

Host Habitat Is the Major Determinant of the Gut Microbiome of Fish

Pil Soo Kim

Kyung Hee University

Na-Ri Shin

Korea Research Institute of Bioscience and Biotechnology

Jae-Bong Lee

National Institute of Fisheries Science

Min-Soo Kim

Kyung Hee University

Tae Woong Whon

Kyung Hee University

Dong-Wook Hyun

Kyung Hee University

Ji-Hyun Yun

Kyung Hee University

Mi-Ja Jung

Kyung Hee University

Joon Yong Kim

Kyung Hee University

Jin-Woo Bae (✉ baejw@khu.ac.kr)

Kyung Hee University <https://orcid.org/0000-0001-6433-5270>

Research Article

Keywords: Fish, Fish gut microbiota, Freshwater fish, Seawater fish, Vertebrate gut microbiota

Posted Date: March 22nd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-332643/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Microbiome on July 31st, 2021. See the published version at <https://doi.org/10.1186/s40168-021-01113-x>.

Abstract

Background: Our understanding of the gut microbiota of animals is largely based on studies of mammals. To better understand the evolutionary basis of symbiotic relationships between animal hosts and indigenous microbes, it is necessary to investigate the gut microbiota of non-mammalian vertebrate species. In particular, fish have the highest species diversity among groups of vertebrates, with approximately 33,000 species. In this study, we comprehensively characterized gut bacterial communities in fish.

Results: We analyzed 227 individual fish representing 14 orders, 42 families, 79 genera, and 85 species. The fish gut microbiota was dominated by *Proteobacteria* (51.7%) and *Firmicutes* (13.5%), different from the dominant taxa reported in terrestrial vertebrates (*Firmicutes* and *Bacteroidetes*). The gut microbial community in fish was more strongly shaped by host habitat than by host taxonomy or trophic level. Using a machine learning approach trained on the microbial community composition or predicted functional profiles, we found that the host habitat exhibited the highest classification accuracy. Principal coordinate analysis revealed that the gut bacterial community of fish differs significantly from those of other vertebrate classes (reptiles, birds, and mammals).

Conclusions: Collectively, these data provide a reference for future studies of the gut microbiome of aquatic animals as well as insights into the relationship between fish and their gut bacteria, including the key role of host habitat and the distinct compositions in comparison with those of mammals, reptiles, and birds.

Background

Multicellular eukaryotes appeared 1.2 billion years ago, followed by a long evolutionary history of mutual interactions between multicellular and single-celled organisms [1]. According to Van Valen's "Red Queen hypothesis," evolution is driven by competition among taxa for survival under constantly changing environments [2]. Indeed, co-existence with microbes poses one of the greatest challenges for animals. At the same time, hosts and microbes can establish symbiotic relationships, in which each species benefits from mutualistic interactions [3]. The symbiotic microbiota contributes to animal adaptation to various habitats by providing complementary functional resources (e.g., by digesting indigestible dietary fiber, producing essential vitamins, protecting against enteropathogens, maintaining immune homeostasis, and contributing to intestinal maturation) over a long period of co-evolution [4–10].

In the last decade, numerous studies have explored gut microbial communities of various animal hosts. However, these studies have mostly focused on the gut microbiota of mammals [11], which represent less than 10% of all vertebrate diversity. By contrast, there are more than 33,000 species of fish, representing the greatest species diversity among groups of vertebrates [12, 13]. This focus on a single class of animals allows only limited insight into the vertebrate gut microbiota. To understand the co-evolution of vertebrates and gut microbes, broad analyses of fish are essential. The gut microbiota has been

evaluated in a few model fish species, such as zebrafish [14], guppy [15], and rainbow trout [16], and in economically valuable aquatic animals, such as carp [17], Atlantic salmon [18], sturgeon [19], and Atlantic cod [20]. However, these studies are insufficient to comprehensively understand the composition of the gut microbiota in fish and patterns of co-evolution.

Here, we aimed to resolve long-standing questions about the gut microbiota in fish. For example, is the gut microbiota in fish shaped by the host habitat? Do genetic factors in fish affect the structure of the gut microbiota and, if so, to what extent? How does the gut microbiota of fish differ from those of other vertebrates? To resolve these issues, we comprehensively characterized the gut microbiota of 227 individual fish representing 85 species obtained from lakes, a stream, and seas (i.e., habitats with distinct differences in nutrient availability, salinity, temperature, and depth) (Figs. 1 and 2). We used a clustering approach to find the primary determinants of the structures of the gut microbiome and verified these determinants using unsupervised and supervised machine learning approaches, like PAM clustering and random forest classification. To gain a wider perspective, we compared gut microbial communities in fish and other vertebrates (mammals, reptiles, and birds) by using principal coordinate analysis (PCoA). These data serve as a reference for future studies of the gut microbiota of fish and other aquatic animals. Our findings also support the notion that symbiotic relationships between microbes and vertebrates are important for adaptation and provide insights into the nature of interspecific microbiome variation in various fish species.

Results

Overview of data related to the gut microbiota in fish taxa

We analyzed the bacterial community composition of the intestinal contents of 227 individual fish inhabiting six different environments (23 different sampling spots; Fig. 1a). The collected fish were taxonomically grouped into two classes (ray-finned and cartilaginous fish), 14 orders, 42 families, 79 genera, and 85 species. Overall, 1,014,240 raw 16S rRNA gene sequence reads were obtained from the intestinal contents of fish, and 653,281 high-quality sequence reads were obtained after the removal of low-quality and chimeric sequences. The high-quality sequences were clustered into 3,273 operational taxonomic units (OTUs), with a mean number of OTUs per sample of 91 (± 6 SD), applying a threshold of 97% sequence identity.

To determine whether the sampling depth was sufficient to give an overview of the fish gut microbiota, rarefaction curves were generated for the number of OTUs per individual or species (Additional file 1: Supplementary Fig. S1). The cumulative number of identified OTUs reached a plateau at approximately 150 individuals or 60 fish species. This pattern was not affected by the fish habitat, indicating that the sampling depth was sufficient to capture the global bacterial diversity of the gut microbiota of wild fish.

The fish gut microbiota included 21 bacterial phyla, with three dominant phyla (*Proteobacteria*, *Firmicutes*, and *Cyanobacteria*) accounting for over 70% of all sequence reads (Fig. 1b and Additional file

2: Supplementary Table S1). Notably, the detailed microbial community composition of the fish gut differed considerably from the typical composition of the gut microbiota in vertebrates, mainly composed of *Firmicutes* and *Bacteroidetes* [21–23]. *Proteobacteria* was the most frequent taxon in the fish gut at the phylum level (detected in 219 fish samples), followed by *Firmicutes*, *Cyanobacteria*, and *Planctomycetes*. Although *Fusobacteria* was present in less than 50% of total samples, it was frequently detected in freshwater fish (Figs. 1c and 2).

Host habitat is the major determinant of the gut microbiota of fish

Environmental factors and host genetics shape the gut microbiota of various animal taxa [21, 24, 25]. However, the extent to which these factors contribute to the microbiome composition of fish is unclear. To evaluate the relative importance of various factors, we first examined the similarity of the microbial community using within-sample distances for various clustering scenarios. We found significant variation at both within and between groups; however, the largest differences in microbial communities were obtained for factors related to the host habitat (salinity and sampling sites), and some groups even showed contrasting relationships with these factors (Additional file 1: Supplementary Fig. S2). These results indicate that environmental factors, particularly those associated with properties of the habitat, interact to shape the gut microbiota of fish.

We next performed a clustering analysis using the partitioning around medoids (PAM) clustering algorithm based on the Calinski–Harabasz index and the silhouette score [26] to identify the optimal number of clusters and to evaluate the importance of environmental and genetic factors. The PAM clustering results showed that the gut microbiota of fish could be clustered into two groups, and the clusters were more consistent with variation in the host habitat (freshwater vs. seawater) than host class (Actinopterygii vs. Chondrichthyes) (Additional file 1: Supplementary Fig. S3a). To validate the importance of host habitat in shaping the gut microbiota of fish, we further assessed cluster validity for k -clusters, according to the following categories: habitat (number of variants, $n = 2$; freshwater vs. seawater), sampling site ($n = 6$; Fig. 1a), host order ($n = 8$), host family ($n = 18$), and host genus ($n = 30$). Among various categories, the habitat had the highest proportion of correctly matched constituents (Additional file 1: Supplementary Fig. S3b), indicating that habitat was the primary determinant of the fish gut microbiome. Compared with the former unsupervised learning approach (PAM clustering), we additionally evaluated associations between the various candidate factors and gut microbiota using the R statistic from analysis of similarities (ANOSIM) based on unweighted and weighted UniFrac distances. While all of the factors significantly ($p < 0.001$) affected the microbial structure of the fish gut, habitat and host species had the greatest ability to distinguish among samples (Fig. 3a).

We then performed a comparative analysis of habitat signatures in the gut microbiota of fish. With respect to α -diversity indices, the gut microbiota of freshwater fish exhibited significantly higher values for microbial richness (Shannon index), non-phylogenetic diversity (observed species), and equitability (Simpson evenness) than those of seawater fish, while Faith's phylogenetic diversity was comparable

across habitats (Additional file 1: Supplementary Fig. S4). As expected from the R -values presented in Fig. 3a, a PCoA plot of unweighted UniFrac distance metrics revealed that the gut microbial communities of freshwater fish and seawater fish clustered separately (ANOSIM, $R = 0.471$, $p < 0.001$; Fig. 3b).

Furthermore, distinct clustering of gut microbial communities was apparent when considering a more detailed habitat category, the sampling site, which reflects the type of freshwater and seawater (e.g., stream, lake, coast, or deep sea) and the geographical region (Additional file 1: Supplementary Fig. S5).

Next, we investigated differences in the composition of the gut microbiota with respect to the habitat. Indeed, a phylum-level difference between freshwater fish and seawater fish was detected (Figs. 2 and 3c). To determine whether the taxa with differences in abundance could serve as biologically relevant biomarkers for freshwater and seawater fish, we performed a linear discriminant analysis (LDA) effect size (LEfSe) analysis at all taxonomic levels. The phylum *Proteobacteria* was enriched in seawater fish with a relatively high LDA score ($LDA > 3.8$), while the phyla *Firmicutes* and *Fusobacteria* were significantly enriched in freshwater fish (Additional file 1: Supplementary Fig. S6a). At the family level, *Moraxellaceae*, *Vibrionaceae*, and *Enterobacteriaceae* (all in the class *Gammaproteobacteria*), and *Alcaligenaceae* (*Betaproteobacteria*) were significantly more abundant in seawater fish than in freshwater fish, whereas *Aeromonadaceae* (*Gammaproteobacteria*) was significantly more abundant in freshwater fish. The family *Clostridiaceae* (*Clostridia*) was more abundant in freshwater fish than in seawater fish, whereas *Leuconostocaceae* (*Bacilli*) was more abundant in seawater fish. Furthermore, most OTUs belonging to *Fusobacteriaceae*, showing a higher frequency in freshwater fish than in seawater fish, were assigned to *Cetobacterium* at the genus level (Additional file 1: Supplementary Fig. S6b and Additional file 2: Supplementary Table S1).

We then used a network-based approach to test whether gut microbial communities clustered by fish habitat at the OTU level. In the analysis, a node represents an individual fish and the OTUs are connected to the host fish in which they were detected. In agreement with the compositional differences noted above, in the OTU network-based analysis, the host nodes were more likely to connect to nodes of other hosts sharing the same habitat than to those from different habitats (Fig. 3d).

In ecology, the trophic level is the position occupied by an organism in the food chain. Primary consumers, usually herbivores, occupy lower trophic levels, while predatory species (e.g., carnivores) occupy higher levels [27, 28]. Hence, we further investigated the trophic level of individual fish species to assess the effect of the host dietary gradient on the microbial community. A PCoA plot considering the trophic level, as determined using FishBase [13], revealed a distinct microbial gradient based on the ecological position of fish in the food chain (Additional file 1: Supplementary Fig. S7). These results suggested that the host trophic level is weakly but significantly (ANOSIM, $R = 0.14$, $p < 0.001$) associated with the gut microbial community assemblage.

Host divergence had little influence on the gut microbiota of fish

We observed a statistically significant relationship between the gut microbial community structure and the host genetic variation in the cytochrome c subunit I (*CO1*) gene, although the degree of distinguishability did not exceed that for the habitat, as determined by the *R* value from ANOSIM. The host taxon-dependent variation in the gut microbial community was greater at lower taxonomic levels than at higher taxonomic levels (Fig. 3a). Remarkable variation in both the microbial community composition and structure was observed with respect to the host order (Additional file 1: Supplementary Fig. S8a). Differences in the relative abundances of several bacterial taxa depended on the host order. For example, *Epsilonproteobacteria* was relatively enriched in Tetraodontiformes, whereas the relative abundance of *Gammaproteobacteria* was higher in Lampriformes and Osmeriformes than in other host orders. *Firmicutes*, mainly represented by the class *Clostridia*, was relatively more abundant in Siluriformes, Gadiformes, Cypriniformes, and Osmeriformes than in other host orders. The phylum *Cyanobacteria* was enriched in Perciformes, Rajiformes, Clupeiformes, and Lophiiformes, whereas the phylum *Fusobacteria* was over-represented in Perciformes, Tetraodontiformes, Siluriformes, Cypriniformes, and Lophiiformes. Microbial communities showed significantly greater clustering within the same host order than across different host orders (ANOSIM, $R = 0.20$, $p < 0.001$) (Additional file 1: Supplementary Fig. S8b).

Considering the host species-specificity of the gut microbiota in fish, we next investigated the existence of phylosymbiosis, or a relationship between host phylogeny and the gut microbiota. A scatter plot of weighted UniFrac distances plotted against host genetic relatedness based on variation in the *CO1* gene showed no significant association between similarity in the gut microbial community composition and host phylogenetic distance (Fig. 4a). Regardless of the phylogenetic distance among hosts, the dissimilarities of gut microbial communities between fish taxa were randomly distributed.

Habitat can be linked to host taxonomy because almost all fish inhabit specialized niches. To test whether this association confounds the distinguishability of the habitat/host taxonomy based on the gut microbiota, we next compared the microbial communities in Perciformes and Cypriniformes. In the current study, the order Perciformes was represented by 10 species (at least three individuals per species) caught in freshwater or seawater, and the order Cypriniformes included various freshwater species collected at multiple sampling sites (a stream and lakes). PCoA and UPGMA trees revealed that gut microbial communities from a single host order showed clearer clustering by habitat than by host species (Additional file 1: Supplementary Figs. S9 and S10), further supporting the greater role of the environment than host genetic factors in shaping the gut microbial community in fish.

In addition, we used a machine learning algorithm (the random forest classifier) to evaluate the predictive value of the microbial composition for host habitat or taxonomy. Owing to a large class imbalance in the number of variants per factor, precision-recall curves (PRC) were generated, and classification accuracy was calculated based on the area under the curve (AUC). The classification ability of the composition of the fish gut microbial community was better for discriminating the host habitat (freshwater vs. seawater) than for discriminating the sampling site or host taxonomy (Fig. 4b).

Functional profiling of microbial communities

We used the PICRUSt pipeline [29] to investigate whether the habitat-dependent differences in the microbial taxonomic composition are related to differences in functional profiles. Kyoto Encyclopedia of Genes and Genomes (KEGG) ortholog groups (KOs) predicted from the 16S rRNA gene sequences were assigned to broad functional categories based on the BRITE hierarchy. PCoA based on KOs predicted by PICRUSt revealed that the host habitat significantly affects the functional gene distribution (ANOSIM, $R = 0.37$, $p < 0.001$) (Fig. 5a). Most gene functions were related to metabolism (49.5%), environmental information processing (17.6%), and genetic information processing (14.5%) (Additional file 1: Supplementary Fig. S11 and Additional file 3: Supplementary Table S2). Gene families in the following categories were enriched in seawater fish: membrane transport, xenobiotic biodegradation and metabolism, amino acid metabolism, lipid metabolism, and transport and catabolism. By contrast, gene families in the following categories were enriched in freshwater fish: nucleotide metabolism, carbohydrate metabolism, metabolism of cofactors and vitamins, energy metabolism, translation, replication and repair, and cell motility (Fig. 5b).

We then used a machine learning approach to examine whether the functional profiles of the gut microbiota could be used to predict the environment or host taxon. The AUCs of PRC calculated using the functional profiles showed better prediction accuracy for the host habitat than for other factors, consistent with the results of the random forest classifier analysis based on microbial taxonomic profiles (Fig. 5c).

Comparison of the gut microbiota of fish and other vertebrate species

Lastly, we examined the impact of vertebrate evolution on gut microbes by comparing the microbial taxonomic profiles among fish and other vertebrate species. We compared the microbiota data for fish obtained in the current study with data for humans (Human Gut Microbiota Project [HMP] data [30, 31]), 66 aquatic mammals [32], 39 non-human mammals [33], 41 iguanas and snakes (reptiles) [34, 35], and 124 wild birds [36]. PCoA of the Bray–Curtis dissimilarity index and the binary Jaccard index revealed that the fish gut microbiome clustered separately from the other microbiomes (PERMANOVA, $p < 0.001$ and $p < 0.001$, respectively). Furthermore, the gut bacterial communities from other animals were clearly separated (Additional file 4: Supplementary Table S3). Clustering by phylogenetic relationships among hosts was clearly evident in a PCoA of the binary Jaccard index at the family level (Fig. 6), and significant differences were observed (ANOSIM, $R = 0.64$, $p < 0.001$). As shown in Fig. 6a, the gut microbiome of fish showed slight overlap with the gut microbiomes of aquatic mammals (dolphins and sea lions) and snakes, while the avian gut microbiome was distinct from those of other vertebrates. When examining the microbial composition of each host group, the components of the gut microbiota of each host differed at the microbial phylum level. *Proteobacteria* was most abundant in fish, *Firmicutes* was enriched in birds and reptiles, and *Bacteroidetes* was enriched in humans (Fig. 6b–d).

Discussion

We characterized the gut microbial communities of various wild fish. Few studies have focused on the fish gut microbiota, despite the importance of fish in the evolutionary history of vertebrates and the tremendous species diversity, accounting for nearly half of all vertebrate species [13]. The gut microbiota of vertebrates is host-specific and arose as a result of co-evolution between hosts and microbes [37, 38]. Even in invertebrate species (e.g., shrimp or insect species), the gut microbiota is distinguished by the presence of specific commensal bacterial consortia [39, 40]. As expected, we found that the gut microbiota of wild fish is a host-specific and deterministic microbial assemblage. Furthermore, we showed that the gut microbiota is primarily determined by the fish environment, rather than by genetic factors.

The gut microbiota of most vertebrates, including amphibians, reptiles, mammals, and birds, is dominated by the phyla *Firmicutes* and *Bacteroidetes* [22, 31, 41, 42]. Indeed, a bloom of *Proteobacteria* is considered a sign of dysbiosis or instability of the gut microbial community in mammals [43]. Many commensal *Proteobacteria* can become pathobionts, infecting the host under specific conditions and facilitating inflammation [44, 45]. However, in the current study, *Proteobacteria* dominated the gut microbiota of the majority of fish, in agreement with recent studies of the gut microbiota of fish [14, 15]. These compositional differences at the phylum level can be explained by a partial projection of the vast diversity of marine *Proteobacteria* associated with the unsegmented digestive system of fish, unlike that of mammals [46, 47]. A stochastic assemblage of environmental microbes in the fish gut microbiota is unlikely because the predominant bacterial taxa in the ocean or other aquatic habitats, such as SAR11 (*Pelagibacter ubique* HTCC1062), SAR116 (*Puniceispirillum marinum* IMCC1322), and SAR86, were absent or not abundant in the fish gut (Additional file 5: Supplementary Table S4) [48, 49]. These findings prompt the question of whether *Proteobacteria* outcompete other environmental bacterial taxa in the aquatic habitat or whether they have been selected by the host itself [50–52].

We found that host habitat was the predominant determinant of the fish gut microbial community. Assessments of the discriminative structuring factors of the gut microbiota using both unsupervised and supervised learning approaches, such as PAM clustering, ANOSIM, and random forest classifier analysis, supported the importance of habitat, and this was particularly apparent in fish with a similar genetic background (e.g., Perciformes and Cypriniformes). Nevertheless, various other factors that were examined, including host taxonomy and trophic level, contributed to the fish gut microbial community (Fig. 3a). Environmental factors could not explain a large portion of the total variance (Fig. 3b; variation explained by PC1 at 7.43%) in the fish gut microbiota. Hence, intrinsic genetic factors are also important, and the species-specificity of the gut microbiota is a result of the intrinsic genetic background of the host.

Differences in the microbial compositional with respect to salinity can be explained in terms of host adaptation to the environment. The dominant taxa might reflect the affinity of the host for gut bacteria that contribute to the maintenance of immune function and metabolic activity. For example, the high proportion of *Fusobacteria* in freshwater fish might be associated with vitamin B12 (cobalamin). *Cetobacterium somerae* (order *Fusobacteriales*) is widely distributed in various freshwater fish, and its prevalence is negatively correlated with the dietary availability of vitamin B12 [53, 54]. Different

environmental conditions affect vitamin B12 availability, and freshwater fish harbor more vitamin B12-synthesizing bacteria, such as *C. somerae*, to satisfy their dietary needs. The importance of metabolic properties is consistent with the predicted functions of gut bacteria in freshwater fish, which showed a relatively high abundance of genes related to the metabolism of cofactors and vitamins. This suggests that basic nutrient availability in the environment drives selection of the fish gut microbiota to account for the nutritional deficiencies of the host. Based on the performance of classifiers, better results are obtained when using functional profiles as a training trait than when using taxonomic profiles. Environmental factors (e.g., habitat type) result in functional redundancy, with the host physiology governed by the ability to adapt.

We also observed high similarity between the gut microbiota of hosts that share feeding preferences. The average trophic level of seawater fish collected in the current study was higher than that of freshwater fish. Seawater fish show carnivorous and herbivorous dietary preferences, while freshwater fish tend to show omnivorous dietary preferences [28]. In particular, the family *Enterobacteriaceae* was significantly enriched in seawater fish, consistent with results for other carnivorous fish [15, 55]. Further, a bloom of marine-associated bacteria, such as *Enterobacteriaceae* and *Moraxellaceae*, is correlated with a low-fiber or animal-based diet in humans [56, 57]. By contrast, *Clostridium* and *Aeromonadaceae* were predominant in freshwater fish in the current study. Several *Clostridium* species are well-known cellulose-degraders associated with herbivorous vertebrates [58, 59]. *Aeromonas* is dominant in fish feeding on detritus of plant origin and in omnivorous freshwater fish (intermediate trophic level) [15, 60]. Differences in the gut microbiota are not simply a consequence of the host diet or feeding preference, as divergence between the gut microbiota of freshwater and seawater fish can also be a cause of the functional potential of hosts (Fig. 5).

In a comparison between fish and other vertebrates, including Reptilia, Avia, and Mammalia, we detected clearly distinct structures of each gut microbiota (Fig. 6). This was observed despite analogous taxon with similar metabolic or biological roles, i.e., a relatively high proportion of *Enterobacteriaceae* and *Moraxellaceae* (*Proteobacteria*) in animal-based diet vertebrates [56, 57] and the dominance of *Clostridium* species (*Firmicutes*) in plant-based diet vertebrates [58, 59]. Unlike the gut microbial composition of fish, the dominant gut bacteria of terrestrial mammals and humans belong to the phyla *Firmicutes* and *Bacteroidetes*. *Firmicutes* is the sole prominent microbial phylum in the guts of reptiles and birds. This difference at the microbial phylum level can be explained by evolved differences between fish and other vertebrates in the selectivity of the gut environment [15, 61]. Early fish arose 600 million years ago and became ancestors of all extant vertebrate clades [12]. Since the appearance of early vertebrates, they have evolved a number of physiological adaptations for survival in various environments. During this process, symbiotic gut microbes and host species co-evolved to survive in the continuously changing environment. It is difficult to experimentally simulate gut microbial selection and colonization during vertebrate evolution; however, surveys and experiments involving extant vertebrate species can provide insight into the contribution of various environmental and genetic factors to the gut microbiota.

Our species-wide study included an unprecedented number of fish; however, it had several limitations. Since sample collection focused on East Asia (the Korean peninsula), the taxa are not representative of the total species diversity of fish. The analysis should be extended to other fish species and several other factors, especially microbial communities in the surrounding environment. The detailed characterization of ecological niches and metabolic differences among fish will improve our understanding of the fundamental assemblage of the gut microbial consortium in fish. Furthermore, we analyzed the 16S rRNA gene to evaluate the bacterial composition and predicted functional profiles using the PICRUSt pipeline. These analyses indicated that some taxa are linked to specific biological activities of the fish host. Additional studies involving shotgun metagenomic sequencing could yield a more comprehensive dataset for detailed analyses of the determinants of the specific consortia of gut microbes in fish and their roles.

Conclusions

In summary, our results provide a comprehensive view of the fish gut microbiota. In particular, we found that host habitat (freshwater vs. seawater) has a dominant role in shaping wild fish gut microbial communities over host taxonomy and trophic level. Moreover, the microbial functional profiles predicted from 16S rRNA gene sequences were predominantly determined by host habitat. We further demonstrate that random forest classifiers trained on microbial community composition or functional features showed better prediction accuracy for the host habitat than for other factors. In addition, the fish gut microbiome in a PCoA plot clustered separately from those of other vertebrates, such as mammals, reptiles, and birds. Our findings improve our understanding of the long-term co-evolution of vertebrates and their indigenous microbial communities.

Methods

Sample collection

Gut samples from 227 seawater and freshwater fish were collected at 23 sites in Korea between June 2013 and October 2013 (Figs. 1a and 2 and Additional files 6 and 7: Supplementary Tables S5-6). Seawater fish were caught by the fisheries resource research vessel Tamgu-20 of the National Institute of Fisheries Science (NIFS), Korea. During a seasonal fisheries resource investigation of the deep sea of East Sea, near the seas of Ulleung-do and Dok-do, and West Sea, 175 seawater fish were caught by bottom trawling, mid-water (pelagic) trawling, and trammel. Freshwater fish were collected in collaboration with the Inland Fisheries Research Institute (NIFS) by using cast net and fish traps. All procedures for the collection and handling of seawater and freshwater fishes were approved by the NIFS and performed under the supervision of authorized and experienced members of the NIFS staff. The seawater fish were handled in a laboratory facility on the fishing vessel and the freshwater fish were handled at appropriate facilities near the fishing sites. All fish were stunned and dissected immediately after catching. Approximately 1.0–1.5 cm of the rectum was collected using sterile instruments, and the samples, including the luminal content and mucosa, were stored at -80°C until analyses. An

accompanying fish taxonomist identified the fish host species by briefly assessing fish morphological characteristics during sample collection. The fish host species were re-identified in the laboratory by a molecular phylogenetic analysis (*vide infra*).

DNA extraction and pyrosequencing of bacterial 16S rRNA genes

The gut specimens were squeezed out with sterile instruments to collect the luminal content. The gut samples were cut laterally to remove the mucus layer of the fish gut by visual inspection. A cover glass was used to separate the mucus layer from the gut samples. The luminal content and mucus layer were pooled and transferred to a sterile conical tube containing 6.5 mM dithiothreitol for mucus degradation [62]. After incubation for 1 h at 37°C, the pellet was collected by centrifugation and re-suspended in 750 µl lysis buffer (500 mM NaCl, 50 mM Tris-HCl, pH 8.0, 50 mM EDTA, and 4% sodium dodecyl sulfate). To maximize microbial cell lysis before DNA extraction, the re-suspended pellets were homogenized by shaking in a sterile screw-cap tube containing zirconia beads (2.3 and 0.1 mm diameter) and glass beads (0.5 mm diameter) using FastPrep-24 (MP Biomedicals, Santa Ana, CA, USA) for 50 s at 6.0 m/s. Genomic DNA from the homogenized samples was then extracted by the standard phenol-chloroform extraction method using the UltraClean Microbial DNA Isolation Kit (MOBIO, London, UK). The hypervariable regions V1–V3 of the 16S rRNA gene were amplified from the extracted genomic DNA of the sampled fish guts by using a sample-specific barcoded bacterial primer set [63] and Ex-Taq premix (Takara Bio, Kyoto, Japan). The polymerase chain reaction (PCR) conditions were as follows: 94°C for 10 min; followed by 29 cycles of 94°C for 60 s, 50°C for 30 s, and 72°C for 1 min 30 s; followed by a final extension step at 72°C for 10 min. Four independent PCR products for each sample were pooled and purified using the QIAquick PCR Purification Kit (Qiagen, Hilden, Germany). The concentration of purified PCR products was determined using the Quant-it PicoGreen dsDNA Assay Kit (Life Technologies, Carlsbad, CA, USA). The quality and quantity of DNA were checked using a Bioanalyzer 2100 instrument (Agilent, Santa Clara, CA, USA) and a DNA 1000 Lab Chip (Agilent). The pooled DNA was then amplified by emulsion PCR before 454 pyrosequencing using a GS FLX Titanium instrument (Roche, Basel, Switzerland) by a certified service provider (MacroGen, Seoul, Korea), according to the manufacturer's instructions.

Identification of fish host species and phylogenetic analysis of fish

To identify the fish host species, genomic DNA was extracted from the fish flesh collected aseptically from each specimen. Tissue fragments were suspended in 750 µl lysis buffer and homogenized by using FastPrep-24 (MPbio) with glass beads (0.5 mm diameter) for 40 s at 6.0 m/s. DNA was extracted using a standard phenol-chloroform extraction method. The *CO1* gene was amplified by using AccuPower PCR Premix (Bioneer, Daejeon, Korea) and the *CO1* gene primer cocktail set 3 [64]. The PCR conditions were as follows: initial denaturation at 95°C for 3 min; followed by 30 cycles of 94°C for 30 s, 52°C for 40 s, and 72°C for 1 min; followed by a final extension step at 72°C for 10 min [64]. The PCR products were

sequenced using the BigDye Terminator Cycle Sequencing Ready Reaction Kit (Applied Biosystems, Foster City, CA, USA), according to the manufacturer's instructions. The reaction products were analyzed using an automated DNA analyzer system (PRISM 3730XL DNA Analyzer, Applied Biosystems). Sequence fragments were assembled using SeqMan (DNASTAR, Madison, WI, USA). The assembled *CO1* gene sequences were then compared with other *CO1* gene sequences in the nucleotide collection (nr/nt) of the GenBank database by searches using the Basic Local Alignment Search Tool (BLAST) [65]. The *CO1* gene sequences were aligned using the multiple alignment program CLUSTAL W (v. 1.4), and a phylogenetic tree was generated by using MEGA 6 [66, 67] using the maximum-likelihood algorithm with 1000 bootstrap replicates [68].

Sequence analysis

The raw 16S rRNA sequences generated using the GS FLX Titanium platform were processed using QIIME (v. 1.8.0) [69]. All raw sequences with average quality scores below 25 and those shorter than 200 bp or longer than 1000 bp were removed. The quality-filtered sequences were denoised using the QIIME denoising algorithms [70]. The sequences were then clustered into OTUs at a 97% sequence similarity threshold using UCLUST [71] in QIIME. The OTUs were generated by searches against the Greengenes reference database from August 2013 using a subsampled open-reference method [72]. Before further analysis, chimeric sequences were detected by comparing with a reference database using USEARCH (v. 7.0.1090) [71] and were removed. A representative sequence for each OTU was picked and aligned with the Greengenes reference database by using PyNAST [73]. The alignments were used for phylogenetic tree construction using the FastTree algorithm [74]. An even-depth rarefied OTU table matrix (600 sequences) was constructed to calculate various diversity indices [75]. The Ribosomal Database Project classifier against the Greengenes reference database was used at a minimal confidence of 60% [76] for the taxonomic assignment of representative OTUs. The calculation of α -diversity indices (phylogenetic diversity, observed species count, Chao1 richness estimators, and the Shannon and Simpson indices) and β -diversity indices (Bray–Curtis dissimilarity and UniFrac weighted and unweighted metrics), as well as PCoA, were performed using QIIME pipelines. The calculated coordination was visualized using a web-based visualization tool, Plotly (<http://plot.ly>). To check for the presence of transient environmental bacteria in the gut microbiota, the full dataset was BLAST-searched against SAR11 (GenBank accession no. CP000084), SAR86 (JX530677), and SAR116 (CP001751) sequences. PICRUSt (<http://picrust.github.io>) [29] was used to examine the functional profiles of the fish gut microbial community based on the 16S rRNA gene composition. For the PICRUSt analysis, an OTU table was constructed by closed-reference OTU picking against the May 2013 Greengenes database using QIIME. The OTU table was converted into the PICRUSt format and normalized by the 16S rRNA gene copy number to correct for the over- and under-estimation of microbial abundance. The normalized dataset was analyzed using the KO dataset [77]. Detailed microbiome analytical scripts and computational environments are provided online (Additional file 8: Supplementary Method).

Comparison of the gut microbiota among various animals

Gut microbiota data from various organisms were used for meta-analysis. The human gut microbiota dataset was downloaded from the NIH HMP (<http://hmpdacc.org/>) [31]. The aquatic mammalian gut microbiota [32] data were obtained from the NCBI Sequence Read Archive (SRA). The land and marine iguana gut microbiota data [34] were downloaded from the Dryad data package [78]. Non-human mammalian gut microbiota [33], snake gut microbiota [35], and wild avian gut microbiota [36] data were obtained from the Qiita database (<https://qiita.ucsd.edu/>), as pre-processed data. Closed-reference OTU picking methods (*pick_closed_reference_otus.py*) were used to cluster the OTUs against the same reference sequences (gg_13_8) using the QIIME pipeline (v. 1.8.0). After discarding the unaligned sequences, an even-depth rarefied OTU table was generated and used for subsequent analyses. A non-phylogenetic β -diversity metric (the binary Jaccard index) was calculated and visualized by PCoA.

OTU network-based analysis

For an OTU network-based analysis, OTU network maps were constructed using QIIME and visualized using Cytoscape (v. 3.0.1) [79, 80]. Briefly, the OTU table generated at the 97% sequence similarity threshold was converted to the Cytoscape format (*make_otu_network.py*). In the converted OTU network maps, the samples and OTUs were set to represent network nodes connected by edges, which represented OTU abundance in the samples. The edge-weighted spring embedded model was derived to arrange network constituents.

Statistical analysis

All statistical analyses were performed using GraphPad Prism (v. 5.0; GraphPad, San Diego, CA, USA). The significance of differences between groups was assessed using two-tailed Mann–Whitney U -tests. To compare the β -diversity indices among multiple groups, one-way analysis of variance was used, followed by Duncan's post-hoc tests. For multiple comparisons, p-values were corrected by the Benjamini–Hochberg false discovery rate (FDR) procedure, and $FDR < 0.05$ was considered statistically significant. ANOSIM and PERMANOVA tests with the β -diversity matrix were performed using the QIIME pipeline (*compare_categories.py*). Statistical significance for both tests was determined based on 10,000 permutations. Assessment models to identify the discriminative factors shaping the fish gut microbiota were constructed using random forest classifiers in Weka v. 3.8.3 open source software (<http://www.cs.waikato.ac.nz/~ml/weka/index.html>) developed at Waikato University, New Zealand [81, 82]. The random forest classifiers were trained using individually generated input tables of the relative OTU abundance and discriminative variables with 10-fold cross-validation. To determine the optimal number of clusters for evaluating the cohesiveness of clusters with various metadata, the Calinski–Harabasz index (CH index) and the silhouette score were calculated for each set of clusters generated by PAM clustering [26] (<https://enterotype.embl.de/enterotypes.html#>). The differentially abundant taxonomic and functional features were also confirmed using LEfSe in the Galaxy server (<http://huttenhower.sph.harvard.edu/galaxy/>) [83, 84]. The significance threshold of the α parameter for the Kruskal–Wallis test for classes was set to 0.05. The threshold for the logarithmic LDA score for taxonomic features was 3.8, and that for functional features was 3.0.

List Of Abbreviations

ANOSIM, analysis of similarities;

AUC, area under the curve;

BLAST, Basic Local Alignment Search Tool;

CO1, cytochrome c oxidase subunit I;

FDR, false discovery rate;

KEGG, Kyoto Encyclopedia of Genes and Genomes;

KO, KEGG orthology;

LDA, linear discriminant analysis;

LEfSe, linear discriminant analysis effect size;

OTU, operational taxonomic unit;

PCoA, principal coordinate analysis;

PERMANOVA, permutational multivariate analysis of variance;

PICRUSt, Phylogenetic Investigation of Communities by Reconstruction of Unobserved States;

PRC, precision-recall curves.

Declarations

Ethics approval

All experiments were approved by the Institutional Animal Care and Use Committee of Kyung Hee University and performed in accordance with the protocol KHUASP(SE)-15-087.

Consent for publication

Not applicable.

Availability of data and material

The obtained 16S rRNA gene sequences for the fish gut microbiota and the CO1 gene sequences for collected fish were submitted to the European Nucleotide Archive (ENA) of EMBL-EBI and NCBI GenBank

databases under the accession numbers PRJEB31232 (16S rRNA gene sequences) and MK560532-MK560758 (CO1 gene sequences), respectively.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by a grant from the National Institute of Biological Resources (NIBR), funded by the Ministry of Environment (MOE) of the Republic of Korea (NIBR202002108 to J-WB) and a grant from the Mid-Career Researcher Program (NRF-2020R1A2C3012797 to J-WB) through the National Research Foundation of Korea (NRF), and KRIBB Research Initiative Program (KGM5232113 to N-RS).

Authors' contributions

PSK, J-BL, and J-WB planned and designed the research and experiments. PSK, J-BL, D-WH, and JYK undertook the field work and processing of samples. PSK, N-RS, M-SK, TWW, D-WH, J-HY, M-JJ, JYK, and J-WB performed the experiments and analyzed the data. PSK, N-RS, and J-WB wrote the paper. All authors read and approved the final manuscript.

Acknowledgments

We would like to thank Dr. Hae Won Lee, Dr. Jae Hyeong Yang, and Dr. Mi Young Song at the National Institute of Fisheries Science (NIFS) for their outstanding assistance in collecting fish specimens, as well as the captain and crew members of the fisheries resource research vessel Tamgu-20 of the NIFS for their assistance during expeditions.

References

1. Butterfield NJ, Knoll AH, Swett K: **A bangiophyte red alga from the Proterozoic of arctic Canada.** *Science* 1990, **250**(4977):104-107.
2. Van Valen L: **A new evolutionary law.** *Evol Theor* 1973, **1**:1-30.
3. Hentschel U, Steinert M, Hacker J: **Common molecular mechanisms of symbiosis and pathogenesis.** *Trends Microbiol* 2000, **8**(5):226-231.
4. Bäckhed F, Ding H, Wang T, Hooper LV, Koh GY, Nagy A, Semenkovich CF, Gordon JI: **The gut microbiota as an environmental factor that regulates fat storage.** *Proc Natl Acad Sci USA* 2004, **101**(44):15718-15723.
5. Bäckhed F, Ley RE, Sonnenburg JL, Peterson DA, Gordon JI: **Host-bacterial mutualism in the human intestine.** *Science* 2005, **307**(5717):1915-1920.
6. Robosky LC, Wells DF, Egnash LA, Manning ML, Reily MD, Robertson DGJTS: **Metabonomic identification of two distinct phenotypes in Sprague-Dawley (CrI: CD (SD)) rats.** *Toxicol Sci* 2005,

87(1):277-284.

7. Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI: **An obesity-associated gut microbiome with increased capacity for energy harvest.** *Nature* 2006, **444**(7122):1027-1131.
8. Stecher B, Robbiani R, Walker AW, Westendorf AM, Barthel M, Kremer M, Chaffron S, Macpherson AJ, Buer J, Parkhill J: ***Salmonella enterica* serovar typhimurium exploits inflammation to compete with the intestinal microbiota.** *PLoS Biol* 2007, **5**(10):e244.
9. Round JL, Mazmanian SK: **The gut microbiota shapes intestinal immune responses during health and disease.** *Nat Rev Immunol* 2009, **9**(5):313-323.
10. Hoffmann C, Hill DA, Minkah N, Kirn T, Troy A, Artis D, Bushman F: **Community-wide response of the gut microbiota to enteropathogenic *Citrobacter rodentium* infection revealed by deep sequencing.** *Infect Immun* 2009, **77**(10):4668-4678.
11. Colston TJ, Jackson CR: **Microbiome evolution along divergent branches of the vertebrate tree of life: what is known and unknown.** *Mol Ecol* 2016, **25**(16):3776-3800.
12. Nelson JS: **Fishes of the World:** John Wiley & Sons; 2006.
13. Froese R, Pauly D: **FishBase. Species list: World Wide Web electronic publication.** In.; 2019.
14. Roeselers G, Mittge EK, Stephens WZ, Parichy DM, Cavanaugh CM, Guillemin K, Rawls JF: **Evidence for a core gut microbiota in the zebrafish.** *ISME J* 2011, **5**(10):1595-1608.
15. Sullam KE, Essinger SD, Lozupone CA, O'CONNOR MP, Rosen GL, Knight R, Kilham SS, Russell JA: **Environmental and ecological factors that shape the gut bacterial communities of fish: a meta-analysis.** *Mol Ecol* 2012, **21**(13):3363-3378.
16. Navarrete P, Magne F, Mardones P, Riveros M, Opazo R, Suau A, Pochart P, Romero J: **Molecular analysis of intestinal microbiota of rainbow trout (*Oncorhynchus mykiss*).** *FEMS Microbiol Ecol* 2010, **71**(1):148-156.
17. Ye L, Amberg J, Chapman D, Gaikowski M, Liu W-T: **Fish gut microbiota analysis differentiates physiology and behavior of invasive Asian carp and indigenous American fish.** *ISME J* 2013.
18. Hovda MB, Lunestad BT, Fontanillas R, Rosnes JT: **Molecular characterisation of the intestinal microbiota of farmed Atlantic salmon (*Salmo salar* L.).** *Aquaculture* 2007, **272**(1):581-588.
19. Geraylou Z, Souffreau C, Rurangwa E, D'Hondt S, Callewaert L, Courtin CM, Delcour JA, Buyse J, Ollevier F: **Effects of arabinoxylan-oligosaccharides (AXOS) on juvenile Siberian sturgeon (*Acipenser baeri*) performance, immune responses and gastrointestinal microbial community.** *Fish & shellfish immunology* 2012, **33**(4):718-724.
20. Wilson B, Danilowicz BS, Meijer WG: **The diversity of bacterial communities associated with Atlantic cod *Gadus morhua*.** *Microb Ecol* 2008, **55**(3):425-434.
21. Ley RE, Lozupone CA, Hamady M, Knight R, Gordon JI: **Worlds within worlds: evolution of the vertebrate gut microbiota.** *Nat Rev Microbiol* 2008, **6**(10):776-788.
22. Waite DW, Taylor MW: **Characterizing the avian gut microbiota: membership, driving influences, and potential function.** *Front Microbiol* 2014, **5**:223.

23. Nishida AH, Ochman H: **Rates of gut microbiome divergence in mammals.** *Mol Ecol* 2018, **27**(8):1884-1897.
24. Benson AK, Kelly SA, Legge R, Ma F, Low SJ, Kim J, Zhang M, Oh PL, Nehrenberg D, Hua K: **Individuality in gut microbiota composition is a complex polygenic trait shaped by multiple environmental and host genetic factors.** *Proc Natl Acad Sci USA* 2010, **107**(44):18933-18938.
25. Campbell JH, Foster CM, Vishnivetskaya T, Campbell AG, Yang ZK, Wymore A, Palumbo AV, Chesler EJ, Podar M: **Host genetic and environmental effects on mouse intestinal microbiota.** *ISME J* 2012, **6**(11):2033.
26. Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, Fernandes GR, Tap J, Bruls T, Batto J-MJn: **Enterotypes of the human gut microbiome.** 2011, **473**(7346):174.
27. Lindeman RL: **The trophic-dynamic aspect of ecology.** *Ecology* 1942, **23**(4):399-417.
28. Stergiou KI, Karpouzi VS: **Feeding habits and trophic levels of Mediterranean fish.** *Rev Fish Biol Fisher* 2002, **11**(3):217-254.
29. Langille MG, Zaneveld J, Caporaso JG, McDonald D, Knights D, Reyes JA, Clemente JC, Burkepille DE, Thurber RLV, Knight R: **Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences.** *Nat Biotechnol* 2013, **31**(9):814-821.
30. Methé BA, Nelson KE, Pop M, Creasy HH, Giglio MG, Huttenhower C, Gevers D, Petrosino JF, Abubucker S, Badger JH: **A framework for human microbiome research.** *Nature* 2012, **486**(7402):215.
31. Huttenhower C, Gevers D, Knight R, Abubucker S, Badger JH, Chinwalla AT, Creasy HH, Earl AM, FitzGerald MG, Fulton RSJN: **Structure, function and diversity of the healthy human microbiome.** 2012, **486**(7402):207.
32. Bik EM, Costello EK, Switzer AD, Callahan BJ, Holmes SP, Wells RS, Carlin KP, Jensen ED, Venn-Watson S, Relman DA: **Marine mammals harbor unique microbiotas shaped by and yet distinct from the sea.** *Nat Commun* 2016, **7**:10516.
33. Muegge BD, Kuczynski J, Knights D, Clemente JC, González A, Fontana L, Henrissat B, Knight R, Gordon JI: **Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans.** *Science* 2011, **332**(6032):970-974.
34. Lankau EW, Hong PY, Mackie RI: **Ecological drift and local exposures drive enteric bacterial community differences within species of Galapagos iguanas.** *Mol Ecol* 2012, **21**(7):1779-1788.
35. Costello EK, Gordon JI, Secor SM, Knight R: **Postprandial remodeling of the gut microbiota in Burmese pythons.** *ISME J* 2010, **4**(11):1375-1385.
36. Peralta-Sánchez JM, Martín-Platero AM, Wegener-Parfrey L, Martínez-Bueno M, Rodríguez-Ruano S, Navas-Molina JA, Vázquez-Baeza Y, Martín-Gálvez D, Martín-Vivaldi M, Ibáñez-Álamo JD: **Bacterial density rather than diversity correlates with hatching success across different avian species.** *FEMS Microbiol Ecol* 2018, **94**(3):fiy022.
37. Ley RE, Hamady M, Lozupone C, Turnbaugh PJ, Ramey RR, Bircher JS, Schlegel ML, Tucker TA, Schrenzel MD, Knight R: **Evolution of mammals and their gut microbes.** *Science* 2008, **320**(5883):1647-1651.

38. Vital M, Gao J, Rizzo M, Harrison T, Tiedje JM: **Diet is a major factor governing the fecal butyrate-producing community structure across *Mammalia*, *Aves* and *Reptilia*.** *ISME J* 2015, **9**(4):832.
39. Yun J-H, Roh SW, Whon TW, Jung M-J, Kim M-S, Park D-S, Yoon C, Nam Y-D, Kim Y-J, Choi J-H: **Insect gut bacterial diversity determined by environmental habitat, diet, developmental stage, and phylogeny of host.** *Appl Environ Microbiol* 2014, **80**(17):5254-5264.
40. Rungrassamee W, Klanchui A, Maibunkaew S, Chaiyapechara S, Jiravanichpaisal P, Karoonuthaisiri N: **Characterization of intestinal bacteria in wild and domesticated adult black tiger shrimp (*Penaeus monodon*).** *PloS One* 2014, **9**(3):e91853.
41. Hong P-Y, Wheeler E, Cann IK, Mackie RI: **Phylogenetic analysis of the fecal microbial community in herbivorous land and marine iguanas of the Galápagos Islands using 16S rRNA-based pyrosequencing.** *ISME J* 2011, **5**(9):1461-1470.
42. Colombo BM, Scalvenzi T, Benlamara S, Pollet N: **Microbiota and mucosal immunity in amphibians.** *Front Immunol* 2015, **6**:111.
43. Shin N-R, Whon TW, Bae J-W: **Proteobacteria: microbial signature of dysbiosis in gut microbiota.** *Trends Biotechnol* 2015, **33**(9):496-503.
44. Morgan XC, Tickle TL, Sokol H, Gevers D, Devaney KL, Ward DV, Reyes JA, Shah SA, LeLeiko N, Snapper SB: **Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment.** *Genome Biol* 2012, **13**(9):R79.
45. Walujkar SA, Dhotre DP, Marathe NP, Lawate PS, Bharadwaj RS, Shouche YS: **Characterization of bacterial community shift in human Ulcerative Colitis patients revealed by Illumina based 16S rRNA gene amplicon sequencing.** *Gut Pathog* 2014, **6**(1):22.
46. Guillaume J, Kaushik S, Bergot P, Metailler R: **Nutrition and feeding of fish and crustaceans:** Springer Science & Business Media; 2001.
47. Sunagawa S, Coelho LP, Chaffron S, Kultima JR, Labadie K, Salazar G, Djahanschiri B, Zeller G, Mende DR, Alberti A: **Structure and function of the global ocean microbiome.** *Science* 2015, **348**(6237):1261359.
48. Morris RM, Rappé MS, Connon SA, Vergin KL, Siebold WA, Carlson CA, Giovannoni SJ: **SAR11 clade dominates ocean surface bacterioplankton communities.** *Nature* 2002, **420**(6917):806.
49. Dupont CL, Rusch DB, Yooseph S, Lombardo M-J, Richter RA, Valas R, Novotny M, Yee-Greenbaum J, Selengut JD, Haft DH: **Genomic insights to SAR86, an abundant and uncultivated marine bacterial lineage.** *ISME J* 2012, **6**(6):1186.
50. Hosokawa T, Kikuchi Y, Nikoh N, Shimada M, Fukatsu T: **Strict host-symbiont cospeciation and reductive genome evolution in insect gut bacteria.** *PLoS Biol* 2006, **4**(10):e337.
51. Rousk J, Bååth E, Brookes PC, Lauber CL, Lozupone C, Caporaso JG, Knight R, Fierer N: **Soil bacterial and fungal communities across a pH gradient in an arable soil.** *ISME J* 2010, **4**(10):1340.
52. Nicholson JK, Holmes E, Kinross J, Burcelin R, Gibson G, Jia W, Pettersson S: **Host-gut microbiota metabolic interactions.** *Science* 2012, **336**(6086):1262-1267.

53. Sugita H, Miyajima C, Deguchi Y: **The vitamin B12-producing ability of the intestinal microflora of freshwater fish.** *Aquaculture* 1991, **92**:267-276.
54. Tsuchiya C, Sakata T, Sugita H: **Novel ecological niche of *Cetobacterium somerae*, an anaerobic bacterium in the intestinal tracts of freshwater fish.** *Lett Appl Microbiol* 2008, **46**(1):43-48.
55. Baldo L, Pretus JL, Riera JL, Musilova Z, Nyom ARB, Salzburger W: **Convergence of gut microbiotas in the adaptive radiations of African cichlid fishes.** *ISME J* 2017, **11**(9):1975.
56. De Filippo C, Cavalieri D, Di Paola M, Ramazzotti M, Poullet JB, Massart S, Collini S, Pieraccini G, Lionetti P: **Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa.** *Proc Natl Acad Sci USA* 2010, **107**(33):14691-14696.
57. David LA, Maurice CF, Carmody RN, Gootenberg DB, Button JE, Wolfe BE, Ling AV, Devlin AS, Varma Y, Fischbach MA: **Diet rapidly and reproducibly alters the human gut microbiome.** *Nature* 2014, **505**(7484):559.
58. Xue Z, Zhang W, Wang L, Hou R, Zhang M, Fei L, Zhang X, Huang H, Bridgewater LC, Jiang Y: **The bamboo-eating giant panda harbors a carnivore-like gut microbiota, with excessive seasonal variations.** *MBio* 2015, **6**(3):e00022-00015.
59. Liu H, Guo X, Gooneratne R, Lai R, Zeng C, Zhan F, Wang W: **The gut microbiome and degradation enzyme activity of wild freshwater fishes influenced by their trophic levels.** *Sci Rep* 2016, **6**:24340.
60. Nayak SK: **Role of gastrointestinal microbiota in fish.** *Aquac Res* 2010, **41**(11):1553-1573.
61. Rawls JF, Mahowald MA, Ley RE, Gordon JI: **Reciprocal gut microbiota transplants from zebrafish and mice to germ-free recipients reveal host habitat selection.** *Cell* 2006, **127**(2):423-433.
62. Lim YW, Evangelista JS, Schmieder R, Bailey B, Haynes M, Furlan M, Maughan H, Edwards R, Rohwer F, Conrad D: **Clinical insights from metagenomic analysis of sputum samples from patients with cystic fibrosis.** *J Clin Microbiol* 2014, **52**(2):425-437.
63. Shin N-R, Lee J-C, Lee H-Y, Kim M-S, Whon TW, Lee M-S, Bae J-W: **An increase in the *Akkermansia* spp. population induced by metformin treatment improves glucose homeostasis in diet-induced obese mice.** *Gut* 2014, **63**(5):727-735.
64. Ivanova NV, Zemplak TS, Hanner RH, Hebert PD: **Universal primer cocktails for fish DNA barcoding.** *Mol Ecol Notes* 2007, **7**(4):544-548.
65. Johnson M, Zaretskaya I, Raytselis Y, Merezhuik Y, McGinnis S, Madden TL: **NCBI BLAST: a better web interface.** *Nucleic Acids Res* 2008, **36**(suppl 2):W5-W9.
66. Tamura K, Stecher G, Peterson D, Filipowski A, Kumar S: **MEGA6: molecular evolutionary genetics analysis version 6.0.** *Mol Biol Evol* 2013, **30**(12):2725-2729.
67. Thompson JD, Higgins DG, Gibson TJ: **CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.** *Nucleic Acids Res* 1994, **22**(22):4673-4680.
68. Felsenstein J: **Evolutionary trees from DNA sequences: a maximum likelihood approach.** *J Mol Evol* 1981, **17**(6):368-376.

69. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI *et al*: **QIIME allows analysis of high-throughput community sequencing data.** *Nature methods* 2010, **7**(5):335-336.
70. Reeder J, Knight R: **Rapidly denoising pyrosequencing amplicon reads by exploiting rank-abundance distributions.** *Nature methods* 2010, **7**(9):668-669.
71. Edgar RC: **Search and clustering orders of magnitude faster than BLAST.** *Bioinformatics* 2010, **26**(19):2460-2461.
72. Rideout JR, He Y, Navas-Molina JA, Walters WA, Ursell LK, Gibbons SM, Chase J, McDonald D, Gonzalez A, Robbins-Pianka A: **Subsampled open-reference clustering creates consistent, comprehensive OTU definitions and scales to billions of sequences.** *PeerJ* 2014, **2**:e545.
73. Caporaso JG, Bittinger K, Bushman FD, DeSantis TZ, Andersen GL, Knight R: **PyNAST: a flexible tool for aligning sequences to a template alignment.** *Bioinformatics* 2010, **26**(2):266-267.
74. Price MN, Dehal PS, Arkin AP: **FastTree: computing large minimum evolution trees with profiles instead of a distance matrix.** *Mol Biol Evol* 2009, **26**(7):1641-1650.
75. McDonald D, Clemente JC, Kuczynski J, Rideout JR, Stombaugh J, Wendel D, Wilke A, Huse S, Hufnagle J, Meyer F: **The Biological Observation Matrix (BIOM) format or: how I learned to stop worrying and love the ome-ome.** *GigaScience* 2012, **1**(1):7.
76. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, Andersen GL: **Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB.** *Appl Environ Microbiol* 2006, **72**(7):5069-5072.
77. Kanehisa M, Goto S: **KEGG: kyoto encyclopedia of genes and genomes.** *Nucleic Acids Res* 2000, **28**(1):27-30.
78. Wheeler EL, Hong P, Mackie RI: **Data from: Ecological drift and local exposures drive gastrointestinal bacterial community differences among Galápagos iguana populations.** In.: Dryad Data Repository; 2012.
79. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T: **Cytoscape: a software environment for integrated models of biomolecular interaction networks.** *Genome Res* 2003, **13**(11):2498-2504.
80. Navas-Molina JA, Peralta-Sánchez JM, González A, McMurdie PJ, Vázquez-Baeza Y, Xu Z, Ursell LK, Lauber C, Zhou H, Song SJ: **Advancing our understanding of the human microbiome using QIIME.** *Method Enzymol* 2013, **531**:371-444.
81. Breiman L: **Random forests.** *Mach Learn* 2001, **45**(1):5-32.
82. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH: **The WEKA data mining software: an update.** *ACM SIGKDD Explor Newslett* 2009, **11**(1):10-18.
83. Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, Huttenhower C: **Metagenomic biomarker discovery and explanation.** *Genome Biol* 2011, **12**(6):R60.

84. Goecks J, Nekrutenko A, Taylor J: **Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences.** *Genome Biol* 2010, 11(8):R86.

Figures

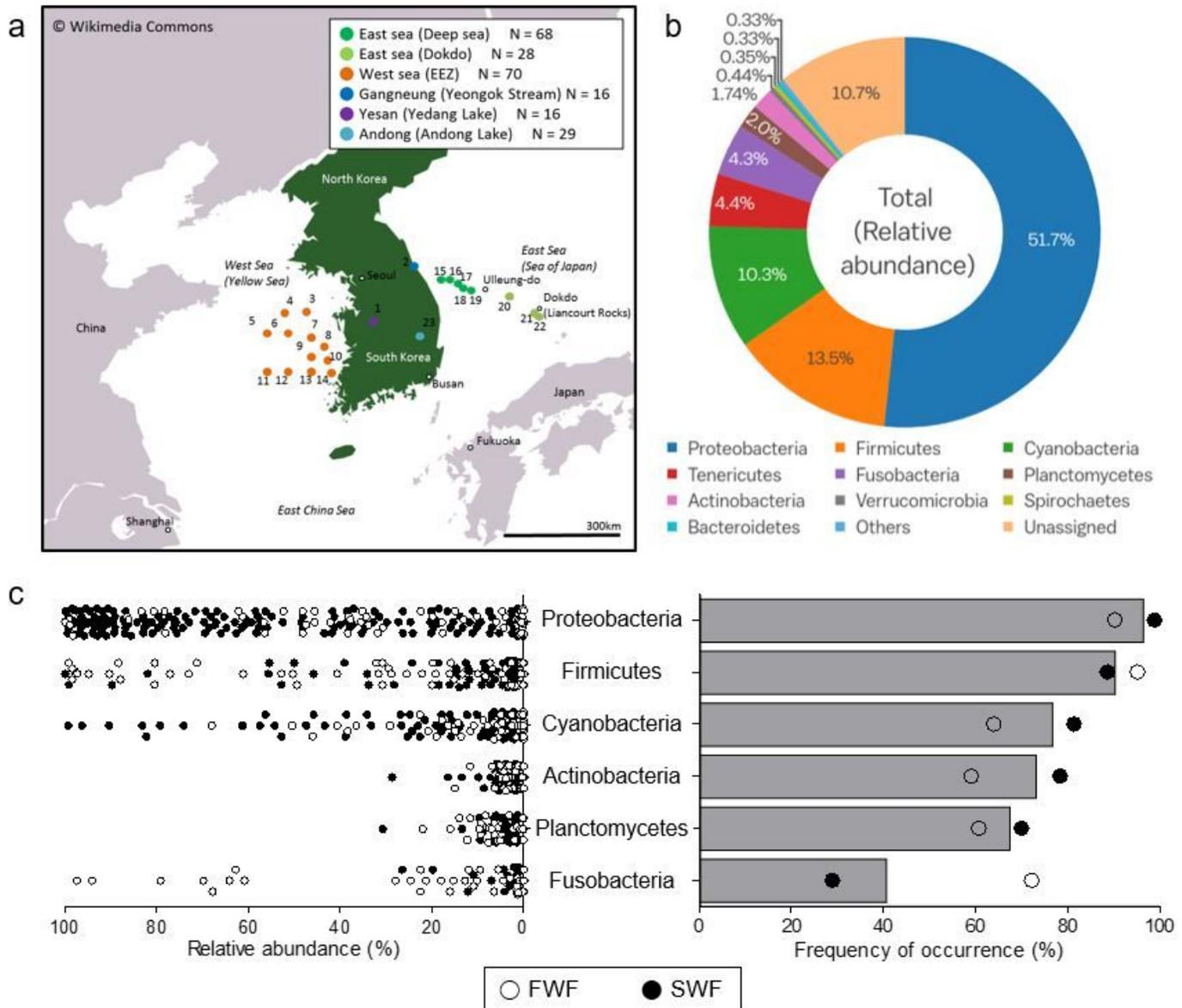


Figure 1

Overview of the data. (a) Regional map showing the approximate locations of 23 sampling sites (227 fish). (b) Pie chart of the relative abundance of bacterial phyla (>0.3%) in the gut microbiota in all fish samples. (c) Dot plot of the overall distribution of the relative abundance (left) and frequency of occurrence (right) of taxa in total fish (bar) and freshwater fish or seawater fish (dot) at the bacterial phylum level. FWF, freshwater fish; SWF, seawater fish. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the

part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

Fig. 2

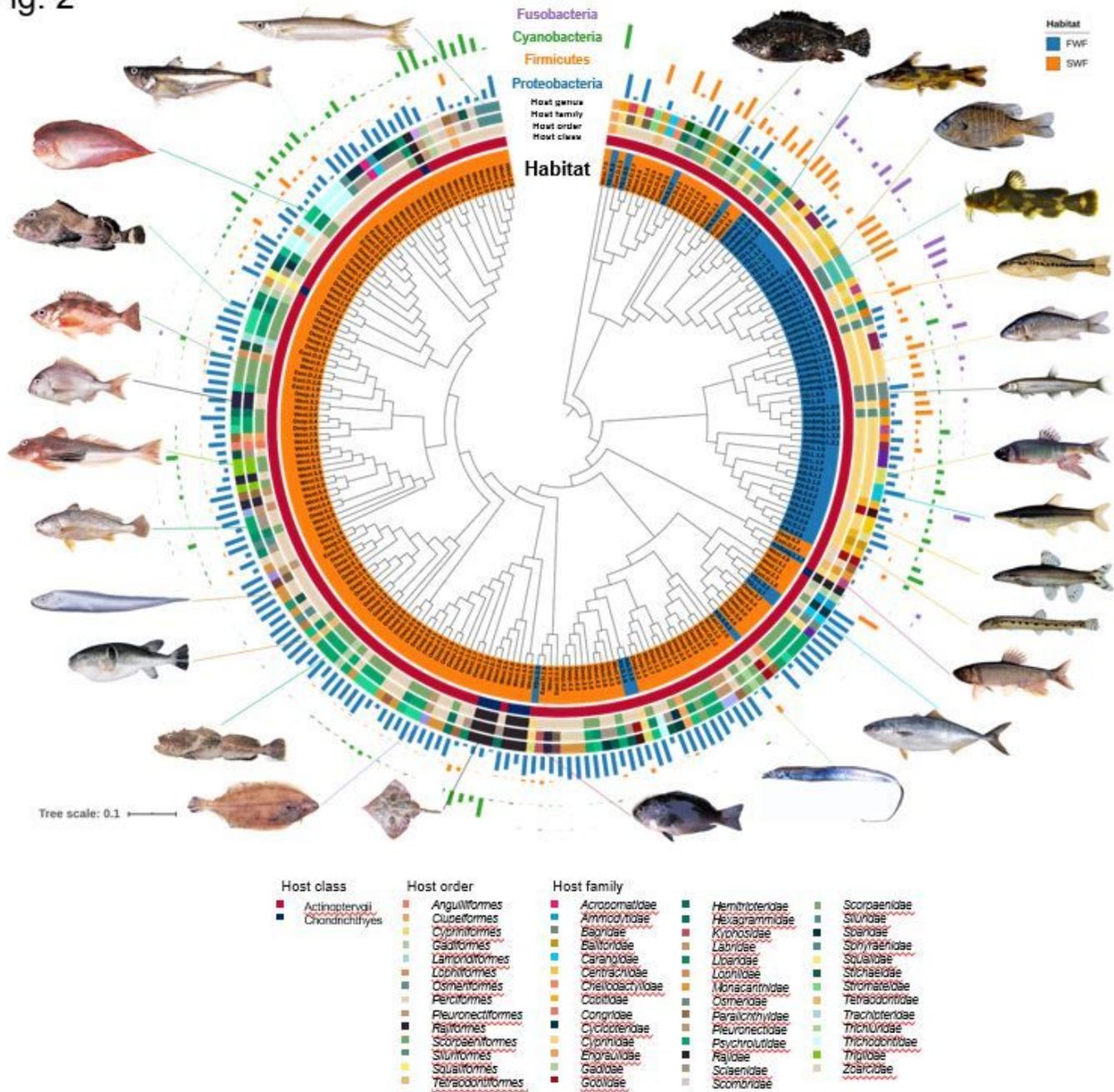


Figure 2

Overview of the gut microbiome of fish. UPGMA tree was constructed with 227 fish samples. Phylum-level profiling of the gut microbiome composition and host taxonomic attributes are presented with representative photos of fish species included in this study. Detailed color index of bacterial genera was abridged. Fish photos courtesy of the National Institute of Fisheries Science (NIFS), Korea.

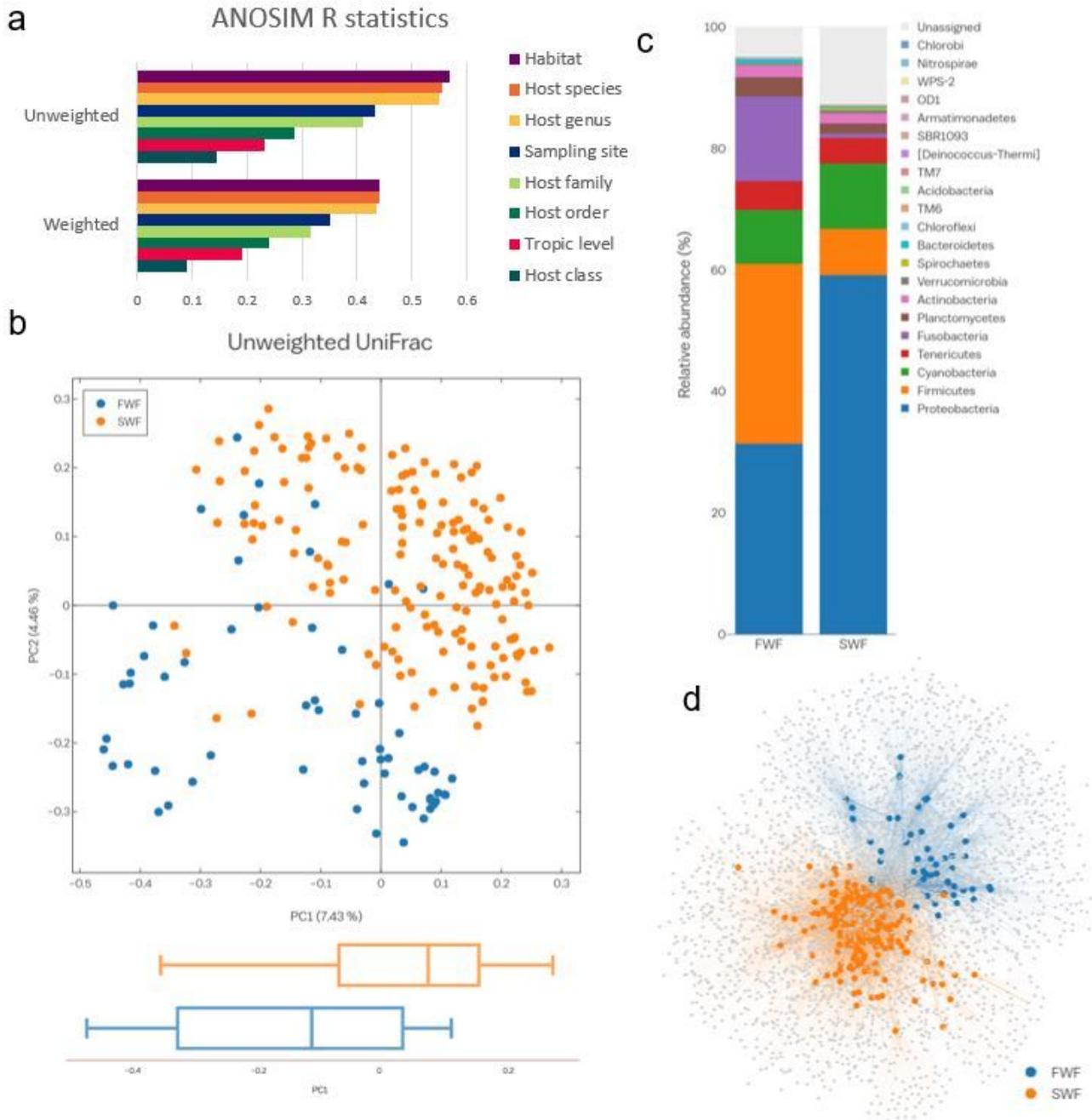


Figure 3

Fish gut microbiota is determined by the host habitat. (a) Analysis of the contributions of host environmental or genetic factors to the fish gut microbiota. Variation was determined by between-sample unweighted or weighted UniFrac distances. The size effect and statistical significance were calculated by ANOSIM using the R “vegan” package in the QIIME pipeline. (b) PCoA of unweighted UniFrac distances for 227 fish samples (ANOSIM, $R = 0.47$, $p < 0.001$) and boxplots illustrating PC1 coordinates of freshwater and seawater fish. The center line shows the median, the boxes cover the 25th to 75th percentiles, and the whiskers extend to 1.5× the interquartile range (c) Bar charts of the relative abundance of bacterial phyla in the gut microbiota of fish from different habitats. (d) OTU network-based analysis of the microbial communities in fish from different habitats. The edges connecting nodes

representing fish samples (circles) to species-level OTUs in a particular sample are colored according to the host habitat type (edge-weighted spring embedded model in Cytoscape v. 3.0.1). FWF, freshwater fish; SWF, seawater fish.

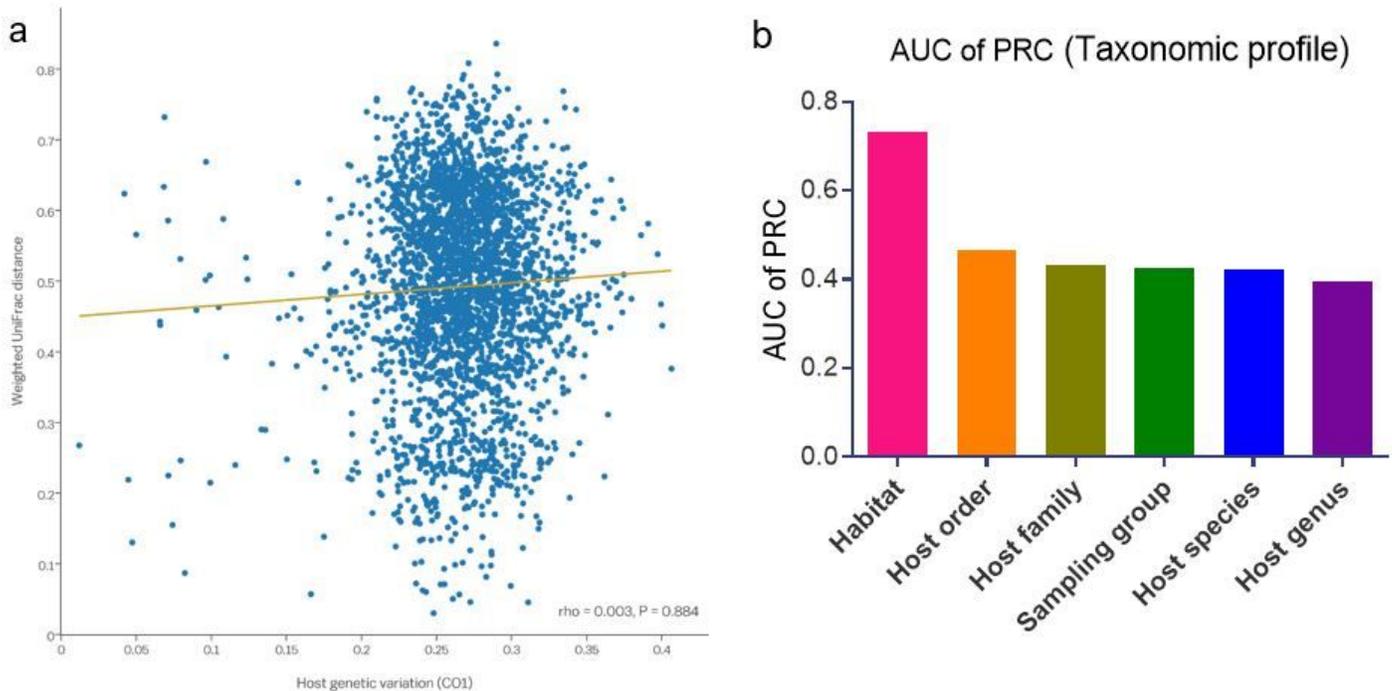


Figure 4

Limited evidence for an association between the fish gut microbiota and host genetic factors. (a) Pairwise comparison of phylogenetic distances between the fish gut microbiota based on weighted UniFrac distances and host genetic variation (CO1 gene). The relationship was not statistically significant ($p = 0.884$ and $\sigma = 0.003$, Spearman correlation). (b) AUC of PRC for random forest classifiers for various discriminative factors of the taxonomic profiles of the fish gut.

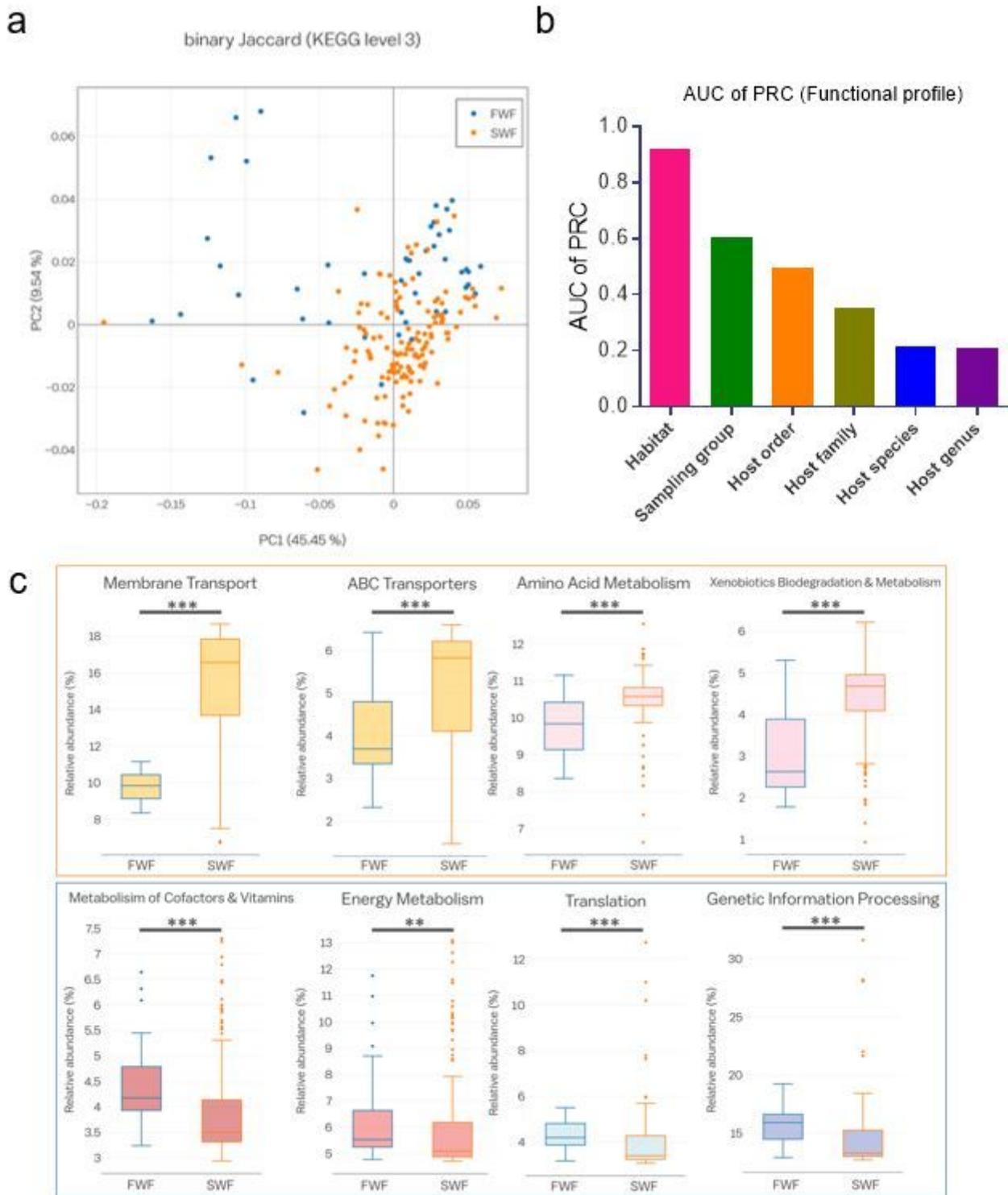


Figure 5

KEGG categories derived from the 16S rRNA sequences of the fish gut microbiome by PICRUSt. (a) PCoA of the binary Jaccard dissimilarity of the functional profiles (ANOSIM, $R = 0.37$, $p < 0.001$). (b) Box-and-whisker plots of the relative abundance of the selected KOs for samples from two different habitats determined by the LefSe analysis (LDA score > 3.0). The center line shows the median, the boxes cover the 25th to 75th percentiles, the whiskers extend to $1.5\times$ the interquartile range, and the outer points are

outliers. Asterisks indicate significant differences between freshwater and seawater fish according to a two-tailed Mann–Whitney U-test. ** $p < 0.01$; *** $p < 0.001$. (c) Bar plot of the AUC of PRC for random forest classifiers for various discriminative factors of functional profiles of the fish gut. FWF, freshwater fish; SWF, seawater fish.

Fig. 6

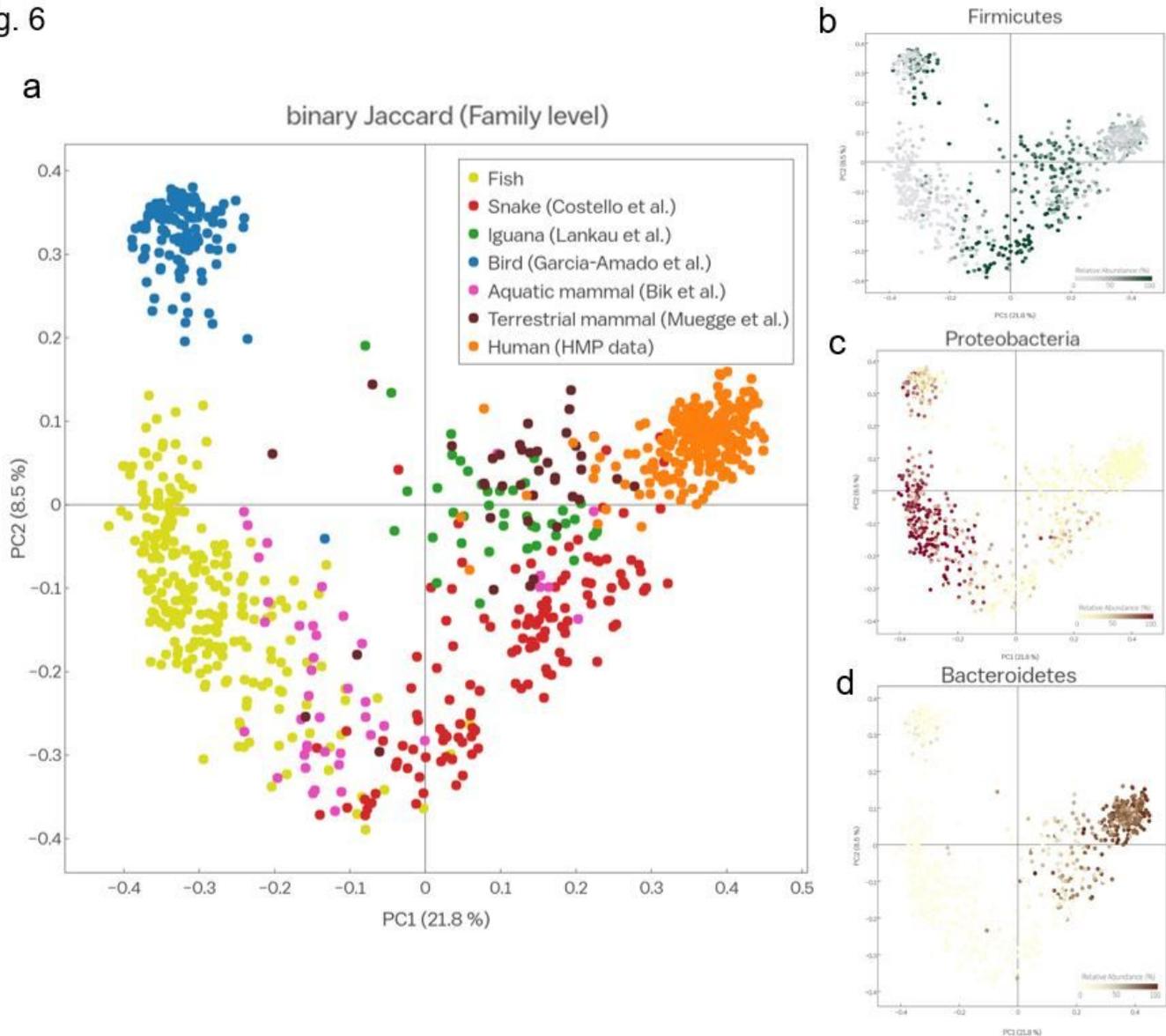


Figure 6

Fish gut microbiota clustered separately from the microbiotas of other vertebrates. (a) PCoA of the binary Jaccard indices of the gut microbiota from fish and various vertebrates (iguanas, snakes, aquatic mammals, birds, terrestrial mammals, and humans; sequences were obtained from the NCBI SRA and the Qiita server). (b-d) PCoA plots colored by the relative abundance of the phyla Proteobacteria (b), Firmicutes (c), and Bacteroidetes (d). Color intensity is proportional to the relative abundance (%) of each phylum.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Additionalfile1SupplementaryFigures2021.docx](#)
- [Additionalfile2SuppleTableS1Relativeabundance.xlsx](#)
- [Additionalfile3SuppleTableS2fxnprofile.xlsx](#)
- [Additionalfile4SuppleTableS3betameta.xlsx](#)
- [Additionalfile5SuppleTableS4blastresults.xlsx](#)
- [Additionalfile6SuppleTableS5samplingsite.xlsx](#)
- [Additionalfile7SuppleTableS6collectedfish.xlsx](#)
- [Additionalfile8SuppleMethodscript.txt](#)