virtually all ecosystems are impacted by anthropogenic stressors and are part of a stressor gradient that can be determined for all potential contaminant (Jarque et al., 2015); the identification of such contaminant pressure on respective environments and the minimization of other confounding chemical influences at the less impacted "reference sites" remains primordial in the initial analysis design. Some guidance regarding experimental design in field-based omics has been published recently (Marjan et al., 2017a) and the insight generated from these studies should be considered further when designing experiments in the field.

6. Conceptual and technical challenges

6.1 Confounding/explicative factors

Linking exposure levels to impacts on fish populations and biomarker responses require establishing associations with health outcomes and all ecologically relevant factors. In ecotoxicology the variability that is a result of environmental changes can be an important aspect of, or even the focus of, a study, the investigated biotic or abiotic factor been then rather considered as explicative or confounding:

- <u>One important consideration was to recognize the role of an</u> <u>individual's sex in omics-based assessments</u> (Le Manach et al. 2016; Qiao et al., 2016). Pina and Barata (2011) pointed out that careful attention was needed in the interpretation of such data, and omics profiles at different developmental and reproductive stages needed to be further elucidated and defined for ecologically-relevant non-model species. As we move from chemical toxicity testing in the laboratory into field applications, we quickly realize that the complexity and diversity of responses increase dramatically. The individual's health status, diet, and reproductive status are important biological variables that can influence an omics response.

- In terms of environmental investigations, there have been efforts to characterize the variability of omics in fish, with the goal of identifying the influence of age or specific times of the year (or specific stages of reproduction) to omics phenotypes (Ivanisevic et al., 2016). Dreier et al. (2016) assessed transcript variability at different stages maturity in largemouth bass and identified the reproductive stage that showed the lowest variability in the transcriptome, the highest variability being observed when the fish were undergoing atresia. Regardless of the biological reason, these types of studies demonstrate that an individual's reproductive statue can significantly influence the phenotypic variability, and it may be most useful, if not strongly mandatory, to conduct studies on fishes of the same age class, of the same sex and that are not reproductively active (such as YOY juvenile fish), in order to lower the individual phenotypic variability that rises as a potential experimental issue.

- Additionally, if susceptibility to pollutants has a genetic basis this could provide the potential for evolutionary adaptation via the selection of less susceptible individuals. Indeed, in a contaminant-exposed population, fish with a resistance allele may have greater reproductive success, increasing within few generations the frequency of such allele in that specific allele in this population (known as local genetic adaptation). Furthermore, chemicals that alter the contributions of individual parental fish to subsequent generations could alter the genetic composition of a population and/or reduce the genetic diversity of populations, without necessarily affecting population size. In this context, one should also investigate the effect of such a selective process on the metabolome determinism using genetic diversity markers as metadata in order to test its influence on local phenotypic (metabol)omics signatures. In parallel, knowledge of population-genetic structure can then inform on how populations are exposed to contaminants and at the local scale which changes in population genetic diversity are likely to occur.

- Most, if not all, <u>organisms shelter a large and diverse number</u> of micro-organisms that provide them significant benefits from symbiotic relationships. Thereby, they can contribute to traits that the hosts have not evolved on their own, such as the synthesis of essential amino acids and vitamins or process otherwise indigestible components in the diet (Gill et al., 2006). Microbial symbionts, from intestines or other epithelia such as epidermis (Reverter et al., 2017b), can also promote efficiently the stress resistance mechanisms or contribute to detoxification processes (Gressel 2018; Ceja-Navarro et al., 2015). However, the taxonomical and functional diversity and composition of the microbiome might also critically influence the capability and the efficiency of the microbial community to positively or aversively contribute to their host resistance to environmental stress or toxicity (Evariste et al., 2019; Li et al., 2019; Rolli et al., 2015). In a pilot experiment, we had observed the significant alteration that a balneation exposure to *Microcystis* extracts can induce on the gut microbiome of the medaka fish (Duperron et al., 2019). Taken together, these considerations suggest that the fish microbiome may contribute to the host resistance to cyanobacterial proliferation-associated stresses. This contribution may also locally vary according to the specificity of the microbiome composition, and aim at being judiciously appreciated and considered as potential explicative/confounding factor.

Apart from biology, there also remains a number of abiotic factors that can influence the individual omics responses. For example, hydraulics, water quality, the physical structure of the habitat, water dynamic, and other variables more obvious such as temperature and dissolved oxygen are expected to affect in some way the physiological and biological parameters of organisms. Recognizing that such abiotic factors other than investigated contaminants (e.g., cyanotoxins) may drive the omics response of individuals from natural environments; combining landscape features (as metadata) with omics characters may prove useful moving forward for the convincing demonstration of explicative variables determining local omics signatures.

6.2 Bringing multi-omics at the organism for the development of an individual (eco)toxicology approach

Such as precision medicine and single-cell, two modern and disruptive concepts that both take into account individual variability and population characteristics to provide efficient personalized care and understanding (Collins and Varmus 2015; Patel et al., 2014), one could expect that a "precision ecotoxicology" approach aims at being now undertaken in order to widen our ecotoxicological knowledge. It may significantly contribute to explore the great diversity of individual toxicological response, and making possible to better understand their causal consequences at the population and community levels. It may rely on both biological individuality and population knowledge to provide a customized assessment of an ecotoxicological situation.

One of the goals of such precision ecotoxicology would be to use the ever-growing understanding of individual biology processes for a representative set of population sentinels to provide an accurate and tailored diagnosis. This strategy includes the use of decision-making processes supported by artificial intelligent tools based on the reductionist biomarker/signature extraction from holistic omics-driven investigations of the bio-indicative compartments. To this end, machine learning and multi-omics technologies revolutionize the way we acquire and process data. Such artificial intelligence algorithms dissect the big data corpus to learn their structure and associations within, often without the need of specific knowledge on processes that generates them, and allow an unprecedented system-level view of any organism. Using pertinent environmental and toxicological descriptors integrated as metadata, this innovative approach may rapidly support the development of operational tools for environmental diagnostic (Kim & Tagkopoulos 2018). However, there is also still a clear need for methods and disruptive concept for processing, normalizing and integrating the volume of heterogeneous multi-omics data to a cohesive corpus that can be used as a training grounds from further analyses.

6.3 Recent expansion of epigenomics for ecotoxicology

Epigenomics constitute the next frontier for understanding how mechanisms of temporal and spatial control of gene activity (through DNA methylation and histone modification processes, for example) operate during adaptive responses to external stimuli. Most importantly, since epigenetic marks constitute dynamic and potentially reversible modifications, they represent outstanding candidates for developing fast and sensible bio-monitoring programs in a diverse environmental field such as ecotoxicology (Asselman et al., 2016). Nevertheless, the lack of epigenetic knowledge in ecologically and environmentally relevant organisms has hampered the application of these tools in a broader range of ecology. Fortunately, that scenario is now changing thanks to the growing availability of complete reference genome sequences along with the development and democratisation of highthroughput DNA sequencing and adapted bioinformatic pipelines.

6.4 Clonal and toxicological diversity hidden in each singular bloom

One key specificity of the ecotoxicological investigation of cyanobacterial blooms consists in the fact that these biological vectors of toxicity can produce a very large set of bio-active components, a sort of which being potent toxicants. As the second side of the same coin, the ecotoxicological evaluation of cyanobacterial proliferation event may basically begin by the parallel evaluation of the chemical diversity of the produced metabolites (both in term of family of components and of chemical variants within the same family), in order be able to identify the chemical source of the observed adverse effects on exposed organisms. The synthesis of these various bio-active components is orchestrated under genomic control, and the genotypes of the cyanobacterial cells that constitute a natural bloom being often hyper-variable, the production of some specific cyanotoxin can not be deduced from a simple taxa identification and require an expert characterization by genomics, metabolomics, toxicology or both approaches (Kim Tiam et al., 2019).

So far, the consensual list of potent cyanotoxins (Janssen 2019) comprises a few molecular families, including microcystins, cylindropsermopsins, anatoxins-a, anatoxinsa(s) and saxitoxins (for freshwater cyanobacteria), but, other bio-active compounds families, such as anabaenopeptins or cyanopeptolins, are suspected to be also very toxic and represent potential emerging toxins, that could soon enrich the cyanotoxin list. However, for some technical restrictions (such as the availability of reference material and purified molecules) the toxicological status of these compounds remains an open debate (Schwarzenberger et al., 2010; Lenz et al., 2019). In addition, in our hand the toxicological investigation of different chemotypes of cyanobacteria clonal cultures shows that some strains that do not produce any known cyanotoxins, but that produce many other various metabolites, can induce as important metabolite dysregulation than a microcystin-producing strain on medaka chronically exposed by balneation (Le Manach et al. 2018). This observation various still-unidentified suggests that components produced by cyanobacteria may constitute a serious environmental hazard that remains to be characterized.

6.5 Need for technical platforms and operational pipelines for the use of omics in environmental monitoring

The ability to generate high-throughput omics data and to build intelligent systems based on large-scale data and convoluted knowledge has revolutionized the way we conduct modern biology. It has been suggested that the various omics modalities, such as transcriptomics, proteomics, metabolomics, and epigenomics, may be used to augment traditional methods, or may be combined with other nextgeneration technologies to address limitations and provide a better characterization of the chemical risk (Van Agellen et al., 2010). However, since then, progresses for the omics in environmental risk assessment have been relatively slow perhaps compared to other societal applications, but we must not forget about the technological advancements performed in the recent years and the new knowledge that we have gained. Recent prospectives have proposed a general path forward for incorporating omics data into an environmental monitoring framework (Martynuik 2018; Sauer et al., 2017). To this end, the community must first establish Good Laboratory Practices (GLPs), Standard Operating Procedures (SOPs), and a consensus on how to best process and store data. Bioinformatics and statistical tools also need to be standardized, and entire pipelines, from the laboratory bench to quantitative conclusions, need to be packaged and made distributable. But this is no easy task - credible use of these technologies will require massive efforts toward validation and standardization.

7. Conclusions

During the last decade, the proof of principle of omics technologies has been performed in terms of potential application to environmental toxicology, specifically the assessment of environmental pollution impacts in non-model organisms. Omics advances in ecotoxicological studies not only generate new knowledge regarding mechanisms of toxicity and environmental effect, improving the relevance and immediacy of laboratory toxicological assessment, but can also provide a new paradigm for ecotoxicology by linking ecological models to mechanism-based, systems biology approaches.

Increasing the use of omics technology in chemical risk assessment and environmental monitoring may now require to reconsider the way we generate, collect, organize and treat ecotoxicological data, and are set to transform the process of decision-making for environmental protection taking advances of artificial intelligence technology.

Acknowledgements

We would like to thanks Q. Qiao, S. Le Manach and B. Sotton for their collaboration on data acquisition and treatment. We are also greatful to A. Paris for its support and prolific discussions.

The NMR and the MS spectra were respectively acquired at the Plateau technique de Résonance Magnétique Nucleaire and the Plateau technique de spectrométrie de masse bioorganique, Muséum National d'Histoire Naturelle, Paris, France. This work was supported by grants from CNRS (Défi ENVIROMICS "Toxcyfish" project).

This work was supported by the fruitful discussions generated with the community of the GDR "Ecotoxicologie Aquatique", within the thematic axe on "the use of omics form aquatic ecotoxicology".

The author declares no conflict of interest

References

- Ankley, Gerald T., et al. "Adverse outcome pathways: a conceptual framework to support ecotoxicology research and risk assessment." *Environmental Toxicology and Chemistry: An International Journal* 29.3 (2010): 730-741.
- Asselman, J., De Coninck, et al., & De Schamphelaere, K. A. (2016). Bisulfite sequencing with *Daphnia* highlights a role for epigenetics in regulating stress response to *Microcystis* through

preferential differential methylation of serine and threonine amino acids. *Environmental science & technology*, *51*(2), 924-931.

- Astudillo-García, C., Hermans, S. M., Stevenson, et al., Lear, G. (2019). Microbial assemblages and bioindicators as proxies for ecosystem health status: potential and limitations. *Applied Microbiology and Biotechnology*, 1-15.
- Bertucci, A., Pierron, F., Thébault, J., Klopp, C., Bellec, J., Gonzalez, P., & Baudrimont, M. (2017). Transcriptomic responses of the endangered freshwater mussel *Margaritifera margaritifera* to trace metal contamination in the Dronne River, France. *Environmental Science and Pollution Research*, 24(35), 27145-27159.
- Biales, A. D., Kostich, M. S., et al., & Bencic, D. C. (2016). Initial development of a multigene omics-based exposure biomarker for pyrethroid pesticides. *Aquatic toxicology*, *179*, 27-35.
- Boccard, J., Tonoli, D., Strajhar, P., Jeanneret, F., Odermatt, A., & Rudaz, S. (2019). Removal of batch effects using stratified subsampling of metabolomic data for in vitro endocrine disruptors screening. *Talanta*, *195*, 77-86.
- Bouslimani, A., Melnik, A. V., Xu, Z., Amir, A., da Silva, R. R., Wang, M., ... & Dorrestein, P. C. (2016). Lifestyle chemistries from phones for individual profiling. *Proceedings of the National Academy of Sciences*, *113*(48), E7645-E7654.
- Cappello, T., Pereira, P., Maisano, M., Mauceri, A., Pacheco, M., & Fasulo, S. (2016). Advances in understanding the mechanisms of mercury toxicity in wild golden grey mullet (*Liza aurata*) by ¹H NMR-based metabolomics. *Environ. Poll.*, *219*, 139-148.
 Catherine, A., Selma, M., Mouillot, D., Troussellier, M., & Bernard, C. (2016). Patterns and multi-scale drivers of phytoplankton species richness in temperate peri-urban lakes. *Science of the Total Environment*, *559*, 74-83.
- Catherine, A., Bernard, C., Spoof, L., & Bruno, M. (2017). Microcystins and nodularins. *Handbook of cyanobacterial monitoring and cyanotoxin analysis. Wiley*, 109-126.
- Ceja-Navarro, J. A., Vega, F. E., Karaoz, U., Hao, Z., Jenkins, S., Lim, H. C., ... & Brodie, E. L. (2015). Gut microbiota mediate caffeine detoxification in the primary insect pest of coffee. *Nature communications*, *6*, 7618.
- Codd G., Lindsay J., Young F., Morrison L.F., Metcalf J.S. (2005) Harmful cyanobacteria. From mass mortalities to management measures. Huisman J., Matthijs C.P. & Visser M. (Eds), Springer Verlag, Berlin, 241 p.
- Collins, F.S., Varmus, H. (2015) A new initiative on precision medicine. *N. Engl. J. Med.* 372, 793–795.
- Conolly, R. B., Ankley, G. T., et al., & Watanabe, K. H. (2017). Quantitative adverse outcome pathways and their application to predictive toxicology. *Environ. Sci. & technol.*, 51(8), 4661-4672.
- Davis, J. M., Ekman, D. R., Teng, Q., Ankley, et al., & Jorgenson,
 Z. G. (2016). Linking field-based metabolomics and chemical analyses to prioritize contaminants of emerging concern in the
 Great Lakes basin. *Environmental toxicology and chemistry*, 35(10), 2493-2502.
- Degli-Esposti, D., Almunia, C., Guery, M. A., et al., Geffard, O. (2019). Co-expression network analysis identifies gonad-and embryo-associated protein modules in the sentinel species Gammarus fossarum. *Scientific reports*, *9*(1), 7862.
- Duperron, S., Halary, S., Habiballah, M., et al., & Marie, B. (2019). Response of Fish Gut Microbiota to Toxin-Containing Cyanobacterial Extracts: a Microcosm Study on the Medaka (*Oryzias latipes*). *Environmental Science & Technology Letters*.

- Evariste, L., Barret, M., Mottier, A., et al., Pinelli, E. (2019). Gut microbiota of aquatic organisms: A key endpoint for ecotoxicological studies. *Environ. Poll.*. 248, 989-999.
- Gil-Solsona, R., Nácher-Mestre, J., Lacalle-Bergeron, L., et al., Pérez-Sánchez, J. (2017). Untargeted metabolomics approach for unraveling robust biomarkers of nutritional status in fasted gilthead sea bream (*Sparus aurata*). *PeerJ*, *5*, e2920.
- González, I., Déjean, S., Martin, P. G., & Baccini, A. (2008). CCA: An R package to extend canonical correlation analysis. *Journal of Statistical Software*, 23(12), 1-14.
- González-Ruiz, V., Schvartz, D., et al., & Rudaz, S. (2019). An Integrative Multi-Omics Workflow to Address Multifactorial Toxicology Experiments. *Metabolites*, *9*(4), 79.
- Gressel, J. (2018). Microbiome facilitated pest resistance: potential problems and uses. *Pest management science*, 74(3), 511-515.
- Griffitt, R. J., Brown-Peterson, N. J., et al., Brouwer, M. (2012). Effects of chronic nanoparticulate silver exposure to adult and juvenile sheepshead minnows (*Cyprinodon variegatus*). *Environmental Toxicology and Chemistry*, *31*(1), 160-167.
- Hamilton, P. B., Cowx, I. G., Oleksiak, M. F., et al., & Tyler, C. R. (2016). Population-level consequences for wild fish exposed to sublethal concentrations of chemicals–a critical review. *Fish and Fisheries*, *17*(3), 545-566.
- Hervé, M. R., Nicolè, F., & Lê Cao, K. A. (2018). Multivariate analysis of multiple datasets: a practical guide for chemical ecology. *Journal of chemical ecology*, *44*(3), 215-234.
- Hook, S. E., Gallagher, E. P., & Batley, G. E. (2014). The role of biomarkers in the assessment of aquatic ecosystem health. *Integr. Environ. assess. manag.t*, *10*(3), 327-341.
- Ivanisevic, J., Stauch, K. L., Petrascheck, M., et al., Boska, M. D. (2016). Metabolic drift in the aging brain. *Aging (Albany NY)*, 8(5), 1000.
- Janssen, E. M. L. (2019). Cyanobacterial peptides beyond microcystins–A review on co-occurrence, toxicity, and challenges for risk assessment. *Water research* 151, 488.
- Kim, M., & Tagkopoulos, I. (2018). Data integration and predictive modeling methods for multi-omics datasets. *Molecular omics*, 14(1), 8-25.
- Kim Tiam, S., Gugger, M., Demay, J., et al., Marie, B. (2019). Insights into the Diversity of Secondary Metabolites of *Planktothrix* Using a Biphasic Approach Combining Global Genomics and Metabolomics. *Toxins*, *11*(9), 498.
- Lê Cao, K. A., Rohart, F., Gonzalez, I., & Le Cao, M. K. A. (2018). R package 'mixOmics'.
- Le Manach S., Khenfech N., Huet H., et al., Marie B. (2016) Gender-specific toxicological effects of chronic exposure to pure microcystin-LR or complex *Microcystis aeruginosa* extracts on adult medaka fish. Environm. Sci. Technol. 50:8324-34.
- Le Manach S., Sotton B., Huet H., et al., Marie B. (2018) Physiological effects caused by microcystin-producing and nonmicrocystin producing *Microcystis aeruginosa* on medaka fish: a proteomic and metabolomic study on liver. Environmental Pollution 234, 523-537.
- Lenz, K. A., Miller, T. R., & Ma, H. (2019). Anabaenopeptins and cyanopeptolins induce systemic toxicity effects in a model organism the nematode Caenorhabditis elegans. *Chemosphere*, *214*, 60-69.
- Li, X., Brejnrod, A. D., Ernst, M., Rykær, M., et al., Sørensen, S. J. (2019). Heavy metal exposure causes changes in the metabolic

health-associated gut microbiome and metabolites. *Environment international*, *126*, 454-467.

- Liang, X., Feswick, A., Simmons, D., & Martyniuk, C. J. (2018). Environmental toxicology and omics: A question of sex. *Journal* of Proteomics, 172, 152-164.
- Liquet, B., Lê Cao, K. A., Hocini, H., & Thiébaut, R. (2012). A novel approach for biomarker selection and the integration of repeated measures experiments from two assays. *BMC bioinformatics*, 13(1), 325.
- Malbrouck, C., Kestemont, P., 2006. Effects of microcystins on fish. Environ. Toxicol. Chem. 25, 72-86.
- Malécot, M., Marie, A., Puiseux-Dao, S., & Edery, M. (2011).
 iTRAQ-based proteomic study of the effects of microcystin-LR on medaka fish liver. *Proteomics*, 11(10), 2071-2078.
- Marie, B., Huet, H., Marie, A., et al., Edery, M. (2012). Effects of a toxic cyanobacterial bloom (*Planktothrix agardhii*) on fish: insights from histopathological and quantitative proteomic assessments following the oral exposure of medaka fish (*Oryzias latipes*). *Aquatic toxicology*, *114*, 39-48.
- Marjan, P., Martyniuk, C. J., Fuzzen, M. L., et al., Servos, M. R. (2017a). Returning to normal? Assessing transcriptome recovery over time in male rainbow darter (*Etheostoma caeruleum*) liver in response to wastewater-treatment plant upgrades. *Environ. Toxicol. chem.*, *36*(8), 2108-2122.
- Marjan, P., Bragg, L. M., MacLatchy, D. L., Servos, M. R., & Martyniuk, C. J. (2017b). How does reference site selection influence interpretation of omics data?: Evaluating liver transcriptome responses in male rainbow darter (*Etheostoma caeruleum*) across an urban environment. *Environ. Sci. & technol.*, *51*(11), 6470-6479.
- Martyniuk, C. J. (2018). Are we closer to the vision? A proposed framework for incorporating omics into environmental assessments. *Environ. toxicol pharmaco.*, *59*, 87-93.
- Mehl, F., Marti, G., Merle, P., et al., Boccard, J. (2015). Integrating metabolomic data from multiple analytical platforms for a comprehensive characterisation of lemon essential oils. *Flavour and fragrance j.* 30(2), 131-138.
- Mezhoud, K., Puiseux-Dao, S., & Edery, M. (2011). Toxins and stress in fish: proteomic analyses and response network. *Toxicon*, *57*(7-8), 959-969.
- Patel, A. P., Tirosh, I., Trombetta, J. J., et al., Louis, D. N. (2014). Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science*, *344*(6190), 1396-1401.
- Oberg AL, Vitek O. 2009. Statistical design of quantitative mass spectrometry-based proteomic experiments. J Proteome Res 8:2144–2156.
- O'neil, J. M., Davis, T. W., Burford, M. A., & Gobler, C. J. (2012). The rise of harmful cyanobacteria blooms: the potential roles of eutrophication and climate change. *Harmful algae*, *14*, 313-334.
- Ortega, V.A., Katzenback, B.A., Stafford, J.L., Belosevic, M., Goss, G.G. (2015). Effects of polymer-coated metal oxide nanoparticles on goldfish (*Carassius auratus* L.) neutrophil viability and function. Nanotoxicology 9, 23–33.
- Pavagadhi, S., & Balasubramanian, R. (2013). Toxicological evaluation of microcystins in aquatic fish species: Current knowledge and future directions. *Aquatic Toxicology*, *142*, 1-16.
- Qiao Q., Le Manach S., Huet H., et al., Marie B. (2016a) An integrated omic analysis of hepatic alteration in medaka fish

chronically exposed to cyanotoxins with possible mechanisms of reprotoxicity. Environm. Poll. 6: 32459.

- Qiao Q., Le Manach S., Sotton B., et al., Marie B. (2016b) Deep sexual dimorphism in adult medaka fish liver highlighted by multi-omic approach. Scientific Reports 6: 32459.
- Reverter, M., Sasal, P., Banaigs, B., et al., Tapissier-Bontemps, N. (2017a). Fish mucus metabolome reveals fish life-history traits. *Coral Reefs*, *36*(2), 463-475.
- Reverter, M., Sasal, P., Tapissier-Bontemps, N., Lecchini, D., & Suzuki, M. (2017b). Characterisation of the gill mucosal bacterial communities of four butterflyfish species: a reservoir of bacterial diversity in coral reef ecosystems. *FEMS microbial. Ecol.*, 93(6), fix051.
- Rohart, F., Gautier, B., Singh, A., & Lê Cao, K. A. (2017). mixOmics: An R package for omics feature selection and multiple data integration. *PLoS computational biology*, *13*(11), e1005752.
- Rolli, E., Marasco, R., Vigani, G., et al., Pierotti Cei, F. (2015). Improved plant resistance to drought is promoted by the rootassociated microbiome as a water stress-dependent trait. *Environ. microbiol*, *17*(2), 316-331.
- Sauer, U. G., Deferme, L., Gribaldo, L., et al., Gant, T. W. (2017). The challenge of the application of omics technologies in chemicals risk assessment: Background and outlook. *Regulatory Toxicology and Pharmacology*, *91*, S14-S26.
- Schwarzenberger A, Zitt A, Kroth P, Mueller S, Von Elert E. (2010) Gene expression and activity of digestive proteases in Daphnia: effects of cyanobacterial protease inhibitors. BMC Physiol. 4; 10:6.
- Simmons DB, Sherry JP (2013) Plasma proteome profiles from white sucker (*Catostomus commersonii*) sampled from the Athabasca River, upstream and downstream of oil sands development. Proceedings, SETAC North America 34th Annual Meeting, Nashville, TN, USA, November 17–21, 2013, p WP005.
- Simmons, D. B., Benskin, J. P., Cosgrove, J. R., et al., Sherry, J. P. (2015). Omics for aquatic ecotoxicology: Control of extraneous variability to enhance the analysis of environmental effects. *Environmental toxicology and chemistry*, *34*(8), 1693-1704.
- Sotton, B., Paris, A., Le Manach, S., et al., Marie B. (2017) Metabolic changes in Medaka fish induced by cyanobacterial exposures in mesocosms: an integrative approach combining proteomic and metabolomic analyses. *Scientific reports*, 7(1), 4051.
- Sotton B., Le Manach S., Paris A., et al., Marie B. (2017b) Global metabolome changes induced by cyanobacterial blooms in three representative fish species. Sci. Tot. Environm. 591: 333-42.
- Sotton, B., Paris, A., Le Manach, S., et al., Marie, B. (2019). Specificity of the metabolic signatures of fish from cyanobacteria rich lakes. *Chemosphere*, *226*, 183-191.
- Tang, R., Ding, C., Ma, Y., Wang, J., Zhang, T., & Wang, X. (2017). Time-dependent responses of earthworms to soil contaminated with low levels of lead as detected using ¹H-NMR metabolomics. *RSC Advances*, 7(54), 34170-34181.
- Van Aggelen, G., Ankley, G. T., et al. Yu, L. (2010). Integrating omic technologies into aquatic ecological risk assessment and environmental monitoring: hurdles, achievements, and future outlook. *Environmental health perspectives*, *118*(1), 1–5.
- Wei, F., Sakata, K., Asakura, T., Date, Y., & Kikuchi, J. (2018). Systemic homeostasis in metabolome, ionome, and microbiome of wild yellowfin goby in estuarine ecosystem. *Scientific reports*, 8(1), 3478.