Host Species and Body Site Explain the Variation in the Microbiota Associated to Wild Sympatric Mediterranean Teleost Fishes

M. Ruiz-Rodríguez, M. Scheifler, S. Sanchez-Brosseau, E. Magnanou, N. West, M. Suzuki, S. Duperron, Y. Desdevises

Host species and body site explain the variation in the microbiota associated to wild sympatric Mediterranean teleost fishes

M. Ruiz-Rodríguez[1], M. Scheifler[1], S. Sanchez-Brosseau[1], E. Magnanou[1], N. West[2], M. Suzuki[2], S. Duperron[3], Y. Desdevises[1]

[1]Sorbonne Université, CNRS, Biologie Intégrative des Organismes Marins, BIOM, Observatoire Océanologique, F-66650 Banyuls/Mer, France.

[2]Sorbonne Université, CNRS, Observatoire Océanologique de Banyuls, FR3724, F-66650 Banyuls/Mer, France.

[3]Muséum National d'Histoire Naturelle, CNRS, Molécules de Communication et Adaptation des Micro-organismes, MCAM, Muséum National d'Histoire Naturelle, F-75000 Paris, France.

Corresponding author:

M. Ruiz-Rodríguez. ORCID 0000-0002-4202-5180

E-mail address: magdalenaruizr@gmail.com

Tel: +33 66 22 60 351

ABSTRACT

Microorganisms are an important component shaping the evolution of hosts and as such, the study of bacterial communities with molecular techniques is shedding light on the complexity of symbioses between bacteria and vertebrates. Teleost fish are a heterogeneous group that live in a wide variety of habitats, and thus a good model group to investigate symbiotic interactions and their influence on host biology and ecology. Here we describe the microbiota of thirteen teleostean species sharing the same environment in the Mediterranean Sea, and compare bacterial communities among different species and body sites (external mucus, skin, gills and intestine). Our results show that *Proteobacteria* is the dominant phylum present in fish and water. However, the prevalence of other bacterial taxa differs between fish and the surrounding water. Significant differences in bacterial diversity are observed among fish species and body sites, with higher diversity found in the external mucus. No effect of sampling time nor species individual were found. The identification of indicator bacterial taxa further supports that each body site harbours its own characteristic bacterial community. These results improve current knowledge and understanding of symbiotic relationships among bacteria and their fish hosts in the wild, since the majority of previous studies focused on captive individuals.

Keywords: Bacteria, Metabarcoding, Mediterranean Sea, Microbiota, Teleost fish

INTRODUCTION

Symbiosis among microorganisms and hosts are driving forces of evolution. Among microorganisms involved in symbioses with animals, bacteria are the most widespread and ubiquitous, shaping the evolution of their hosts by contributing both positively (basic functions) and negatively (pathogenicity) to their fitness [1]. In-depth knowledge of the taxonomic composition of the microbiota, and their relationships with their hosts, can provide insights into both the function and dysfunction of the host organisms [2], due to the important role that host-microbe interactions play in the physiology and performance of animals [3].

Fish constitute about half of living vertebrate species [4], but the majority of studies on microbiota have been conducted in mammals [5, 6]. High throughput 16S rRNA gene sequencing is being increasingly employed to investigate the fish microbiota (e.g., [7, 8, 9,]), producing high-resolution descriptions of their community structure and diversity [2, 10], and providing further understanding of the consequences of symbioses with bacteria [11]. Despite progress, important gaps remain in our current knowledge regarding the factors that shape fish microbiota [12], such as environmental or phylogenetic variables, whose influences could vary among species and / or habitats. Furthermore, most of the information available is focused on the gut of captive species used in aquaculture, i.e., fish of economic significance and commercial interest [8, 13], or from model organisms such as the zebrafish [14].

Wild teleost fishes exhibit a broad variety of morphologies, physiologies, ecologies and natural histories [11, 15]. Therefore, they represent a good model group to investigate the factors that shape host-bacterial assemblages in nature [16]. Fishes are in intimate contact with the surrounding water, and bacterial colonisation from the environment is thought to be one of the primary mechanisms of microbiota acquisition for fishes [16, 17]. Hence, environmental factors influencing water microbial reservoirs, such as water salinity, could ultimately influence fish microbiota [11, 18]. The water environment also hosts high concentrations of potentially pathogenic microorganisms [19] that may enter the fish body. Therefore, the dynamics and diversity of the microbial communities originating from the surrounding water could influence significantly the individuals' fitness [20].

The external surfaces of fishes, such as the skin and gills, are coated in a mucus secretion that hosts an indigenous microbiota dominated by bacteria [10, 21, 22] that acts as a protective barrier against pathogens [23-26]. The digestive tract of fish also receive water

and food that are populated with microorganisms that will undoubtedly affect the resident microbiota.

Besides the surrounding water, host-related factors may also shape fish-associated microbial communities [18, 27]. In some teleost species, the skin bacterial diversity is very different from that of the surrounding bacterioplankton, and varies among different parts of the host body [28]. In addition, different species reared in the same water environment vary in their gut [18, 29] and gills [9] microbiota.

We hypothesize that wild teleost microbiota could be influenced by abiotic environmental factors and host-related factors. If host-related factors are more important than abiotic factors in determining the fish microbiota, we would expect to find specific microbiota from each part of the body, and in different fish species, independently of when the fishes were captured. To test this hypothesis, and evaluate the importance of these different factors, multiple individuals of 13 species of teleost were collected at the same location in the shallow Mediterranean Sea on four different occasions and four body sites were sampled (mucus, skin, gills, and gut). The microbial community composition of these samples was characterized by 16S rRNA gene sequencing and their alpha and beta diversity compared. We also tried to determine indicator bacterial taxa that could be representative for each teleost species and body site, being both unique to a given group (exclusivity) and occurring in all sample units within a group (fidelity) [30, 31].

This study is novel in characterizing the bacteria associated with wild fish, while the majority of previous studies addressed bacterial assemblages in captive fish.

MATERIAL & METHODS

Fish sampling

Fishes were captured in the NW Mediterranean Sea, Bay of Banyuls (Gulf of Lion, Banyuls-sur-mer, SE France, 42°29’4.618’’N, 3°8’, 35.39’’E) on 4 different days in 2017: June 21, June 26, July 18, and October 4, and several physicochemical variables for each sampling were recorded (Table S1). We captured a total of 59 individuals belonging to 5 families and 13 species (see Table 1). For all fish individuals except the Family Gobiidae, one gill net was placed between 0 and 6 m depth during the entire night period, and at sunrise, it was recovered with the fish (42°29’15.073’’N, 3°7’, 49.688’’E). Fish were collected dead, handled with gloves and stored into individual plastic bags.

They were immediately brought from the vessel to the laboratory. They were kept at 4 °C until dissection, within the next 48 h, mostly within 6 h. Individuals from the Family Gobiidae were captured with one net placed for just 2 hours close to the seashore.

All thirteen species share nearly the same environment since they were captured in the same place at the bay, and most of these species have similar feeding habits. In general, they are omnivorous, eating mainly small crustaceans, invertebrates and small fishes, but they can also ingest, to different degrees, other particles such as algae and phytoplankton [33]. There is an exception for the species *S. salpa*, in which juveniles are omnivorous (mainly carnivores), but adults are exclusively herbivores [33]. The adult stage is reached at about three years, when the body size is longer than 20 cm [34]. The length of individuals belonging to *S. salpa* that were captured were shorter than 20 cm for four of the five individuals (the lengths were 16.8, 17.2, 15.5 and 14 cm). The fifth individual was of adult size (23.8 cm.) This is the only individual, from all fish investigated here, whose feeding would be exclusively vegetarian. For the other 12 species, no significant differences in feeding habits according to age are described [33].

Once in the laboratory, each sample was taken as follows: (i) the skin mucus was scraped with a sterile scalpel from the entire body surface; (ii) when the mucus was completely removed, a 3 cm^2 piece of skin was cut from the central part of the body, close to the lateral line; (iii) then, one gill arch was cut from each side of the body; finally, (iv) 5 cm of the intestine distal part was cut. We cut the final part of the gut because, due to defecation, it reflects the bacterial community of the entire intestinal tract [35, 36]. Because of possible differences in bacterial communities among different parts of the same organ [28], samples were always taken from the same part of the body in all the different species, sterilizing the material by flaming before taking each part of the same individual. Samples were immediately placed into sterile Eppendorf tubes and kept at -80 ºC until DNA extraction. In the case of the species *Scorpaena notata*, there was no mucus on the skin, so the skin piece was cut directly.

At each sampling site, 2 litres of seawater were taken from the same place where the nets were installed and the water stored in a sterile glass bottle. Briefly, in the lab, the water was vacuum filtered through a 47 mm diameter membrane filter with a pore size of 0.2 µm (Fisherbrand, Thermo Fisher Sientific, France). This filter was kept at -80 ºC prior to DNA extraction.

DNA sequencing

A total DNA extraction was performed using the Kit Quick-DNA Fecal/Soil Microbe MiniPrep Kit (Zymo Research, Orange, California), following the manufacturer's instructions. The V4-V5 regions from the 16S rRNA-encoding gene of isolated DNA were amplified by means of PCR, using the universal primers 515F-Y (5'-GTGYCAGCMGCCGCGGTAA) and 926R (5'-CCGYCAATTYMTTTRAGTTT]) [37]. The PCR mix contained 5 µl of KAPA 2G, 0.2 µl of each primer, 3.6 µl of ultrapure water and 1 µl of DNA for a final volume of 10 µl. After 3 min of initial denaturation at 95 °C, the following conditions were applied: 22 cycles of 95 °C for 45 s (denaturation), 50 °C for 45 s (annealing) and 68 °C for 90 s (extension), ending with a final extension at 68 °C for 5 min. For each sample, three PCRs were performed in the same conditions, to increase the DNA quantity, while minimizing cycle numbers to avoid PCR bias [38]. The product of each PCR was verified by agarose gel electrophoresis and triplicate reactions were pooled.

A second PCR was performed to attach the Illumina adapters and 8-bp barcodes for multiplexing. They were added in a reaction mix in which barcode sequences were individually added for each sample. The mix for this second PCR contained 12.5 µl of KAPA 2G, 0.5 µl of each barcode primer (Nextera Index Sequences in http://seq.liai.org/204-2/), 10.5 µl ultrapure water and 1 µl of DNA for a final volume of 25 µl. PCR conditions were as follows: initial denaturation at 98 °C for 30 s, 8 cycles of 98 °C for 10 s, 60 °C for 20 s, 72 °C for 30 s, and a final extension at 72 °C for 2 min. PCR products were purified using the USB ExoSAP-IT PCR Product Cleanup (Thermofisher, France) kit by incubating the samples at 37 °C for 30 min, followed by 15 min at 85 °C. The concentration of samples was normalized with the SequalPrep Normalization Plate (96 well) Kit (Thermofisher, France). Thereafter, all the samples with different barcode sequences belonging to each run were pooled, and the pooled products were quantified by using the Quant-iT^{TM} PicoGreen (Thermofisher, France). The normalized amplicons were concentrated by using the Wizard SV Gel and PCR Clean up Kit (Promega, France) and amplicons were sequenced one run in an Illumina® MiSeq sequencer using the 2 × 250 bp protocol by FASTERIS SA, Switzerland.

Sequence analyses

Sequence analysis was performed by using a combination of tools, including the USEARCH v9.0.2132 [39] program, Qiime V.1.9.1 [40] and *bash* scripts. Briefly, from demultiplexed fastq files obtained from the sequencing center, forward and reverse reads were merged to create consensus sequences in a single fastq file using *usearch9 -fastq_mergepairs*. Assembled reads were quality trimmed by the *usearch9 -fastq_filter* command. Primer sequences were removed from raw amplicon sequences using *strip_primers_exclude.py*. At this point, sequences were dereplicated and denoised using *usearch9 -unoise* (minampsize 8). As a certain number of artifactual short sequences remained, sequences less than 360 bp were removed with a combination of *bash* scripts and the *filter_fasta.py* command from Qiime V.1.9.1. OTU tables were generated by *usearch9 -usearch_global* against the primer-stripped raw sequences with a 97 % identity to OTU representatives ordered by abundance. OTUs were identified using Qiime V.1.9.1 *assign_taxonomy.py -m rdp* and a previously described modified version of the Greengenes [41] August 2013 database [9]. The taxonomy assignments were corrected to contain full paths from phylum to species. Final OTU tables were generated using biom-format. Sequences matching "Archaea", "Eukaryota", "Unassigned", "Chloroplasts" and "Mitochondria" were discarded using *filter_taxa_from_otu_table.py*. Details of data analysis, and the analysis environment are described in the supplemental methods. Further details on the bioinformatics methods, as well as the samples characteristics, including the number of reads and of sequences, are in the Electronic Supplementary Material. All the resultant sequences were deposited in the Sequence Read Archive (SRA, http://www.ncbi.nlm.nih.gov/sra) (Accession number: PRJNA531247).

Statistical analyses

R 3.5.0 [42] was used to calculate both α-diversity (Shannon diversity index) and β-diversity matrices of distances among samples. Matrices of genera were used since they contained the most detailed information on the total community, because in many cases species assignment was not possible using the rdp classifier and the sequenced region. However, in some of the OTUs, the genus classification was not possible, and in those cases we used the most precise taxonomic level that we could identify (for example the family, or the class). The Shannon diversity index based on OTUs was estimated by the R *vegan* package [43]. Due to the high variation of the index values, it was standardized

before subsequent analyses. This technique is very useful and provides simplification, since it modifies the scale and considers the mean as zero (e.g., [44]). This variable was fitted on a normal distribution and met the normality criteria. Differences in Shannon's diversity between fish species and body parts were analysed using one-way ANOVA in Statistica 7.1 [45]. In addition, a General Linear Mixed Model (GLMM) was applied to identify which variables (fish species, body part and sampling date, the last being considered as a random factor) influence bacterial richness.

To evaluate the effects that the species, body site, individual fish or sampling date would have in explaining the variability of the bacterial communities' composition in teleosts (i.e., β-diversity), we performed a PERMANOVA analysis. First, a distance matrix on resultant OTUs was constructed with the function *vegdist* in the R package "vegan", which was considered as the dependent variable. Here, bacterial communities of each sample are compared with all the rest, by means of the Bray-Curtis dissimilarity index for pairs of samples [43]. The index values ranging from 0 to 1: a value of 0 (i.e., no distance) indicated that two samples share all the bacterial taxa in the same abundances, and 1 (i.e., maximum distance) indicated that they did not share any of the detected OTUs. This distance matrix was considered as the dependent variable in further analysis. Then, the *adonis* function of the "vegan" R package was used with all the previously mentioned explanatory factors, considering the sampling date as the stratification variable. Bray-Curtis distances were also represented by means of a non-metric multidimensional scaling (NMDS), implemented in the R package *phyloseq* [46]. Because for some of the groups we have less than 3 replicates (Table 1), we performed the analyses with and without those groups, obtaining virtually the same results. We show here the results with $N > 3$ for each category, and in the supplementary material, the results obtained with the whole dataset (Tables S2, S3 and S4).

Finally, we performed an analysis to determine indicator species of the community in which the patterns of species distribution were compared among different categories (all body sites within each fish species), by using the *Indicspecies* R package [47, 48]. The results display only bacterial taxa that are significantly ($p < 0.05$) prevalent in a given category while absent in the rest. A target taxon is considered as "indicator" if it is exclusive to a given group (exclusivity) and occurs in all sample units within a group (fidelity). In this way, the analysis is based on the abundance within particular groups (exclusivity), and on the relative frequency of species within groups (fidelity; [31, 49]).

However, groups may be composed of a single or multiple categories (such as the case where indicator species are present in all of the samples of both categories but absent from the rest). An indicator value index is assigned between a species and each group, identifying the group with the highest association value. Then, randomization methods (permutation tests) are used to test the statistical significance of values [31, 49].

RESULTS

Bacterial taxonomic composition of the fish and water samples

We obtained a total of 2,964,227 sequences assigned to bacteria (i.e., filtering out reads belonging to Archaea, Eukarya and unidentified reads). After filtering, the remaining OTUs belonged to 181 genera, spread over 108 families, 67 orders, 37 classes and 13 phyla. The most abundant bacterial group in the fish samples was *Proteobacteria*, when considering both body site (Table 2) or host species (Table 3), and this group was also the most abundant in the water column. The second most abundant group in the water samples was *Bacteroidetes* (34.1 %), while in fishes this group was not highly represented, except in *S. salpa* (16.1 %). In fishes, the other abundant bacterial phyla were *Fusobacteria* and *Firmicutes*, that were rare in the surrounding water (0.7 and 0.3 % respectively). Of the most abundant taxa (> 1% of total sequences) *Proteobacteria* showed a relative abundance of around 80 % in the mucus, skin and gills, although its abundance was only 37.6 % in the intestine. The other most abundant taxa in the intestine were Tenericutes, *Firmicutes* (25.9 %), and *Fusobacteria* (13.7 %), and these groups were either absent from (Tenericutes), or showed lower abundances in the other three body sites.

In detail, the majority of the OTUs retrieved from fishes belonged to the family *Vibrionaceae* (*Proteobacteria*, class *Gammaproteobacteria*), and were mainly assigned to the genera *Vibrio* and *Photobacterium*, followed by the genus *Cetobacterium* (*Fusobacteria*; Table 2). The relative abundances of these three genera were similar in mucus, skin and gills, although in the latter tissue, *Cetobacterium* were more abundant than *Photobacterium* (Table 2). In the intestine, the most abundant group was the order *Clostridiales* (*Firmicutes*), and the Genus *Cetobacterium* (*Fusobacteria*). Within the family *Vibrionaceae*, the most abundant genera in the intestine included *Aliivibrio* and *Photobacterium*. *Vibrio* was clearly less abundant in the intestine compared to external

surfaces. The proportion of each bacterial group (for groups that represent more than 1% of the total abundance) varied among different species (Table 3), although *Photobacterium* and *Vibrio* were predominant in the majority of the species.

Flavobacteriales (Bacteroidetes phylum) were abundant in the seawater samples (30 %) as were the *Rhodobacterales* (family *Rhodobacteraceae*) and *Rickettsiales* (family *Pelagibacteriaceae*) orders of the Alphaproteobacteria class, showing respectively relative abundances of 18.7% and 14.5%. The class *Gammaproteobacteria* constituted 7.2 %, with *Vibrio* representing 1.5 % of total reads of this class and the *Synechococcaceae* family (Cyanobacteria) accounted for 4.2 of the total sequences.

Alpha-diversity of bacterial communities

There were significant differences in levels of Shannon α-diversity when comparing all the different fish species and water ($F_{12,174} = 2.16$, $p = 0{,}015$, Figure 1). However, when removing the water from the analyses, differences among species were no longer significant ($F_{11,171} = 1.54$, $p = 0.12$). This result implied that differences were among the water and the rest of species, with water displaying a higher α-diversity than all the fish species (Fig. 1). When separating by the water and different body sites, we found significant differences among groups, with again water showing the highest α-diversity, and then the mucus ($F_{4,82} = 20.39$, $p < 0.0001$, Fig. 2). Post-hoc tests revealed no significant differences among the water and the mucus samples ($p = 0.28$), while both categories were different from the other three groups (all $p < 0.001$). Therefore, in this case, when removing the water from the analyses, significant differences remained due to differences between the mucus and the rest of body sites ($F_{3,179} = 23.13$, $p < 0.0001$). No differences were found when comparing the skin, gills and intestine (all $p > 0.1$, Fig. 2). Results with all the dataset are presented in Table S2.

The GLMM results suggest that fish species ($F_{11,165} = 1.82$, $p = 0.05$) and body site ($F_{3,165} = 23.84$, $p < 0.0001$), but not the sampling day ($F_{3,165} = 0.85$, $p = 0.47$), influenced the Shannon diversity index. Results were the same with the entire dataset (Table S3). The α-diversity comparisons among the body sites in each species, and in all the species for each body site, are presented in Fig. S1 and S2.

Beta-diversity

The PERMANOVA results based on Bray-Curtis distances (i.e., β-diversity, Fig. S3) determined that both factors, fish species and body site, explained the variation in bacterial community composition among the samples (Table 4). This means that individual fish within a given species have significantly more similar bacterial communities than with fish from other species, which is also the case with the different body sites. However, individuals and sampling date did not have any significant effect on the structure of the microbiota, which means that individuals from the same species share their microbiota, independently of the sampling date (Table 4). This result clearly discards the possibility of any contamination among fishes at the same sampling date. Tests done with the whole dataset were similar (Table S4).

Indicator bacteria

Several indicator bacterial groups were significant (all $p < 0.05$, listed in Supplementary material, Table S5). Results displayed a high variation among different fish species regarding indicator microbial species. *Sarpa salpa* hosts the highest bacterial diversity, and also the highest number of indicator bacteria (95 in total, see Table S5). In addition, only this species hosted the phylum *Bacteroidetes*, which was relatively abundant in the water. No indicator bacterium was found in *Gobius cruentatus*, nor in the water samples. In all fish species except *S. salpa*, the gut samples did not share any indicator bacteria with the other three body sites. For the other body sites, gills and mucus shared indicator bacteria in five host species, the gills with the skin in two species, and the mucus with the skin in seven species. In six fish species, the mucus, skin and gills shared several indicator bacterial taxa (Table S5).

The body site with the highest number of indicator bacteria was the mucus (Table S5). However, in the species *S. notata* that lacks external mucus, indicator bacteria from skin were by far the most numerous in this part of the body (18 taxa against nine in *S. salpa*, and one in *G. bucchichi*). In the rest of the species, no indicator bacteria were found exclusively in the skin.

An overview of indicator bacteria from all fish species showed that mucus had the highest number of exclusive bacterial taxa (30 taxa), followed by skin (15 taxa, Fig. 3). Although there were many bacterial taxa that were shared between the different categories, the

highest number was shared by mucus, skin and gills, whereas the gut shared a lower number of indicator bacteria with the other three body sites (Fig. 3).

DISCUSSION

Our results confirm that *Proteobacteria* was the predominant phylum in water and teleost fish in the Mediterranean Sea, as it was indicated by previous studies [50]. However, the prevalence of *Bacteroidetes*, *Fusobacteria* and *Firmicutes* differed among the water and fish samples.

Both environmental and/or host-associated factors may shape fish microbiota at different levels [9, 51, 52, 53]. Previous studies found that the mucus microbiota may be highly variable and dynamic, depending on the environmental conditions including seasonality [54, 55]. However, other studies suggested that genetic factors have greater effect than environmental factors in shaping mucus bacterial assemblages [56, 57]. In our study, we have found that individuals from the same species shared significantly more bacterial communities than they did with individuals of the other species. This was observed despite being captured in the same place and having overall similar feeding habits, which suggests species-specific assemblages. These results are in accordance with previous works claiming that the autochthonous microbes are not a passive reflection of their habitat communities [11], i.e., fish have their own microbiota. We did not detect any effect of the sampling date, although the variation in abiotic factors was moderate among our four samplings (see Table S1). Therefore, other factors may influence teleosts' microbiota, such as seasonal environmental variation throughout the year (e.g. related to temperature or salinity changes), which should be tested in future studies.

Bacterial communities from a given body site were significantly more similar among themselves than with communities from other body sites. Although the majority of bacteria from mucus, skin and gills belong to the Family *Vibrionaceae*, there are differences in the relative prevalence of genera among the gills with respect to mucus and skin. The highest bacterial diversity in the body was found in the outermost external layer, i.e., the mucus, which is in intimate contact with water, thus a continuous exchanging of bacteria among them could be expected. The mucus acts as a protective barrier for fishes, since it may impede the penetration of potentially pathogenic microorganisms inside the body [24]. Our findings that the mucus had a significantly higher diversity than the rest

of the body (including the skin), is in accordance with this protection hypothesis, suggesting that environmental bacteria remain trapped here. Moreover, the fact that *S. notata*, the only fish species without mucus from our sample, had the highest diversity of indicator species in the skin, also supports this hypothesis since in this case, the skin is the direct interface between the fish and the environment. However, we cannot exclude that since the mucus is rich in polysaccharides (e.g. [58, 59]), it may also constitute a food source that attracts heterotrophic bacteria, which may in turn lead to a more diverse community.

Pathogenic bacteria can enter the host fish through the skin, gills or gastrointestinal tract, and the integrity of these physical and immunological barriers determine the outcome of host-pathogen interactions (reviewed in [13]). Balanced and complex interplays within the mucus layer are thus key to disease resistance [60], and are essential for supporting host health and fitness [61].

Differences in the relative abundance of bacteria between different sites on the body are more pronounced between the gut and the other three external body parts (mucus, skin and gills, Table 2). For instance, the phylum *Proteobacteria* was found at lower relative prevalence in the intestine than in the other three parts, and relative abundance of *Firmicutes* and *Fusobacteria* also differ. In addition, we found that indicator bacteria can be shared between the gills, mucus and skin (see Table S5); however, almost no indicator bacteria were shared by the gut and those three external body sites. Therefore, our results suggest that, although we found that body site explains a significant fraction of the variance observed in bacterial communities over all the body, it is mostly because the intestine differs markedly from the rest of the body sites. Overall, these results were not unexpected given the very different physico-chemical conditions (oxygen, pH and organic substrate levels) occurring in guts compared to external body parts. Bacterial community composition has been previously shown to vary in the gut depending on the life cycle stage, diet, environment, and region of the gastrointestinal tract, and even varies greatly among individuals within the same species [2, 5, 10, 16, 61, 62]. In our study, we only examined the distal part of the gut, and individuals from the same species were adults, so we did not find any significant within-species effect, while the fish species explained most of the variance in bacterial assemblages.

There are some groups that predominate in the gut of marine fish across different studies, such as *Proteobacteria*, *Firmicutes*, *Actinobacteria* and *Bacteroidetes*. Also,

Fusobacteria, *Clostridia*, *Bacilli* and *Verrucomicrobia* have been identified (reviewed in [2, 10]), as well as *Mycoplasma* [5], a group that we found exclusively in the gut samples of teleosts. There is some congruence in the identity of bacterial taxa we found in the gut (including *Vibrio*, *Clostridium* and *Mycoplasma*) with other marine fish belonging to different species, including some from aquaculture (e.g. Atlantic salmon, [5]). The presence of similar bacterial taxa in the gut-associated microbiota across different fish species, populations or geographic locations, suggests that these microbes are important contributors to host gut functions, such as digestion, nutrient absorption and immune response [63]. In general, although the intestine ecosystem is expected to harbour a dense population of microbes, sequence-based analyses have demonstrated that it comprises the lowest phylogenetic diversity compared to external parts of the body [2]. This was also observed in the present study when comparing α-diversity (Shannon diversity index) to the external mucus layer. Previous works also found a higher bacterial diversity in external organs such as skin or gills than in the gut [56, 60], which could be due to a reflection of environmental diversity in the exterior of the body, whereas the gut may offer more stable habitats leading to a more specialized microbial communities [60].

There was no evidence that sampling day had an effect on the microbiota, which is in accordance with the difference between fish and water microbiota. That is, the identity of the fish species was more important in the structuring of bacterial communities than environmental conditions at the moment of their capture. This observation suggests that our results are not biased by the sampling date, and that fishes display their characteristic microbiota independently of the external conditions, at least between these two time points. Different samplings in the same location and with the same species could be performed throughout the year to verify an effect of seasonality on the bacterial diversity of our fish community as described in other aquatic systems (e.g. [54]).

In conclusion, we found that the thirteen Mediterranean teleost species studied here, living in the same environment and sharing most ecological traits, differed in their bacterial microbiota composition. This suggests that host taxonomic status mainly shapes fish microbiota. In addition, we found a characteristic microbial community in different parts of the body, indicating that microbiota are also influenced by local characteristics of their animal-associated microhabitats. The absence of a significant effect of individuals on the bacterial communities that they host suggests that the level of intra-specific variation is significantly lower than the level of inter-tissue and inter-species variations

in shaping fish microbiota. Interestingly, the mucus cover showed the highest bacterial diversity, which supports the hypothesis that it is a barrier between the fish and its environment. That is, bacteria may be retained in this layer but do not reach the skin, and therefore do not penetrate into the body. This physical barrier may thus help to impede pathogenic infections.

Microbial communities associated with fishes are key factors in host physiology, ecology and evolution [1, 64]. Therefore, the knowledge of the factors shaping microbiota may help predict how changes in abiotic or biotic conditions affect bacterial assemblages and their functions. Our results help to clarify these factors in wild fish, highlighting the importance of fish species and body sites. Further studies should investigate the possible functions of bacterial taxa on their hosts, and also investigate environment-induced variations through time-series samplings.

Acknowledgements

This work was funded by Sorbonne Université, programme Emergence 2016 (project SU-16-R-EMR-22-MICROFISH), which included a post-doctoral contract for MRR. Fish sampling was performed by the "Service des Moyens à la Mer (FR3724) de l'Observatoire Océanologique de Banyuls/Mer (OOB)". We are grateful to the BIO2MAR platform (http://bio2mar.obs-banyuls.fr) for providing technical help and support and access to instrumentation.

Ethics statement

All applicable international, national, and/or institutional guidelines for the use of animals were followed. The Observatoire Océanologique de Banyuls sur Mer holds the authorization from the "Direction interrégionale de la Mer Méditerrannée" for fishing and handling wild Mediterranean teleosts. Wild fish were caught (see above for details) by competent persons on the research vessel "Nereis II" and in accordance with the European Union Regulations concerning the protection and welfare of experimental animals (European directive 91/492/CCE).

455 **References**

456 1. Mcfall-Ngai M, Hadfield MG, Bosch TC et al (2013) Animals in a bacterial world,
457 a new imperative for the life sciences. Proc Natl Acad Sci USA 110:3229-36.

458 2. Ghanbari M, Kneifel W, Domig KJ (2015) A new view of the fish gut
459 microbiome: Advances from next-generation sequencing. Aquaculture 448:464-475.

460 3. Kohl KD (2018) A microbial perspective on the grand challenges in comparative
461 animal physiology. mSystems 3:e00146-17.

462 4. Nelson JS, Grande TC, Mark WVH (2016) Fishes of the World, 5th Edition. John
463 Wiley & Sons, New York.

464 5. Ciric M, Waite D, Draper J, Jones JB (2018) Characterisation of gut microbiota
465 of farmed Chinook salmon using metabarcoding. BioRxiv 288761.
466 https://www.biorxiv.org/content/early/2018/03/26/288761.full.pdf+html

467 6. Barko PC, McMichael MA, Swanson KS, Williams DA (2018) The
468 gastrointestinal microbiome: a review. J Vet Intern Med 32: 9-25.

469 7. Rosado D, Pérez-Losada M, Severino R, Cable J, Xavier R (2018)
470 Characterization of the skin and gill microbiomes of the farmed seabass (*Dicentrarchus*
471 *labrax*) and seabream (*Sparus aurata*). Aquaculture 500:57–64. doi:
472 10.1016/j.aquaculture.2018.09.063

473 8. Tarnecki AM, Burgos FA, Ray CL, Arias CR (2017) Fish intestinal microbiome:
474 diversity and symbiosis unraveled by metagenomics. J Appl Microbiol
475 http://doi.wiley.com/10.1111/jam.13415

476 9. Reverter M, Sasal P, Tapissier-Bontemps N, Lecchini D, Suzuki M (2017)
477 Characterisation of the gill mucosal bacterial communities of four butterflyfish species:
478 a reservoir of bacterial diversity in coral reef ecosystems. FEMS Microbiol Ecol 93(6).
479 doi: 10.1093/femsec/fix051.

480 10. Llewellyn MS, Boutin S, Hoseinifar SH, Derome N (2014) Teleost microbiomes:
481 the state of the art in their characterization, manipulation and importance in aquaculture
482 and fisheries. Front Microbiol 5:207.
483 http://journal.frontiersin.org/article/10.3389/fmicb.2014.00207/abstract

484

485 11. Sullam KE, Essinger SD, Lozupone CA, O'Connor MP, Rosen GL, Knight R,
486 Kilham SS, Russell JA (2012) Environmental and ecological factors that shape the gut
487 bacterial communities of fish: a meta-analysis. Mol Ecol 21:3363-78.

488 12. Eichmiller JJ, Hamilton MJ, Staley C, Sadowsky MJ, Sorensen PW (2016)
489 Environment shapes the fecal microbiome of invasive carp species. Microbiome 4:44.

490 13. Li X, Ringø E, Hoseinifar SH, Lauzon HL, Birkbeck H, Yang D (2018). The
491 adherence and colonization of microorganisms in fish gastrointestinal tract. Rev Aquacult
492 https://doi.org/10.1111/raq.12248

493 14. Stagaman K, Burns AR, Guillemin K, Bohannan BJM (2017) The role of adaptive
494 immunity as an ecological filter on the gut microbiota in zebrafish. ISME J 11:1630-1639.

495 15. Wong S, Rawls JF (2012) Intestinal microbiota composition in fishes is influenced
496 by host ecology and environment. Mol Ecol 21:3100-3102.

497 16. Nayak SK (2010) Role of gastrointestinal microbiota in fish. Aquacult Res
498 41:1553-1573.

499 17. Xing M, Hou Z, Yuan J, Liu Y, Qu Y, Liu B (2013) Taxonomic and functional
500 metagenomic profiling of gastrointestinal tract microbiome of the farmed adult turbot
501 (*Scophthalmus maximus*). FEMS Microbiol Ecol 86:432-443.

502 18. Li T, Long M, Gatesoupe FJ, Zhang Q, Li A, Gong X (2015) Comparative analysis
503 of the intestinal bacterial communities in different species of carp by pyrosequencing.
504 Microb Ecol 69:25-36.

505 19. Ellis AE (2001) Innate host defense mechanisms of fish against. Dev Comp
506 Immunol 25:827-839.

507 20. Chen CY, Chen PC, Weng FCH, Shaw GTW, Wang D (2017) Habitat and
508 indigenous gut microbes contribute to the plasticity of gut microbiome in oriental river
509 prawn during rapid environmental change. PLoS ONE 12:e0181427

510 21. Esteban MA (2012) An overview of the immunological defenses in fish skin.
511 ISRN Immunol 853470:1-29. doi: 10.5402/2012/853470

512 22. Legrand TPRA, Catalano SR, Wos-Oxley ML, Stephens F., Landos M, Bansemer
513 MS, Stone DAJ, Qin JG, Oxley APA (2018) The inner workings of the outer surface: skin

514 and gill microbiota as indicators of changing gut health in yellowtail kingfish. Front
515 Microbiol 8:2664. doi: 10.3389/fmicb.2017.02664

516 23. Marshall WS, Bellamy D (2010) The 50year evolution of in vitro systems to reveal
517 salt transport functions of teleost fish gills. Comp Biochem Physiol A Mol Integr Physiol
518 155:275-280. doi: 10.1016/j.cbpa.2009.11.016

519 24. Peatman E, Lange M, Zhao H, Beck BH (2015) Physiology and immunology of
520 mucosal barriers in catfish (*Ictalurus* spp.). Tissue Barriers 3:e1068907. doi:
521 10.1080/21688370.2015.1068907

522 25. Derome N, Gauthier J, Boutin S, Llewellyn M (2016) Bacterial opportunistic
523 pathogens of fish. In: Hurst CJ (ed) The Rasputin effect: when commensals and symbionts
524 become parasitic, Springer International Publishing, Cham, pp 81-108.

525 26. Boutin S, Audet C, Derome N (2013) Probiotic treatment by indigenous bacteria
526 decreases mortality without disturbing the natural microbiota of *Salvelinus fontinalis*.
527 Canadian J Microbiol 59:662–670.

528 27. Ye L, Amberg J, Chapman D, Gaikowski M, Liu WT (2014) Fish gut microbiota
529 analysis differentiates physiology and behavior of invasive Asian carp and indigenous
530 American fish. ISME J 8:541-551.

531 28. Chiarello M, Villéger S, Bouvier C, Bettarel Y, Bouvier T (2015) High diversity
532 of skin-associated bacterial communities of marine fishes is promoted by their high
533 variability among body parts, individuals and species. FEMS Microbiol Ecol 91:fiv061.

534 29. Sullam KE, Rubin BE, Dalton CM, Kilham SS, Flecker AS, Russell JA (2015)
535 Divergence across diet, time and populations rules out parallel evolution in the gut
536 microbiomes of Trinidadian guppies. ISME J 9:1508-1522.
537 http://dx.doi.org/10.1038/ismej.2014.1231.

538 30. McGeoch MA, Chown SL (1998) Scaling up the value of bioindicators. Trend
539 Ecol Evol 13:46-47.

540 31. Legendre P, Legendre L (2012) Numerical ecology, 3rd English edition. Elsevier
541 Science BV, Amsterdam.

32. Chiba SN, Iwatsuki Y, Yoshino T, Hanzawa N (2009) Comprehensive phylogeny of the family Sparidae (Perciformes: Teleostei) inferred from mitochondrial gene analyses. Genes Genet Syst 84:153-170.

33. Louisi P (2015) Europe and Mediterranean marine fish identification guide. Eugen Ulmer, Paris.

34. Pallaoro A, Dulcic S, Matic–Skoko M, Kraljevic M, Jardas I (2008) Biology of the salema *Sarpa salpa* (L. 1758) (Pisces, Sparidae) from the Middle Eastern Adriatic. J Appl Ichthyol 24:276-281.

35. Savage DC (1977) Microbial ecology of the gastrointestinal tract. Annu Rev Microbiol 31:107-133.

36. Vaahtovuo J, Toivanen P, Eerola E (2001) Study of murine fecal microflora by cellular fatty acids analysis; effect of age and mouse strain. A Van Leeuw 80:35-42.

37. Parada AE, Needham DM, Fuhrman JA (2016) Every base matters: Assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. Environ Microbiol 18:1403–14.

38. Suzuki M, Giovannoni SJ (1996) Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. Appl Environ Microbiol 62:625-

39. Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26:2460-2461.

40. Caporaso JG, Kuczynski J, Stombaugh J et al (2010) QIIME allows analysis of high-throughput community sequencing data. Nat Methods 7:335-6. http://www.nature.com/doifinder/10.1038/nmeth.f.303

41. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, Andersen GL (2006) Greengenes, a Chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Env Microbiol 72:5069-5072.

42. R Core Team (2018). The R Project for Statistical Computing. https://www.r-project.org

43. Oksanen J, Blanchet FG, Friendly M et al (2018) Vegan: Community Ecology Package. R package version 2.5-2. https://CRAN.R-project.org/package=vegan

571 44. Sokal RR, Rohlf FJ (2015) Biometry. The principles and practice of statistics in
572 biological research. 4th edition. W.H. Freeman and Company, New York.

573 45. Statsoft Inc (2005) STATISTICA (data analysis software system), version 7.1.

574 46. McMurdie PJ, Holmes S (2013) phyloseq: An R package for reproducible
575 interactive analysis and graphics of microbiome census data. Plos One 8: e61217.

576 47. De Cáceres M, Legendre P (2009) Associations between species and groups of
577 sites: indices and statistical inference. Ecology 90:3566-3574.

578 48. De Cáceres M, Legendre P, Moretti M (2010) Improving indicator species
579 analysis by combining groups of sites. Oikos 119:1674-1684.

580 49. Dufrêne M, Legendre P (1997) Species assemblages and indicator species: the
581 need for a flexible asymmetrical approach. Ecol Monogr 67:345-366.

582 50. Crespo BG, Pommier T, Fernández-Gómez B, Pedrós-Alio P (2013) Taxonomic
583 composition of the particle-attached and free-living bacterial assemblages in the
584 Northwest Mediterranean Sea analyzed by pyrosequencing of the 16S rRNA. Microbiol
585 Open 2:541-552.

586 51. Semova I, Carten JD, Stombaugh J, Mackey LC, Knight R, Farber SA, Rawls JF
587 (2012) Microbiota regulate intestinal absorption and metabolism of fatty acids in the
588 zebrafish. Cell Host Microbe 12:277-288.

589 52. Navarrete P, Magne F, Araneda C, Fuentes P, Barros L, Opazo R, Espejo R,
590 Romero J (2012) PCR-TTGE analysis of 16S rRNA from rainbow trout (*Oncorhynchus*
591 *mykiss*) gut microbiota reveals host-specific communities of active bacteria. PLoS One
592 7:e31335.

593 53. Franchini P, Fruciano C, Frickey T, Jones JC, Meyer A (2014) The gut microbial
594 community of Midas cichlid fish in repeatedly evolved limnetic–benthic species pairs.
595 PLoS One 9:e95027.

596 54. Larsen AM, Bullard SA, Womble M, Arias CR (2015) Community structure of
597 skin microbiome of gulf killifish, *Fundulus grandis*, is driven by seasonality and not
598 exposure to oiled sediments in a Louisiana salt marsh. Microb Ecol 70:534-544.

599 55. Schmidt VT, Smith KF, Melvin DW, Amaral-Zettler LA (2015) Community
600 assembly of a euryhaline fish microbiome during salinity acclimation. Mol Ecol 24:2537-
601 2550.

602 56. Smith CJ, Danilowicz BS, Meijer WG (2007) Characterization of the bacterial
603 community associated with the surface and mucus layer of whiting (*Merlangius*
604 *merlangus*). FEMS Microbiol Ecol 62:90-97.

605 57. Larsen A, Tao Z, Bullard SA, Arias CR (2013) Diversity of the skin microbiota
606 of fishes: evidence for host species specificity. FEMS Microbiol Ecol 85:483-494.

607 58. Qin C, Huang K, Xu H (2002) Isolation and characterization of a novel
608 polysaccharide from the mucus of the loach, *Misgurnus anguillicaudatus*. Carbohydr.
609 Polym. 49:367–371.

610 59. Zhang C, Huang KX (2005) Characteristic immunostimulation by MAP, a
611 polysaccharide isolated from the mucus of the loach, *Misgurnus anguillicaudatus*.
612 Carbohydr. Polym. 59:75–82.

613 60. Lowrey L, Woodhams DC, Tacchi L, Salinas I (2015) Topographical mapping of
614 the rainbow trout (*Oncorhynchus mykiss*) microbiome reveals a diverse bacterial
615 community with antifungal properties in the skin. Appl Env Microbiol 81:6915–6925.

616 61. Gómez GD, Balcázar JL (2018) A review on the interactions between gut
617 microbiota and innate immunity of fish. FEMS Immunol Med Microbiol 52:145-54.
618 http://www.ncbi.nlm.nih.gov/pubmed/18081845

619 62. Gallet A, Koubbi P, Léger N, Scheifler M, Ruiz-Rodríguez M, Suzuki MT,
620 Desdevises Y, Duperron S (2019) Low-diversity bacterial microbiota in Southern Ocean
621 representatives of lanternfish genera *Electrona, Protomyctophum* and *Gymnoscopelus*
622 (family Myctophidae). PLoS ONE 14(12): e0226159.
623 https://doi.org/10.1371/journal.pone.0226159

624 63. Roeselers G, Mittge EK, Stephens WZ et al (2011) Evidence for a core gut
625 microbiota in the zebrafish. ISME J 5:1595-608.
626 http://www.ncbi.nlm.nih.gov/pubmed/21472014

627 64. Tremaroli V, Backhed F (2012) Functional interactions between the gut
628 microbiota and host metabolism. Nature 489:242-249.
629 https://doi.org/10.1038/nature11552.

630 **Table 1**. Total number of DNA samples sequenced, by category: fish species, body site,
631 and water. Note that *Spicara maena* is considered as belonging to the family Sparidae
632 based on phylogenetic evidence [32].

Fish Family	Fish species	No. of individuals	Mucus	Skin	Gills	Gut
Gobiidae	*Gobius bucchichi*	5	3	5	5	3
	Gobius cruentatus	2	2	2	2	2
	Gobius niger	3	2	3	3	3
Labridae	*Symphodus tinca*	5	5	5	5	3
Scorpaenidae	*Scorpaena notata*	5	-	5	5	4
Serranidae	*Serranus scriba*	5	5	5	5	3
Sparidae	*Diplodus annularis*	5	5	5	5	3
	Diplodus vulgaris	5	2	5	5	2
	Oblada melanura	5	4	4	5	3
	Pagellus bogaraveo	4	3	4	4	2
	Pagellus erythrinus	5	5	5	5	5
	Sarpa salpa	5	5	5	5	3
	Spicara maena	5	5	5	5	2
Total of each category		59	46	58	59	38
Water samples (1 per sampling)		4				
Total samples			**205**			

633

634 **Table 2**. Percentage of the most abundant bacterial taxa (abundance higher than 1 % in the total community) found in the Mediterranean teleost
635 fish captured in the present work, in total and in each body site.

Taxonomy					**Relative abundance (% of reads)**					
Phylum	**Class**	**Order**	**Family**	**Genera**	**Total**	**Mucus**	**Skin**	**Gills**	**Intestine**	**Water**
Firmicutes	Clostridia	Clostridiales	Lachospiraceae	*Clostridium*	2.2	-	1.5	-	8.6	-
			Peptostreptococcaceae		3.4	1.7	2.3	-	12.8	-
Fusobacteria	Fusobacteriia	Fusobacteriales	Fusobacteriaceae	*Cetobacterium*	10.5	5.7	6.2	18.2	13.2	0.68
				Propionigenium	1.5	2.7	1.4	1.5	-	0.08
Proteobacteria	Betaproteobacteria				2.3	-	-	8.6	-	1.06
	Gammaproteobacteria	Alteromonadales	Shewanellaceae	*Shewanella*	5.5	6.3	6.5	7.3	-	-
		Oceanospirillales	Endozoicomonaceae	*Endozoicomonas*	2.1	-	1	6.4	-	-
		Vibrionales	Vibrionaceae	*Aliivibrio*	3.2	3.1	5.8	2.2	12.4	0.03
				Enterovibrio	1.4	1.5	1.7	-	2.2	0.1
				Photobacterium	15.3	21.9	20.4	15.5	9.7	0.35
				Vibrio	29.5	42.5	38	25.1	1.5	6.2
Tenericutes	Mollicutes	Mycoplasmatales	Mycoplasmataceae		3.1	-	-	-	15.4	-

636

637 **Table 3.** Percentage of the most abundant bacterial taxa (abundance higher than 1 % in the total community) in each fish species. Legend: DA:
638 *Diplodus annularis*; DV: *Diplodus vulgaris*; GB: *Gobius bucchichi*; GC: *Gobius cruentatus*; GN: *Gobius niger*; OM: *Oblada melanura*; PB:
639 *Pagellus bogaraveo*; PE: *Pagellus erythrinus*; SM: *Spicara maena*; SN: *Scorpaena notata*; SSa: *Sarpa salpa*; SSc: *Serranus scriba*; ST: *Simphodus*
640 *tinca*.

641

Taxonomy					**Relative abundance (% of reads)**													
Phylum	**Class**	**Order**	**Family**	**Genera**	**Total**	**DA**	**DV**	**GB**	**GC**	**GN**	**OM**	**PB**	**PE**	**SM**	**SN**	**SSa**	**SSc**	**ST**
Firmicutes	Clostridia	Clostridiales	Clostridiaceae	*Clostridium*	2.2	0.36	0.83	0.16	0.24	3.43	0.39	0.37	4.08	0.6	21.87	1.65	5.72	0.25
			Peptostreptococcaceae		3.4	1.81	10.68	11.81	0.52	7.82	0.52	2.72	12.34	0.23	1.58	0.32	0.26	3.93
Fusobacteria	Fusobacteriia	Fusobacteriales	Fusobacteriaceae	*Cetobacterium*	10.5	5.34	8.71	5.03	16.21	2.81	10.27	24.86	19.25	23.98	3.93	8.84	15.78	10.41
				Propionigenium	1.5	3.02	1.11	3.67	0.37	1.79	0.49	2	0.12	0.82	0.02	0.72	0.36	7.77
Proteobacteria	Betaproteobacteria				2.3	0.64	4.52	4.6	0.03	9.04	0.01	6.71	0.02	0	0.42	18.38	-	-
	Gammaproteobacteria	Alteromonadales	Shewanellaceae	*Shewanella*	5.5	6.64	3.93	4.43	2.08	1.88	8.97	6.56	10.91	7.69	1.84	2.42	1.11	18.72
		Oceanospirillales	Endozoicimonaceae	*Endozoicomonas*	2.1	0.01	0.01	11.41	1.71	7.51	0.14	-	-	-	25.07	-	0.01	0.01
		Vibrionales	Vibrionaceae	*Aliivibrio*	3.2	0.73	13.64	0.74	1.12	0.48	7.94	1.16	1.16	6.34	2.36	8.17	5.18	2.2
				Enterovibrio	1.4	1.12	1.77	0.21	0.3	12.19	4.21	0.43	0.66	0.46	2.67	0.75	0.83	0.57
				Photobacterium	15.3	16.09	11.52	9	10.87	5.76	16.71	19.88	33.16	24.41	19.94	26.78	25.89	8.84
				Vibrio	29.5	56.94	42.83	25.5	48.14	27.76	46.55	32.82	19.3	34.77	18.82	31.38	43.77	35.62
			Vibrionaceae		1.7	7.3	0.44	0.23	0.1	0.19	3.54	2.46	0.11	0.7	0.71	0.55	0.97	6.24
Tenericutes	Mollicutes	Mycoplasmatales	Mycoplasmataceae		3.1	0.01	0.01	23.21	18.31	19.34	0.26	0.05	-	-	0.75	0.07	0.12	5.42

642

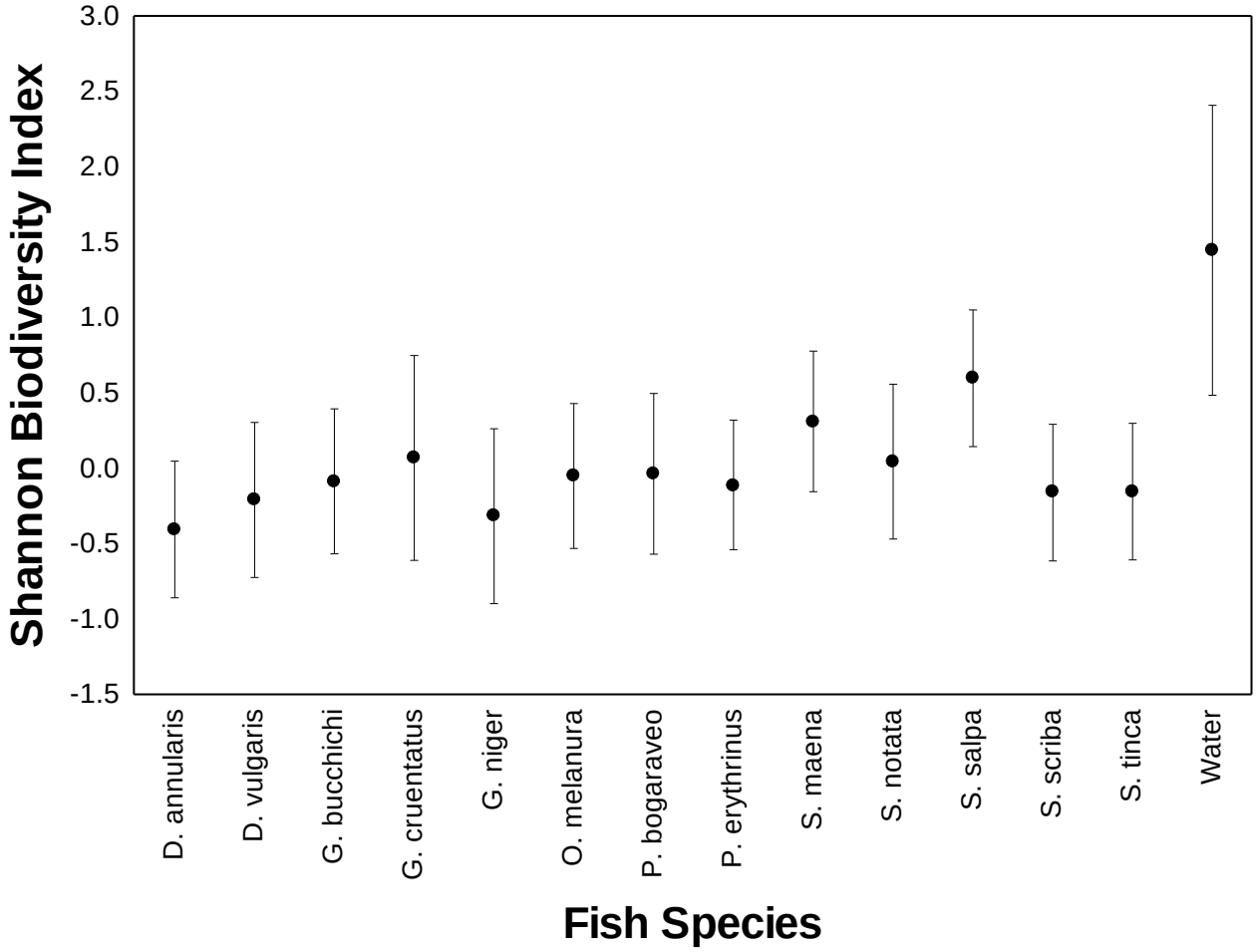
643 **Table 4**. Results of PERMANOVA on factors explaining bacterial communities
644 composition (β-diversity). Significant results are in bold.

	F	D. f.	p-value
Species	2.74	12,169	**0.001**
Body site	4.68	3,169	**0.001**
Individual	1.63	1,169	0.173
Sampling event	1.72	1,169	0.117

645

646 **Figure 1**. Standardized Shannon diversity index of species and water. Vertical bars
647 denote 0.95 confidence intervals.

648



649

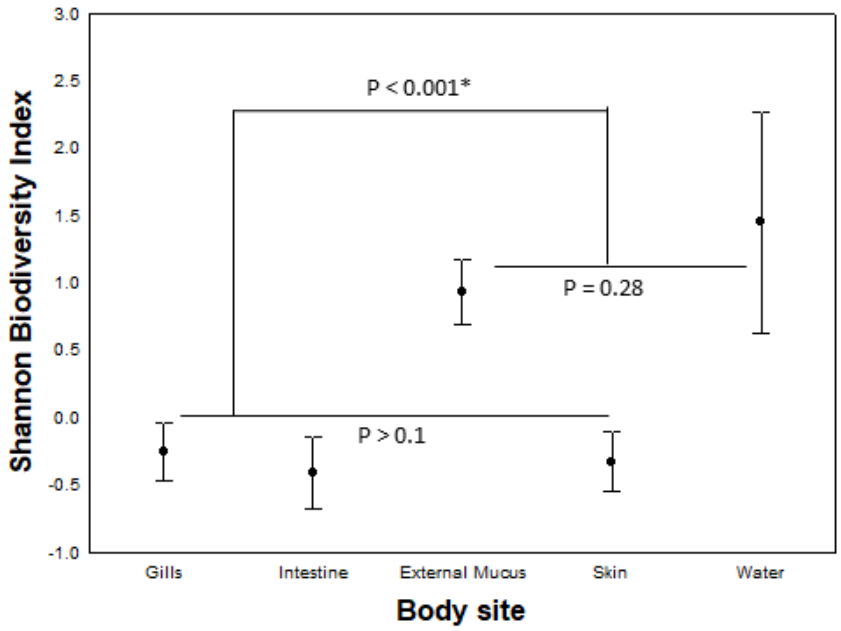
650

651

652

653 **Figure 2**. Standardized Shannon diversity index of different body parts and water.
654 Vertical bars denote 0.95 confidence intervals.

655



656

Figure 3. Venn diagram representing how many indicator bacterial taxa are exclusive or shared among the different body sites across all the fish species. More information on the specific indicator bacteria is given in Table S5.

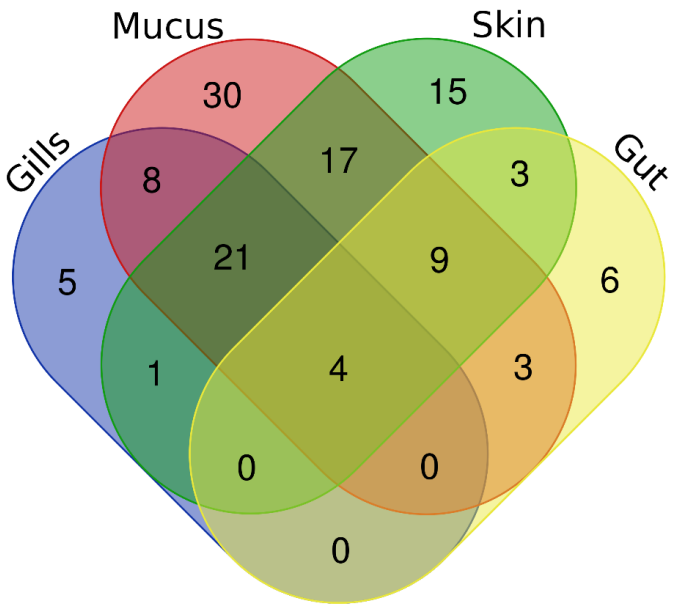


Figure 1

Click here to access/download;Figure;Fig 3.tif

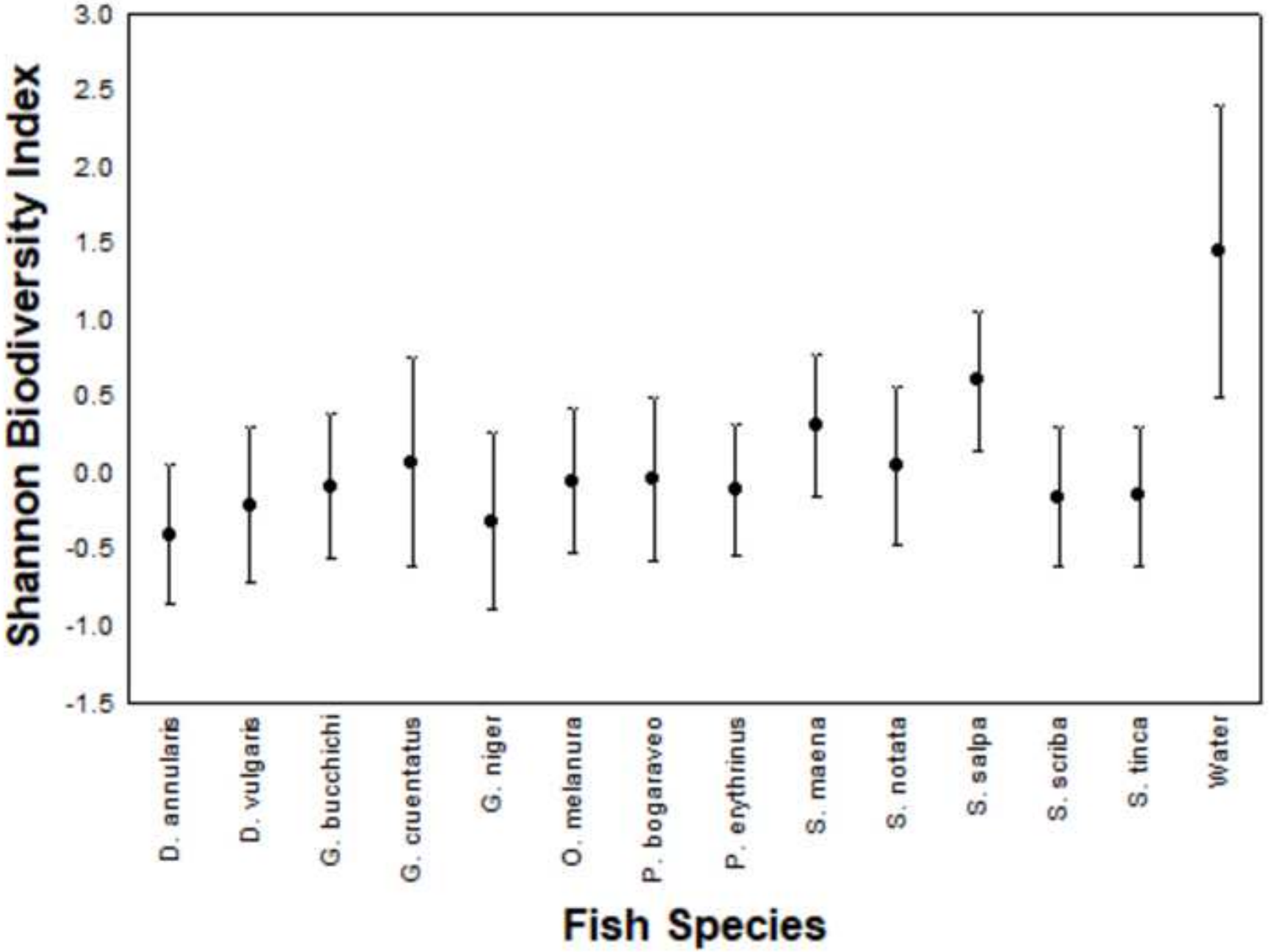


Figure 2

Click here to access/download;Figure;Fig 2.tif

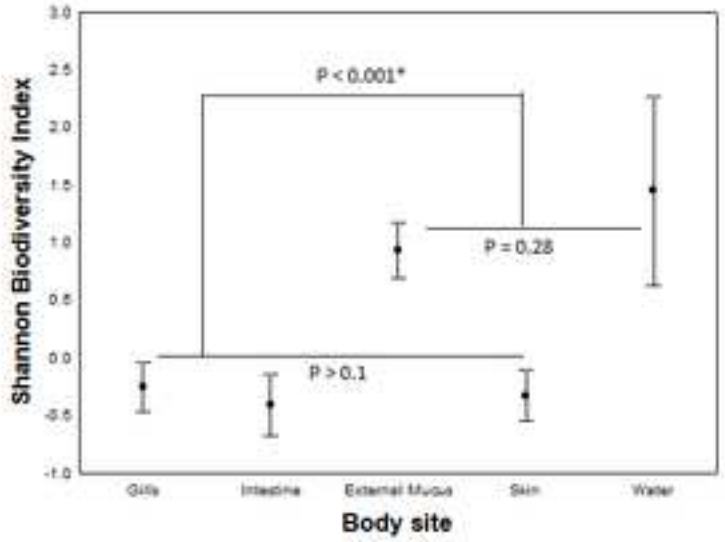


Figure 3

Click here to access/download;Figure;Fig 5.png

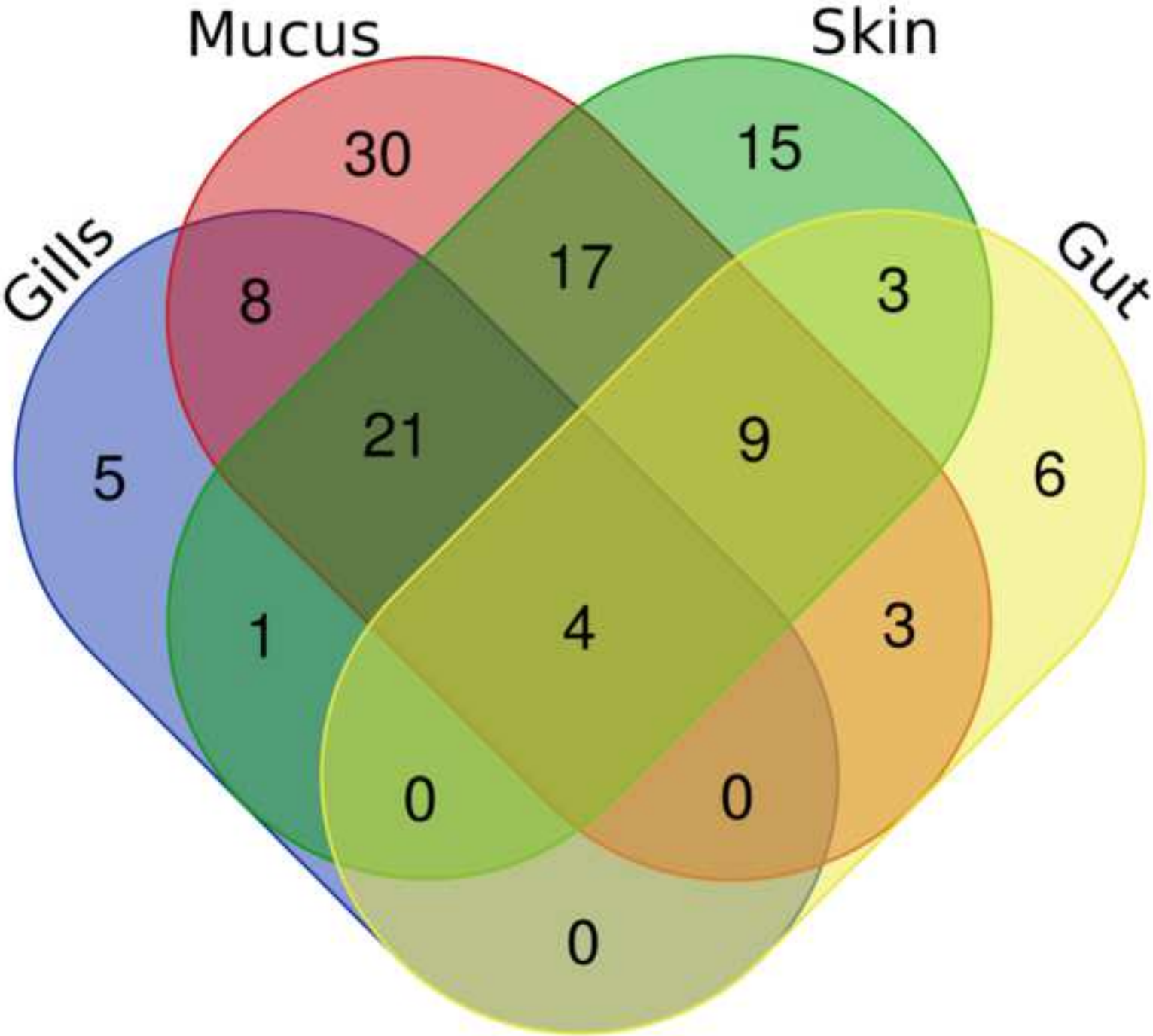


Table S1. Abiotic descriptors for each sampling day obtained from the station SOLA (http://sooob.obs-banyuls.fr/fr/les_sites_d_observation/le_site_sola.html), located in the Bay of Banyuls-sur-Mer, close to the fish sampling sites. All captures were performed in 2017.

	June 21		**June 27**		**July 18**		**October 4**	
	Surface	**Seabed**	**Surface**	**Seabed**	**Surface**	**Seabed**	**Surface**	**Seabed**
Temperature ºC	20.92	16.319	27.97	15.986	22.235	14.5	19.378	18.015
Salinity (PSU)	37.896	38.223	37.798	38.012	37.952	38.457	38.098	38.228
Oxygen (ml/l)	5.29	5.64	5.07	5.78	4.84	5.33	5.24	5.27
pH	8.14	8.14	8.11	8.14	8.07	8.08	8.03	8.07
NH_4 (µmol/l)	0	0.01	0.09	0.04	0.052	0.034	0.014	0.019
NO_3 (µmol/l)	0.08	0.04	0.18	0.21	0.08	0.54	0.11	0.02
NO_2 (µmol/l)	0.01	0.01	0.01	0.01	0.02	0.09	0.02	0.01
PO_4 (µmol/l)	0.01	0.01	0.01	0.01	0.01	0.06	0.01	0.01
$SiOH_4$ (µmol/l)	0.6	0.86	0.5	0.6	0.55	1.52	0.76	0.8

Table S2. Results of ANOVA testing on differences in α-diversity (Shannon diversity index) among different species, body sites and water, including all the samples, i.e., with categories in which N < 3.

	F	df	p	Fisher LSD Post-hocs	
Species with water	1.82	13, 191	0.04		
Species without water	1.21	12, 188	0.28		
Body sites with water	23.44	4, 200	< 0.0001	Water *vs*. mucus	p = 0.24
				Water *vs*. Mucus *vs*. Gills / Skin / Gut	p < 0.001
Body sites without water	18.07	3, 197	< 0.0001	Mucus *vs*. Gills / Skin / Gut	p < 0.001
				Skin *vs*. Gills *vs*. Gut	All p > 0.36

Table S3. Results of the GLMM analysis considering all the dataset (i.e., including those categories with N < 3).

	F	df	p
Fish species	1.83	12,185	0.04
Body site	29.17	3,185	<0.0001
Sampling day	0.75	3,185	0.50

Table S4. Results of the PERMANOVA analysis considering all the dataset (i.e., including those categories with N < 3).

	F	D. f.	p-value
Species	2.50	13,169	**0.001**
Body site	5.06	3,169	**0.001**
Individual	1.57	1,169	0.193
Sampling day	1.91	1,169	0.054

Table S5. Distribution of indicator bacterial groups among all the body sites, in each fish species. Legends: G: Gills, I: Intestine, M: Mucus, S: Skin

Fish species	No. bacterial taxa	No. indicator bacterial taxa	Distribution in the body	Bacterial taxa
Diplodus annularis	59	12	G	*Unidentified Betaproteobacteria, Neorickettsia*
			I	*Unidentified Peptostreptococcaceae*
			M + S	*Pseudoalteromonas*
			G + M	*Bacillus, Unidentified Saprospiraceae, Tenacibaculum, Hyphomonadaceae, Croceitalea, Verrucomicrobium, Winogradskyella,*
			G + M + S	*Clostridium*
Diplodus vulgaris	66	3	G	*Bacillales*
			M	*Unidentified Deltaproteobacteria*
			M + S	*Unidentified Ruminococcaceae*
Gobius bucchichi	43	17	G	*Polynucleobacter, Unidentified Betaproteobacteria*
			I	*Unidentified Rhodobacteraceae, Actibacter, Unidentified Planctomycetes, Loktanella, Lutimonas, Unidentified Desulfobulbaceae, Thiohalorhabdales, Aliivibrio, Lawsonia, Unidentified Desulfovibrionaceae*
			M	*Flexibacter*
			S	*Spironema*
			M + S	*Bacillus*
			G + M	*Francisella*
			G + M + S	*Shewanella*
Gobius cruentatus	63	0		
Gobius niger	38	3	G	*Unidentified Betaproteobacteria*
			I	*Roseobacter, Unidentified Phyllobacteriaceae*
Oblada melanura	69	6	I	*Unidentified Mycoplasmataceae*
			M	*Winogradskyella, Actibacter, Hyphomonadaceae, Polaribacter, Tenacibaculum*
Pagellus bogaraveo	30	6	M	*Kiloniellales, Unidentified Betaproteobacteria, Croceitalea*
			M + S	*Unidentified Flavobacteriaceae, Unidentified Bacillales, Saprospiraceae*
Pagellus erythrinus	24	7	G	*Unidentified Betaproteobacteria*
			M + S	*Bacillus, Planomicrobium*

			G + M	*Moritella*
			G + S	*Propionigenium*
			G + M + S	*Shewanella, Psychrilyobacter*
Spicara maena	31	5	M	*Eudoraea, Planomicrobium, Unidentified Flavobacteriaceae, Unidentified Planococcaceae*
			M + S	*Unidentified Bacillales*
Scorpaena notata	66	24	S	*Unidentified Acidimicrobiales, Halomicronema, Unidentified Trueperaceae, Unidentified Chromatiales, Unidentified Saprospiraceae, Unidentified Phyllobacteriaceae, Unidentified Rhodobacteraceae, Kiloniellales, Eudoraea, Lutimonas, Amaricoccus, Sulfitobacter, Unidentified Anaerolineae, Unidentified Pseudanabaenaceae, Unidentified Alphaproteobacteria, Unidentified Unidentified Sphingomonadales, Thiothrix, Unidentified Acidimicrobiales (2)*
			G + S	*Unidentified Flavobacteriaceae, Unidentified Planococcaceae, Unidentified Alphaproteobacteria, Endozoicomonas, Psychroserpens, Pseudoalteromonas*
Sarpa salpa	189	95	G	*Unidentified Chlamydiia, Unidentified Burkholderiales*
			I	*Unidentified Rickettsiales*
			M	*Agrococcus, Pseudonocardia, Fulvivirga, Unidentified Chloroflexi, Rivularia, Unidentified Sphingomonadaceae, Unidentified Bacteriovoracaceae, Unidentified Verrucomicrobiaceae, Unidentified Rhizobiales, Unidentified Spirochaetaceae, Unidentified Kiloniellales, Unidentified Chroococcales, Unidentified Actinomycetales, Unidentified Myxococcales, Unidentified Nocardioidaceae, Enterococcus, Unidentified Trueperaceae, Unidentified Chromatiales, Unidentified Gaiellales, Devosia, Unidentified Acidimicrobiales, Lutibacterium, Fluviicola, Thiohalorhabdales, Halorhodospira, Loktanella, Lysobacter, Leptolyngbya, Unidentified Bacteroides, Planctomyces, Turicibacter, Alcanivorax, Cytophaga*
			S	*Salegentibacter, Coccinimonas, Roseivirga, Odoribacter, Anaerofilum, Unidentified Mollicutes, Unidentified Desulfobulbaceae, Coraliomargarita, Paludibacter*
			M + S	*Piscirickettsiaceae, Polaribacter, Unidentified Porphyromonadaceae, Unidentified Flavobacteriales, Desulfovibrio, Winogradskyella, Eudoraea, Unidentified Bacillales, Bacillus, Clostridium, Croceitalea, Planctomycete, Unidentified Flammeovirgaceae, Unidentified Spirobacillales, Grimontia, Unidentified Saprospiraceae, Marixanthomonas, Unidentified Deferribacteraceae, Anaerotruncus, Dinoroseobacter, Unidentified Desulfarculaceae*
			G + M	*Unidentified Hyphomonadaceae, Unidentified Verrucomicrobiaceae, Unidentified Anaplasmataceae*

			G + M + S	*Unidentified Flammeovirgaceae, Aliivibrio, Unidentified Flavobacteriaceae, Unidentified Alteromonadales, Tenacibaculum, Unidentified Pirellulaceae, Unidentified Rhodobacteraceae, Unidentified Planctomycete, Unidentified Hyphomicrobiacea, Enterovibrio*
			I + M + S	*Unidentified Macellibacteroides, Unidentified Cerasicoccaceae, Unidentified Opitutae, Unidentified Bacteroidales, Unidentified Rikenellaceae, Unidentified Ruminococcaceae, Unidentified Alphaproteobacteria, Unidentified Mollicutes, Subdoligranulum, Unidentified Victivallaceae*
Serranus scriba	30	7	G	*Pseudoalteromonas, Arcobacter*
			M	*Psychrobacter, Methylobacterium*
			G + M + S	*Propionigenium, Shewarella, Moritella*
Simphodus tinca	29	6	M	*Endozoicomonas*
			G + M	*Octadecabacter, Unidentified Bacillales*
			G + M + S	*Moritella, Propionigenium, Unidentified Vibrionaceae*

```
# Sequences were previously demultiplexed into forward and reverse reads
# Example :
# 171222_SN1126_A_L001_JIC-2-171_R2.fastq.gz and 171220_SN234_A_L001_JIC-1-193_AdapterTrimmed_R2.fastq.gz
# In the mapping file fields 16 and 17 contains prefixes and suffixes given by the sequencing center first field is the sample code

# Unzip files

ls *.gz | awk '{print "gzip -df "$1}' | /bin/sh

# Rename files to have new sample name. Forward first

cat map.txt |sed '1d' | awk '{print "mv "$16"_"$17"_R1.fastq "$1"_R1.fastq"}' |/bin/sh
cat map.txt |sed '1d' | awk '{print "mv "$16"_"$17"_R2.fastq "$1"_R2.fastq"}' |/bin/sh

# Usearch 9 merges reverse and forward from all reads, renames samples based on file name

usearch9 -fastq_mergepairs *R1*.fastq -fastq_maxdiffs 0 -fastqout merged.fastq -relabel @ 2> fastqmergepairs.log &

# Usearch 9 filters bad quality reads

usearch9 -fastq_filter merged.fastq -fastq_maxee 1.0 -fastaout merged.fasta -fasta_cols 0 &

# T. Walters https://gist.github.com/walterst/2fce207ff38ad04c0bcbb2e8531ac230 script to remove primers

strip_primers_exclude.py map.txt merged.fasta merged1.fasta log

# Denoise with Usearch 9,
        # 1 dereplicate

usearch9 -fastx_uniques merged1.fasta -fastaout uniques.fasta -sizeout

        # 2 denoise

usearch9 -unoise uniques.fasta -tabbedout uniques_unoise.txt -fastaout uniques_denoised.fasta

# Remove sequences shorter than 360 bp
# Make single line fasta

mv uniques_denoised.fasta uniques_denoised

cat uniques_denoised |  awk '{if ($1 ~ />.*/){printf ("%s","\n");printf ("%s",$0);printf ("%s","\n")}else {printf ("%s",$0)}}' | sed '1d' >  uniques_denoised.fasta"

# Identify reads shorter than 360 bp

cat uniques_denoised.fasta | gawk -F ""  '{if($1~/>/){printf("%s ",$0)}else{printf("%s\n",NF)}}' | awk '{if($NF<360)print $1}' >less360
```

Qiime1 removes reads less than 360 bp identified above

filter_fasta.py -n -s less360 -f uniques_denoised.fasta -o uniques_denoised_plus360.fasta

Make OTU table

usearch9 -usearch_global merged1.fasta -db uniques_denoised_plus360.fasta -strand plus -id 0.97 -otutabout otu_table.txt

Qiime1 assigns taxonomy with rdp classifier and a modified green genes 13_8 database (Reverter *et al.*, 2017)

assign_taxonomy.py -m rdp -i uniques_denoised_plus360.fasta -t /Volumes/data/greengenesrdp/gg_13_8_otus/taxonomy/99_otu_taxonomy_species.txt -r /Volumes/data/greengenesrdp/gg_13_8_otus/rep_set/99_otus.fasta -o gg_13_8Assigned_tax_species --rdp_max_memory=4000 &

Fix names in taxonomy table since it is different from the OTU table (i.e. contains read count)

cat gg_13_8Assigned_tax_species/uniques_denoised_plus360_tax_assignments.txt | awk '{{sub(/;.*;/,"",$1)};print $1"\t"$2"\t"$3}' > tax_assignments

Add missing phylogenetic levels to taxonomy

cat tax_assignments |gawk -F "\t" '{if($2!~/p__/){printf("%s\t",$1);printf("%s;p__;c__;f__;g__;s__\t",$2);printf("%s\n",$3)}else{print $0}}'| gawk -F "\t" '{if($2!~/c__/){printf("%s\t",$1);printf("%s;c__;f__;g__;s__\t",$2);printf("%s\n",$3)}else{print $0}}' | gawk -F "\t" '{if($2!~/f__/){printf("%s\t",$1);printf("%s;f__;g__;s__\t",$2);printf("%s\n",$3)}else{print $0}}'|gawk -F "\t" '{if($2!~/g__/){printf("%s\t",$1);printf("%s;g__;s__\t",$2);printf("%s\n",$3)}else{print $0}}'|gawk -F "\t" '{if($2!~/s__/){printf("%s\t",$1);printf("%s;s__\t",$2);printf("%s\n",$3)}else{print $0}}'>tax_assignments_fixed

biom http://biom-format.org/ , convert table to biom format

biom convert --table-type="OTU table" -i otu_table.txt -o otu_table.biom --to-json &

biom, add taxonomy

biom add-metadata --sc-separated taxonomy --observation-header OTUID,taxonomy --observation-metadata-fp tax_assignments_fixed -i otu_table.biom -o otu_table_final.biom
filter taxa

Qiime remove chloroplasts, mitochondria, archaea and unclassified taxa from table

filter_taxa_from_otu_table.py -i otu_table_final.biom -o otu_table_final_f.biom -n c__Chloroplast,f__mitochondria, k__Archaea, Unclassified

Summarize table

biom summarize-table -i otu_table_final_f.biom -o summary.txt; less summary.txt

System configuration

uname -a > version
system_profiler SPSoftwareDataType >> version
pip freeze >> version
conda list >> version

Version :
Darwin woese.obs-banyuls.fr 15.6.0 Darwin Kernel Version 15.6.0: Tue Apr 11 16:00:51 PDT 2017; root:xnu-3248.60.11.5.3~1/RELEASE_X86_64 x86_64
Software:

System Software Overview:

System Version: OS X 10.11.6 (15G1510)
Kernel Version: Darwin 15.6.0
Boot Volume: woese
Boot Mode: Normal
Computer Name: woese
User Name: Marcelino Suzuki (suzuki)
Secure Virtual Memory: Enabled
System Integrity Protection: Enabled
Time since boot: 111 days 6:45

appnope==0.1.0
backports.shutil-get-terminal-size==1.0.0
biom-format==2.1.5
biopython==1.68
bitarray==0.8.1
burrito==0.9.1
burrito-fillings==0.1.1
bz2file==0.98
checkm-genome==1.0.5
click==6.6
cogent==1.5.3
decorator==4.0.11
DendroPy==4.1.0
EMIRGE==0.61.1
emperor==0.9.51
enum34==1.1.6
funcsigs==1.0.2
future==0.16.0
gdata==2.0.18
h5py==2.6.0
ipython==5.1.0
ipython-genutils==0.1.0
joblib==0.9.3
khmer==2.0
matplotlib==1.4.3
mmtf-python==1.0.5
mock==2.0.0

msgpack-python==0.4.8
natsort==3.5.0
nose==1.3.7
numpy==1.10.4
pandas==0.18.1
pathlib2==2.2.0
pbr==1.10.0
pexpect==4.2.1
pickleshare==0.7.4
PICRUSt==1.1.0
Pillow==4.0.0
poretools==0.6.0
prompt-toolkit==1.0.9
ptyprocess==0.5.1
Pygments==2.1.3
pygobject==3.10.0
pynast==1.2.2
pyparsing==2.0.3
pyqi==0.3.2
pysam==0.10.0
python-dateutil==2.3
pytz==2016.10
qcli==0.1.1
qiime==1.9.1
qiime-default-reference==0.1.3
quast==4.4
reportlab==3.3.0
ruffus==2.6.3
scandir==1.4
scikit-bio==0.2.3
scipy==0.17.1
ScreamingBackpack==0.2.333
screed==0.9
seaborn==0.7.1
simplegeneric==0.8.1
simplejson==3.8.1
six==1.10.0
traitar==1.1.2
traitlets==4.3.1
wcwidth==0.1.7
# packages in environment at /usr/local/miniconda3/envs/bioinfo2:
#

appnope	0.1.0	py27_0	
backports	1.0	py27_0	
bamtools	2.4.0	3	bioconda
bcftools	1.6	1	bioconda
biom-format	2.1.5	py27_1	bioconda
biopython	1.68	py27_0	bioconda
bitarray	0.8.1	py27_0	

boost	1.60.0	py27_0	
bowtie	1.2.0	py27_0	bioconda
burrito	0.9.1	py27_0	bioconda
burrito-fillings	0.1.1	py27_0	bioconda
bz2file	0.98	py27_0	
bzip2	1.0.6	hd86a083_4	
checkm-genome	1.0.5	py27_0	bioconda
click	6.6	py27_0	bioconda
cogent	1.5.3	py27_0	bioconda
curl	7.45.0	2	bioconda
decorator	4.0.11	py27_0	
dendropy	4.1.0	py27_0	bioconda
emirge	0.61.1	py27_0	bioconda
emperor	0.9.51	py27_0	bioconda
enum34	1.1.6	py27_0	
freetype	2.5.5	1	
funcsigs	1.0.2	py27_0	
future	0.16.0	py27_1	
gdata	2.0.18	py27_0	
get_terminal_size	1.0.0	py27_0	
google-sparsehash	2.0.3	1	bioconda
h5py	2.6.0	np110py27_7	conda-forge
hdf5	1.8.17	2	
hmmer	3.1b2	2	bioconda
htslib	1.3.2	0	bioconda
icu	54.1	0	
infernal	1.1.2	0	bioconda
ipython	5.1.0	py27_1	
ipython_genutils	0.1.0	py27_0	
java-jdk	8.0.92	1	bioconda
jbig	2.1	0	
jemalloc	3.6.0	1	bioconda
joblib	0.9.3	py27_0	bioconda
jpeg	9b	0	bioconda
khmer	2.0	py27_1	bioconda
libgcc	4.8.5	1	
libpng	1.6.17	0	
libtiff	4.0.6	3	
lordec	0.8	0	atgc-montpellier
matplotlib	1.4.3	np110py27_3	
mkl	11.3.3	0	
mmtf-python	1.0.5	py27_0	bioconda
mock	2.0.0	py27_0	
msgpack-python	0.4.8	py27_0	
natsort	3.5.0	py27_0	
nose	1.3.7	py27_1	

numpy	1.10.4	py27_2	
openssl	1.0.2k	0	
pandas	0.18.1	np110py27_0	
path.py	10.0	py27_0	
pathlib2	2.2.0	py27_0	
pbr	1.10.0	py27_0	
pear	0.9.6	3	bioconda
perl	5.22.0.1	0	conda-forge
perl-threaded	5.22.0	pl5.22.0_12	bioconda
pexpect	4.2.1	py27_0	
pickleshare	0.7.4	py27_0	
picrust	1.1.0	py27_0	bioconda
pillow	4.0.0	py27_0	
pip	9.0.1	py27_1	
poretools	0.6.1a1	py27_5	bioconda
prompt_toolkit	1.0.9	py27_0	
ptyprocess	0.5.1	py27_0	
pygments	2.1.3	py27_0	
pygobject	3.10.0	py27_1	vgauthier
pygtk	2.24.0	py27_1	vgauthier
pynast	1.2.2	py27_0	bioconda
pyparsing	2.0.3	py27_0	
pyqi	0.3.2	py27_0	bioconda
pyqt	4.11.4	py27_4	
pysam	0.10.0	py27_1	bioconda
python	2.7.13	0	
python-dateutil	2.3	py27_0	bioconda
pytz	2016.10	py27_0	
qcli	0.1.1	py27_0	bioconda
qiime	1.9.1	np110py27_0	bioconda
qiime-default-reference	0.1.3	py27_0	bioconda
qt	4.8.7	4	
quast	4.4	boost1.60_1	bioconda
readline	6.2	2	
reportlab	3.3.0	py27_0	
ruffus	2.6.3	py27_0	bioconda
samtools	1.3.1	5	bioconda
scandir	1.4	py27_0	
scikit-bio	0.2.3	np110py27_0	bioconda
scipy	0.17.1	np110py27_1	
screamingbackpack	0.2.333	py27_0	bioconda
screed	0.9	py27_0	bioconda
seaborn	0.7.1	py27_0	
seqtk	1.2	0	bioconda
setuptools	27.2.0	py27_0	

simplegeneric	0.8.1	py27_1	
simplejson	3.8.1	py27_0	bioconda
sip	4.18	py27_0	
six	1.10.0	py27_0	
sqlite	3.13.0	0	
tbb	4.4_20150728	0	bioconda
tk	8.5.18	0	
traitar	1.1.2	<pip>	
traitlets	4.3.1	py27_0	
trimmomatic	0.36	3	bioconda
vsearch	2.4.0	0	bioconda
wcwidth	0.1.7	py27_0	
wheel	0.29.0	py27_0	
xz	5.2.2	1	
zlib	1.2.11	hf3cbc9b_2	