MR. FRANÇOIS-ÉTIENNE SYLVAIN (Orcid ID : 0000-0002-1763-0434)

Article type : Original Article

Amazon fish bacterial communities show structural convergence along widespread hydrochemical gradients

Running title: Convergent microbiotas along hydrochemical gradients

Sylvain, François-Étienne[1]*; Holland, Aleicia[2]; Audet-Gilbert, Émie[1]; Val, Adalberto Luis[3]; Derome, Nicolas[1]

1. Institut de Biologie Intégrative et des Systèmes, Université Laval, 1030 avenue de la Médecine, Québec (QC), G1V 0A6, Canada
2. La Trobe University, School of Life Science, Department of Ecology, Environment and Evolution, Albury/Wodonga Campus, Vic, Australia
3. Instituto Nacional de Pesquisas da Amazônia (INPA), Laboratório de Ecofisiologia e Evolução Molecular, Manaus, AM, 69067-375, Brazil

*__Corresponding author__:

François-Étienne Sylvain

Email : francois-etienne.sylvain.1@ulaval.ca

Abstract

The world's richest freshwater fish community thrives in gradients of contrasting environments in Amazonia, ranging from ion-poor acidic black waters, to ion-rich circumneutral white waters. These hydrochemical gradients structure Amazonian fish assemblages via ecological speciation events. Fish bacterial communities contain an important genetic heritage essential for their hosts' survival, and are also involved in adaptive divergence via niche adaptation processes, but the extent to which they evolve in response to hydrochemical gradients in Amazonia is unknown. Here we investigated bacterial communities (gut and skin mucus) of two ecologically and phylogenetically divergent host species (*Mesonauta festivus* and *Serrasalmus rhombeus*) distributed throughout these hydrochemical gradients. The goal was to characterize intra and interspecific Amazonian fish microbiome variations across multiple scales. Using a 16S metabarcoding approach, we investigated the microbiota of 43 wild *M. festivus*, 32 *S. rhombeus*, and seven water samples, collected at seven sampling sites encompassing both water colors. Taxonomical structures of bacterial communities from both host species were significantly correlated to the environmental continua of magnesium, sodium, dissolved organic carbon, calcium, dissolved O_2, pH, potassium, hardness and chloride. Analysis of discriminating features in community structures across multiple scales demonstrated intra and interspecific structural parallelisms in the response to the hydrochemical gradients. Together, these parallelisms suggest the action of selection on bacterial community structures along Amazonian hydrochemical gradients. Functional approaches along with

reciprocal transplant experiments will provide further insights on the potential contribution of Amazonian fish microbiomes in host adaptation and ecological speciation events.

Keywords: Fish ; Microbiota ; Microbiome ; Amazon ; Gut ; Black water

Introduction

The Amazon Basin contains the world's richest freshwater Teleost fish assemblage, housing more than 3000 species (Van der Sleen and Albert 2018). This colossal fish diversity is found throughout a spectrum of hydrochemically contrasted black and white water environments (Sioli, 1984). These two water colors are also found in several countries worldwide (e.g. in Australia, United-States, Indonesia, Canada). White waters are eutrophic (nutrient-, oxygen- and ion-rich), turbid, and have a circumneutral pH (Dal Pont, 2017). Black waters are oligotrophic (nutrient-, oxygen- and ion-poor) and transparent, yet tinted and acidified by tannins and humic acids from surrounding forest vegetation (pH 3.8 - 5.0) (Johannsson *et al*. 2017). Thus, black waters are seemingly far more challenging for fish physiology (e.g. for osmoregulation and ionoregulation processes) than white waters (Val & Almeida-Val, 1995). Despite this, more than 1000 fish species inhabit black water streams in the Amazon, and are thus adapted to these extreme habitats.

A few studies have shown that water color gradients structure Amazonian fish assemblages via the adaptive divergence of fishes to water colors, ultimately leading to ecological speciation events (Cooke *et al*. 2012a,b,c,d, 2014; Beheregaray *et al*. 2015). Ecological gradients and ecotones are commonly known to generate biodiversity via divergent natural selection (Endler 1973; Smith *et al.* 1997, 2001). Bacterial communities of Amazonian fishes

may play prominent roles in their hosts' adaptive divergence to water color via niche adaptation processes; fish microbiotas play key roles to support physiological functions essential to the hosts' survival and metabolism (nutriment assimilation, immune system modulation, ion absorption, amino acid synthesis, detoxification and more) in accordance with the specific selection pressures in their environment (reviewed in Ghanbari *et al.* 2015). Furthermore, fish bacterial communities contain an important genetic heritage essential for their hosts' survival - studies on other vertebrates have shown that there are about 150 times more unique genes in the gut bacterial community than in the host's genome (Zhu *et al*. 2010). Thus, there is a potentially extensive functional genes pool, fundamental to the survival and adaptation of host-microbiota metaorganisms (also termed "holobionts"), that is comprised in the genomes of host-associated bacteria (Margulis and Fester 1991). While the nuclear and mitochondrial genomes of several wild Amazonian fish species have been documented (Cooke *et al*. 2012a,b,c,d, 2014), their resident bacterial genomes have never been investigated and the extent to which fish bacterial communities change along the hydrochemical gradients in Amazonia is still unknown.

Here we investigated the bacterial communities (gut and skin mucus) of two ecologically and phylogenetically divergent host species (flag cichlids, *Mesonauta festivus*; and black piranhas, *Serrasalmus rhombeus*) distributed throughout these hydrochemical gradients (Pires *et al*. 2015; Van der Sleen and Albert, 2018). The main objective of our study was to characterize Amazonian fish microbiota variations across multiple scales: among different species, different tissues, from both water colors, and from replicate collecting sites within each water color. We used a metabarcoding approach targeting the V3-V4 region of the 16S gene to analyze the gut and skin mucus bacterial communities of 43 wild flag cichlids, 32 black piranhas, and seven water samples, collected at seven sampling sites encompassing both water colors. We chose these two fish species for (1) their abundance at all sampling sites in both black and white waters; (2) their contrasting ecologies (black piranhas are

piscivorous and feed opportunistically on all encountered fish, while flag cichlids are detritivorous and mostly feed on dead plant fragments and periphyton) (Van der Sleen and Albert, 2018); and (3) their divergent phylogeny (the *Serrasalmidae* diverged from the *Cichlidae* family 229.9 million years ago (Hedges *et al*. 2015)). Our results enabled us to respond to two important questions:

Question #1: how do fish bacterial communities change along the hydrochemical gradients?

To date, the responses and resilience of fish bacterial communities to physicochemical variations have been assessed in controlled conditions: Boutin *et al*. (2013) studied the impact of anoxia on the microbiota of *Salvelinus fontinalis*, while Sylvain *et al*. (2016) studied the impact of pH variations on the microbiota of *Colossoma macropomum*. In both cases, the selective pressure (anoxia, acidity) drove parallel changes in microbiota taxonomic composition (between replicate treatments), both in terms of diversity and structure. However, the response of fish bacterial communities to a gradient of physicochemical environments in the wild have never been studied (but see Llewellyn *et al*. 2015). It is paramount to study wild fish bacterial consortia, as they evolve in complex environments shaped by water physicochemical parameters, trophic networks and biotic/abiotic disturbances, which are documented to drive microbiota compositional shifts (reviewed in Ghanbari *et al.* 2015) in controlled conditions. Furthermore, wild Amazonian fishes are extensively exploited by local fishing communities, which have an annual *per capita* fish consumption rate exceeding 30 kg - among the highest in the world (FAO, 2011). Thus, it is important to understand the forces driving the structure of wild Amazonian fish microbiotas, as these microbial consortia potentially play crucial roles in the survival, the growth, the local adaptation and the geographical repartition of native fishes - the primary protein source in the region.

We hypothesized that a significant response of Amazonian fish microbiotas to water color would be observed from both gut and skin mucus bacterial communities. Indeed, water physicochemical parameters such as pH, salinity, conductivity, [NO_3-N] and [PO_4^3-P] are known drivers of gut microbiota structure in several fish species (Sylvain *et al*. 2016; Llewellyn *et al*. 2015; Giatsis *et al*. 2015), while pH and dissolved oxygen are known drivers of fish skin mucus microbiota (Boutin *et al*. 2013; Sylvain *et al*. 2016). To test our hypothesis, we used a multivariate statistical approach to explore differences in community structure and diversity at different scales (i.e. tissue, species and water color scales). Then, we investigated further the effect of the hydrochemical gradients on community structure by exploring the correlations between specific water physicochemical parameters and bacterial taxa.

Question #2: are there intra and interspecific bacterial markers of convergent evolution associated with water color?

The numerous processes driving the establishment, function and maintenance of bacterial communities are still poorly understood (Prosser *et al*. 2007; Widder *et al*. 2016). It remains unclear whether the development of bacterial communities is inherently predictable (controlled by deterministic factors), or unpredictable (controlled by stochastic factors) (Langenheder and Szekely, 2011; Pagaling *et al*. 2014; Koskella *et al.* 2017). Multiple (thus unpredictable) states may also be possible, stemming from feedbacks between species or species-environment interactions (Scheffer *et al*. 2001). In these cases, the observed state is determined by stochastic factors. Parallel/convergent community structures in similar ecosystems inform us on the predictability of possible community states. Studies investigating parallel evolution on fish microbiota in natural conditions are still scarce,

and, at first sight, their results appear contradictory and diet-specific (Sullam *et al.* 2015; Baldo *et al.* 2017; Sevellec *et al.* 2018).

The physiological tolerance and geographical distribution of black piranhas and flag cichlids at various sites in both water colors make them great models to explore intra and interspecific structural parallelisms in the response of their bacterial communities to the widespread hydrochemical gradients in Amazonia. We hypothesized that intra and interspecific convergence in bacterial community structure exist among fish found in the same water color, as we expect the same bacterial taxa to be positively selected in similar environments with comparable selection pressures (Conte *et al*. 2012; Bailey *et al*. 2015, 2017). To test this hypothesis, we investigated the discriminating features in community structures across multiple scales - including identification of core taxa, a multivariate dispersion analysis, along an analysis of interspecific convergent biomarkers - to evidence intra and interspecific structural parallelisms in the response of bacterial communities to the hydrochemical gradients.

Overall, our results showed a significant response of bacterial communities from both host species to the hydrochemical gradients, comprising a continuum of 10 environmental physicochemical parameters. Based on the structural responses of the sampled communities to the hydrochemical gradients, we identified several candidate intra and interspecific bacterial markers of convergent evolution associated with water color.

Materials and Methods

Ethics Statement

This study was carried out in accordance with the recommendations of the Ethics Committee for the Use of Animals of the *Instituto Nacional de Pesquisas da Amazonia* (INPA). The protocol (number 026/2015 as of Dec 18th, 2015) was approved by the Ethics Committee for the Use of Animals of INPA.

Fish sampling

The fish sampling was done between 05/11/2015 and 25/11/2015 at seven different locations of the Brazilian Amazon close to Manaus (AM, Brazil): (1) Catalão Lake (white water); (2) two sites in the Solimões River near Jacurutu Island (white water); (3) two sites in the Manacapuru River (mixed and white waters); and (4) two sites in Anavilhanas National Park (black water). Characterization of water physicochemical parameters can be found in Table 1. Map of sampling sites is in Supplementary Material (Suppl. Fig. 1).

Fishing was done using a combination of different methods: small seine net fishing (for *M. festivus*), fixed gillnet (for *S. rhombeus*), and line fishing (for *M. festivus* and *S. rhombeus*). If specimens of the targeted species caught via seine net fishing and fixed gillnet were in contact with other fishes – e.g. if multiple species were caught in a same seine net, or if two fishes were in adjacent net mesh – these specimens were discarded to avoid cross-contamination of the skin mucus bacterial communities. We aimed to collect 10 adult fish specimens per species at each sampling site. 10 flag cichlids (FC) and 4 black piranhas (BP)

were sampled at site BWS1, 3 FC and 7 BP at site BWS2, 10 FC at sites MPS1 and MPS2, 6 BP at site WWS1, 5 BP at site WWS2, 10 FC and 10 BP at site Catalão (N total = 75 fish). Most of the studies on fish gut microbiota so far have only used 3-5 fish per site or species (reviewed in Ghanbari *et al*. 2015). Net-fishing can affect skin mucus bacterial composition by inducing physiological stress on caught fishes (by modifying mucus secretion and composition) (Smith and Ramos, 1976; Shephard, 1994). Thus, when gillnet-fishing, we checked the nets every 15 minutes to minimize the time fish spent in the net and to reduce sources of contamination from the net itself. After capture, the skin mucus of all fishes was immediately sampled by gently rubbing a sterile cotton swab on ≈ 50% of the total surface (upper half) of the right side of each fish. The same area was sampled on each fish to standardize the sampling zone. Fishes and their respective skin mucus samples were then quickly stored in individual plastic bags (to prevent contamination with other fishes), in a large cooler filled with ice and liquid Nitrogen until arrival at the *Laboratório de Ecofisiologia e Evolução Molecular* of INPA. There were about 4 hours between fish collection/flash freezing and dissection at the laboratory, and all fishes were processed the same way.

2 L of water were also sampled at each site in sterile Nalgene™ bottles to characterize the bacterioplankton community. The water samples were taken at 30 cm below the water surface. The two species collected in this study are found at these depths (Van der Sleen and Albert, 2018). The water samples were collected just before we left each of the sampling sites to minimize time between collection and filtration. They were immediately stored in a large cooler full of ice and liquid nitrogen until arrival at the laboratory. There were about 4 hours between collection and filtration.

At the laboratory, all fishes were weighed, measured and dissected. Standard Student's T tests showed that there was no significant difference (p-values > 0.05) between the average

length and weight of both flag cichlids and black piranhas in white water and black water. All fish were dissected with sterile instruments, under a flame, to isolate a section comprising midgut (right after pyloric caeca) and hindgut. Complete midgut and hindgut were pooled together in one sample for DNA extractions for each fish. Fish were sampled during dry season, which corresponds to a period of fasting for Amazonian fishes (Lowe-McConnell 1975, 1987). Thus, the isolated gut sections of both host species were small, generally empty of intestinal content, and therefore, gut samples mostly represented mucosal flora rather than flora associated with digesta. Gut samples were kept at -80°C along with skin mucus and water samples until DNA extraction. Water samples were filtered on 0.2 µm membranes (Nucleopore©) using a Masterflex Easy-Load® II peristaltic pump from Cole-Parmer®. Post-filtration, the membranes were stored dry at -80°C.

Preparation of 16S amplicon libraries

DNA extraction of skin mucus samples and 0.2 µm membranes from water samples was performed using DNeasy® Blood and Tissue Kit from QIAGEN according to the manufacturer's instructions. DNA extraction of gut samples was performed using QIamp® Fast DNA Stool Mini Kit according to the manufacturer's instructions. Extracted DNA from guts, skin mucus, and water was stored at -80°C until amplification. The fragment V3-V4 (≈ 500 base pairs) of the 16S rRNA gene was amplified by polymerase chain reaction (PCR) using a two-step dual-indexed PCR approach specifically designed for Illumina instruments by the Plateforme d'Analyses Génomiques (IBIS, Université Laval, Quebec City, Canada). Additional information on the PCR approach are in Supplementary data. PCR program: (1) 30 sec 98°C; (2) 10 sec 98°C; (3) 30 sec 64°C; (4) 20 sec 72°C; (5) 2 min at 72°C; 35 amplification cycles total. PCRs were done in triplicates to reduce PCR bias and to increase precision in the assessment of bacterial community composition and diversity. Amplified DNA was purified according to the manufacturer's instructions with AMPure beads

(Beckman Coulter Genomics) to eliminate primers, dimers, proteins and phenols. All three PCR products for each sample were kept separate for post-PCR DNA purification but were pooled together before sequencing. Post-PCR DNA concentration and quality were assessed on Nanodrop and by electrophoresis on [1.5%] agarose gels. After purification, Multiplex Sequencing was performed using the MiSeq platform from Illumina® **(**Illumina), by the *Plateforme d'analyses génomiques* at the *Institut de Biologie Intégrative et des Systèmes* (IBIS) of *Université Laval*. The kit used was the MiSeq Reagent Kit v3 (2 x 300 = 600 cycles). The estimated overlap was approximately 100-150 bp. All proceedings followed the manufacturer's protocols. A total of 157 samples were sequenced (43 gut and 43 skin mucus samples of flag cichlids, 32 gut and 32 skin mucus samples of black piranhas, and 7 water samples). Of these, 153 samples were sequenced successfully, and considered for downstream analysis: four samples were removed because of exceptionally low sequencing depth (< 500 reads).

Data availability

The sequence files are available from the Sequence Read Archive (http://www.ncbi.nlm.nih.gov/sra), BioProjectID: PRJNA470725, SRA accession number: SRP145296. The scripts used for the statistical analysis of sequence variants (dada2 pipeline), the output ASV table, taxonomy data, mapping (metadata) file and R markdowns used during this project are freely available from the Open Science Network platform (URL: https://osf.io/x5uje/).

Processing of 16S sequences

The analysis of amplicon sequences was done at the *Institut de Biologie Intégrative et des Systèmes* (IBIS) at *Université Laval*. After sequencing, 2,904,635 sequences were obtained (mean of 18,500 sequences per sample). The demultiplexed fastq sequence files were processed through QIIME2 (Bolyen *et al*. 2019), and the dada2 tool (Callahan *et al*. 2016) was used for Amplicon Sequence Variant (ASV) picking. Quality control of reads was processed through the filterAndTrim() function by using the following parameters : 270 for the read truncation length, 2 as the phred score threshold for total read removal, and a maximum expected error of 2 for forward reads and 4 for reverse reads. The filtered reads were then fed to the error rate learning, dereplication and ASV inference steps using the functions learnErrors(), derepFastq() and dada(). Chimeric sequences were removed using the removeBimeraDenovo() function with the “consensus” method parameter. Taxonomic classification was done through the assignTaxonomy() function using the SILVA v. 132 reference database. A rarefaction analysis of observed "species" counts and Shannon diversity (according to sampling depth) for each sample type (Supp. Fig. 2) showed that four samples had conspicuously lower sampling depth than the others: MPWS1 = 26 reads, WWS1 = 199 reads, BWS1 = 381 reads, 10.F = 474 reads. These four samples (< 500 reds) were discarded, bringing total number of samples to 153, and lowest sampling depth to 5223 reads. The average Good's coverage index for all samples was 0.9917 ± 0.0005 (S.E.).

Statistical analysis of sequence variants

Principal Components Analysis (PCoA) based on weighed Unifrac distances (Lozupone *et al*. 2010) were performed with R (Fig. 1) using the package phyloseq (McMurdie and Holmes 2014) to visualize sample clustering at different scales (1) tissue: gut versus skin mucus (2) species: flag cichlids versus black piranhas, and (3) water color: black versus white water. Then, the significance of these clusterings were assesed with p-values from PERMutational

ANalyses Of VAriances (PERMANOVAs), computed with 10 000 permutations using the *vegan* package (Dixon, 2003; Oksanen *et al*. 2019) from R and a distance matrix of weighed Unifrac indexes between samples (Fig. 1).

To investigate the diversity of bacterial taxa in each group, Faith's Phylogenetic diversity was calculated for each sample (Fig. 2) using the R package twbattaglia/btools (Battaglia, 2019). Diversity boxplots for each group were plotted on R v.3.5.2 using package ggplot2 (Ginestet, 2011). Student's T tests were used to compare the means of each group. Then, the taxonomic structure of each group was assessed with stacked barplots built on phyloseq (McMurdie and Holmes 2014), from the relative abundance of the 12 most abundant classes (Fig. 2).

To assess the effect of specific physicochemical parameters on the relative abundance of each bacterial ASV, we decomposed the effect of water color in its physicochemical parameters and computed Spearman correlations between all ASVs and physicochemical parameters. Correlations kept for downstream analysis had a significant Spearman correlation value < 0.05 after Bonferroni correction. Two correlations networks (skin mucus and gut communities) were constructed using Cytoscape version 3.2.1 (Shannon *et al*. 2003) to plot significant interactions (Fig. 3). Then, to assess overall community structure in both tissues, we computed Spearman correlations between all ASVs, and without hydrochemical parameters. Again, all significant interactions were plotted using Cytoscape. Network metrics (degree, betweenness centrality, topological coefficient, and neighborhood connectivity) were computed to assess the modularity of the networks representing "Overall community interactions" using the Network Analyzer Tool on Cytoscape. The degree of each node (in this study, a node corresponds to an ASV or a hydrochemical parameter) represents the number of interactions of the node. The betweenness centrality of a node reflects the

amount of control exerted by this node over the interactions of other nodes in the network (Yoon *et al*. 2006). The topological coefficient metric represents a relative measure for the extent to which a node shares its neighbors with other nodes (Stelzl *et al*. 2005). The neighborhood connectivity of a node *n* is defined as the average connectivity (degree) of all neighbors of *n* (Maslov and Sneppen 2002). Boxplots for each network metric were then plotted with ggplot2 (Ginestet, 2011).

To detect discriminative features of bacterial consortia at multiple scales (Fig. 4), we used a core microbiota approach. The core microbiota is defined as the group of organisms shared across multiple samples obtained from the same sample type (Turnbaugh *et al*. 2007). Describing the core microbiota is important for understanding the stable and consistent components across an array of complex and variable microbial assemblages (Shade and Handelsman 2011). The core microbiota potentially comprises keystone taxa playing crucial roles in the functionality of a specific sample type (Compant *et al*. 2019). In our study, we used the package microbiome (Lahti and Shetty 2017) to detect core ASVs at multiple scales (Fig. 4). Technically, we defined the core microbiota of a sample type as the reads from the ASVs present in at least 70% of all the samples from that sample type. Then, the core ASVs of each level were subtracted to the input data.frame to compute the core ASVs for the subsequent analysis level - e.g. the core ASVs from the skin mucus of both host species (analysis level: tissue) were discarded before the computation of the core ASVs from the skin mucus of flag cichlids (analysis level: species).

Then, we compared the dispersion (i.e. variance) in the relative abundance of core ASVs versus all other (non-core) ASVs (Fig. 5), using the betadisper function in the vegan package. Betadisper computes multivariate homogeneity of groups dispersions, a multivariate analogue of Levene's test for homogeneity of variances. The function handles non-euclidean distances between objects and group centroids by reducing the distances to principal coordinates (Oksanen *et al*. 2019). Then, the results of the dispersion analysis (Fig. 5) were plotted using with ggplot2 (Ginestet, 2011).

LEfSe tests (Linear Discriminant Analysis Effect Size) (Segata *et al*. 2011) were conducted to identify the taxonomic groups which abundances varied significantly between fishes from white and black water habitats (Fig. 6). The *class* used for the test was *Water color*, and no *subclass* was used. The threshold for the LDA parameter was the default 2.0. Boxplots were built on R for the bacterial taxonomic groups for which a significant and shared (between both host species) response was detected. P-values in the boxplots were calculated with Student's T tests, by comparing the relative abundance of the selected bacterial taxonomic groups in both species microbiota, versus their abundance in water bacterioplankton, for each water color.

Additional information about the statistical approaches used for ASV analysis are in Supplementary data.

Water chemistry parameters

PH, temperature and dissolved oxygen were measured at each site. For other measurements, 2 L of water were sampled at each site and brought back to the laboratory for further analysis. These 2 L of water were sampled at the same time and the same depth

than the 2 L of water used to characterize bacterioplankton structure. Ionic compositions (Na^+, Ca^{2+}, Mg^{2+}, K^+) of waters (N = 3) were analyzed using flame atomic absorption spectroscopy (Perkin-Elmer model 3100). Cl^- was measured using the colorimetric method described by Clarke (1950) (N = 3). Hardness was calculated from the Ca^{2+} and Mg^{2+} concentrations and alkalinity was measured using the method APHA Standard Methods for Examination of Water and Wastewater method 2320 (1992). Samples for DOC determination were filtered using 0.45 um Sartorius filters and were analyzed using a total carbon analyzer (Apollo 9000 combustion TOC analyzer: ©Teledyne Tekmar) (N = 3). The TOC machine was calibrated as per manufacturer's instructions, using primary standard grade potassium hydrogen phthalate (KHP). Water parameters for each sampling site are found in Table 1.

Results

Question #1: how does the structure of fish bacterial communities change along the hydrochemical gradients?

We first assessed sample clustering at different scales. The PCoA analysis showed that the types of tissue (gut or skin mucus), host species (flag cichlids or black piranhas), and water color (black or white) were all significant drivers of bacterial communities' structure (all PERMANOVA p-values < 0.009 **; Fig. 1). The two main axes of the PCoAs on Fig. 1 indicate that water color explained 36.5 % of the variance for skin mucus samples of flag cichlids, 34.6 % of the variance for skin mucus samples of black piranhas, 37.8 % of the variance for gut samples of flag cichlids, and 55.1 % of the variance for gut samples of black piranhas. Then, we investigated whether the significant differences observed in Fig. 1 were related to the phylogenetic diversity (i.e. the composition of the samples) or to the structure

(i.e. the relative abundance of bacterial taxa) of each community. The data may suggest a potential increase in phylogenetic diversity from black to white water samples (Fig. 2), however, this difference was non-significant (p-values > 0.05) for both fish species and for both tissues. Thus, it appears that the differences observed in Fig. 1 are related to differential patterns of relative abundances of bacterial taxa between water colors, rather than differential phylogenetic assemblages. In the skin mucus of both hosts, the most abundant taxa comprise the *Gammaproteobacteria*, the *Mollicutes*, the *Alphaproteobacteria* and the *Actinobacteria* (Fig. 2). From black to white waters, we observe a significant decrease in relative abundance of *Alphaproteobacteria* ($p < 0.004$ **) and a significant increase in relative abundance of *Mollicutes* ($p < 0.001$ ***). In the gut community, both host species show a high relative abundance of *Proteobacteria*, especially *Gammaproteobacteria*, *Deltaproteobacteria*, and *Alphaproteobacteria*. Other relatively abundant classes in the fishes' gut include *Fusobacteria*, *Bacteroidia*, and *Clostridia*. From black to white waters, both host species show a significant increase in relative abundance of *Gammaproteobacteria* ($p < 0.02$ *) and *Alphaproteobacteria* ($p < 0.001$ ***) in the guts, along a significant decrease in relative abundance of *Fusobacteria* ($p < 0.02$ *).

Then, to assess the correlation of specific physicochemical parameters with the relative abundance of each bacterial ASV, we decomposed the effect of water color in 10 physicochemical parameters and computed Spearman correlations between all ASVs and physicochemical parameters. The Spearman correlation networks from Fig. 3 reveal that community structure of fish skin mucus is correlated to, at least, nine physicochemical parameters within the hydrochemical gradients of Amazonia. The parameters that were significantly correlated with the abundance of bacterial ASVs in skin mucus were (from most to least number of interactions): magnesium, sodium, DOC, O_2, calcium, hardness, pH, potassium, and chloride. Contrastingly, the gut community was only correlated to - with significantly weaker average Spearman correlations - one parameter: environmental pH. The

low average Spearman correlation index observed between gut ASVs and hydrochemical parameters, and the low number of parameters (only one) significantly correlated to gut ASVs both suggest that gut bacterial communities are less sensitive to the hydrochemical gradients than skin mucus communities. Then, to assess overall community structure in both tissues, we computed Spearman correlations and built interaction networks from all ASVs - without hydrochemical parameters (Fig. 3). We show that the structure of interactions is tissue specific - an observation corroborated by four network metrics: network average degree, betweenness centrality, topological coefficient, and neighborhood connectivity. These metrics highlight two discriminative architectural particularities of these networks: (1) there is an increased connectivity between ASVs of the gut community; and (2) there is a higher modularity of community interactions in the gut tissue, compared to the skin mucus environment.

Question #2: are there intra and interspecific bacterial markers of convergent evolution associated with water color?

To detect discriminative features of bacterial consortia at multiple scales, we used a core microbiota analysis approach (Fig. 4). The core microbiota analysis showed the existence of core ASVs at each analysis level: core taxa existed for each tissue, each species, and each water color. Furthermore, we observed that the taxa that composed the core microbiota characteristic of specific water colors represented in average 17.0 % of the sequences in that water color. Thus, these taxa are not only ubiquitous, but also relatively abundant in the selected samples. Finally, our results from the last row of barplots in Fig. 4 show intraspecific site parallelism - they suggest that the relative abundance of core taxa is conserved among specimens of a specific host species collected at various sites within the same water color. To verify the hypothesis that the abundance of core ASVs was significantly more stable than other (non-core) ASVs at the individual host level (rather than at the site level), we compared

the dispersion (i.e. variance) of the relative abundance of core ASVs versus all other ASVs (Fig. 5) by calculating the multivariate homogeneity of groups dispersions for each ASV category. We showed that in all cases, core ASVs had a lower abundance dispersion than all other ASVs (in 9 cases out of 12 this difference was significant) (Fig. 5). Thus, our results highlight the presence of ubiquitous and relatively abundant core taxa, specific to a water color, for which the abundances are positively selected in the environment (fish tissue) where they are core. In this way, these taxa are candidate intraspecific bacterial markers of convergent evolution in relation to the hydrochemical gradients in Amazonia.

Then, we investigated further these communities to detect the presence/absence of interspecific bacterial markers associated with specific water colors. LEfSe tests (Linear Discriminant Analysis Effect Size) (Segata *et al*. 2011) were conducted to detect discriminative features (at all taxonomic levels possible) in community structure between hosts from the different water colors. The analysis highlighted a host species-specific response to the hydrochemical gradients (see Suppl. Fig. 3, 4, 5, 6). Furthermore, we identified four bacterial taxa in the skin mucus, and one taxa in the gut environment, that showed a significant and parallel response (shared between both host species) to the hydrochemical gradients in this study (Fig. 6). In skin mucus, these taxa included *Reyranella massiliensis*, *Roseiarcus fermentans*, BD1-7 clade, and *Flavobacteriales*. In fish guts, only the *Betaproteobacteriales* showed a significant and parallel response. Significant enrichment (niche preference) of these taxa in host-associated habitats, in comparison with their relative abundance in bacterioplankton, was observed for all the aforementioned taxa, except *Betaproteobacteriales*. While the BD1-7 clade, *Flavobacteriales*, and the *Betaproteobacteriales* were abundant in the bacterioplankton community, two taxa found in fish skin mucus (*Reyranella massiliensis* and *Roseiarcus fermentans*) were not found (0 reads detected) in the bacterioplankton community. Thus, the significant and parallel response (on both host species) of these two taxa to the hydrochemical gradients, along with

their strong and consistent niche preference for host-fish associated habitats, suggest that they may represent potential interspecific bacterial markers of convergent evolution associated with water color.

Discussion

The goal of this study was to characterize Amazonian fish microbiome variations across multiple scales: among different tissues, host species, water colors, and from replicate collecting sites within each water color. We observed a significant response of bacterial communities from both host species to the hydrochemical gradients, comprising 10 environmental physicochemical parameters. The tissue-specific response to the gradients was associated with a differential sensitivity to the hydrochemical parameters, along tissue-specific community architecture and modularity. Based on the structural response of the sampled communities to the hydrochemical gradients, we identified potential intra and interspecific bacterial markers of convergent evolution associated with water color.

Fish bacterial communities in hydrochemical gradients

Water physicochemical parameters are known to be important drivers of free living (bacterioplankton) and host-associated bacterial communities (Cheaib *et al*. 2018; Giatsis *et al.* 2015; Sylvain *et al*. 2016). Our results concur, as they show a significant effect of environmental factors on the bacterial community structure of two different fish host species, which showed a similar response independently (Fig. 1 and Fig. 2). We observed that skin and gut microbiota did not show the same response to the water color gradient (Fig. 3): while magnesium, sodium, DOC, O_2, calcium, hardness, pH, potassium, and chloride were drivers of the skin mucus community, only pH was significantly correlated to the relative abundance of gut ASVs. Thus, the structure of bacterial communities from each tissue is influenced by a

different range of physicochemical parameters. As observed in Sylvain *et al*. 2016, the fish skin mucus and gut bacterial habitats show contrasting responses to variations of water physicochemical parameters. In the gut samples, factors other than hydrochemical parameters may influence the structure of the gut community, such as fish diet and immune response, both of these being known to modulate both microbiota structure and activity (Sylvain and Derome 2017; Hevia *et al*. 2015). A recent study from Araujo *et al*. (2017) has shown that gene expression patterns, including genes involved in the host immune response, differ depending on water colors in the Amazonian fish *Triportheus albus*. The impact of the isolated and stable nature of the gut environment, combined with the impact of fish diets (piscivorous diet for black piranhas and detritivorous diet for flag cichlids) and immune response, may surpass the influence of environmental hydrochemical parameters as drivers of gut bacterial community structure.

Our network analysis (Fig. 3) shows three intrinsic differences between gut and skin mucus community structure: (1) a significantly lower average of Spearman correlations between gut ASVs and hydrochemical parameters than between skin mucus and such abiotic parameters; (2) a higher connectivity between gut bacterial taxa than between skin mucus ASVs; and (3) a higher modularity of interactions in the gut community than in the skin mucus community (Fig. 3). This tissue-specific difference in connectivity between bacterial taxa might be related to the high stability of the buffered fish gut environment (Payne 1978). Several studies show that the nature and number of perturbations to an ecological community affect the interaction network architecture of the community (reviewed in Ponisio *et al*. 2019), its response to disturbances, and its resilience (Violle *et al*. 2010; Shade *et al*. 2012). The skin mucus bacterial taxa are more exposed to environmental disturbances than gut bacterial taxa, due to their direct contact with the surrounding environment. This increased exposure to disturbances might weaken the network connectivity of skin mucus bacterial taxa, thus decreasing skin community robustness to pressures such as

environmental physicochemical parameter variations. This might explain the increased sensitivity of the skin mucus microbiota to hydrochemical parameters, and thus to water color in general. In contrast, the high connectivity in the gut interaction network might confer to the gut microbiota greater resilience to variable the hydrochemical parameters found in streams of the Amazon basin. This higher resilience to hydrochemical parameters could explain the lower average Spearman correlation values observed between physicochemical parameters and bacterial taxa in the gut, than those measured in the skin mucus network (Fig. 3). Strong associations between connectivity and resilience have been reported in the literature (Field and Parrott 2017). In an ecological community, architectural properties of interaction networks can affect community robustness (McCann *et al*. 1998; Pimm 1984; Rozdilsky and Stone, 2001). Tébault and Fontaine (2010) have shown that highly connected and nested community architecture promotes stability and resilience in mutualistic interaction networks. Furthermore, the modularity observed in the gut interaction network is a testimony of the stable nature of the gut environment, also associated with mature bacterial communities mostly composed of specialists rather than generalists (Ponisio *et al*. 2019).

Parallelisms and community adaptation

Biological communities overcome short-term or stochastic pressures by acclimation (i.e. physiological adaptation), whereas continuous or predictable pressures can be met by community adaptation (i.e. genetically determined response) (Bradshaw and Hardwick 1989; Davison and Pearson 1996; DeAngelis *et al*. 2010). The effect of selection by environmental conditions – an important deterministic factor influencing bacterial community development – can be modeled and predicted (Martiny *et al*. 2006; Langenheder and Szekely, 2011). Thus, parallel bacterial community structures in similar ecosystems can be indicative on the predictability of community assemblages. Studies investigating parallel evolution on fish microbiota in natural conditions have shown contradictory results. For instance, Sullam *et al.*

(2015) did not detect any parallel evolution in the microbiota of Trinidadian guppies (*Poecilia reticulata*), neither in the wild nor following dietary manipulation. In the same manner, Sevellec *et al.* (2018) compared the intestinal microbiota of five sympatric pairs of dwarf (limnetic) and normal (benthic) lake whitefish (*Coregonus clupeaformis*), and did not detect a clear evidence for parallelism as similar microbiota taxonomic composition was only observed for two out of five, lake specific, sympatric species pairs. Contrastingly, in a study on African cichlid fishes, Baldo *et al.* (2017) did detect significant parallelism, as distinct herbivore species from different lakes converged both in terms of key compositional and functional community aspects associated with plant fiber degradation. Our results showed two levels of parallelism.

First, they demonstrate the existence of predictable features of community structure (core bacterial taxa associated with water color) between fish specimens from the same species, but found in different habitats of the same water color. In population genetics, parallelism is often considered a hallmark of positive selection, meaning that selection is the predominant evolutionary force shaping genetic variation (McElroy *et al*. 2014) as opposed to stochastic forces such as genetic drift. In addition, when measuring quantitative traits such as gene expression levels (Derome *et al*. 2006), traits under selection are not only expressed in parallel, but also exhibit a significantly reduced variance, as we observed in this study for the relative abundance of the core taxa associated with water color. Thus, the environment-specific core taxa (Fig. 4) – as opposed to all other non-core taxa – are potentially positively selected according to the water color factor. Positive selection for a certain trait in a given environment can shed light on the adaptation mechanisms of species exploiting this environment.

Second, our results showed interspecific parallel features of community structure between host fishes from different species found in the same water color. To identify genes under selection in eukaryote hosts, one has to search for genes/markers, that exhibit significantly greater genetic differentiation among populations than expected under neutrality (reviewed by Holderegger 2008). The allelic frequencies at such outlier loci can then be correlated to ecological factors. In parallel, in a bacterial-communitywide approach, we can infer that taxa under selection should exhibit significantly greater differences in relative abundance among populations living in different environments than expected under neutrality. In our case, the LEfSe analysis (Fig. 6) identified key taxa that are discriminative features significantly associated with water colors for two ecologically and phylogenetically contrasting host species. Four taxonomic groups showed a parallel response to the water color factor in the skin mucus of both host species, and may represent potential bacterial markers of convergent evolution associated with water color. Two of these four groups (*Reyranella massiliensis* and *Roseiarcus fermentans*) were especially interesting, as they were undetectable in environmental water, and therefore exhibited a very strong niche preference for fish-associated habitats. Thus, we believe these two taxa represent key bacterial markers that could be investigated further, using a functional approach, to explore the potentially adaptive role of bacterial taxa in facilitating environmental adaptation to water colors in Amazonia.

An extensive characterization of the taxonomic structure of the bacterial communities enabled us to identify the tissue-specific, species-specific and water-color specific (Fig. 4 and Fig. 6) bacterial taxa. Our results are correlative and, thus, we cannot infer causative roles of bacterial taxa in fish adaptation to water color. The taxa identified as positively selected according to the water color factor could also be commensal bacteria “recruited” through neutral source-sink processes, with an extremely low abundance in water bacterioplankton at the time of sampling (thus undetectable in our water samples) coupled

with a strong preference for fish-associated habitats. Shedding light on microbe-mediated adaptation to water color may indicate a physiological adaptation of Amazonian fish-microbe holobionts to the high heterogeneity of hydrochemical parameters found in their vast distribution area. Indeed, genetic variation in bacterial communities can produce phenotypic variation of the metaorganism (a host and its microbiota) and may impact the distribution and ecological tolerance of the host (McMullin *et al*., 2000). For example, bacterial taxa in pea aphids provide selectable allelic variation (e.g., color, thermotolerance) that enables the host to thrive in a vast range of different environmental conditions (Oliver *et al*., 2009; Tsuchida *et al*., 2010; Moran and Yun, 2015).

Future research on Amazonian fish bacterial assemblages should build on the results of this present study to determine whether the intraspecific and interspecific bacterial markers associated with water colors play roles in the adaptive divergence of fish in Amazonia. Reciprocal transplant experiments (e.g. using a F1 generation of fish with 'seeded' microbiotas exposed to different water colors in the laboratory) would enable hypothesis tests on whether the recruitment of these bacterial markers results from neutral source-sink processes or local adaptation (Kawecki and Ebert 2004). A functional (e.g. metatranscriptomics) approach would reveal whether such taxonomic parallelism underlies functional differentiation between black and white water habitats. The metatranscriptomic approach, which allows the simultaneous characterization of both host and bacterial community transcriptomes, would shed light on the contribution of the bacterial markers in the functional interactions of the host-microbiota system. One of the interspecific bacterial markers identified in our study (*Reyranella massiliensis*) is especially interesting as it is known to interact with aromatic carbon compounds (Pagnier *et al*. 2012), such as the humic and fulvic acids which are abundant in black water environments (Sioli 1984), and which are known to play key roles in the osmotic equilibrium of fishes in acidic and ion-poor black water habitats (Duarte *et al*. 2016).

Conclusion

This study showed that wild Amazonian fish bacterial communities are significantly restructured by the hydrochemical gradients encompassing the black and white waters of Amazonia. We partitioned the effects of water color into several physicochemical parameters, which impacted differently gut and skin mucus bacterial communities, probably due to the nested and stable nature of the gut environment, contrasting with the exposed and variable nature of the skin mucus habitat. We also identified several potential intra and interspecific bacterial markers of convergent evolution associated with water color. Specifically, our results suggested that the two bacterial taxa *Reyranella massiliensis* and *Roseiarcus fermentans* are interesting candidates to investigate further the potential implication of bacterial taxa in the adaptive divergence of fish holobionts in the Amazon. Functional approaches will provide further evidence for the contribution of fish microbiota in host adaptive divergence in contrasting environmental conditions.

Acknowledgements

We gratefully acknowledge help from all members of the Derome and Val laboratory. Thank you to ICMBIO/Instituto Chico Mendes de Conservação da Biodiversidade for in situ support for fish collection and for issuing the permit to transport biological samples (permit number 29837/11). This research was part of the ADAPTA project at INPA and was supported by *Ressources Aquatiques Québec*, the *Fonds de recherche du Québec – Nature et technologies* (FRQNT) grant to FES, INCT ADAPTA (CNPq/FAPEAM) and INPA/MCTI grants to ALV (ORCID: 0000-0002-3823-3868), and the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery grant (grant #6333) to ND (ORCID: 0000-0002-2509-6104). Finally, we would like to thank three anonymous reviewers for their constructive comments.

This research is included in the Thesis of the first author (Sylvain 2017), François-Étienne Sylvain from Université Laval.

Additional acknowledgments are in the Supplementary data.

References

Araújo, J. D. A., Ghelfi, A., & Val, A. L. (2017). Triportheus albus Cope, 1872 in the Blackwater, Clearwater, and Whitewater of the Amazon: A Case of Phenotypic Plasticity?. Frontiers in Genetics. 8, 114.

Bailey, S. F., Blanquart, F., Bataillon, T., Kassen, R. (2017). What drives parallel evolution?. BioEssays. 39: 1-9 e201600176.

Bailey, S.F., Rodrigue, N., Kassen, R. (2015). The effect of selection environment on the probability of parallel evolution. Molecular Biology Evolution. 32: 1436–48.

Baldo, L., Pretus, J. L., Riera, J. L., Musilova, Z., Bitja Nyom, A. R., Salzburger, W. (2017), Convergence of gut microbiotas in the adaptive radiations of African cichlid fishes. ISME J. 11:1975-1987.

Battaglia, T. (2019). btools: A suite of R function for all types of microbial diversity analyses. R package version 0.0.1.

Beheregaray, L.B., Cooke, G. M., Chao, N. L. Landguth, E.L. (2015). Ecological speciation in the tropics: insights from comparative genetic studies in Amazonia. Frontiers in Genetics. 5.

Bolyen et al. (2019). QIIME 2: Reproducible, interactive, scalable, and extensible microbiome data science. PeerJ Preprints.

Boutin, S., Bernatchez, B., Audet, C., Derome, N. (2013). Network Analysis Highlights Complex Interactions between Pathogen, Host and Commensal Microbiota. Plos One. 8(12): e84772.

Bradshaw, A. D., Hardwick K (1989) Evolution and stress—genotype and phenotype components. Biol. J. Linn. Soc. 37:137–155.

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J., Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. Nature methods. 13(7): 581-3.

Cheaib, B., Boulch, M.L., Mercier, P.L., Derome, N. (2018). Taxon-Function Decoupling as an Adaptive Signature of Lake Microbial Metacommunities Under a Chronic Polymetallic Pollution Gradient. Frontiers in Microbiology. 9:869.

Clarke, F.E. (1950). Determination of chloride in water - improved colorimetric and titrimetric methods. Analytical Chemistry. 22: 553-555.

Compant, S., Samad, A., Faist, H., Sessitsch, A. (2019). A review on the plant microbiome: Ecology, functions, and emerging trends in microbial application. Journal of Advanced Research.

Conte, G.L., Arnegard, M.E., Peichel, C.L., Schluter, D. (2012). The probability of genetic parallelism and convergence in natural populations. Proceedings Royal Society London B Biological Science. 279: 5039–47.

Cooke, G. M., Chao, N. L., Beheregaray, L. B. (2012a). Divergent natural selection with gene flow along major environmental gradients in Amazonia: insights from genome scans, population genetics and phylogeography of the characin fish Triportheus albus. Molecular Ecology. 21: 2410–2427.

Cooke, G. M., Chao, N. L., Beheregaray, L. B. (2012b). Natural selection in the water: freshwater invasion and isolation by adaptation to water colour in the Amazonian pufferfish. Journal of Evolutionary Biology. 251305–1320.

Cooke, G. M., Chao, N. L., Beheregaray, L. B. (2012c). Marine incursions, cryptic species and ecological diversification in Amazonia: the biogeographic history of the croaker genus Plagioscion(Sciaenidae). Journal of Biogeography. 39: 724–738.

Cooke, G. M., Chao, N. L., Beheregaray, L. B. (2012d). Five cryptic species in the Amazonian catfish Centromochlus existimatus identified based on biogeographic predictions and genetic data. PLoS ONE. 7: e48800.

Cooke, G. M., Landguth, E. L., Beheregaray, L. B. (2014). Riverscape genetics identifies replicated ecological divergence across an Amazonian ecotone. Evolution. 68: 1947–1960.

Dal Pont, G., Domingos, F.X.V., Fernandes-de-Castilho, M., Val, A.L. (2017). Potential of the Biotic Ligand Model (BLM) to Predict Copper Toxicity in the White-Water of the Solimes-Amazon River. Bulletin of Environmental Contamination and Toxicology. 98: 27-32.

Davison, I. R., Pearson G. A. (1996). Stress tolerance in intertidal seaweeds. Journal Phycology. 32:197–211.

DeAngelis, K. M.,W., Silver, L., Thompson, A.W., Firestone, M. K. (2010). Microbial communities acclimate to recurring changes in soil redox potential status. Environmental Microbiology. 12:3137–3149.

Dixon, P. (2003). VEGAN, a package of R functions for community ecology. Journal of Vegetation Science. 14: 927-930.

Duarte, R.M., Smith, S.D., Val, A.L., Wood, C.M. (2016). Dissolved organic carbon from the upper Rio Negro protects zebrafish (Danio rerio) against ionoregulatory disturbances caused by low pH exposure. Sci. Rep. 6: 20377.

Endler, J. A. (1973). Gene flow and population differentiation. Science. 179: 243–250.

FAO (2011). The State of World Fisheries and Aquaculture. Consulted online on October 15, 2017. [URL]: http://www.fao.org/fishery/facp/BRA/en#CountrySector-Statistics

Field, R.D., Parrott, L. (2017). Multi-ecosystem services networks: A new perspective for assessing landscape connectivity and resilience. Ecological Complexity. 32: 31-41.

Ghanbari, M., Kneifel, W., Domig, K.J. (2015). A new view of the fish gut microbiota: Advances from next-generation sequencing. Aquaculture. 448:464-75.

Giatsis, C., Sipkema, D., Smidt, H., Heilig, H., Benvenuti, G., Verreth, J. et al. (2015). The impact of rearing environment on the development of gut microbiota in tilapia larvae. Scientific Reports. 5.

Ginestet, C. (2011). ggplot2: Elegant Graphics for Data Analysis. Journal of the Royal Statistical Society Series a-Statistics in Society. 174: 245-245.

Hedges, S. B., Marin, J., Suleski, M., Paymer, M., & Kumar, S. (2015). Tree of life reveals clock-like speciation and diversification. Molecular biology and evolution. 32(4): 835-45.

Hevia, A., Bernardo, D., Montalvillo, E., Al-Hassi, H.O., Fernandez-Salazar, L., Garrote, J.A. et al. (2015). Human colon-derived soluble factors modulate gut microbiota composition. Frontiers in Oncology.

Holderegger, R. (2008). Land ahead: using genome scans to identify molecular markers of adaptive relevance. Plant Ecology and Diversity. 1(2):273-283.

Johannsson, O.E., Smith, D.S., Sadauskas-Henrique, H., Cimprich, G., Wood, C.M., Val, A.L. (2017). Photo-oxidation processes, properties of DOC, reactive oxygen species (ROS), and their potential impacts on native biota and carbon cycling in the Rio Negro (Amazonia, Brazil). Hydrobiologia. 789: 7-29.

Kawecki, T.J., Ebert, D. (2004). Conceptual issues in local adaptation. Ecology Letters. 7(12): 1225-1241.

Koskella, B., Hall, L.J., Metcalf, C.J.E. (2017). The microbiome beyond the horizon of ecological and evolutionary theory. Nature ecology & evolution. 1(11): 1606.

Lahti, L., Shetty S. (2017). Tools for microbiome analysis in R. Version 1.5.28. URL: http://microbiome.github.com/microbiome.

Langenheder, S., and Szekely, A.J. (2011) Species sorting and neutral processes are both important during the initial assembly of bacterial communities. ISME J 5: 1086–1094.

Llewellyn, M.S., Boutin, S., Hoseinifar, S.H., Derome, N. (2015). Teleost microbiotas: the state of the art in their characterization, manipulation and importance in aquaculture and fisheries. Frontiers in Microbiology. 5:207.

Llewellyn, M.S., McGinnity, P., Dionne, M., Letourneau, J., Thonier, F., Carvalho, G.R. et al. (2016). The biogeography of the atlantic salmon (Salmo salar) gut microbiota. Isme Journal. 10: 1280-1284.

Lowe-McConnell, R.H. (1975). Fish Communities in Tropical Freshwaters. Their Distribution, Ecology and Evolution. Longman, London.

Lowe McConnell, R.H. (1987). Ecological studies in the in Tropical Fish communities. Cambridge University Press, Cambridge.

Lozupone, C., Lladser, M. E., Knights, D., Stombaugh, J., & Knight, R. (2010). UniFrac: an effective distance metric for microbial community comparison. The ISME journal. 5(2): 169–172.

Margulis, L., Fester, R. (1991). Symbiosis as a Source of Evolutionary Innovation: Speciation and Morphogenesis. The MIT Press. Cambridge, Massachissetts.

Martiny, J.B., Bohannan, B.J., Brown, J.H., Colwell, R.K., Fuhrman, J.A., Green, J.L., et al. (2006). Microbial biogeography: Putting microorganisms on the map. Nature Reviews Microbiology. 4: 102–112.

Maslov, S., Sneppen, K. (2002). Specificity and stability in topology of protein networks. Science 296: 910-913.

McCann, K., Hastings, A., Huxel, G. R. (1998). Weak trophic interactions and the balance of nature. Nature. 395: 794–798.

McElroy, K. E., Hui, J.G., Woo, J. K., Luk, A. W., Webb, J. S., Kjelleberg. S., Rice, S. A., Thomas, T. (2014). Strain-specific parallel evolution drives short-term diversification during Pseudomonas aeruginosa biofilm formation. Proc Natl Acad Sci U S A. 111:E1419-27.

McMurdie, P.J., Holmes, S. (2014). Shiny-phyloseq: Web Application for Interactive Microbiome Analysis with Provenance Tracking. Bioinformatics. 31(2): 282–283.

Moran, N.A., Yun, Y. (2015). Experimental replacement of an obligate insect symbiont. Proc. Natl. Acad. Sci. USA. 112: 2093-2096.

Oksanen, J.F., Blanchet, G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Henry, M., Stevens, H., Szoecs, E., Wagner, H. (2019). vegan: Community Ecology Package. R package version 2.5-4. https://CRAN.R-project.org/package=vegan

Oliver, K.M., Degnan, P.H., Hunter, M.S., Moran, N.A. (2009). Bacteriophages encode factors required for protection in a symbiotic mutualism. Science. 325.

Pagaling, E., Strathdee, F., Spears, B.M., Cates, M.E., Allen, R.J., and Free, A. (2014). Community history affects the predictability of microbial ecosystem development. ISME J. 8: 19–30.

Pagnier, I., Croce, O., Robert, C., Raoult, D., La Scola, B. (2012). Genome sequence of Reyranella massiliensis, a bacterium associated with amoebae. Journal of bacteriology. 194(20): 5698.

Payne, A.I. (1978). Gut ph and digestive strategies in estuarine grey mullet (Mugilidae) and tilapia (Cichlidae). Journal of Fish Biology. 13: 627-629.

Pimm, S. L. (1984). The complexity and stability of ecosystems. Nature. 307: 321–326.

Pires, T.H.S., Campos, D.F., Ropke, C.P., Sodre, J., Amadio, S., Zuanon, J. (2015). Ecology and life-history of Mesonauta festivus: biological traits of a broad ranged and abundant Neotropical cichlid. Environmental Biology of Fishes. 98: 789-799.

Ponisio, L.C., Valdovinos, F.S., Allhoff, K.T., Gaiarsa, M.P., Barner, A., Guimarães, P.R. Jr., Hembry, D.H., Morrison, B., Gillespie, R. (2019). A Network Perspective for Community Assembly. Front. Ecol. Evol. 7:103.

Prosser, J.I., Bohannan, B.J., Curtis, T.P., Ellis, R.J., Firestone, M.K., Freckleton, R.P., et al. (2007). The role of ecological theory in microbial ecology. Nature Reviews Microbiology. 5: 384–392.

Rozdilsky, I. D., Stone, L. (2001). Complexity can enhance stability in competitive systems. Ecology Letters. 4: 397–400.

Scheffer, M., Carpenter, S., Foley, J.A., Folke, C., and Walker, B. (2001). Catastrophic shifts in ecosystems. Nature. 413: 591–596.

Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W.S. et al. (2011). Metagenomic biomarker discovery and explanation. Genome Biology. 12: R60.

Sevellec, M., Derome, N., Bernatchez, L.B. (2018). Holobionts and ecological speciation: the intestinal microbiota of lake whitefish species pairs. Microbiome. 6: 47.

Shade, A., Handelsman, J. (2011). Beyond the Venn diagram: the hunt for a core microbiome. Environmental microbiology. 14(1):4-12.

Shade, A., Peter, H., Allison, S.D., Baho, D.L. Berga, M., Bürgmann, H., et al. (2012). Fundamentals of Microbial Community Resistance and Resilience. Frontiers in Microbiology. 3 : 417.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D. et al. (2003). Cytoscape: A software environment for integrated models of biomolecular interaction networks. Genome Research. 13: 2498-2504.

Shephard, K.L. (1994). Functions for fish mucus. Reviews in Fish Biology and Fisheries. 4(4): 401-429.

Sioli, H. (1984). The Amazon Limnology and Landscape Ecology of a Mighty Tropical River and its Basin. Dr. W. Junk Publishers: Dordrecht.

Smith, A.C., Ramos, F. (1976). Occult haemoglobin in fish skin mucus as an indicator of early stress. Journal of Fish Biology. 9(6): 537-541.

Smith, T.B., Schneider C.J., Holder K. (2001). Refugial isolation versus ecological gradients. Genetica 112 383–398.

Smith, T.B., Wayne, R.K., Girman, D.J., Bruford, M.W. (1997). A role for ecotones in generating rainforest biodiversity. Science. 276: 1855–1857.

Stelzl, U., et al. (2005). A human protein-protein interaction network: a resource for annotating the proteome. Cell. 122: 957-968.

Sullam, K.E, Rubin, B.E., Dalton, C.M., Kilham, S.S., Flecker, A.S, Russell, J.A. (2015). Divergence across diet, time and populations rules out parallel evolution in the gut microbiomes of Trinidadian guppies. ISME J. 9(7):1508-22.

Sylvain, F.E. (2017). Caractérisation des systèmes hôte-microbiote ichthyens dans des contextes environnementaux contrastés du bassin amazonien. [Master's Thesis]. [Quebec (QC)]: Université Laval

Sylvain, F.E., Cheaib, B., Llewellyn, M., Correia, T.G., Fagundes, D.B., Val, A.L. et al. (2016). pH drop impacts differentially skin and gut microbiota of the Amazonian fish tambaqui (Colossoma macropomum). Scientific Reports. 6.

Sylvain, F.E., Derome, N. (2017). Vertically and horizontally transmitted microbial symbionts shape the gut microbiota ontogenesis of a skin-mucus feeding discus fish progeny. Scientific Reports. 7:14.

Thébault, E., Fontaine, C. (2010). Stability of ecological communities and the architecture of mutualistic and trophic networks. Science. 329(5993):853-856.

Tsuchida, T., Koga, R., Horikawa, M., Tsunoda, T., Maoka, T., Matsumoto, S., Simon, J.C., Fukatsu, C. (2010). Symbiotic bacterium modifies aphid body colorScience, 330: 1102-1104.

Turnbaugh, P.J., Ley, R.E., Hamady, M., Fraser-Ligett, C., Knight, R., and Gordon, J.I. (2007). The human microbiome project: exploring the microbial part of ourselves in a changing world. Nature. 449, 804–810.

Val, A.L., Almeida-Val, V.M.F. (1995). Fishes of the Amazon and Their Environment: Physiological and Biochemical Aspects. Springer-Verlag Berlin Heidelberg. 224 p.

Van der Sleen, A., Albert, J.S. (2018). Field guide to the fishes of the Amazon, Orinoco and Guianas. Oxford University Press.

Violle, C., Pu, Z., Jiang, L. (2010). Experimental disturbance of the importance of competition under disturbance. PNAS. 107(29): 1295-9.

Widder, S., Allen, R.J., Pfeiffer, T., Curtis, T.P., Wiuf, C., Sloan, W.T., et al. (2016). Challenges in microbial ecology: Building predictive understanding of community function and dynamics. ISME J. 10: 2557–2568.

Yoon, J., Blumer, A., Lee, K. (2006). An algorithm for modularity analysis of directed and weighted biological networks based on edge-betweenness centrality. Bioinformatics. 22: 3106-8

Zhu, B., Wang, X., Li, L. (2010). Human gut microbiome: the second genome of human body. Protein Cell. 1(8): 718-25.

Data accessibility

The datasets generated and analyzed during the current study can be found in the Sequence Read Archive (http://www.ncbi.nlm.nih.gov/sra) repository, BioProjectID: PRJNA470725, SRA accession number: SRP145296. The scripts used for the statistical analysis of sequence variants (dada2 pipeline), the output ASV table, taxonomy data, mapping (metadata) file and R markdown used during this project are freely available from the Open Science Network platform (URL: https://osf.io/x5uje/).

Author Contributions

FS, ND and AV designed the experiment. FS, AH, ND and AV organized sampling expeditions. FS, AH and ND sampled fish during field expeditions. FS, AH, EG and ND processed samples in the laboratory (fish dissections, DNA extractions and PCRs). FS and EG performed 16S sequence analysis. FS, and ND wrote the manuscript. All authors reviewed the manuscript.

Tables and figures

Tables

Table 1: Water physicochemical parameters at each sampling location

Water	Locations	Site	pH	O2 (%)	Temperature (°C)	Sodium (mg/L)	Chloride (mg/L)	Calcium (mg/L)	Magnesium (mg/L)	Potassium (mg/L)	DOC (mg/L)	Hardness (mg/L CaCO3)
Black	Anavilhanas National Park	BWS1	4.70	89.32	31.20	0.64	0.03	0.01	0.07	1.19	7.20	0.00
		BWS2	4.65	88.50	30.30	0.68	0.02	0.00	0.07	1.19	7.79	0.00
Mixed	Manacapuru River Tributary	MPS1	5.85	76.00	33.00	4.14	12.81	0.11	0.15	1.44	5.81	1.00
White	Manacapuru River	MPS2	6.45	91.00	32.50	10.81	23.07	0.93	0.99	2.16	3.78	6.00
	Solimões River	WWS1	6.90	81.00	32.70	6.02	7.49	0.75	0.67	2.06	3.52	5.00
		WWS2	6.86	76.00	30.70	5.96	7.69	0.74	0.64	2.02	3.54	4.00
	Lake Catalao	Catalao	6.71	42.00	31.50	12.56	3.19	2.22	1.96	2.14	7.73	14.00

*DOC = dissolved organic carbon

Figures

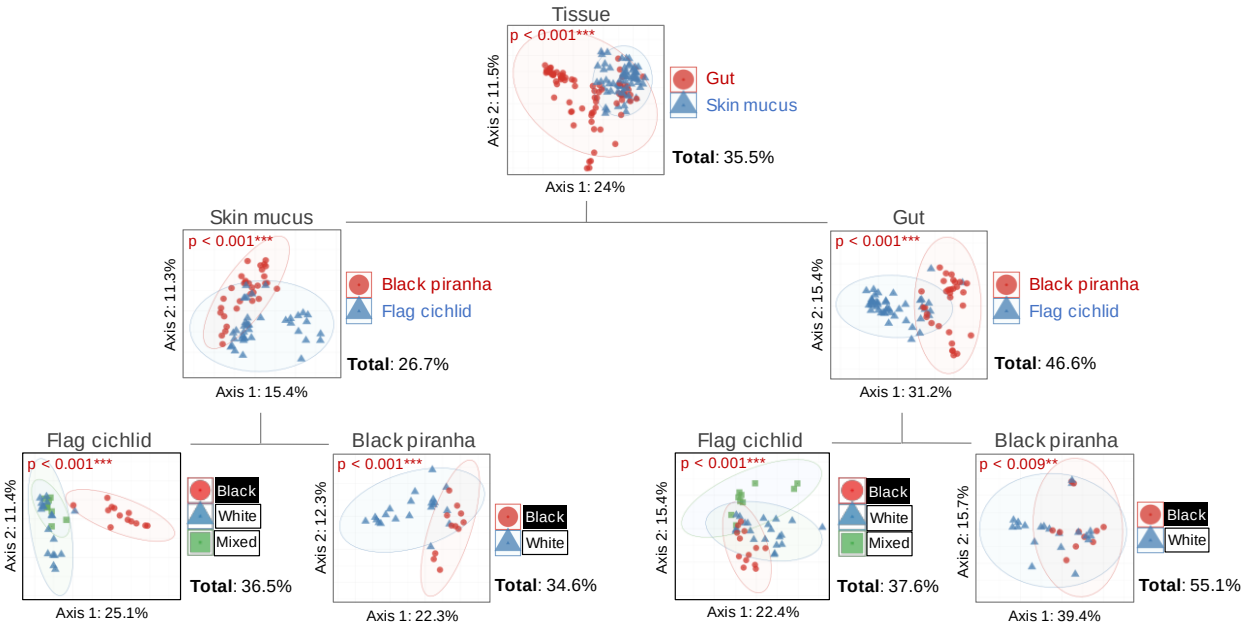

Figure 1: Analysis of bacterial communities at multiple scales evidenced tissue-, species-, and water color-specific bacterial assemblages. PCoA analysis were computed from weighed Unifrac distances. P-values are results of PERMANOVAs computed with 10 000 permutations.

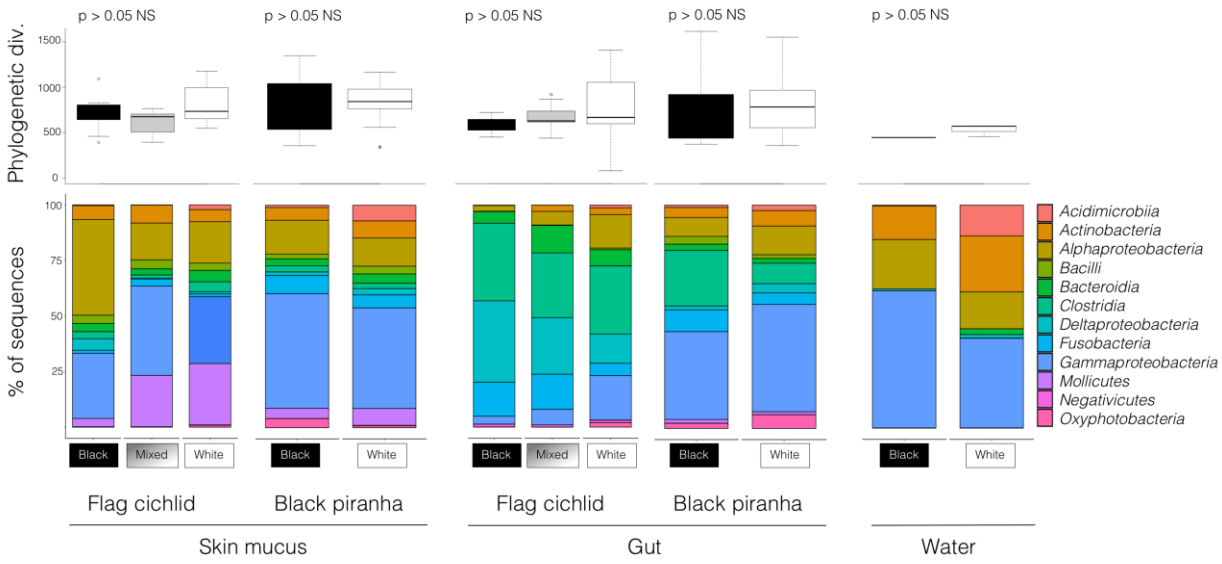

Figure 2: Hydrochemical gradients of Amazonia impact the taxonomic structure, rather than the phylogenetic diversity of fish bacterial communities. No significant difference between water colors was observed for the Faith's phylogenetic diversity metric, although there seems to be a slight (non-significant) increase in diversity from black to white water habitats. The phylogenetic diversity was computed from the most precise taxonomic annotation possible of all individual ASVs. Stacked barplots representing the relative abundance of the 12 most abundant classes show variations in the taxonomic architecture of the bacterial communities from black to white water.

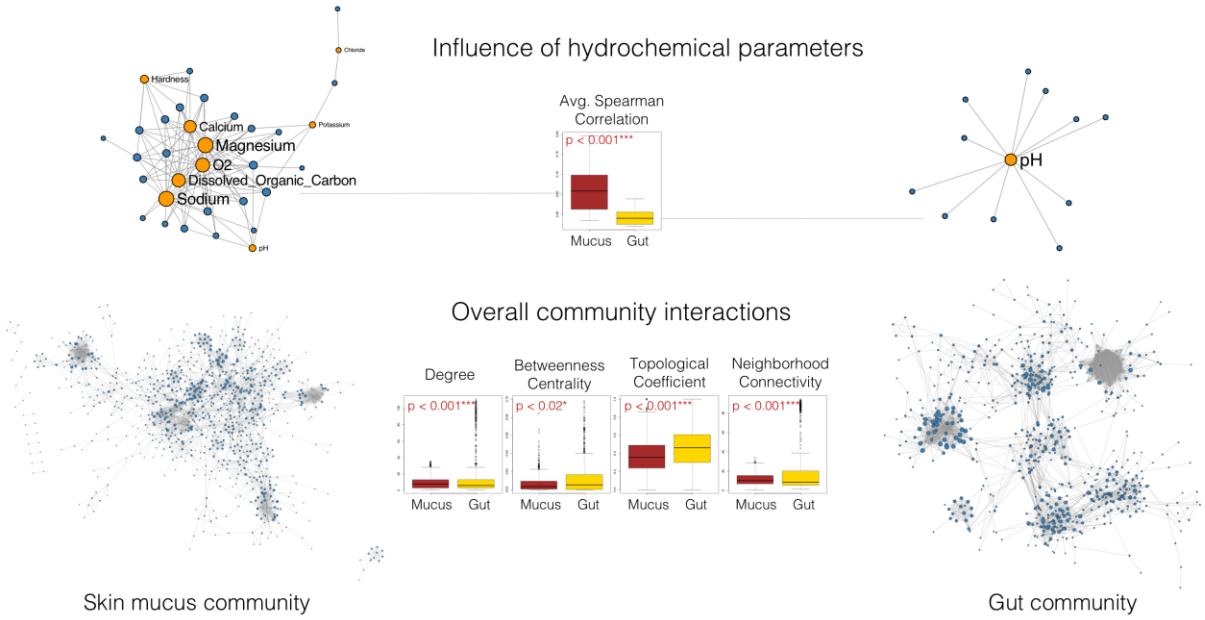

Figure 3: Spearman correlation networks highlight tissues-specific sensitivity to hydrochemical parameters, along tissue-specific community architecture. Orange nodes represent hydrochemical parameters while blue nodes represent bacterial taxa (i.e. specific ASVs). The two upper networks show significant interactions between ASVs and hydrochemical parameters. The two lower networks show interactions between all ASVs (without hydrochemical parameters). Edges between nodes represent significant Spearman correlations between both nodes ($p < 0.05$ after Bonferroni correction).

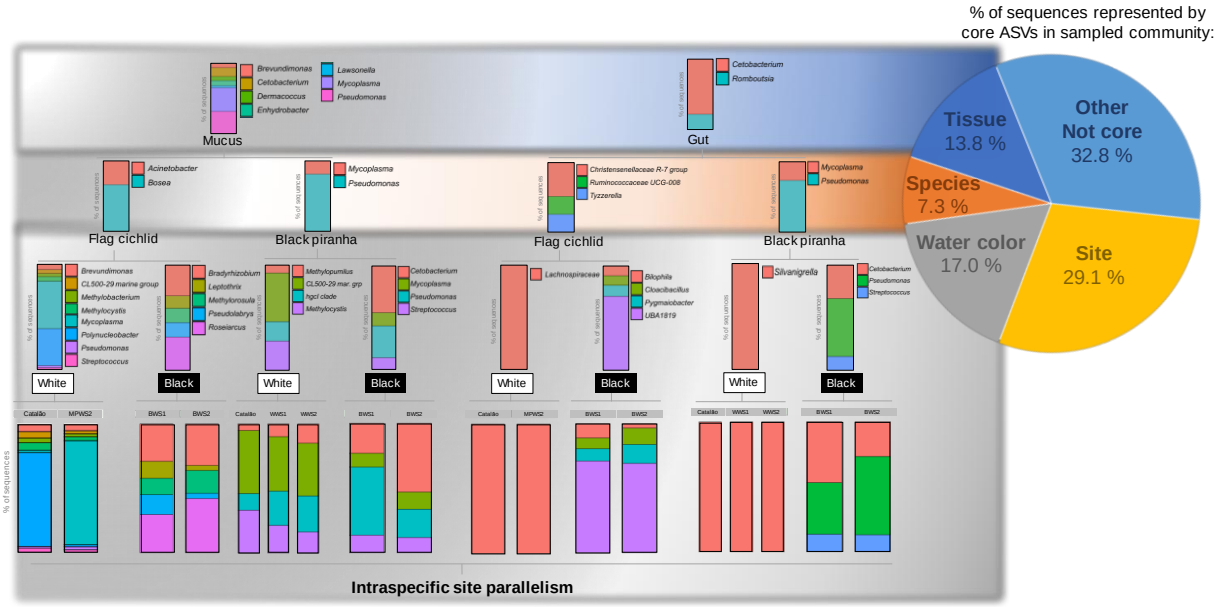

Figure 4: A core microbiota analysis identifies ubiquitous and relatively abundant ASVs at the tissue-, species-, and water color level. Core ASVs from a sample type were present in at least 70% of all the samples from that sample type.

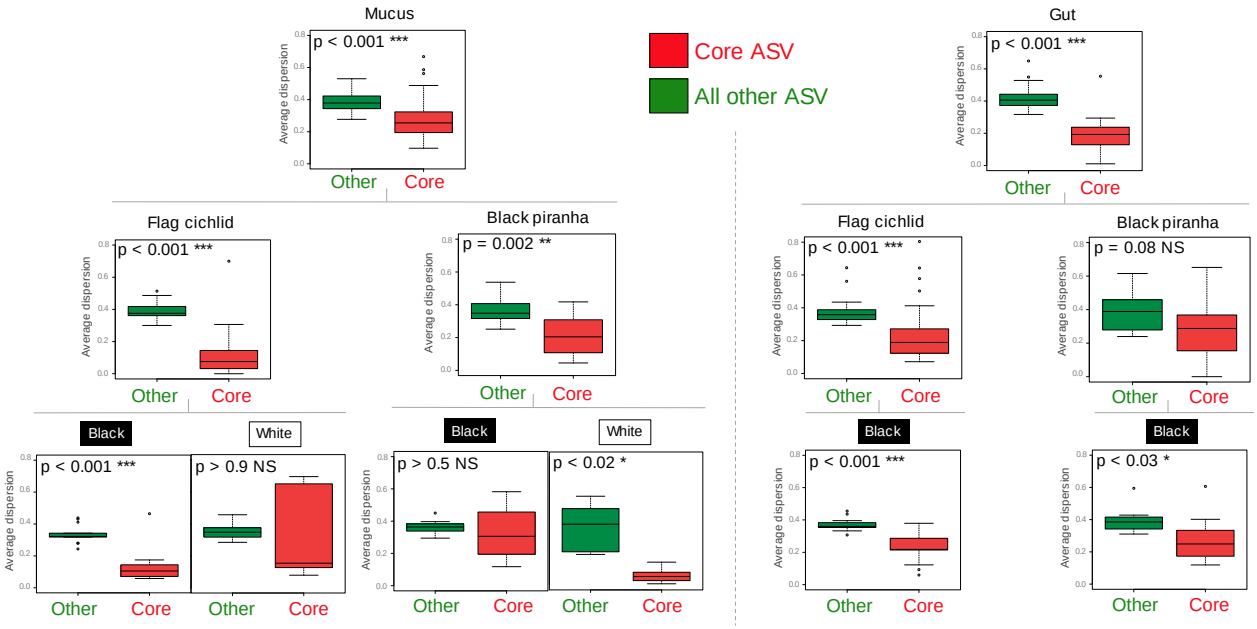

Figure 5: A multiscale dispersion (i.e. variance) analysis shows that the dispersion of the relative abundances of core ASVs is always lower than the abundance dispersion of all other (non-core) ASVs. In nine groups out of 12, this difference is significant. Core ASVs from a sample type were present in at least 70% of all the samples from that sample type. The comparison of abundance dispersion in white water environments for the guts of both species could not be computed, as only one core taxon was identified in those communities (see Fig. 4).

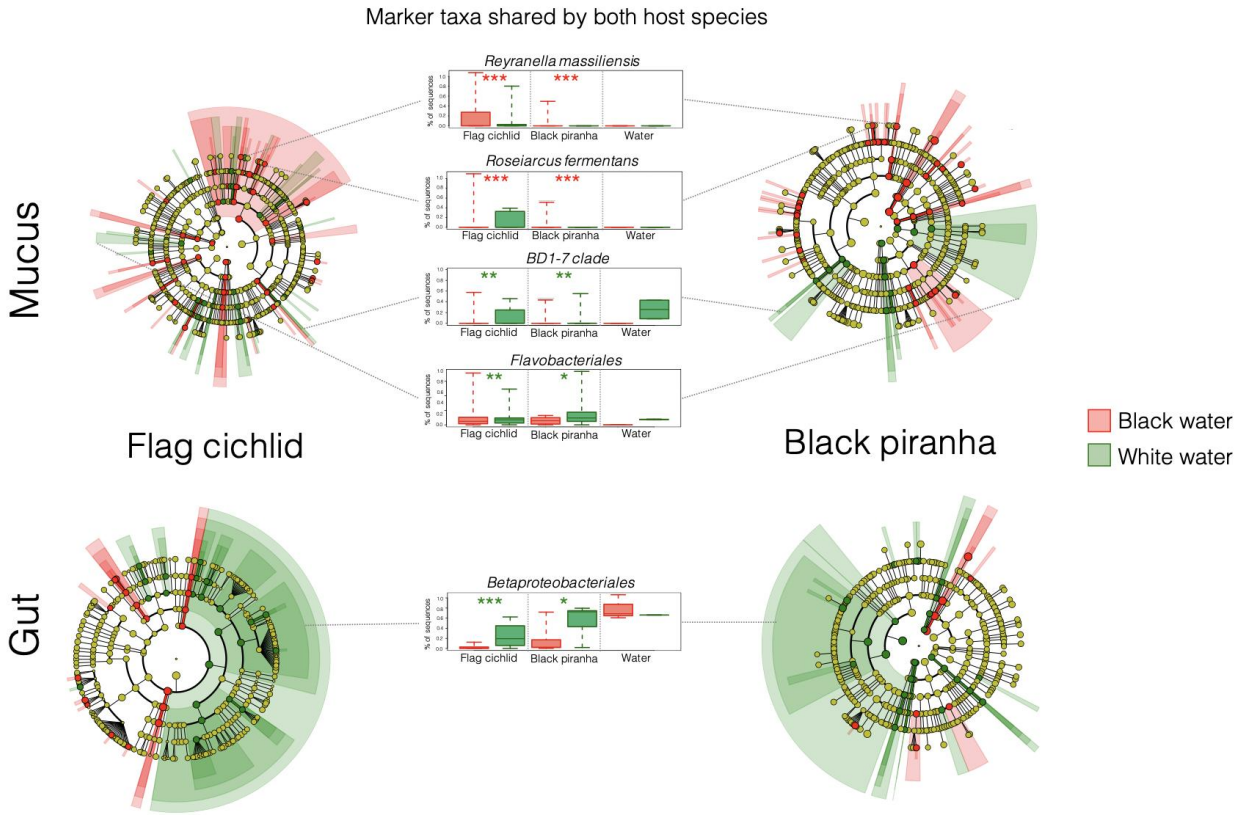

Figure 6: LEfSe analyses show a significant and parallel response (i.e. a response shared by both host species) of five taxa to the hydrochemical gradients. The figure represents four circular phylogenetic trees where discriminative features of different water colors are highlighted in red (significantly more abundant in black waters) or green (white waters). The boxplots represent the relative abundance of taxa showing a parallel response between the two host species. The four circular phylogenetic trees with their detailed legends including all discriminative taxa are in Suppl. Fig. 3, 4, 5, 6.