Promising potential of high-throughput molecular phenotyping of freshwater fishes for environmental assessment

Benjamin MARIE[1,], Julie LELOUP[2], Aurélie BONIN[3], Alison GALLET[1], Maud MOUCHET[4]*

[1] 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.
[2] 7618 iEES-Paris Sorbonne Université - 4 place Jussieu - 75252 Paris Cedex 05, France.
[3] UMR 5553, CNRS/Université Joseph Fourrier, Laboratoire d'Écologie Alpine (LECA), BP 53, 38041 Grenoble Cedex-9, France.
[4] UMR 7204, CNRS, SU, MNHN Centre d'Ecologie et des Sciences de la Conservation (CESCO), 43 rue Buffon - CP 135, 75231 Paris Cedex 05, France.

* Corresponding author: benjamin.marie@mnhn.fr

Abstract:

The recent democratisation of high-throughput molecular phenotyping allows the rapid expansion of promising untargeted multi-dimensional approaches (*e.g.* epigenomics, transcriptomics, proteomics, metabolomics, as well as microbiome metabarcoding), that now represent innovative perspectives for environmental assessments. Indeed, when developed for ecologically relevant species, these emerging "omics" analyses may present valuable alternatives for the development of novel generations of ecological indicators, that in turn could provide early warnings of eco(toxico)logical impairments.

This pilot study investigates the bio-indicative potential of different multi-metric tools based on different high-throughput molecular phenotyping approaches (*i.e.* metabarcoding of the intestine microbiome, and liver metabolomics by nuclear magnetic resonance (NMR) and liquid chromatography coupled to high resolution mass spectrometry (LC-HRMS) on two sentinel fish species (*Perca fluviatilis* and *Lepomis gibbosus*) from a set of eight water bodies of the peri-urban area of Paris (France). We show that the LC-MS metabolome dataset allows remarkably clear separation of individuals according to the species but also according to their respective sampling lakes. Interestingly, the similar variations of *Perca* and *Lepomis* metabolomes occur locally indicating that local environmental constraints drive the observed metabolome variations beyond their obvious genetic differences. Thus, the development of such reliable molecular phenotyping for environmental monitoring constitutes a promising and innovative bio-indicative tool for environmental assessment.

Keywords: microbiome metabarcoding, untargeted metabolomics, environmental assessment, biomarkers, sentinel species

1. INTRODUCTION

In the past decades, the use of bio-indicator aquatic organisms has been widely adopted in order to monitor, maintain or improve the quality of water bodies worldwide.[1] In parallel, the difficulties and the rational limits of direct chemical monitoring for providing sufficient information to adequately assess the risks from anthropogenic chemicals in the environment have been increasingly recognized.[2]

Recent years have offered intense implementation of molecular techniques in a wide variety of ecology research fields. Modern bio-monitoring and bio-assessment may greatly benefit from the specific development of high-throughput methods that potentially provide reliable, high quantity and quality standardized biomolecular data in a cost- and time-efficient way.[3-4]

Fishes, in particular, present unique features that make them especially relevant for bio-indication purposes. They indeed occupy in almost any aquatic habitat, as they are capable of experiencing different and variable environmental conditions.[5] Fishes are also responsive to numerous abiotic (*e.g.* temperature, water velocity, sediment load, hypoxia) or biotic (*e.g.* famine, predation, parasitism) pressures, together with anthropogenic stressors such as contaminants, that represent additional constraints that fish may experience in disturbed ecosystems.[6] Environmental stressors are known to induce biological variations, in relation to the specific physiological, developmental and reproductive capabilities of the different fish species allowing them to colonize and occupy ecosystems.[7] Thus, fishes appear to be valuable bio-indicators and their individual or community responses to environmental pressures, therefore, might have a high ecological relevance.

In Europe, for example, according to the Water Framework Directive (WFD), local ecological quality can be assessed by comparing the community structure (of fishes, but also of diatoms, sea-grasses, macro-invertebrates, etc.) in impacted sites to those observed in reference sites, using various biological quality indices.[8] However, experts face several important challenges originating from the main limitation of different diversity indexes based on fish guild: they remain especially time-consuming, requires costly sampling procedures, and can even exhibit low responsiveness to field variation.[9-10] One other important drawback of fish diversity-based bio-indicative indexes is that they also present a high sensitivity to species migration, human transplantation or fishing activities, that critically compromise their predictive power.[11] On the other hand, fish bio-indication can be

also considered through various measurable biological traits (*e.g.* population structure, genetic or functional diversity, physiological or biological parameters). Some of them have been already implemented, including morphometrical features,[12] punctual biochemical markers,[13] or genetic diversity.[7] However, these attempts have also emphasized their relative limitations in terms of operational predication ability.[14]

The recent democratisation of high-throughput molecular phenotyping allows the rapid expansion of promising untargeted multi-dimensional approaches (*e.g.* epigenomics, transcriptomics, proteomics, metabolomics, as well as microbiome composition by metabarcoding). These constitute many early responding biological traits, and represent promising perspectives for environmental assessments.[15-16] In this context, it is very likely that emerging "omics"-based analyses, developed for ecologically relevant fish species, may present valuable alternative for the development of new ecological evaluation tools, that could provide even earlier warning of eco(toxico)logical impairs.[17]

To this end, the present pilot study investigates the bio-indicative potential of different multi-metric tools based on different high-throughput molecular phenotyping approaches (*i.e.* metabarcoding of the intestine microbiome and liver metabolomics by NMR and LC-MS) on two sentinel fish species (*Perca fluviatilis* and *Lepomis gibbosus*) from a set of eight water bodies of the peri-urban area of Paris (France). We hypothesize that molecular phenotyping approaches can be exploited for environmental diagnosis, and provide valuable multi-metric characters indicative of ecological conditions for both evaluation or prediction purposes, together with a prospective mechanistic understanding that allows cause-effect relationship investigation. We compare here the potential phenotypic bio-indication of two sentinel species, as these organisms could present different or comparable sensitivity and responsiveness to local environmental constraints, that may reinforce the proof of concept of their respective interest for "omics"-based bio-indication.

2. MATERIALS AND METHODS

2.1. Fish sampling

103 Field sampling campaigns were performed during late summer 2015 (7-10th September) in
104 eight peri-urban pounds around the Paris' area (Île-de-France region, France), chosen for
105 their respective eutrophication levels and the presence or the absence of recurrent
106 cyanobacterial blooms: Cergy-Pontoise (Cer), Champs-sur-Marne (Cha), Maurepas (Mau),
107 Rueil (Rue), Verneuil (Ver), Varennes-sur-Seine (Var), Fontenay-sur-Loing (Fon), and Triel
108 (Tri) pounds (Supplementary figure S1).[18-19] These sites were sampled by the Hydrosphère
109 company (www.hydrosphere.fr) with an electric fishing device (FEG 8000, EFKO, Leutkirch,
110 Germany) for capturing fish alive. The investigation of the fish guild indicates that only the
111 perch (*Perca fluviatilis*) and pumpkinseed sunfish (*Lepomis gibbosus*) were present in all or
112 almost all these pounds (supplementary table S1) and were thus selected as sentinel species
113 for further molecular phenotyping analyses by metabarcoding and metabolomics.
114 Briefly, alive caught fishes (n=5-10 young-of-the-year per pounds and per species) were
115 directly measured (12.0±4.8 cm), weighed (9.3±2.6 g), briefly euthanized by neck dislocation
116 and then liver and intestine of each individual was shortly sampled, flash-frozen in liquid
117 nitrogen and kept at -80°C until analyses, in accordance with European animal ethical
118 concerns and regulations.
119 In every lake, sub-surface chlorophyll-*a* equivalent concentrations attributed to the four
120 main phytoplankton groups (Chlorophyta, Diatoms, Cyanobacteria and Cryptophyta) were
121 measured with an *in-situ* fluorometer (Fluoroprobe II, Bbe-Moldenke, Germany). Sub-surface
122 water samples filtered on 20-μm mesh size were also collected for phytoplanktonic
123 community analysis and further metabolomics characterisation, and then kept at -80°C until
124 analysis.
125
126 **2.2. Intestinal microbiome sequencing and OTU analyses**
127 Intestinal microbiome DNA was extracted from each fish individual (4-5 replicate fishes per
128 pound), stored and filtered following Mouchet et al..[20] After DNA concentration and purity
129 checking, a 295 pb fragment of the 16S rRNA-encoding gene corresponding to the V5-V6
130 variable region was amplified using 789-F and 1080-R primers and sequenced on an Illumina
131 MiSeq platform (2x250 bp, Fasteris, Switzerland). Raw reads analysis was processed through
132 the FROGS pipeline (Find Rapidly OTU with Galaxy Solution) implemented on a galaxy
133 instance (2.3.0) (https://galaxy.migale.inra.fr/). Sequences were dereplicated before
134 clustering using SWARM algorithm (2.1.5) with, at first, a denoising step using an

135 aggregation distance equal to 3.[21] The taxonomic assignation of each cluster was performed
136 using the BLAST tools. OTU table was rarefied, normalized by Total Sum Scaling using
137 MicrobiomeAnalyst platform.[22]
138
139 **2.3. Liver metabolite extraction and metabolomics analyses**
140 The liver extraction was performed on 132 individuals (comprising 78 *Perca* and 54 *Lepomis*)
141 with methanol/chloroform/water (ratio 2/2/1.8 – 22 $mL.g^{-1}$ at 4°C) and the polar fraction
142 was analyzed on a 600-MHz NMR spectrometer equipped with a 5-mm cryoprobe (Advance
143 III HD Bruker, Germany) with a noesygppr1d pulse sequence as previously described.[19] ^{1}H-
144 NMR spectra were treated with the Batman R package for deconvolution, peak assignment
145 and quantification of 222 putative metabolites.[23]
146 The liver extracted polar phase was additionally injected (2 µL) on C_{18} column (Polar
147 Advances II 2.5 pore - Thermo), then eluted at a 300 $\mu L.min^{-1}$ flow rate with a linear gradient
148 of acetonitrile in 0.1% formic acid (5 to 90 % in 21 min) with a ultra-high performance liquid
149 chromatography (UHPLC) system (ELUTE, Bruker). Consecutively, the individual metabolite
150 contents were analyzed using an electrospray ionization hybrid quadrupole time-of-flight
151 (ESI-Qq-TOF) high-resolution mass spectrometer (Compact, Bruker) at 2 Hz speed on positive
152 MS mode on the 50–1500 *m/z* range. The feature peak list was generated from recalibrated
153 MS spectra (< 0.5 ppm for each sample, as an internal calibrant of Na formate was injected
154 at the beginning of each sample analysis) within a 1-15 min window of the LC gradient, with
155 a filtering of 5,000 count of minimal intensity, a minimal occurrence in at least 50% of all
156 samples, and combining all charge states and related isotopic forms using MetaboScape 4.0
157 software (Bruker).
158 Additionally, five pools of six different individuals randomly selected for *Perca* and *Lepomis*
159 (quality check samples) were similarly eluted then analyzed on positive autoMS/MS mode at
160 2-4 Hz on the 50-1500 *m/z* range for further metabolite annotation. Molecular networks
161 were performed with the GNPS (http://gnps.ucsd.edu) and/or MetGem
162 (http://metgem.github.io) softwares, as previously described[24] and metabolite annotation
163 was assayed using GNPS, Mona, HMDB and Massbank MS/MS spectral libraries.
164 Phytoplankton samples of each water pound, concentrated with a 20-μm mesh size, were
165 also extracted with 75% methanol (2 min sonication, 5 min centrifugation at 15,000 g - 4°C)

166 and then similarly was analyzed on LC-HRMS on autoMS/MS positive mode, as earlier
167 described, and the metabolite list was annotated with the same pipelines.
168
169 **2.4. Data matrix treatment**
170 The resulting count table of 87 OTUs (microbiome metabarcoding), and the intensity data
171 tables of the 222 metabolites (^{1}H NMR) and 1252 analytes (LC-MS) were further treated for
172 quantile normalization and Pareto's scaling, data representation by PCA and heatmap with
173 hierarchical clustering and then analyzed to investigate the influences of "Species" (*Perca* or
174 *Lepomis*) "Lakes" (Che, Cha, Fon, Mau, Rue, Tri, Var or Ver) parameters on the datasets by
175 partial least square – differential analysis (PLS-DA) using the MixOmics R Package,[25] the
176 MetaboAnalyst 4 tool [26] and the MicrobiomeAnalyst platform.[22] The PLS model
177 performances were determined by accuracy, performance (R^2) and predictability (Q^2) values,
178 and classification error.
179
180
181 **3. RESULTS AND DISCUSSION**
182 **3.1. Global informativeness of the three molecular datasets on *Perca* and *Lepomis***
183 Figure 1 illustrates the global relative distribution of OTUs, metabolites observed by NMR
184 and by LC-MS for each fish (comprising both *Perca* and *Lepomis*) on a heatmap with
185 hierarchical classifications and individual plot principal component analyses (PCA). Based on
186 the relative distribution of the 87 identified OTUs within fish individuals (44 *Perca* and 34
187 *Lepomis*), no reliable discrimination could be observed according to the species and lakes
188 (Fig. 1A and D). Although the 222 potential metabolites quantified by NMR on 128 fish (78
189 *Perca* and 54 *Lepomis*) present no obvious distinction for neither the "species" nor "lakes"
190 variables (Fig. 1B and E), the 1,252 analytes semi-quantified from the liver metabolomes by
191 LC-MS clearly present more discriminant patterns between species and lakes (Fig. 1C and F).
192 Additional investigations on those datasets were then performed using supervised models.
193 While unsupervised PCA were first used to evaluate the global dispersion between species
194 and sampling groups, a supervised model such as partial least square differential analysis
195 (PLS-DA) (Supplementary figure S2; supplementary tables S2, S3 and S4) allows to maximize
196 the separation between sample classes (*i.e.* "species" and "lakes") and to extract
197 information on discriminating features (variable importance on projection - VIP).

198 Interestingly, these differential analyses indicate that the microbiome metabarcoding
199 dataset allows a faint but noteworthy discrimination among the intestinal microbiomes
200 observed within individuals of the two species (R^2 = 0.54), but with only a limited
201 predictability of the models (Q^2 = 0.13 and classification error = 0.32) (Supplementary figure
202 S2A), while no reliable microbiome discrimination could be retrieved for the "lake" variable
203 according to PLS-DA investigation (Supplementary figure S2D). Concerning the ^{1}H NMR liver
204 metabolomics dataset, the PLS-DA model provide a comparable discrimination between
205 species (R^2 = 0.51 and Q^2 = 0.38) (Supplementary figure S2B), but could neither allow reliable
206 "lake" discrimination or prediction (Supplementary figure S2E). On the other hand, a very
207 strong discrimination for the "species" was achieved by PLS-DA for the LC-MS liver
208 metabolomics dataset (R^2 = 0.96 and Q^2 = 0.96, classification error = 0) (Supplementary
209 figure S2C). Only a slight distinction among the "lakes" can be observed for the LC-MS
210 dataset (R^2 = 0.34 and Q^2 = 0.24; classification error = 0.84) (Supplementary figure S2F),
211 whereas both the heatmap with hierarchical clustering (Figure 1C) and the individual plot
212 from the un-supervised PCA (Figure 1F) show remarkable grouping of fishes from the same
213 species collected in the same lakes. This surprising lack of performance of the PLS-DA
214 performed with grouped LC-MS data from the two fish species, *Perca* and *Lepomis*, is very
215 likely due to species variation among the different lakes. Indeed, this hypothesis is further
216 explored by investigating separately the *Perca* and *Lepomis* locality differences on their
217 respective intestinal microbiomes and liver metabolomes.
218
219 **3.2. Comparison of *Perca* and *Lepomis* molecular signature variation among lakes**
220 Considering separately the 16S rDNA metabarcoding datasets of the two fish species, the
221 intestine microbiome of *Perca* and *Lepomis* shows only faint specificity within the different
222 sampling lake groups, as investigated by both un-supervised and supervised multivariate
223 analyses and hierarchical classification (Supplementary figure S3). Indeed, the different
224 performance scoring retrieved for PLS-DA remains distinctly low, and does not support an
225 important structuration of the data according to the "lake" variable, whatever the species
226 considered. However, in the present case, the limited sample number per sampling lake
227 (n=4-5) could constitute a significant limitation to the informativeness of the intestinal
228 microbiome for local discrimination (supplementary figure S3).
229 The intestinal microbiome constitutes a remarkable interface between the organism and the

230 surrounding environment that is involved in various fish biological processes including
231 digestion, immunity or ecotoxicology.[27] The analysis of its microbial composition, in terms of
232 genotype occurrence, density and taxonomy represent remarkable biological traits for the
233 fish host that aim at characterizing the functionality and the ecology of this specific intestinal
234 microbial ecosystem.
235 Although the intestinal microbiome is the object of numerous investigations in microbial
236 ecology and related host disease/health studies,[28] this compartment has been rarely
237 included in environmental biomonitoring programs. However, changes in the microbial
238 intestinal composition may denote significant environmental variations that have potentially
239 altered the host homeostasis and may be influenced by the specific responsiveness of the
240 different fish species. However, microbial diversity estimated from 16S-rRNA sequences
241 does not allow taking into consideration that different genotypes can present functional
242 redundancy within the intestinal microbial ecosystem,[29] or alternatively, that
243 undistinguishable OTUs (regarding only V5-V6 16S-rRNA fragment) can mask different sub-
244 taxa with distinct physiological capacities, as already exemplified for *Vibrio* sp..[30] Taken
245 together, these technical issues can represent noticeable limitations for the use of microbial
246 diversity retrieved from OTU sequencing for environmental assessment, and may require
247 additional function-based approaches for microbial community exploration.[27] Also, one of
248 the most pregnant limitation for the democratisation of microbiome metabarcoding
249 analyses is maybe its financial cost, that still remains too important for its integration in the
250 development of affordable biomonitoring solutions.
251
252 Concerning the ^{1}H NMR metabolomics datasets of the two fish species taken separately, the
253 liver metabolome of *Perca* and *Lepomis* shows only a limited specificity in relation to the
254 "lakes". Indeed, according to the relatively low performances of the hierarchical
255 classifications and the discriminant analyses, the NMR liver metabolomes present restricted
256 discrimination on "lakes" for both species and could also support only a limited separation
257 predictability of the supervised models (Supplementary figure S4; supplementary table S3).
258 The ^{1}H NMR metabolomics allow the reliable quantification of main liver metabolites
259 (belonging mostly to amino acids, sugars, TCA metabolites, etc.), by detecting characteristic
260 chemical groups, with regards to their respective position within the molecules.[31] These
261 primary metabolites are key contributors to the cellular metabolism and are supposed to be

262 responsive to variable physiological conditions of fish encountering peculiar biological
263 outcomes and/or environmental constraints.[32] However, this approach remains focused on a
264 few principal metabolites involved in main cellular pathways or regulation processes, and
265 not being specific of subtle molecular regulation processes. Moreover, ^{1}H NMR presents only
266 low sensitivity and limited discrimination capability for molecules presenting similar
267 chemical structures, such as lipids, that exhibit an aggregated and unspecific signal on NMR
268 spectra.[33] For this reason, it is supposed to present lower discriminating power than less
269 quantitative, but more selective and more sensitive, mass spectrometry-based
270 metabolomics approaches.
271
272 In contrast, the LC-MS dataset of the two fish species considered separately clearly shows
273 reliable discrimination of the metabolomes of fishes originating from the different sampling
274 lakes (Figure 2; supplementary figure S5; supplementary figure S6). As illustrated with the
275 un-supervised PCA and heatmap with hierarchical classification (Figure 2), the liver
276 metabolomes analyzed by LC-MS present distinct local signatures that support reliable
277 discriminant analysis models for both *Perca* and *Leptomis*, with significant performance
278 scoring of the PLS-DA (R^2_{Perca} = 0.46 and Q^2_{Perca} = 0.35; $R^2_{Lepomis}$ = 0.50 and $Q^2_{Lepomis}$ = 0.38)
279 (supplementary figure S5; supplementary figure S6). Taken together, these observations
280 indicate that the locality seems to globally influence the LC-MS metabolome composition for
281 both species.
282 However, as previously observed when comparing LC-MS datasets of both fish species
283 together, the *Perca* and *Lepomis* liver metabolomes of the Fontenay-sur-Loing lake (Fon)
284 exhibited a quite species-specific trajectory, with regards to the relative composition of their
285 metabolomes (Figure 2A and 2D). This suggests that, although the specific "lake" signature
286 of the metabolome appears in good agreement between the two species for most localities,
287 in some environments, local specificities are variable from one species to the other one, as
288 represented here for the local species-specificity of the *Perca* and *Lepomis* metabolomes in
289 Fon lake.
290 Interestingly, the hierarchical classifications performed respectively on the *Perca* and
291 *Lepomis* LC-MS metabolome datasets show very similar lake relationships, grouping together
292 the fish from the lakes of Cer and Mau, from Cha, Fon and Rue, and from Tri, Var and Ver
293 (for *Perca*) or Tri and Ver for *Lepomis*, as this latter fish species was not retrieved in Var

294 (Figure 2B and 2E). Then, we re-explored the fish metabolome discrimination considering
295 together fish from the "Cer/Mau", "Cha/Fon/Rue" and "Tri/Var/Ver" (or "Tri/Ver" for
296 Lepomis) groups, by PCA (supplementary figure S7), and PLS-DA (Figure 2C and 2F). For both
297 *Perca* and *Lepomis* the metabolomes present an even better discrimination scoring (*Perca*:
298 accuracy = 0.77, R^2 = 0.79, Q^2 = 0.70; *Lepomis*: accuracy = 0.70, R^2 = 0.73, Q^2 = 0.60) and
299 discriminant analysis predictive error (Class. $error_{Perca}$ = 0.04; Class. $error_{Lepomis}$ = 0.10) for
300 these three groups than when considering the different lakes separately. Remarkably, the
301 observation of the metabolite semi-quantification of the best VIP presents largely similar
302 variations among these three groups of lakes for *Perca* and *Lepomis* (Table 1; supplementary
303 figure S8).
304 In addition, the LC-MS metabolite annotation, supported by molecular networking analysis
305 (supplementary figure S9; table 1), shows that fish collected from the Cer/Mar lake group
306 seem to present greater energetic and anti-oxidant/detoxification reserves (*e.g.* adenosine
307 and oxidized glutathione) than in other lake groups. In addition, the molecular networking
308 annotation process highlights that most, if not all, of the known metabolites (presenting
309 structural identity or analogy hits) are shared between the two species (*e.g.* saccharides,
310 nucleic acids, carnitines, glutathiones, lipids) when most of the species-specific cluster
311 metabolites remains uncharacterized, as corresponding molecules present no match within
312 public chemical databases (supplementary figure S9).[34] This representation remarkably
313 illustrates the portion of the specific liver molecular metabolism that remains to be
314 characterized for these two species.
315
316 In general, various genetic or phenotypic factors (such as development stages,
317 contamination levels, predator/parasite pressures, food availability, etc.) could influence and
318 explain local discrepancies of the metabolome of fishes collected from different
319 environments.[17] However, in the present case, the fact that *Perca* and *Lepomis*
320 metabolomes, analyzed in parallel, present similar metabolite variations clearly indicates
321 that local environmental constraints drive such observed phenotypic co-variations (beyond
322 the obvious genetic distinction between *Perca* and *Lepomis*). Then, these present results
323 claim for the great value of LC-MS-based metabolomic imprints for molecular phenotyping
324 on sentinel species and the exploration of environmental effects.[35]
325 Although LC-MS-based metabolomics remains a selective analytical approach (according to

both LC separation and ionisation respective properties of each molecule), it presents un-precedent analytical capabilities in terms of sensitivity, dynamic range and number of characterized components and it now offers unprecedented perspectives for the investigation for molecular phenotyping applied to environmental sciences.[36] We assume that a single LC-MS fingerprint analysis does not embed the whole metabolite picture of the biological compartment (the fish liver, in the present case), especially because of its selectivity performances.[37] However, it provides a specific and precise measurement of a large number of components that can endorse high-throughput and in-depth phenotyping. For comparison, ^{1}H NMR metabolomics allows the global quantification of the CH_2-CH_2 or CH_2-CH=CH fatty acid bounds,[38] whereas LC-MS metabolomics can discriminate and characterize hundreds of different lipids belonging to various sub-classes,[39] providing undoubtedly more informative and discriminant features. In addition, multi-omics integration tools, such as Mixomics,[40] provide innovative solutions for the comparison of heterogeneous dataset matrices obtained from a single set of samples or individuals, and may contribute to investigate metabolomics biomarker relevance (supplementary figures S10-S11).

3.3. Use of LC-MS metabolomics fingerprint for environmental assessment?

Although the value of LC-MS metabolomics for investigating the impacts of environmental stressors or contaminants, and their respective modes of action, has been well-explored in medical sciences[41] or in ecotoxicology laboratory-based studies on aquatic models,[42-45] such methods have been used only faintly in field research so far.[46-47] Apart from a limited number of evidence on the utility of NMR-based metabolomics in environmental fish studies,[17;48-49] few other examples indicate that field-based LC-MS metabolomics constitute an emerging and powerful, but still underused, approach for increasing our understanding of *in situ* biological, physiological, ecological or ecotoxicological processes.[50-52]

Our analysis constitutes one of the first attempts to push forward the potential of high-throughput molecular phenotyping, and especially through LC-MS-based metabolomics, for environmental assessment. Indeed, this organismal molecular phenotyping supported by multi-variable chemometric investigation offers remarkably rich biological information that serves at describing specific phenotypic plasticity. Moreover, this organismal variability/responsiveness can further be confronted to local environmental factors in order

to search for correlation/causality relationships (supplementary figure S12). This effort pushes a step further into the objective environmental omics-assisted assessment relying on field data modelling using artificial intelligence decision-supporting tools.[17]

As described by Pompfret and co-workers,[53] environmental metabolomics exhibits very promising perspectives for operational bio-monitoring applications, because of its reliability, its reproducibility, and its high predictive potential. However, these authors also point out that the responsiveness and robustness of the bio-indicative object, that is characterized through the analytical prism of the metabolomics, remains crucial and has to be carefully evaluated and tested with an appropriate experimental design.[53] On the other hand, freshwater fishes are already considered as good indicators of the ecological status of aquatic ecosystems, as various fish biodiversity indexes have already been proposed and are still under-evaluation at local or large scales.[8]
However, for ethical concerns, fish bio-monitoring would also gain at being less invasive and deleterious for the organisms. To this end, non-lethal mucus sampling has been investigated by LC-MS and has demonstrated the remarkable informativeness of this approach for environmental studies.[51] In parallel, recent attempts have also been made concerning the interest of planktonic or benthic microbial communities characterized by metabarcoding for water quality bio-indication,[16] but these efforts still remain explorative.

The main challenge we face remains the development of more integrative approaches connecting chemical, biological and ecological evaluations, in the context of anthropized natural environments experiencing multi-stressor pressures. A major caveat of the use of fish environmental metabolomics may now be the lack of dedicated databases[54] fulfilled by studies considering together different species, populations, development stages, seasons and environments. This pre-requisite knowledge could be used to provide baseline reference data that would support machine learning or artificial neural network tools for the training of decision-making models.

ASSOCIATED CONTENT

Supporting information

390 Map of sampled lakes (Supplementary figure S1); individual plots of PLS-DA performed with intestine
391 metabarcoding of 16S rDNA (A and D), Liver NMR (B and E) and LC-MS (C and F) metabolomics
392 datasets for *Perca* and *Lepomis*, discriminated according to the "species" (A-C) and "sampling lakes"
393 (D-F) parameters (Supplementary figure S2); individual plots of PCA, heatmap with hierarchical
394 clustering and PLS-DA testing the difference among the different lakes according to the intestinal
395 microbiomes of *Perca* (A-C) and *Lepomis* (D-F) presenting insufficient accuracy, predictability and
396 quality performances for sampling lake discrimination (Supplementary figure S3); individual plots of
397 PCA, heatmap with hierarchical clustering and PLS-DA testing the differences among lakes according
398 to the liver NMN metabolomes of *Perca* (A-C) and *Lepomis* (D-F) presenting insufficient accuracy,
399 predictability and quality performances for sampling lake discrimination (Supplementary figure S4);
400 individual plots of PLS-DA, testing the differences among lakes according to the LC-MS metabolomes
401 of *Perca*, with corresponding best VIP list and metabolite box-plots (Supplementary figure S5);
402 individual plots of PLS-DA, testing the differences among lakes according to the LC-MS metabolomes
403 of *Lepomis*, with corresponding best VIP list and metabolite box-plots (Supplementary figure S6);
404 Individual plots of PCA for LC-MS metabolomics of *Perca* (A) and *Lepomis* (B) representing the
405 different lake groups highlighted by hierachical classication as shown on Fig. 2 (Supplementary figure
406 S7); box-plots of top-16 best VIP metabolites according to PLS-DA performed on the lake groups
407 indicated by hierarchical classification, as shown on Figure 2, for *Perca* (A) and *Lepomis* (B)
408 (Supplementary figure S8); molecular networking of *Perca* and *Lepomis* metabolomes characterized
409 by GNPS and *t*-SNE algorithms (Supplementary figure S9); Integration of the NMR and LC-MS
410 metabolomics of *Perca* using MixOmics illustrates the comparable PLS-DA patterns of the datasets (a-
411 b) and the global correlation of respective VIPs (correlation score = 0.79; c) (Supplementary figure
412 S10); Integration of the NMR and LC-MS metabolomics of *Lepomis* using MixOmics illustrates the
413 comparable PLS-DA patterns of the datasets (a-b) and the global correlation of respective VIPs
414 (correlation score = 0.83; c) (Supplementary figure S11); phytoplankton composition estimated by
415 BBE measurment of the eight lake sub-surface water (a), corresponding cyanobacteria relative
416 composition for Fon, Tri, Var and Ver (B), and molecular networking of metabolites extracted from
417 the filtered biomass of the respective water of the eight lakes generated with *t*-SNE algorithm, with
418 cyanobacteria peptide clusters indicated in bold and microcystin cluster indicated in red (C)
419 (Supplementary figure S12) (PDF)
420 List and numbers of collected fishes (Supplementary table S1); list of significant VIP discriminating
421 the intestinal microbiomes (Supplementary table S2), the liver NMR metabolomes (Supplementary
422 table S3) and the liver LC-MS metabolomes (Supplementary table S4) of *Perca* and *Lepomis* species
423 determined by PLS-DA; list of significant VIP (>1.5) discriminating the liver LC-MS metabolomes of
424 *Perca* and *Lepomis* from the different lakes determined by PLS-DA (Supplementary table S5) (XLS)

AUTHOR INFORMATION

Corresponding author

Benjamin Marie – 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.

https://orcid.org/0000-0001-9880-5541; Phone : +33 1 40 79 32 12 ; Email : bmarie@mnhn.fr

Authors

Maud Mouchet – UMR 7204 CNRS/SU/MNHN, Centre d'Ecologie et des Sciences de la Conservation (CESCO), 43 rue Buffon - CP 135, 75231 Paris Cedex 05, France.

https://orcid.org/0000-0001-5939-6802

Julie Leloup – UMR 7618 iEES, Paris Sorbonne Université - 4 place Jussieu - 75252 Paris Cedex 05 - France

https://orcid.org/0000-0002-2777-284X

Aurélie Bonin - UMR 5553, CNRS/Université Joseph Fourrier, Laboratoire d'Écologie Alpine (LECA), BP 53, 38041 Grenoble Cedex-9, France.

https://orcid.org/0000-0001-7800-8609

Alison Gallet - 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.

Author contributions

B.M. and M.M. designed experiment. B.M., A.B. and M.M. performed sample analysis. B.M., A.G. and J.L. analysed data. B.M. wrote the manuscript. All authors have given approval to the final version of the manuscript.

Notes

The authors declare not conflict of interest

ACKNOWLEDGMENTS

We would like to thanks B. Sotton for his collaboration on NMR data acquisition and treatment. We thank Kenin Tambosco for his technical assitance. We are also greatful to A. Paris for its support and prolific discussions. This work was supported by grants from CNRS (Défi ENVIROMICS "Toxcyfish"

project) and from Sorbone Université (CARESE-SU J15R323). 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 bio-organique, Muséum National d'Histoire Naturelle, Paris, France. This work benefitted from the French GDR "Aquatic Ecotoxicology" framework which aims at fostering stimulating scientific discussions and collaborations for more integrative approaches.

REFERENCES

(1) Lorenz, C. M. Bioindicators for ecosystem management, with special reference to freshwater systems. In *Trace Metals and other Contaminants in the Environment* (**2003**, Vol. 6, pp. 123-152). Elsevier.

(2) Reyjol, Y.; Argillier, C.; Bonne, W.; Borja, A.; Buijse, A. D.; Cardoso, A. C.; ... Prat, N. Assessing the ecological status in the context of the European Water Framework Directive: where do we go now? *Science of the Total Environment* **2014**, *497*, 332-344.

(3) Schimel, J. Microbial ecology: Linking omics to biogeochemistry. *Nature microbiology* **2016**, *1*(2), 1-2.

(4) Paul, V. J., Freeman, C. J., & Agarwal, V. Chemical ecology of marine sponges: new opportunities through "-Omics". *Integrative and Comparative Biology* **2019**, *59*(4), 765-776.

(5) Harris, J. H. The use of fish in ecological assessments. *Australian Journal of Ecology* **1995**, *20*(1), 65-80.

(6) McClanahan, T. R.; Schroeder, R. E.; Friedlander, A. M.; Vigliola, L.; Wantiez, L.; Caselle, J. E.; ...; Oddenyo, R. M. Global baselines and benchmarks for fish biomass: comparing remote reefs and fisheries closures. *Marine Ecology Progress Series*, **2019**, *612*, 167-192.

(7) Hamilton, P.; Cowx, I.; et al.; Tyler, C. R. Population-level consequences for wild fish exposed to sublethal concentrations of chemicals–a critical review. *Fish and Fisheries* **2016**, *17*(3), 545-566.

(8) Poikane, S.; Ritterbusch, D.; Argillier, C.; Białokoz, W.; Blabolil, P.; Breine, J., ...; Noges, P. Response of fish communities to multiple pressures: development of a total anthropogenic pressure intensity index. *Science of the Total Environment* **2017**, *586*,

502-511.

(9) Birk, S.; Bonne, W.; Borja, A.; Brucet, S.; Courrat, A.; Poikane, S.; ... & Hering, D. Three hundred ways to assess Europe's surface waters: an almost complete overview of biological methods to implement the Water Framework Directive. *Ecological Indicators* **2012**, *18*, 31-41.

(10) Birk, S.; Willby, N. J.; Kelly, M. G., Bonne, W., Borja, A., Poikane, S., & Van de Bund, W. Intercalibrating classifications of ecological status: Europe's quest for common management objectives for aquatic ecosystems. *Science of the Total Environment* **2013**, *454*, 490-499.

(11) Shin, Y. J.; Houle, J. E.; Akoglu, E.; Blanchard, J. L.; Bundy, A.; Coll, M.; ...; Salihoglu, B. The specificity of marine ecological indicators to fishing in the face of environmental change: a multi-model evaluation. *Ecological Indicators* **2018**, *89*, 317-326.

(12) Esin, E. V.; Edosov, A. E. The effect of chronic volcanic pollution on the morphometric characteristics of juvenile Dolly Varden (*Salvelinus malma* W.) on the Kamchatka peninsula. *Hydrobiologia* **2016**, *783*(1), 295-307.

(13) Zhang, Z.; Wang, J.; Pan, Z.; Zhang, Y.; Zhang, X.; Tian, H.; ... ; Ru, S. Distribution of vitellogenin in Japanese flounder (*Paralichthys olivaceus*) for biomarker analysis of marine environmental estrogens. *Aquatic Toxicology* **2019**, *216*, 105321.

(14) Bahamonde, P. A.; Munkittrick, K. R.; Martyniuk, C. J. Intersex in teleost fish: are we distinguishing endocrine disruption from natural phenomena?. *General and comparative endocrinology*, **2013**, *192*, 25-35.

(15) Bahamonde, P. A.; Feswick, A.; Isaacs, M. A.; Munkittrick, K. R.; Martyniuk, C. J. Defining the role of omics in assessing ecosystem health: Perspectives from the Canadian environmental monitoring program. *Environmental toxicology and chemistry*, **2016**, *35*(1), 20-35.

(16) Cordier, T.; Alonso-Sáez, L.; Apothéloz-Perret-Gentil, L.; Aylagas, E.; Bohan, D. A.; Bouchez, A.; ... Keeley, N. Ecosystems monitoring powered by environmental genomics: a review of current strategies with an implementation roadmap. *Molecular Ecology*. **2020**, DOI: 10.1111/mec.15472

(17) Marie, B. 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,* 139701.

(18) Maloufi, S.; Catherine, A.; Mouillot, D.; Louvard, C.; Couté, A.; Bernard, C.; Troussellier, M. (Environmental heterogeneity among lakes promotes hyper β-diversity across phytoplankton communities. *Freshwater Biology, 61*(5), 633-645.

(19) Sotton, B.; Paris, A.; Le Manach, S.; Blond, A.; Duval, C.; Qiao, Q.; ... ; Marie, B. Specificity of the metabolic signatures of fish from cyanobacteria rich lakes. *Chemosphere* **2019**, *226*, 183-191.

(20) Mouchet, M.; Bouvier, C.; Bouvier, T.; Troussellier, M.; Escalas, A.; Mouillot, D. Genetic difference but functional similarity among fish gut bacterial communities through molecular and biochemical fingerprints. *FEMS Microbiology Ecology* **2012**, *79*, 568–580, https://doi.org/10.1111/j.1574-6941.2011.01241.x

(21) Escudié, F.; Auer, L., Bernard, M.; Mariadassou, M.; Cauquil, L.; Vidal, K.; ... Pascal, G. FROGS: find, rapidly, OTUs with galaxy solution. *Bioinformatics* **2018**, *34*(8), 1287-1294.

(22) Chong, J.; Liu, P.; Zhou, G.; Xia, J. Using MicrobiomeAnalyst for comprehensive statistical, functional, and meta-analysis of microbiome data. *Nature Protocols* **2020**, 15, 799-821.

(23) Hao, J.; Liebeke, M.; Astle, W.; De Iorio, M.; Bundy, J. G.; Ebbels, T. M. D. Bayesian deconvolution and quantification of metabolites in complex ^{1}D NMR spectra using BATMAN. *Nature protocols* **2014**, 9(6), 1416–1427.

(24) Kim Tiam, S.; Gugger, M.; Demay, J.; Le Manach, S.; Duval, C.; Bernard, C.; Marie, B. Insights into the Diversity of Secondary Metabolites of *Planktothrix* Using a Biphasic Approach Combining Global Genomics and Metabolomics. *Toxins* **2019**, *11*(9), 498.

(25) Rohart, F.; Gautier, B.; Singh, A.; Lê Cao, K. A. mixOmics: An R package for 'omics feature selection and multiple data integration. *PLoS computational biology* **2017**, *13*(11), e1005752.

(26) Chong, J.; Wishart, D. S.; Xia, J. Using metaboanalyst 4.0 for comprehensive and integrative metabolomics data analysis. *Current protocols in bioinformatics* **2019**, *68*(1):e86.

(27) Duperron, S.; Halary, S.; Gallet, A.; Marie, B. Microbiome-aware ecotoxicology of organisms: relevance, pittfalls and challenges. *Frontiers in Public Health* **2020**, 8, 407.

554 (28) Selber-Hnatiw, S., Rukundo, B., Ahmadi, M., Akoubi, H., Al-Bizri, H., Aliu, A. F., et al.
555 (2017). Human Gut Microbiota: Toward an Ecology of Disease. *Front. Microbiol.* 8,
556 1265. doi:10.3389/fmicb.2017.01265.
557 (29) Andam, C. P. Clonal yet different: understanding the causes of genomic heterogeneity in
558 microbial species and impacts on public health. *mSystems*. **2019**, *4*:e00097-19.
559 (30) Sawabe, T.; Kita-Tsukamoto, K.; Thompson, F.L. Inferring the evolutionary history of
560 vibrios by means of multilocus sequence analysis. *J Bacteriol.* **2007**, 189:7932–6.
561 (31) Martin, J. C.; Maillot, M.; Mazerolles, G.; Verdu, A.; Lyan, B.; Migne, C.; ... ; Manach, C.
562 Can we trust untargeted metabolomics? Results of the metabo-ring initiative, a large-
563 scale, multi-instrument inter-laboratory study. *Metabolomics* **2015**, *11*(4), 807-821.
564 (32) Roques, S.; Deborde, C.; Richard, N., Marchand; Y., Larroquet, L.; Prigent, S.; ...;
565 Fauconneau, B. Proton-NMR Metabolomics of Rainbow Trout Fed a Plant-Based Diet
566 Supplemented with Graded Levels of a Protein-Rich Yeast Fraction Reveal Several
567 Metabolic Processes Involved in Growth. *J. Nutrition* **2020**, *150*(9), 2268-2277.
568 (33) Emwas, A. H.; Roy, R.; McKay, R. T.; Tenori, L.; Saccenti, E.; Gowda, G. A.; ... Wishart, D.
569 S. NMR spectroscopy for metabolomics research. *Metabolites* **2019**, *9*(7), 123.
570 (34) da Silva, R. R.; Dorrestein, P. C.; Quinn, R. A. Illuminating the dark matter in
571 metabolomics. *Proceedings of the National Academy of Sciences* **2015**, *112*(41),
572 12549-12550.
573 (35) Lohr, K. E.; Khattri, R. B.; Guingab-Cagmat, J.; Camp, E. F.; Merritt, M. E.; Garrett, T. J.,
574 Patterson, J. T. Metabolomic profiles differ among unique genotypes of a threatened
575 Caribbean coral. *Scientific reports* **2019**, *9*(1), 1-11.
576 (36) Beyoğlu, D.; Idle, J. R. (2020). Metabolomic and lipidomic biomarkers for premalignant
577 liver disease diagnosis and therapy. *Metabolites* **2020**, *10*(2), 50.
578 (37) Gika, H.; Virgiliou, C.; Theodoridis, G.; Plumb, R. S.; Wilson, I. D. Untargeted LC/MS-
579 based metabolic phenotyping (metabonomics/metabolomics): The state of the art.
580 *Journal of Chromatography B* **2019**, *1117*, 136-147.
581 (38) Marchand, J.; Martineau, E.; Guitton, Y.; Le Bizec, B.; Dervilly-Pinel, G.; Giraudeau, P. A
582 multidimensional ^{1}H NMR lipidomics workflow to address chemical food safety
583 issues. *Metabolomics* **2018**, *14*(5), 60.
584 (39) Dreier, D. A.; Bowden, J. A.; Aristizabal-Henao, J. J.; Denslow, N. D.; Martyniuk, C. J.
585 Ecotoxico-lipidomics: An emerging concept to understand chemical-metabolic

586 relationships in comparative fish models. *Comparative Biochemistry and Physiology*
587 *Part D: Genomics and Proteomics* **2020**, *36,* 100742.
588 (40) Singh, A.; Shannon, C. P.; Gautier, B.; Rohart, F.; Vacher, M.; Tebbutt, S. J.; Lê Cao, K. A.
589 DIABLO: an integrative approach for identifying key molecular drivers from multi-
590 omics assays. *Bioinformatics* **2019**, *35*(17), 3055-3062.
591 (41) Zhang, X. W.; Li, Q. H.; Dou, J. J. Mass spectrometry-based metabolomics in health and
592 medical science: a systematic review. *RSC Advances*, **2020**, *10*(6), 3092-3104.
593 (42) Le Manach, S.; Sotton, B.; Huet, H.; Duval, C.; Paris, A.; Marie, A.; ... ; Marie, B.
594 (Physiological effects caused by microcystin-producing and non-microcystin
595 producing *Microcystis aeruginosa* on medaka fish: A proteomic and metabolomic
596 study on liver. *Environmental pollution* **2018**, *234*, 523-537.
597 (43) Huang, S. S.; Benskin, J. P.; Veldhoen, N.; Chandramouli, B.; Butler, H.; Helbing, C. C.;
598 Cosgrove, J. R. A multi-omic approach to elucidate low-dose effects of xenobiotics in
599 zebrafish (*Danio rerio*) larvae. *Aquatic toxicology* **2017***, 182*, 102-112.
600 (44) Ekman, D. R.; Skelton, D. M.; Davis, J. M.; Villeneuve, D. L.; Cavallin, J. E.; Schroeder, A.;
601 ... Collette, T. W. Metabolite profiling of fish skin mucus: a novel approach for
602 minimally-invasive environmental exposure monitoring and surveillance.
603 *Environmental Science & Technology* **2015**, *49*(5), 3091-3100.
604 (45) Gil-Solsona, R., Nácher-Mestre, J., Lacalle-Bergeron, L., Sancho, J. V., Calduch-Giner, J.
605 A., Hernández, F., & Pérez-Sánchez, J. Untargeted metabolomics approach for
606 unraveling robust biomarkers of nutritional status in fasted gilthead sea bream
607 (*Sparus aurata*). *PeerJ* **2017**, *5*, e2920.
608 (46) Bundy, J. G.; Davey, M. P.; Viant, M. R. Environmental metabolomics: a critical review
609 and future perspectives. *Metabolomics* **2009**, *5*(1), 3.
610 (47) Lankadurai, B. P.; Nagato, E. G.; Simpson, M. J. Environmental metabolomics: an
611 emerging approach to study organism responses to environmental stressors.
612 *Environmental Reviews* **2013***, 21*(3), 180-205.
613 (48) Cappello, T.; Brandão, F.; Guilherme, S.; Santos, M. A.; Maisano, M.; Mauceri, A.; …
614 Pereira, P. Insights into the mechanisms underlying mercury-induced oxidative stress
615 in gills of wild fish (*Liza aurata*) combining ^{1}H NMR metabolomics and conventional
616 biochemical assays. *Science of the Total Environment* **2016**, *548*, 13-24.
617 (49) Wei, F.; Sakata, K.; Asakura, T.; Date, Y.; Kikuchi, J. Systemic homeostasis in

metabolome, ionome, and microbiome of wild yellowfin goby in estuarine ecosystem. *Scientific reports* **2018**, *8*(1), 1-12.

(50) Meador, J. P.; Bettcher, L. F.; Ellenberger, M. C.; Senn, T. D. Metabolomic profiling for juvenile Chinook salmon exposed to contaminants of emerging concern. *Science of The Total Environment* **2020**, *747*, 141097.

(51) Reverter, M.; Sasal, P.; Banaigs, B.; Lecchini, D.; Lecellier, G.; Tapissier-Bontemps, N. Fish mucus metabolome reveals fish life-history traits. *Coral Reefs* **2017**, *36*(2), 463-475.

(52) Goode, K. L.; Dunphy, B. J.; Parsons, D. M. Environmental metabolomics as an ecological indicator: Metabolite profiles in juvenile fish discriminate sites with different nursery habitat qualities. *Ecological Indicators*, **2020**, *115*, 106361.

(53) Pomfret, S. M.; Brua, R. B.; Izral, N. M.; Yates, A. G. Metabolomics for biomonitoring: an evaluation of the metabolome as an indicator of aquatic ecosystem health. *Environmental Reviews*, **2020**, *28*(1), 89-98.

(54) Viant, M. R.; Ebbels, T. M.; Beger, R. D.; Ekman, D. R.; Epps, D. J.; Kamp, H.; ... ; Rocca-Serra, P. Use cases, best practice and reporting standards for metabolomics in regulatory toxicology. *Nature communications* **2019**, *10*(1), 1-10.

Legends of figures

Fig. 1. Visualization of the dataset structuration for intestine metabarcoding 16S rDNA (A), liver NMR (B) and LC-MS (C) metabolomics on heatmaps with hierarchical clustering according to Euclidean distance for *Perca* and *Lepomis* fishes collected during the 7-10th of September, 2015 within eight pounds of the peri-urban Paris' area.

Fig. 2. PCA (A and D), heatmap with hierarchical classification (B and E) and PLS-DA (C and F) of LC-MS metabolome of *Perca* (A-C) and *Lepomis* (D-F) liver according to the "sampling lake" parameter.

Table 1. Annotated VIP (score >2) of respective *Perca* and *Lepomis* LC-MS metabolomics dataset according to PLD-DA performed on lake, and corresponding variation among the here lake groups highlighted by hierarchical classification (according to Figure 2B and 2E).

Fig. 1.

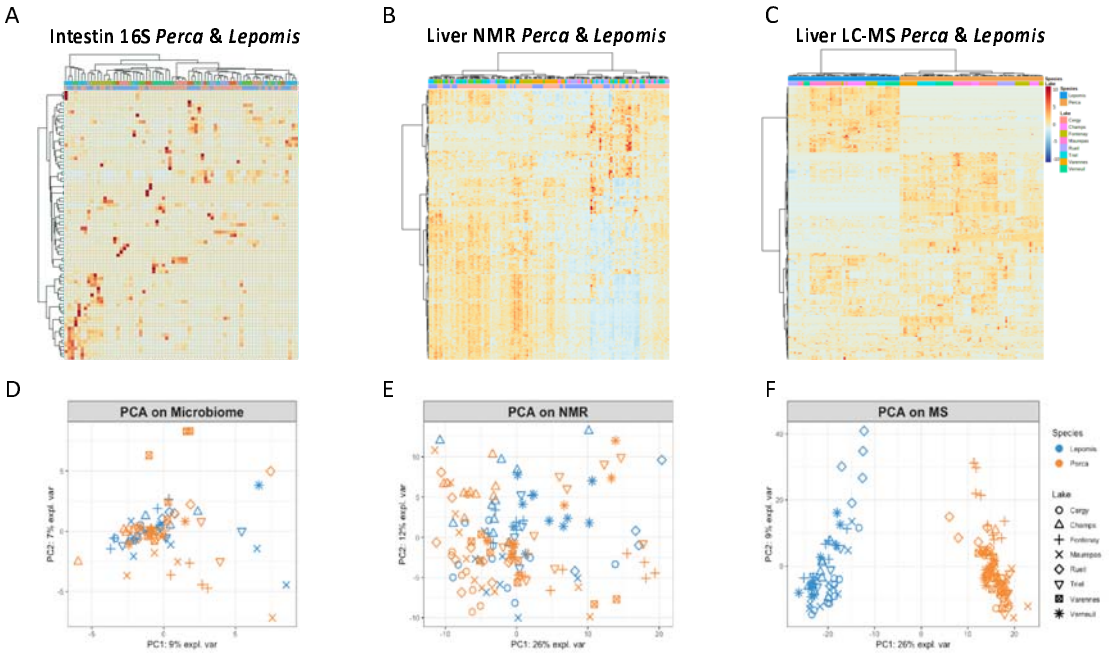

A
Intestin 16S *Perca* & *Lepomis*
B
Liver NMR *Perca* & *Lepomis*
C
Liver LC-MS *Perca* & *Lepomis*
D
PCA on Microbiome
PC1: 9% expl. var
PC2: 7% expl. var
E
PCA on NMR
PC1: 26% expl. var
PC2: 12% expl. var
F
PCA on MS
PC1: 26% expl. var
PC2: 9% expl. var
Species
Lepomis
Perca
Lake
Cergy
Champs
Fontenay
Maumpos
Ruel
Trel
Varennes
Vernouil

Fig. 2.

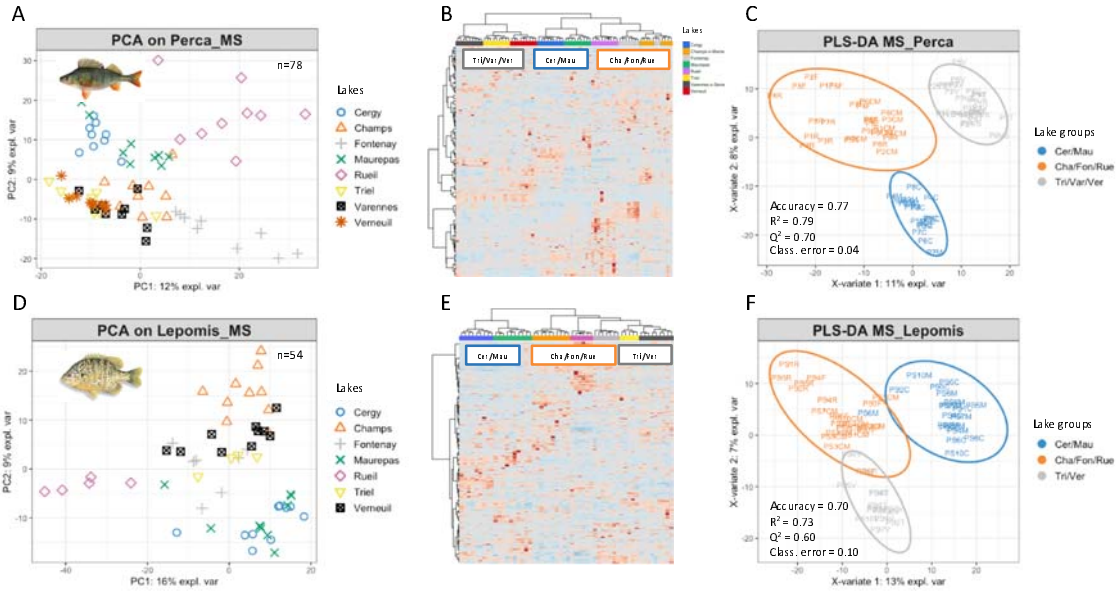
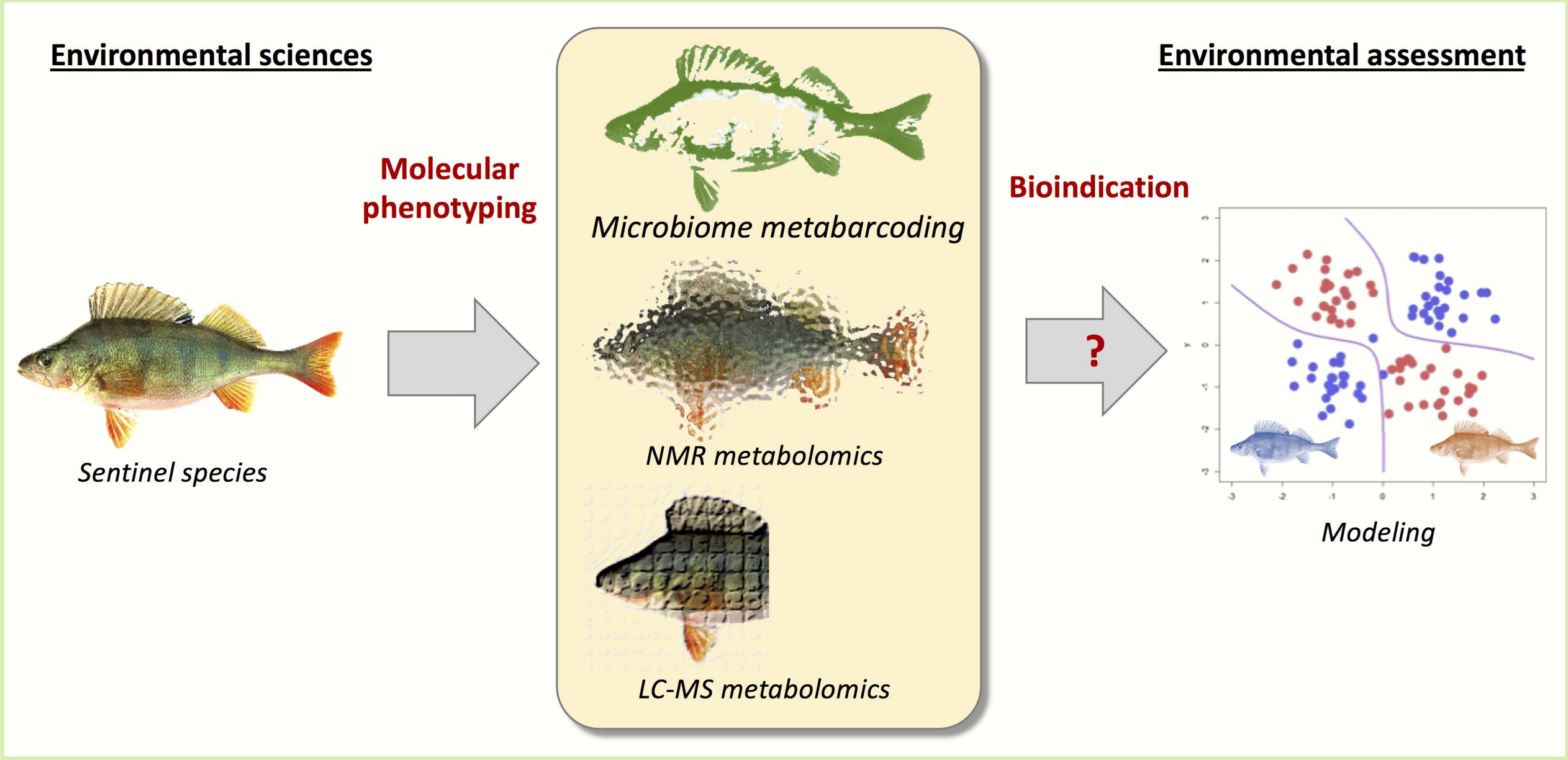

A
PCA on Perca_MS
n=78
PC1: 12% expl. var
PC2: 9% expl. var
Lakes
Cergy
Champs
Fontenay
Maurepas
Rueil
Triel
Varennes
Verneuil
B
Tri/Var/Ver
Cer/Mau
Cha/Fon/Rue
Lakes
C
PLS-DA MS_Perca
X-variate 1: 11% expl. var
X-variate 2: 8% expl. var
Accuracy = 0.77
R² = 0.79
Q² = 0.70
Class. error = 0.04
Lake groups
Cer/Mau
Cha/Fon/Rue
Tri/Var/Ver
D
PCA on Lepomis_MS
n=54
PC1: 16% expl. var
PC2: 9% expl. var
Lakes
Cergy
Champs
Fontenay
Maurepas
Rueil
Triel
Verneuil
E
Cer/Mau
Cha/Fon/Rue
Tri/Ver
F
PLS-DA MS_Lepomis
X-variate 1: 13% expl. var
X-variate 2: 7% expl. var
Accuracy = 0.70
R² = 0.73
Q² = 0.60
Class. error = 0.10
Lake groups
Cer/Mau
Cha/Fon/Rue
Tri/Ver

Table 1.

MW (Da)	Annotation	VIP comp. 1 *Perca*	VIP comp. 1 *Lepomis*
267.09601	Adenosine	8.86 (?>?=?)	15.23 (?>?=?)
612.1507	Glutathione (oxidized)	6.44 (?>?>?)	13.32 (?>?>?)
307.08355	Glutathione (reduced)	7.02 (?>?>?)	2.28 (?>?>?)
383.10736	*Analog*: Cyanidin arabinosine	4.25 (?>?>?)	2.67 (?>?>?)
217.13223	*Analog*: Carnitine	1.55 (?>?>?)	4.18 (?>?>?)
229.08904	Ergothioneine	0.10	4.03 (?>?>?)
626.16763	*Analog*: Glutathione (oxidized)	1.58 (?>?=?)	3.76 (?>?>?)
663.10912	Nicotinamide adenine dinucleotide	0.4	3.61 (?>?=?)
203.1173	Acetyl-carnitine	3.50 (?=?>?)	1.82 (?=?>?)
567.33206	Lysophosphatidylcholine 22:6	2.99 (?>?=?)	0.78
250.02948	*Analog*: N-acetylglutamic acid	2.98 (?>?=?)	2.30 (?>?=?)
429.14971	*Analog*: Adenosine 5'-monophosphate	0.19	2.96 (?>?=?)
463.07428	Adenylo succinic acid	2.87 (?>?=?)	2.90 (?>?>?)
248.14845	*Analog*: N-acetylglutamic acid	0.48	2.78 (?>?>?)
268.0806	*Analog*: Adenosine	1.60 (?>?=?)	2.72 (?>?=?)
265.0952	Tryptophan	0.19	2.65 (?>?=?)
347.06212	Adenosine-5' mono-phosphate	1.96 (?>?=?)	2.55 (?>?=?)


Environmental sciences
Molecular phenotyping
Sentinel species
Microbiome metabarcoding
NMR metabolomics
LC-MS metabolomics
Bioindication
?
Environmental assessment
Modeling