JOURNAL OF Evolutionary Biology

doi: 10.1111/jeb.12374

Microbiome investigation in the ecological speciation context of lake whitefish (*Coregonus clupeaformis*) using next-generation sequencing

M. SEVELLEC, S. A. PAVEY, S. BOUTIN, M. FILTEAU, N. DEROME & L. BERNATCHEZ

Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval, Québec, QC, Canada

Keywords:

Coregonus clupeaformis;
kidney;
microbiome;
microbiota;
next-generation sequencing;
parallel evolution;
speciation.

Abstract

Few studies have applied NGS methods to investigate the microbiome of vertebrates in their natural environment and in freshwater fishes in particularly. Here, we used pyrosequencing of the 16S gene rRNA to (i) test for differences in kidney bacterial communities (i.e. microbiota) of dwarf and normal whitefish found as sympatric pairs, (ii) test the hypothesis of higher bacterial diversity in normal compared with dwarf whitefish and (iii) test for the occurrence of parallelism with the presence and composition of bacterial communities across species pairs inhabiting different lakes. The kidney microbiota of 253 dwarf and normal whitefish from five lakes was analysed combining a double-nested PCR approach with 454 pyrosequencing. Bacteria were detected in 52.6% of the analysed whitefish. There was no overall significant difference among lakes and forms, although the lake × form interaction was found significant. We identified 579 bacterial genera, which is substantially more than previous descriptions using less sensitive techniques of fish bacterial diversity in kidney, pathogenic or not. Ten of these genera contained eighteen pathogenic species. Differences in bacteria composition between whitefish forms were not parallel among lakes. In accordance with the higher diversity of prey types, normal whitefish kidney tissue consistently had a more diverse bacterial community and this pattern was parallel among lakes. These results add to building evidence from previous studies on this system that the adaptive divergence of dwarf, and normal whitefish has been driven by both parallel and nonparallel ecological conditions across lakes.

Introduction

Wild vertebrate species host a considerable bacterial diversity, which may influence their development, physiology, immune system and nutrition (Hooper *et al.*, 2002; Bäckhed *et al.*, 2005; Turnbaugh *et al.*, 2007). Four main types of relationships between bacteria and their hosts have been documented. The first two types are commensal bacteria (Cahill, 1990), which may either have beneficial or neutral effects on the host (Prescott *et al.* 1995). The second type has a symbiotic obligatory relationship with the host, thus allowing a mutual benefit between symbiotic bacteria and host (Perru, 2006). The third type is opportunistic bacteria, which are facultative pathogenic bacteria that may become actively pathogenic when the host immune system is impaired and unable to fight off infection (Berg *et al.*, 2005). The fourth type of relationship pertains to pathogenic bacteria which are responsible for infectious diseases (Falkow, 1997). Species and populations may differ in their susceptibility to infection due to differences in their immune systems (White *et al.*, 2009), which may have evolved in response to selection from exposure to changing microbial communities (Nakajima *et al.*, 2011).

Methods of measuring bacterial communities are rapidly improving. The earliest and most traditional

Correspondence: Maelle Sevellec, Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval, 1030, Avenue de la Médecine, Québec, Québec G1V 0A6, Canada.
Tel.: 418 656 2131 #8455; fax: 1 418 656 7176;
e-mail: maelle.sevellec.1@ulaval.ca

technique is the culture-dependent method. Because of functional interdependency for most of the bacterial community members (Laplante *et al.*, 2013), many bacterial species cannot be cultured, and others vary greatly in their culture requirements. As a result, culture-based approaches may suffer from inconsistencies, low sensitivity and a biased global overview of the bacterial diversity. In recent decades, microbiologists have developed new culture-independent techniques to obtain a better representation of bacterial communities present in host organisms, for example denaturing gradient gel electrophoresis (DGGE) and temperature gradient gel electrophoresis (TGGE) (Muyzer & Smalla, 1998). Despite their usefulness, these methods are also limited by the resolution of band detection with complex bacterial communities and microbes of low abundance may easily be missed (Danilo, 2004).

Advanced culture-independent techniques, such as 16S rRNA massively parallel pyrosequencing, allow a more complete description of complex bacterial communities. The 16S rRNA gene is composed of conserved and 'hypervariable' regions (Amann & Ludwig, 2000; Huse *et al.*, 2008). Thus, it is possible to design nearly universal primers in the conserved regions that capture enough sequence diversity to delineate bacterial genera or even species. This technique has demonstrated its effectiveness in studies of environmental bacterial communities (Huber *et al.*, 2007; Roesch *et al.*, 2007; Ghiglione & Murray, 2011; Kuffner *et al.*, 2012; Roesch *et al.*, 2012; Collin *et al.*, 2013), and in investigating the microbiome of plant, human and mouse (White *et al.*, 2009; Arumugam *et al.*, 2011; Buffie *et al.*, 2011; Lopez-Velasco *et al.*, 2011; Siqueira *et al.*, 2011; Grice & Segre, 2012; Lebeis *et al.*, 2012). In contrast, there have been very few studies using this approach documenting the microbiome of vertebrate species in their natural environment (Yildirim *et al.*, 2010; Lavery *et al.*, 2012; McKenzie *et al.*, 2012; Larsen *et al.*, 2013; Linnenbrink *et al.*, 2013), and, to our knowledge, no study has yet documented the microbiome of any freshwater wild fish species using 16S rRNA pyrosequencing.

Lake whitefish (*Coregonus clupeaformis*) comprises sympatric species pairs referred to as dwarf and normal whitefish that are found in several lakes of the St. John River drainage in the Province of Québec, Canada, and Maine, United States. A recent period of adaptive radiation [post-glacial, 12 000 years before present (YBP)] has led to the parallel phenotypic and ecological evolution in different lakes of the dwarf whitefish derived from the ancestral normal whitefish (Bernatchez, 2004). Dwarf and normal whitefish are partially reproductively isolated in each lake (Gagnaire *et al.*, 2013), differ in genetically based morphological, physiological, behavioural, ecological and life history traits (Fenderson, 1964; Bernatchez *et al.*, 2010) and occupy the limnetic and benthic habitat, respectively. Dwarf and normal whitefish also differ at the immune system level whereby evidence of parallelism of genes relative to immunity was highlighted among whitefish sympatric species forms (St-Cyr *et al.*, 2008; Jeukens *et al.*, 2010).

A recent study also revealed variable patterns of divergence at the MHCIIβ genes between different pairs of dwarf and normal whitefish, although there was no evidence for parallelism in patterns observed among lakes (Pavey *et al.*, 2013). This study also found no parallel patterns in a small subset of genera where pathogenic bacteria have been identified in the literature. To further investigate the possible role of bacteria in the parallel ecological speciation of whitefish, this study considers the *entire* microbiota found in the kidney tissue of dwarf and normal whitefish from different lakes. The kidney of teleost fish, which include whitefish, is known to play several functions, including urinary and a major immune function (Danguy *et al.*, 2011). Furthermore, the presence of bacteria in kidney has been considered evidence of a pathogen infection (Cahill, 1990). Two previous studies performed on the kidney microbial community in salmonids found 10 genera and 27 genera with DGGE technique and culture-dependent technique, respectively (Dionne *et al.*, 2009; Evans & Neff, 2009). Based on previous studies in other groups of organisms, it is expected to detect a greater diversity of genera in using the more sensitive technique of 16S rRNA pyrosequencing on the infected whitefish individuals. In fact, this technique may even be sensitive enough to detect bacterial DNA that is the result of successful immune responses (Pavey *et al.*, 2013). However, throughout the manuscript, we will refer to individuals with bacteria amplified from their kidney tissue as 'infected'. In this context, our first objective is to test for differences in kidney bacterial communities between dwarf and normal whitefish found in sympatry in the same lake. The dwarf whitefish form feed almost exclusively on small zooplankton, whereas normal whitefish feed on a wider diversity of prey types, including zooplankton, but predominantly zoobenthos, molluscs and small fish (Bernatchez *et al.*, 1999, 2010). Because the digestive tract is one of the major infection routes in fish (Ringø & Olsen, 1999), bacteria have the opportunity to colonize kidneys after passing through the intestinal epithelium (Hart *et al.*, 1988; Jutfelt *et al.*, 2006; Knudsen *et al.*, 2008). Thus, these parallel ecological differences in habitat use and diet suggest that dwarf and normal whitefish from a given lake could be exposed to different bacterial communities, whereas whitefish from the same form but from different lakes could be exposed to similar ones. Some of these could be pathogenic and thus potentially imposing differential selection between dwarf and normal whitefish which could have contributed to the parallel divergence of these species pairs. Therefore, our next objectives were to, second, test the hypothesis of higher bacterial taxonomic diversity in normal vs. dwarf, given their broader range of prey types and,

third, test for the occurrence of parallelism at the presence and composition of bacterial communities across species pairs inhabiting different lakes.

Materials and methods

Biological material

Sympatric dwarf and normal whitefish samples were collected in five different lakes (Cliff, East, Témiscouata, Webster and Indian) from the St John River drainage, Québec, Canada, and Maine, United States (Fig. 1). The lakes are geographically and hydrographically isolated from one another (Lu & Bernatchez, 1999). A total of 253 apparently healthy whitefish (from external appearance) were sampled with gill nets between 14 June and 15 July 2010 (Table 1). Fish were dissected on the field in sterile conditions; ventral belly surface of fish was rinsed with ethanol, nondisposable tools were rinsed with ethanol and heated over a blow torch between samples, and kidneys were individually stored in a sterile Eppendorf© tube and flash-frozen in liquid nitrogen. The samples were then transported to the laboratory and kept at −80 °C until further processing.

Bacterial detection using double-nested PCR

To diagnose the presence of bacteria in whitefish kidney, a double-nested PCR was performed. Due to the high concentration of host genomic DNA in kidney tissue relative to potential bacterial DNA, none of our extractions amplify bacterial DNA with repeatability using the standard techniques of DNA amplification (Boutin *et al.*, 2012). Detailed protocols of the DNA

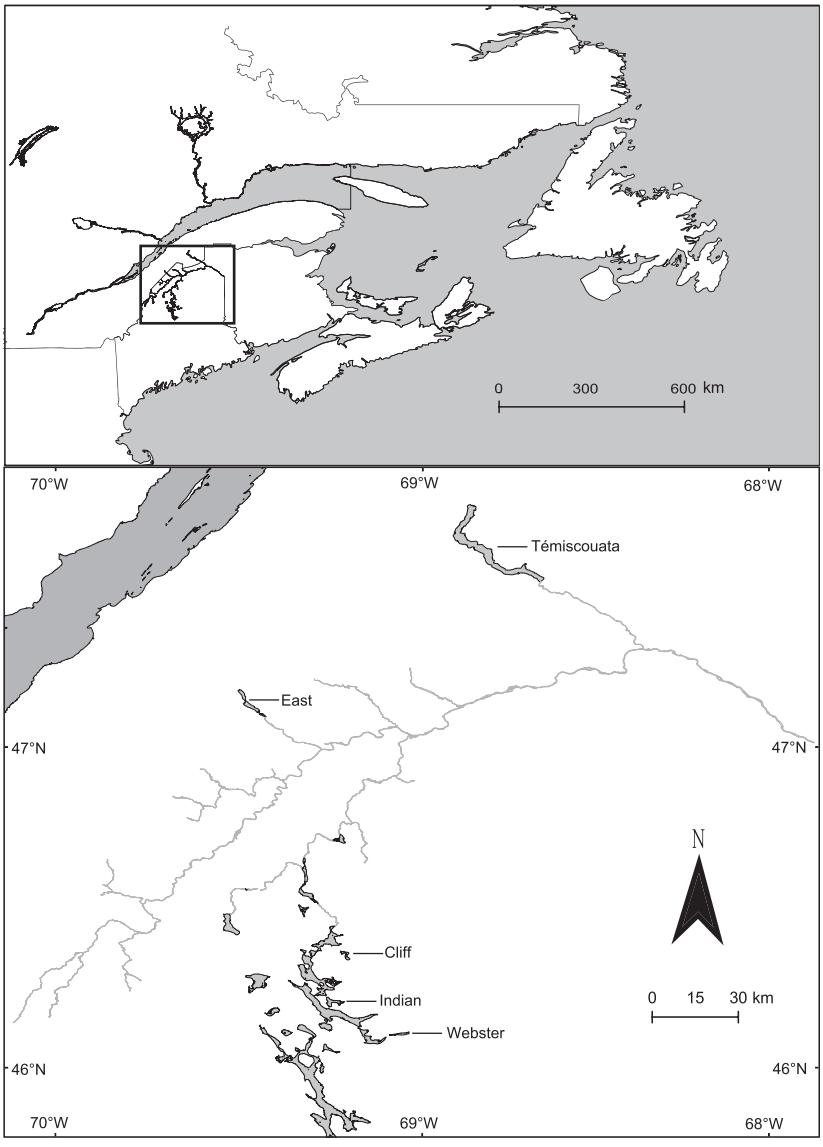

Fig. 1 Map of the study area. The samples come from lakes Témiscouata, East, Cliff, Indian, Webster.

Table 1 Number of infected and noninfected whitefish for each form in each lake according to the results of double-nested PCR.

Lakes	Species	Number of infected fish	Number of healthy fish	Total	% Infected fish	% noninfected fish	Sampling date
Cliff	D	14	20	34	41	59	June 14–16 2010
	N	13	11	24	54	46	
	T	27	31	58	47	53	
East	D	17	7	24	71	29	July 13-14-15 2010
	N	10	16	26	38	62	
	T	27	23	50	54	46	
Indian	D	14	17	31	45	55	June 15–16 2010
	N	23	2	25	92	8	
	T	37	19	56	66	34	
Témiscouata	D	15	11	26	58	42	July 6-7-8 2010
	N	3	10	13	23	77	
	T	18	21	39	46	54	
Webster	D	11	14	25	44	56	June 17 2010
	N	13	12	25	52	48	
	T	24	26	50	48	52	
Total	D	71	69	140	51	49	
	N	62	51	113	55	45	
	T	133	120	253	53	47	

D, dwarf whitefish; N, normal whitefish; and T, total of whitefish per lake.

extraction, double-nested PCR, library construction and 454 pyrosequencing are provided in Boutin *et al.* (2012). In brief, DNA from kidney tissue was extracted with a modified QIAamp DNA minikit protocol for tissues in sterile way, and the double-nested PCR was based on the same principle of the nested PCR described by Yourno (1992), except that three successive amplifications steps were carried out with three different primer pairs (Table 2). The full 16S rDNA (1380 bp) was amplified using primers 1389R and 9F. For the second step, primers 907R and 23F allowed a specific reamplification of the hypervariable region V1-V2-V3-V4-V5 (884 bp). Finally, primers 519R and 63F were used to specifically reamplify the rDNA hypervariable region V1-V2-V3 (456 bp). For each step of amplification, two negative controls [extraction was replaced with 2 μL of sterile nuclease-free water (DEPC-treated Water Ambion®, Ambion, Austin, TX, USA)] and one positive control (extraction was replaced with 2 μL of bacterial cultures in liquid medium) were done. All positive controls were positive, and of 28 negative controls, three only showed fainted bands of contamination. A double-nested PCR was performed on each sample twice independently. In the case of conflicting results (one positive DNA amplification result and one negative DNA amplification result), a final double-nested PCR was performed, serving as a 'tiebreaker'. This double-nested PCR enabled us to significantly increase a very low amount of bacterial DNA while avoiding eukaryotic DNA contamination (Boutin *et al.*, 2012). The library was done during third-step primers of the double-nested PCR summing the A and B adapters required for 454 pyrosequencing and 45 different bar-coded MID-tags (Multiplex identifiers) to the primers 519R and 63F. All the PCR results were purified by AMPure bead calibration method and were sequenced using the GS-20 (Genome Sequencer 20) (Roche, Basel, Switzerland) at the Plateforme d'Analyses Génomiques (Université Laval, Québec, Canada).

Table 2 PCR primers used for the double-nested PCR.

Primer	Sequence	Reference
1389R	5′- ACGGGCGGTGTGTACAAG-3′	Marchesi *et al.* (1998)
907R	5′-CCGTCAATTCCTTTRAGTTT-3′	Lane *et al.* (1985)
519R	5′-GWATTACCGCGGCKGCTG-3′	Turner *et al.* (1999)
9F	5′-GAGTTTGATCCTGGCTCAG-3′	Yoon *et al.* (1998)
23F	5′-TGCAGAYCTGGTYGATYCTGCC-3′	Burggraf *et al.* (1991)
63F	5′-CAGGCCTAACACATGCAAGTC-3′	Marchesi *et al.* (1998)

The double-nested PCR consists in three successive amplification steps conducted with three different primer pairs: 1389R-9F, 907R-23F and 519R-63F, respectively.

Amplicon analysis

First, CLC GENOMICS WORKBENCH 3.1 (CLC Bio, Aarhus, Denmark CLC work bench BIO®) was used to trim

sequences for quality and remove primer sequences and tags (minimum average quality score: 35 for a window of 50, number of differences to the primer sequence = 0, maximum number of differences to the barcode sequence = 0, number of ambiguous base calls = 0, maximum homopolymer length = 8). Second, preprocessing and analysis were carried out with the microbial ecology community software MOTHUR (version 1.22.2) (Schloss *et al.*, 2009) following the protocol of Costello stool analysis (http://www.mothur.org/wiki/Costello_stool_analysis). This allowed identifying and deleting chimeras, and removing smaller sequences that were either smaller than 300 base pairs or that contained pyrosequencing errors. Among the three analytical options available in MOTHUR, the OTU-based analysis protocol was used. We generated alpha diversity results, defined by the diversity in an individual fish kidney sample, and the richness estimators (Mc Cure *et al.*, 2002). The richness or number of species in an individual sample was measured using two indexes: the Chao index was used as the richness estimator and the Simpson index was used as the diversity index (Magurran, 2003). The Chao index is the simplest richness index based on the number of rare species (Magurran, 2003). The Simpson index measures the probability that two randomly selected individuals belong to the same taxa. Consequently, a higher Simpson index value is correlated with a lower diversity (Peet, 1974; Sepkoski, 1988). The OTU-based analysis using MOTHUR was also used for taxonomic identification (using 98% bootstrap score). Taxonomic identification was also performed using the RIBOSOMAL DATABASE PROJECT (using the maximal criterion of 95% bootstrap score available in this method) (Maidak *et al.*, 2001). Unweighted UniFrac tests were performed with the phylotype-based analysis using MOTHUR. Finally, putative fish pathogen bacterial genera were identified according to Austin and Austin (2007). Furthermore, the species of selected putative pathogen genera were investigated using the BLAST algorithm (Altschul *et al.*, 1997). For each putatively pathogenic genus that we described in the MS, we pooled sequencing from all populations and individuals in the study. We considered the top blast hit to be the bacterial species for that sequence. Then, for the globally most abundant species, we performed a literature search to determine whether there are indications of pathogenicity in fishes. We restricted subsequent pathogen analyses to these genera.

Statistical analyses

We constructed a matrix containing the number of bacterial sequences for each bacterial genus in each fish sample from the MOTHUR taxonomy file (stool.final.an.0.02.cons.taxonomy). This matrix was used to perform a principal component analysis per rank (PCA per rank) (Baxter, 1995) using PC-ORD (Mc Cure *et al.*, 2002). Samples were ranked as a function of the number of sequences found for each OTU. Nonmetric Multidimensional Scaling (NMS) analysis was not used in this case because the final stress was above recommended interpretable range (final stress = 25). Therefore, ranked-based PCA was preferred to NMS. As absolute abundance may be influenced by sequence-specific fidelity in the double-nested PCR method, we also performed a nonparametric ranking method of abundance using the METASTATS software (White *et al.*, 2009) to detect differentially abundant OTUs between dwarf and normal whitefish. The OTU by sample abundance matrix was also used for this analysis with standard parameters (p value ≤ 0.05 and number of permutations = 1000).

To determine whether there were statistically significant differences in the proportion of infected fish among lakes and between forms, we used a generalized linear model (GLM) with a binomial family followed by an ANOVA with lakes and forms as factors and also including their interaction. We then used the same GLM procedure to test for differences in the proportions of putative pathogenic bacteria between forms within and among lakes. Then, a GLM with Gaussian family was used to test for differences in both the Simpson and Chao indices and again between forms and among lakes. The Levene test was applied to test for differences in interindividual variance of the Simpson and Chao indices between forms, among lakes and their interaction (Snedecor & Cochran, 1980).

Results

The presence of bacteria was detected in kidneys of 52.6% (133 infected samples) of the analysed whitefish (Table 1). The vast majority of samples produced consistent results between the two independent double-nested PCRs and did not require a tiebreaker (83.6%). There were no overall significant differences in infection levels among lakes (GLM, $P_{\text{Lakes}} = 0.16$) or between forms (GLM, $P_{\text{Forms}} = 0.33$), but there was a significant interaction of the lake and form terms (GLM, $P_{\text{Lakes*Forms}} = 5.5\text{e-}5$) (Table 3). The level of infection in the lakes was 46.6%, 54.0%, 66.1%, 46.2% and 48.0% for Cliff, East, Indian, Témiscouata and Webster lakes, respectively. In East and Témiscouata lakes, the dwarf whitefish infection rate was significantly higher than that of normal whitefish (GLM, $P_{\text{East}} = 0.02$ and $P_{\text{Témiscouata}} = 0.04$), whereas, in Indian Pond, the normal whitefish infection rate was higher than that of dwarf whitefish (GLM, $P_{\text{Indian}} = 0.001$). In Cliff and Webster, the infection rates were similar between forms (GLM, $P_{\text{Cliff}} = 0.330$ and $P_{\text{Webster}} = 0.571$). Therefore, the infection rate in these samples was not globally influenced by a lake effect or form effect separately.

Table 3 Statistic values on the number of infected and noninfected whitefish for lake according the results of double-nested PCR.

	Statistic Tests		
Effect	χ^2	*Z* value	*P* value
GLM + ANOVA			
Lakes	6.6018	–	0.1585
Forms	0.9552	–	0.3284
Lakes*forms	25.7886	–	5.49E-05
GLM			
Cliff	–	0.974	0.3303
Est	–	−2.249	0.0245
Indian	–	3.212	0.0013
Témiscouata	–	−1.97	0.0489
Webster	–	0.566	0.5717

Three effects were tested thanks to a GLM followed by an ANOVA to determine whether there were statistically significant differences in fish infection rate among lakes, between forms and their interaction term. A statistically significant difference between forms within lake was also tested with a GLM.

A total of 79 146 reads were obtained, after trimming, for the entire data set of 133 infected samples. Four samples only containing chimeras or very few sequences were eliminated by CLC Genomics Workbench 3.1 and the MOTHUR preprocessing. Six other samples were not sequenced because their bar-coded MID-tags (Multiplex identifiers), MID 20 and MID 21, did not work, leaving 123 samples. Detailed results for the quality of this 454 pyrosequencing experiment are given in Boutin *et al.* (2012). In brief, the Good's coverage estimator for the set of the lakes was 95.4%, a low occurrence (3%) of chimeric amplicons was observed, 77% of the data set was composed of sequences longer than 300 bases, Chao's curves indicated a deeply and a representative sequencing, and a weighted UniFrac test was carried out on a fragment of our data highlighting a large variation of bacterial communities diversity, which indicates that bias caused by preferential amplification or primer selection during the double-nested PCR, which could obscure initial abundance, does not affect our interpretation.

According to the Simpson index, the bacterial diversity in dwarf whitefish was significantly lower than that observed in normal whitefish diversity in all lakes (GLM, $P_{forms} = 0.01$) (Fig. 2). In contrast, there was no significant difference in bacterial diversity among lakes (GLM, $P_{Lakes} = 0.16$). Similar results were found using the Chao index (GLM, $P_{forms} = 0.04$ and $P_{lakes} = 0.11$) (Fig. S1). Thus, the levels of bacterial diversity between dwarf and normal whitefish were parallel across all lakes. Also, the interindividual variance in diversity measured by the Simpson index was significantly higher in all dwarf whitefish populations compared with normal whitefish samples (Levene Test,

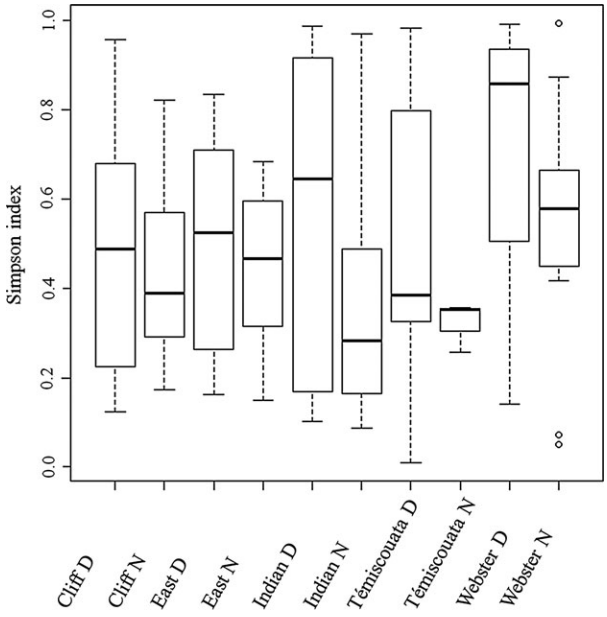

Fig. 2 Plot of bacterial diversity estimated with the Simpson index for all ten populations. D: dwarf whitefish, and N: normal whitefish. The Simpson index is inversely correlated with bacterial diversity. Lower Simpson indices thus mean higher diversity.

$P_{forms} = 0.04$). However, there was no difference in interindividual variance in diversity between forms based on the Chao index (Levene Test, $P_{forms} = 0.80$).

Three phyla were predominantly represented across all samples: *Proteobacteria*, *Actinobacteria* and *Firmicutes* (Fig. 3). Different bacterial communities at the genus level between dwarf and normal whitefish were observed

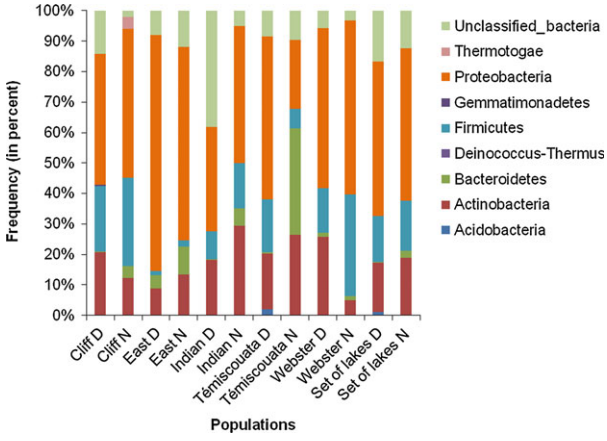

Fig. 3 Relative abundance of phyla members found in kidney whitefish bacterial community for dwarf and normal whitefish in each lake. This taxonomy was constructed with the RIBOSOMAL DATABASE PROJECT (RDP) with a confidence threshold at 95%. D: dwarf whitefish, N: normal whitefish.

within lakes (unweighted UniFrac, $P_{\text{forms Cliff}} < 0.001$, $P_{\text{forms East}} < 0.001$, $P_{\text{forms Indian}} < 0.001$, $P_{\text{forms Témiscouata}} < 0.001$, $P_{\text{forms Webster}} < 0.001$) and among lakes (unweighted UniFrac, $P_{\text{forms all the lakes}} < 0.001$). These differences were also highlighted for both forms among lakes (unweighted UniFrac, $P_{\text{Lakes}} < 0.001$), and the interaction term between lakes and forms was also significant (unweighted UniFrac, $P_{\text{Lakes*forms}} < 0.001$). A total of 579 genera were detected (Table S1), the most frequent ones being (by number of whitefish individuals infected by the same genus): *Propionibacterium*, *Sphingomonas*, *Acinetobacter*, *Clostridium*, *Methylobacterium*, *Pseudomonas*, *Microcella*, *Kocuria*, *Staphylococcus* and *Polynucleobacter* (Fig. S2 and Table S2). The most abundant genera (quantity of reads of the genus present in all whitefish individuals) were *Propionibacterium*, *Sphingomonas*, *Clostridium*, *Acinetobacter*, *Microcella*, *Altererythrobacter*, *Nitrosococcus*, *Croceicoccus*, *Sarcina* and *Polynucleobacter* (Fig. S3 and Table S3).

Two different PCAs per rank were performed (Fig. 4). The first two principal components explained 20.0% of the variance for the analysis comparing all five dwarf whitefish populations for all bacterial genera and 23.5% of the variance for all five normal whitefish

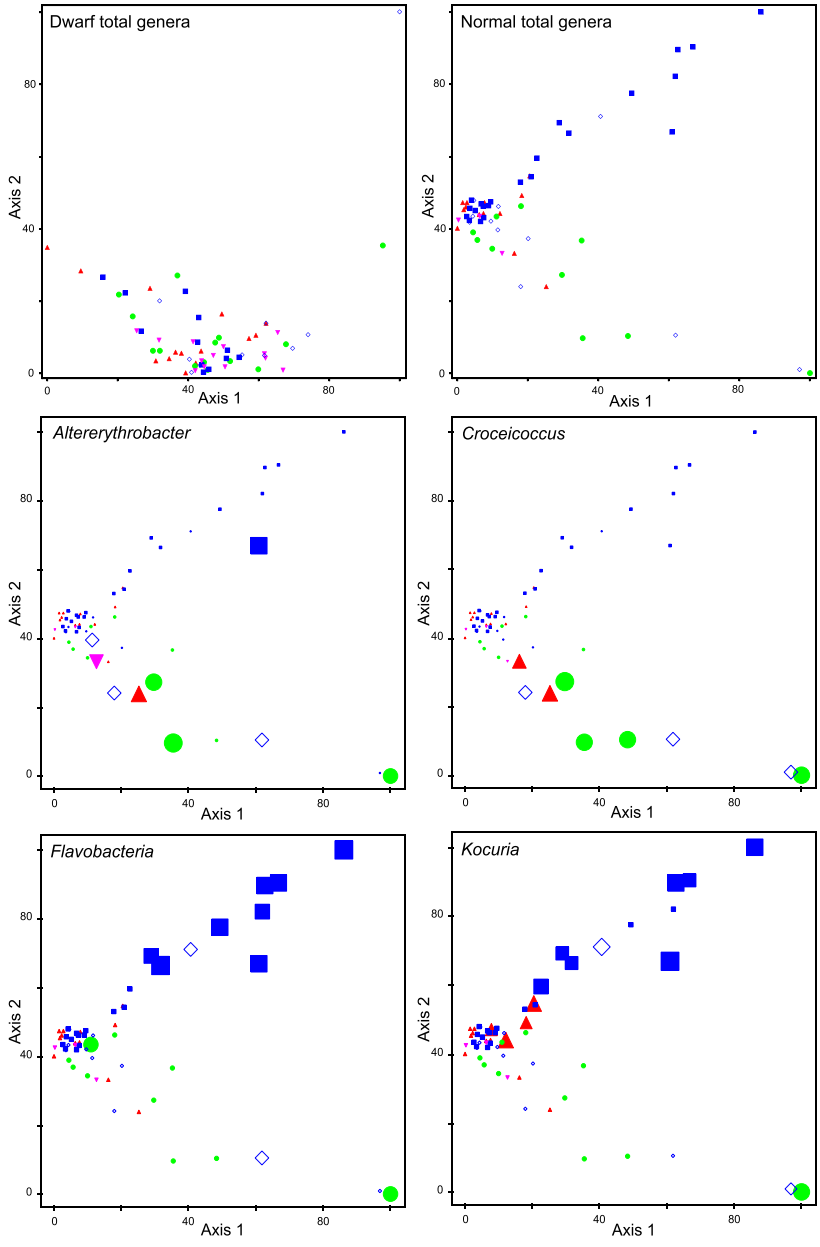

Fig. 4 Principal components analysis per rank (PCA per rank) of Operational Taxonomic Units (OTU) present in whitefish kidney differentiated by lake. Each lake analysed is represented by symbols: Cliff Lake (red triangle), East Lake (green dot), Indian Lake (blue square), Témiscouata Lake (pink inverted triangle) and Webster Lake (blue diamond). The size of the symbol is proportional to OTU abundance. Two different PCAs per rank analyses were performed. The first PCA per rank was performed only with genera present in the dwarf form, named dwarf total genera. The second PCA per rank was performed only with genera present in the normal form, named normal total genera. This last PCA per rank was used to generate the four supplementary PCA per rank displaying only the *Altererythrobacter*, *Croceicoccus*, *Flavobacteria* and *Kocuria* genera. The PCAs per rank were produced by PC-ORD. Each PCA per rank had ten axes that were interpretable, but only axes one and two were used and explained a combined variance of 20% for dwarf total genera PCA per rank and 23.47% for normal total genera PCA per rank.

populations for all bacterial genera. The PCA per rank on all the bacterial genera found in the normal individuals was also used in PC-ORD to generate four figures showing only the *Altererythrobacter*, *Croceicoccus*, *Flavobacteria* and *Kocuria* genera using the same axes calculated for all the genera. Comparing dwarf whitefish populations for these genera revealed no difference between lakes (results not shown), which corroborates the absence of differentiation when considering all bacterial genera. In contrast, results for normal whitefish revealed a more pronounced differentiation between lakes although with some overlap. In particular, Indian Pond and East Lake bacterial communities were the most distinct and clustered in the form of elliptical clouds whereby individual bacterial communities in Indian Pond were more spread along axis 2, whereas those from East Lake were more spread along axis 1. Cliff and Webster bacterial communities overlapped with those observed in Indian Pond and East Lakes, whereas Témiscouata Lake displayed a pattern intermediate between Indian and East Lakes. These differences between the lakes were mainly explained by the genera *Altererythrobacter* and *Croceicoccus* that predominated in normal whitefish from East, Cliff and Webster Lakes. In contrast, *Flavobacteria* and *Kocuria* were the most abundant in some normal whitefish individuals from Indian pond but also of Cliff and Webster lakes.

Among the bacterial genera detected, 10 putative pathogenic bacterial genera were identified and further considered in our analysis because they were both identified as pathogenic genera in Austin and Austin (2007), and the most abundant species in our BLAST analysis were known fish pathogens: *Acinetobacter*, *Aeromonas*, *Chryseobacterium*, *Corynebacterium*, *Flavobacterium*, *Janthinobacterium*, *Micrococcus*, *Pseudomonas*, *Shewanella* and *Staphylococcus*. Seven other putative pathogen genera (*Citrobacter*, *Clostridium*, *Mycobacterium*, *Renibacterium*, *Enterococcus*, *Oxalobacter* and *Streptococcus*) were also detected, but not considered as pathogen in our analysis because their frequency and their abundance were too low or fish pathogen species were not highlighted. Among these 10 putative pathogenic bacterial genera, 18 known fish pathogenic species were found (Table 4). Also in these 10 genera, we found 14 other pathogenic bacteria which are known to infect other species (human, drosophila and plants), but not fish and finally 11 species which had no documented pathogenicity (Table S4).

A total of 73% of the 133 whitefish samples for which bacteria were detected in kidney were infected by at least one putative pathogen (Table 5). There was no significant difference in the occurrence of the pathogenic bacteria between dwarf and normal whitefish within each lake (GLM, $P_{\text{Cliff}} = 0.92$, $P_{\text{East}} = 0.29$, GLM, $P_{\text{Indian}} = 0.15$, $P_{\text{Témiscouata}} = 0.34$, $P_{\text{Webster}} = 0.86$), neither was there significant forms effect or lake effect (GLM, $P_{\text{forms}} = 0.92$, GLM, $P_{\text{Lakes}} = 0.80$).

Within the putative pathogen communities, some taxa were found only in one of the two forms in a

Table 4 Eighteen pathogenic species of 10 identified putative pathogen genera which were identified using the BLAST algorithm.

Species	Reference(s)	Evidence
Acinetobacter junii	Navarrete *et al.* (2009)	Caused outbreak in salmonid fish aquaculture
Aeromonas hydrophila	Dopazo *et al.* (1988)	
Aeromonas salmonicida	Langefors *et al.* (2001)	
Corynebacterium xerosis	Austin *et al.* (1983)	
Flavobacterium psychrophilum	Nematollahi *et al.* (2003)	
Micrococcus luteus	Austin & Stobie (1992)	
Pseudomonas fluorescens	Bruno *et al.* (2013)	
Pseudomonas putida	Altinok *et al.* (2006)	
Pseudomonas chlororaphis	Hatai *et al.* (1975)	Caused outbreak in salmonids
Staphylococcus epidermidis	Gil *et al.* (2000)	Caused outbreak in aquaculture
Shewanella algae	Hau & Gralnick (2007)	
Acinetobacter lwoffii	Li *et al.* (2006)	Experimental fish infections
Aeromonas salmonicida	Langefors *et al.* (2001)	
Chryseobacterium shigense	Zamora *et al.* (2012)	Isolated from diseased salmonid
Flavobacterium hydatis	Bernadet *et al*. (1996); Ekman (2003)	
Flavobacterium succinicans	Bernadet et al. (1996); Ekman (2003)	
Janthinobacterium lividum	Austin *et al*. (1992b)	
Staphylococcus warneri	Gil *et al.* (2000)	
Aeromonas sobria	Olivier *et al.* (1981)	Classified as enterotoxigenic

Table 5 Number of putative pathogens and opportunistic bacteria in whitefish according to 454 sequencing results.

Lakes	Species	Number of fish infected by pathogen bacteria	Number of fish infected by only opportunistic bacteria	Total	% Pathogen infected	% Opportunistic infected
Cliff	D	11	3	14	79	21
	N	10	3	13	77	23
	T	21	6	27	78	22
East	D	10	4	14	71	29
	N	9	1	10	90	10
	T	19	5	24	79	21
Indian	D	7	5	12	58	42
	N	18	4	22	82	18
	T	25	9	34	74	26
Témiscouata	D	9	5	14	64	36
	N	1	2	3	33	67
	T	10	7	17	59	41
Webster	D	6	4	10	60	40
	N	7	4	11	64	36
	T	13	8	21	62	38
Total	D	43	21	64	67	33
	N	45	14	59	76	24
	T	95	35	130	73	27

D, dwarf whitefish; N, normal whitefish; and T, total number of whitefish.
The putative pathogen bacteria were determined according to Austin & Austin (2007).

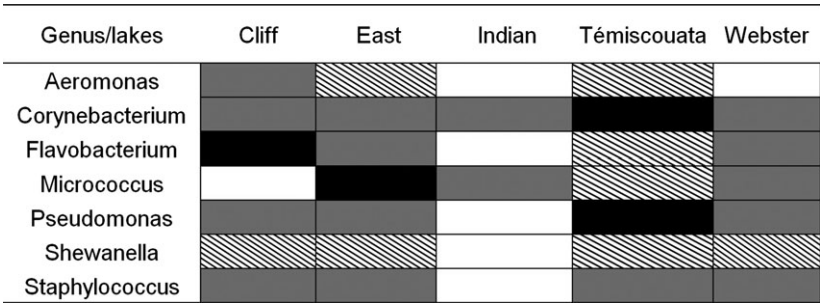

Fig. 5 Differences between dwarf and normal whitefish kidney in selected pathogenic genera. White squares represent putative pathogenic genera only found in normal whitefish and black those only found in dwarf whitefish in each lake. Grey refers to pathogenic genera found in both forms, and hashing refers the genera not found in either form in that lake.

given lake, but this varied between lakes (Fig. 5). Pathogens of the genera *Aeromonas*, *Flavobacterium*, *Micrococcus*, *Pseudomonas*, *Shewanella* and *Staphylococcus* significantly infected normal whitefish. In particular, normal whitefish from Indian Pond were infected by pathogens belonging to five genera: *Aeromonas*, *Flavobacterium*, *Pseudomonas*, *Shewanella* and *Staphylococcus*. Bacterial communities of the normal whitefish of East and Webster were characterized by *Micrococcus* and *Aeromonas*, respectively. For Cliff, dwarf whitefish were infected by *Flavobacterium* and normal whitefish were infected by *Micrococcus*. Finally, only Témiscouata dwarf whitefish were infected by *Corynebacterium* and *Pseudomonas*.

Discussion

The general goal of this study was to investigate the kidney microbiome of sympatric dwarf and normal whitefish pairs to reach three objectives: (i) to test for differences in kidney bacterial communities between dwarf and normal whitefish from the same lake, (ii) to test the hypothesis of higher bacterial taxonomic diversity in normal than dwarf whitefish and (iii) to test for the occurrence of parallelism in those patterns.

To this end, we analysed 253 whitefish samples from five lakes and identified an unprecedented number of bacterial genera in kidneys, including 579 different genera among which 10 were pathogenic bacterial genera

that comprised 18 known pathogenic species. This contrasts with previous studies that considered kidneys to be a sterile organ in healthy fish (Goldschmidt-Clermont *et al.*, 2008; Dionne *et al.*, 2009; Salgado-Miranda *et al.*, 2010). Of course, one must consider that contamination could partly explain the level of bacterial diversity observed in whitefish kidneys. However, as mentioned in Materials & Methods, meticulous care was taken to avoid contamination by working in sterile conditions and avoiding cross-contamination between individuals. Also, the composition of kidney bacterial communities varied importantly between whitefish individuals, which is not compatible with contamination from the same manipulations in the same working environment. In addition, no PCR product was detected for 47% of the analysed individuals, and on every plate there were both positive and negative reactions. The total absence of amplification for 47% of all individuals analysed thus serves as additional negative controls. Thus, although it cannot be entirely ruled out, contamination during the manipulations is very unlikely to explain the main patterns we documented. This rather suggests that bacteria may pass to kidneys through by abrasions, by injuries (Austin, 2006; Larsen *et al.*, 2013) or by the digestive tract which is one of the major infection routes in fish (Ringø & Olsen, 1999). From there, bacterial translocation across the epithelia of the intestine could open a possible route for kidney infection (Hart *et al.*, 1988; Jutfelt *et al.*, 2006; Knudsen *et al.*, 2008), thus offering a possible explanation for the much higher occurrence of bacteria in kidney than previously assumed.

Variation of bacterial communities in whitefish kidney within lake

Regarding our first objective, we observed an overall significant lake effect on the composition of the different taxa. This general pattern can hypothetically be explained by the geographical distance and limited hydrological connectivity between the lakes, which could create different environmental conditions surrounding the lakes and lead to the presence of different bacteria (Lindström *et al.*, 2005; Reche *et al.*, 2005; Laplante *et al.*, 2013). There was also a significant difference for form effect meaning that bacterial community compositions between dwarf and normal whitefish is distinct within and among the lakes. In fact, distinct bacterial communities may exist within a lake because the bacterial diversity and bacterial concentration decrease as a function of depth and due to differences in temperature, oxygen concentration and luminance (Ovreås *et al.*, 1997; Bosshard *et al.*, 2000; Koizumi *et al.*, 2004; Landry *et al.*, 2007). However, although whitefish forms occupy different depths in the water column, there is no barrier preventing contact between dwarf and normal whitefish (Bernatchez *et al.*, 1999). Even if both dwarf and normal whitefish could come in contact with all of these bacterial communities, the probabilities of infection by the different taxa are probably different for the forms. In the absence of other similar studies on bacterial communities in natural populations of freshwater fishes, we cannot compare our observation of an overall relatively modest difference in bacterial communities between sympatric whitefish species pairs to other fish species. However, previous studies on eukaryote parasite communities revealed somewhat similar patterns with respect to benthic diet types having more diverse parasites, for instance between closely related sympatric cichlid species (*Tropheus* sp. Blais *et al.*, 2007), sympatric incipient species of European whitefish (*C. lavaretus;* Knudsen *et al.*, 2003) and threespine sticklebacks [*Gasterosteus aculeatus*; (MacColl, 2009; Natsopoulou *et al.*, 2012)].

Bacterial diversity and whitefish diet

We observed that normal whitefish were characterized in all lakes by a greater average taxonomic diversity compared with dwarf whitefish, thus translating into parallelism in difference of taxonomic diversity. This apparently runs counter to the expectation of abundance of bacteria in the water column, which is expected to decrease with depth (Ovreås *et al.*, 1997; Bosshard *et al.*, 2000; Koizumi *et al.*, 2004; Landry *et al.*, 2007). However, assuming that bacterial transmission through diet may influence the bacterial community observed in kidney, this could perhaps be explained by difference in prey availability among lakes. Indeed, dwarf whitefish (and limnetic whitefish in general) feed almost exclusively on zooplankton, most often on the same taxa in different lakes (Bodaly, 1979; Bernatchez *et al.*, 1999). In contrast, normal whitefish are more generalists and feed on more diverse prey items including zoobenthos, molluscs and fish prey, of which the composition varies between lakes and throughout the year (Bernatchez *et al.*, 1999; Bernatchez, 2004). Another important observation is that we found no differences in kidney bacterial communities between dwarf whitefish from different lakes in our PCA (all dwarf lakes in the same cloud; Fig. 4), which can be interpreted as mirroring evidence for parallelism in the composition of dwarf whitefish microbiota. In sharp contrast, we observed pronounced nonparallel differences in the kidney microbiota of normal whitefish from different lakes. Generally speaking, the zooplankton communities are more similar compared with the benthic and fish prey communities in the studied lakes (Landry *et al.*, 2007; Landry & Bernatchez, 2010). Given that different diets may allow contact with different microbial communities (Gatesoupe & Lésel, 1998), it is thus plausible that parallelism in zooplankton community translates into parallelism in bacterial community of dwarf whitefish, whereas nonparallelism in benthic and fish prey community translates into nonparallelism in bacterial community of normal whitefish.

This again suggests that the diet composition may impact the diversity of the bacterial community found in a given host (Ringø *et al.*, 2006; Zhou *et al.*, 2012). It thus appears that variation in patterns of zooplankton and benthic prey communities across lakes at least partly explain the differences in patterns of bacterial diversity and community composition observed between dwarf and normal whitefish.

Implication of patterns of parallelism and nonparallelism in whitefish diversification

In regard to our third objective, the above results show no strong parallelism in dwarf-normal bacterial community differences or infection rates. Yet, parallelism was observed among dwarf whitefish from different lakes, which had very similar microbiotas among lakes compared with normal whitefish which had wildly different microbiotas among lakes. Indeed, we systematically observed lower bacterial community diversity found in dwarf whitefish across all five lakes. This difference in bacterial diversity could potentially be explained by differences in prey diversity available to both forms. A nonexclusive hypothetical explanation for the observed parallel differences in bacterial diversity between dwarf and normal whitefish pertains to the trade-off made by these forms between resources devoted to growth and those devoted to the immune response. Thus, previous studies in the wild and common rearing conditions showed that normal whitefish have a genetically based faster growth rate than dwarf whitefish (Bernatchez, 2004; Rogers & Bernatchez, 2007). Also, this parallel growth differential is accompanied by parallelism in patterns of gene expression whereby normal whitefish showed significant overexpression of genes involved with growth (protein synthesis, cell growth) (St-Cyr *et al.*, 2008). In contrast, the limnetic life history of the dwarf whitefish requires much energy be expended for constant swimming, for feeding on zooplankton and avoiding predators. Thereby, genes associated with metabolism, muscle contraction and detoxification were overexpressed in dwarf whitefish. According to Matarese and La Cava (2004), many genes have dual functions in both immunity and metabolism. Thus, dwarf whitefish could have a more efficient immune system than the normal whitefish. Besides, genes specific to the immune system are overexpressed in dwarf whitefish (St-Cyr *et al.*, 2008; Jeukens *et al.*, 2010). On the other hand, a recent study revealed no evidence for parallelism at the adaptive immune system in MHCIIβ gene diversity among whitefish sympatric species pairs, suggesting a minor role of pathogenic bacteria in the parallel evolution of whitefish species pairs (Pavey *et al.*, 2013). In that study, specific lake effects associated with different environments appeared more important in explaining MHC variation in this system. However, a parallel evolution of innate immune system, which eliminates bacteria in a nonspecific manner, is possible (Janeway, 2001). Indeed, some of the overexpressed genes documented by St-Cyr *et al.* (2008) and Jeukens *et al.* (2010) were complement C4, complement factor H1 protein, C1q-like adipose-specific protein, implicated in complement system belonging the innate immune system and MHC class I antigen belonging the innate and the specific immune system. This increased expression of genes of the innate immune system could be an adaptation of the dwarf whitefish exposure to pathogen (Goetz *et al.*, 2010; Jeukens *et al.*, 2010), although this remains to be investigated further. Finally, there is another nonexclusive interpretation regarding to the observed dwarf whitefish microbiota parallelism and the trade-off made by these forms between resources devoted to growth and those devoted to the immune response. Certain host genotypes may have the capacity to recruit specific bacterial strains (McKnite *et al.*, 2012). This capacity was recently demonstrated in another salmonidae, the Brook Trout (S. Boutin, C. Sauvage, L. Bernatchez, C. Audet & N. Derome, unpublished), in which three QTLs were related to the relative abundance of three bacterial strains associated with skin and intestine tissue, all of them being documented to synthesize antimicrobial compounds. Therefore, it remains possible that dwarf form may preferentially recruit bacterial strains exerting strong antimicrobial properties to enhance its overall immune capacity against opportunistic pathogens. Although this would likely actively occur in the intestine, that may result in less bacteria infecting the kidney. As a trade-off, the bacterial strains that enhance energetic conversion efficiency may be disfavoured.

Nevertheless, testing for patterns of parallelism at many different levels may help identifying the main factors that are at play in driving the process of parallel adaptive divergence and ultimately reproductive isolation. In the case of whitefish, this strategy has clearly been efficient in achieving this goal. For instance, comparing the limnological settings of each lake allowed identifying the main biotic and abiotic factors (namely the level of oxygen depletion, lake depth, prey size distribution and biomass) that have most likely played a role in the level of adaptive divergence observed between dwarf and normal whitefish from different lakes (Landry *et al.*, 2007; Landry & Bernatchez, 2010). Parallelism in gene expression patterns has also led to identifying the main physiological functions involved in the life history trade-off between growth and survival observed in normal and dwarf whitefish, respectively (St-Cyr *et al.*, 2008). In a recent investigation of respiratory, circulatory and neurological traits across lake whitefish species pairs, Evans *et al.* (2013) found that in each of the species pairs, normal whitefish exhibited larger body size standardized gills compared to dwarf whitefish, a pattern that is suggestive of a common

ecological driver of gill size divergence. Evans *et al.* (2013) also observed a trend towards larger hearts in dwarfs, the more active species of the two species, whereas brain size varied exclusively between the lakes but independent of form. In another recent study, Evans *et al.* (2012) tested for the parallel divergence of traits involved in oxygen transport in dwarf and normal lake whitefish. They found parallel differences in red blood cell morphology between the forms. Taken together, the results of these studies along with the present study imply that the diversification of whitefish has been driven both by parallel and nonparallel ecological conditions across lakes.

Whitefish microbiota

As mentioned above, it has traditionally been assumed that kidney tissue is free of bacteria in healthy fish (Nieto *et al.*, 1984; Cahill, 1990; Goldschmidt-Clermont *et al.*, 2008). However, an increasing number of studies have reported bacteria in healthy fish kidney. According to Dionne *et al.* (2009) and Evans and Neff (2009), 12.1% of juvenile salmon and between 9 and 29% of Chinook Salmon (*Oncorhynchus tshawytscha*) were infected by 10 and 27 genera, respectively. Three other studies on liver and kidney of wild freshwater fish, turbot and five different salmonids, respectively, identified 8, 6 and 19 bacteria genera (Trust & Sparrow, 1974; Toranzo *et al.*, 1993; Sousa & Silva-Souza, 2001). Furthermore, *Flavobacterium psychrophilum* infected kidney of 10.4% of Atlantic Salmon from the Baltic Sea (Ekman *et al.*, 1999). In this study, 579 different genera representing nine phyla were found in kidney from 133 apparently healthy fish. This is a substantial diversity for an internal organ, particularly if one compares with the 10 genera and 27 genera, respectively, reported in two previous studies performed on the kidney microbial community in salmonids (Dionne *et al.*, 2009; Evans & Neff, 2009). However, data from these studies were obtained by culture and DGGE, whereas a double-nested PCR is much more sensitive to detect and specifically amplify DNA than PCR, nested PCR and culture methods (Boutin *et al.*, 2012). Thus, the traditional assumption of kidney sterility may have simply been a factor of the limitations of the techniques available, as differing sensitivity even among modern techniques could largely explain the difference between the results of the present and the previous studies on the kidney bacterial diversity. Moreover, to our knowledge, a single study was carried out on wild freshwater fish with a Sanger DNA sequencing approach where they found 525 OTUs classified in 8 phyla on three zebrafish guts (Roeselers *et al.*, 2011). Other than these studies, some microbiome studies in fish have analysed the digestive system and the skin with pyrosequencing of bar-coded 16S rRNA gene amplicons. Twelve and nine phyla were present in faeces of eight catfish individuals (*Panaque* sp.) and six marine fish, respectively (Di Maiuta *et al.*, 2013; Larsen *et al.*, 2013), and 6058 OTUs were obtained from the seven grass carp (*Ctenopharyngodon idellus*) (Wu *et al.*, 2012). In contrast, the Brook Trout (*Salvelinus fontinalis*) skin of 121 individuals had 16 904 genera distributed among 21 phyla (Boutin *et al.*, 2013). Thus, while high relative to the *a priori* expectations, the kidney diversity appears low compared with those recorded in both gut and skin microbiota. As mentioned in the introduction, kidney of teleost fish, which include whitefish, is known to play several functions, including immune function (Danguy *et al.*, 2011). The anterior kidneys are composed almost exclusively by hematopoietic, lymphoid tissues and melanomacrophage centres where antibody-covered particles arrived through blood, including bacteria, are eliminated by phagocytosis (Press & Evensen, 1999; Agius & Roberts, 2003). It is thus possible that the double-nested PCR is sensitive enough to detect and amplify the bacterial DNA from cells ongoing the process of elimination phagocytosis (Frank, 2002; Pavey *et al.*, 2013).

Another factor that may explain the high level of kidney bacterial diversity in this study is that our sampling was carried out at the end of June and the beginning of July, when water temperatures are relatively high (Mackay & Kalff, 1969; Brunskill & Schindler, 1971). Generally speaking, bacteria abundances increase when water temperatures reach a certain threshold (Larsen *et al.*, 2004; Dionne *et al.*, 2009). For example, outbreaks of *Vibriosis,* a bacteria of the aquatic environment, happen when water temperatures exceed 15 °C (Larsen & Mellergaard, 1981). Studies showed than the immune system efficacy could decline in the presence of high bacteria concentrations in water (Buras *et al.*, 1985; Cahill, 1990). In addition, there may be a reduction in the efficiency of fish immune system as a result of stress factors such as nutritional deficiencies, poorer water quality, overcrowding, parasitism and temperature changes (Cahill, 1990). As such, whitefish may have been more susceptible to the colonization of internal organs from the external environment during the period at which they were sampled. Accordingly, the majority of genera identified in this study were associated with environmental water bacteria predominantly represented by *Proteobacteria*, *Actinobacteria* and *Firmicutes*. *Proteobacteria* and *Actinobacteria* were also predominantly present in all freshwater sites on a study analysing the typical freshwater bacteria (Zwart *et al.*, 2002). Bacteria associated with the aqueous environment such as *Sphingomonas*, *Methylobacterium*, *Lactobacillus*, *Roseomonas*, *Arthrobacter* or *Burkholderia* were also found in Dionne *et al.* (2009) and Evans and Neff (2009).

Among 579 different genera found in our whitefish kidney, 14 genera are also described in the study of Evans and Neff (2009) on Chinook Salmon, and 6 genera in that of Dionne *et al.* (2009) on Atlantic Salmon. Both of these studies were conducted on young anadromous

fish during their freshwater phase. Moreover, the genera *Acinetobacter*, *Aeromonas*, *Citrobacter*, *Pseudomonas and Staphylococcus* that were observed in this study constitute the dominant microbiote of the adult freshwater fish digestive tract (Austin, 2006). Also, *Enterobacter*, *Escherichia*, *Klebsiella*, *Serratia*, *Bacteroides*, *Bacillus* and *Propionibacterium*, which are not considered as pathogens, were also found in whitefish kidney and previously reported in freshwater fish digestive tracts in general (Austin, 2006). Most of the other bacteria that we found in our study are ubiquitous and can be present in soil, water or plants, for instance *Methylobacterium*, *Paracoccus*, *Stenotrophomonas, Bradyrhizobium*, *Altererythrobacter, Croceicoccus*, *Kocuria* or *Burkholderia* (Kaneko *et al.*, 2000; Lidstrom & Chistoserdova, 2002; Kelly *et al.*, 2006; Kwon *et al.*, 2007; Ryan *et al.*, 2009; Xu *et al.*, 2009; Bontemps *et al.*, 2010; Savini *et al.*, 2010).

Among all the 579 genera identified, there were 10 putative pathogen genera known to contain pathogenic species according to the literature (Austin & Austin, 2007). Indeed, the most abundant species representing these genera are known fish pathogens as revealed by our literature search. These genera are *Acinetobacter*, *Aeromonas*, *Chryseobacterium*, *Corynebacterium*, *Flavobacterium*, *Janthinobacterium*, *Micrococcus*, *Pseudomonas*, *Shewanella* and *Staphylococcus*. Bacteria of the genus *Acinetobacter* sp., as *Acinetobacter junii*, and *Acinetobacter lwoffii,* can cause the acinetobacter disease which killed 92% of wild Atlantic Salmon in Norway (Roald & Hastein, 1980). *Aeromonas hydrophila, Aeromonas salmonicida* and *Aeromonas sobria* are known to cause septicaemia or ulcer disease in many freshwater fishes (Austin & Austin, 2007). Some species of *Chryseobacterium* are considered as potentially emerging pathogens, and *C. shigense* was isolated from kidney of diseased rainbow trout (Zamora *et al.*, 2012). *Corynebacterium xerosis* is implicated in bacterial kidney disease (BKD) in salmonid fish (Austin *et al.*, 1983). *Flavobacterium succinicans* and *Flavobacterium hydatis* have been isolated from diseased fresh-water fish (Bernadet *et al.*, 1996). Also *Flavobacterium psychrophilum* is an agent of bacterial cold-water disease and caused septicaemic disease (Nematollahi *et al.*, 2003). *Janthinobacterium lividum, Micrococcus luteus*, *Pseudomonas putida* and *Pseudomonas fluorescens* caused serious diseases in Rainbow Trout (Sakai *et al.*, 1989; Austin *et al.*, 1992b; Austin & Stobie, 1992; Altinok *et al.*, 2006). *Shewanella algae* were responsible of mass mortality of shellfish and fish (Hau & Gralnick, 2007*)*. Also, many studies have reported infection of salmonids and other freshwater fishes by the genus *Staphylococcus* including *S. warneri* and *S. Epidermidis* (Nieto *et al.*, 1984; Austin *et al.*, 1992a; Gil *et al.*, 2000). The majority of the above studies dealt with infection outbreaks in salmonid aquaculture conditions. Yet, this information provides a basis to identify potential pathogenic bacteria that may potentially affect whitefish in natural conditions (Pavey *et al.*, 2013).

The vast majority of the genera that we identified have not previously been described as comprising pathogenic bacteria (567/579). However, relatively little is known about pathogens in natural aquatic systems (Austin & Austin, 1999). It is also possible that some pathogens were not identified as such because it can be difficult to define the limit between pathogen and opportunistic bacteria and that bacteria have a continuous spectrum of pathogenicity (Casadevall & Pirofski, 1999). A box plot of the Simpson index of only the designated pathogenic bacteria did not reveal the same parallel pattern of diversity between dwarf and normal that we found for all bacteria (Fig. S4; Fig. 2). Therefore, the main difference in infection between forms does not directly imply putative pathogenic bacteria. Thus, it is possible that the infection difference may be due to unknown pathogens or more likely opportunistic bacteria. In fact, the majority of genera could be opportunistic bacteria. This kind of bacteria colonizes tissues after a primary infection when the immune system of the host is impaired (Falkow, 1997; Casadevall & Pirofski, 1999). Some members of the most abundant and the most frequent genera, *Propionibacterium*, *Sphingomonas, Polynucleobacter* and *Ralstonia,* are known to be opportunistic (Laskin & White, 1999; McDowell *et al.*, 2011; Youngblut *et al.*, 2013). This hypothesis of primary infection by at least one putative pathogen and second infection by opportunistic could explain the pattern of bacterial diversity we observed (Table 5). In fact, of the 133 fish that were found to be infected, 73% were infected by at least one putative pathogenic genus, whereas, always under the hypothesis of primary infection, 23% of whitefish would have been infected by an unknown pathogen.

Conclusion

In summary, this study represents the very first attempt to document fish microbiome in natural conditions by means of NGS-based approach. Here, our primary goal was to compare the bacterial communities present in kidney of sympatric species pairs of dwarf and normal whitefish to test whether parallel phenotypic and ecological evolution was accompanied by parallelism in their kidney microbiota, and in particular in the putative bacterial pathogenic community. We found bacteria in the kidney of more than half of the fish tested, and a diversity of bacterial genera well beyond previous descriptions of fish bacterial diversity in kidney, both pathogenic and not. Although there were differences in proportion of infected fish between dwarf and normal whitefish, these were not parallel among lakes. Also, there was similar bacterial diversity between dwarf whitefish from different lakes, providing evidence for parallelism in the bacterial diversity for this form, whereas pronounced differences in the bacterial diversity of normal whitefish from different lakes were found.

However, in accordance with the higher diversity of prey types, normal whitefish kidney tissue consistently had more diverse microbiota, and this pattern was parallel among lakes. This again corroborates the expectations based on lake to lake differences and similarities in the benthic and limnetic prey communities, respectively. It is also possible that the trade-off between the immune system and growth functions implies globally weaker immune defences for faster growing normal whitefish relative to dwarf whitefish. Although MHCIIβ patterns are not parallel among lakes (Pavey *et al.*, 2013), this does not exclude parallel evolution in the innate immune system and other components, which will need to be investigated further in future studies.

Acknowledgments

We thank Brian Boyle for his help with the 454 sequencing and Eric Normandeau for his bioinformatics skills and support. SP was supported through a fellowship from Fonds de la recherche en santé du Québec. We are also grateful for the two anonymous referees for their insightful comments and suggestions. This work was supported by a National Sciences and Engineering Council of Canada (NSERC) Discovery grant and a Canadian Research Chair to LB. It is also a contribution to the research programmes of Québec-Océan and RAQ.

References

Agius, C. & Roberts, R.J. 2003. Melano-macrophage centres and their role in fish pathology. *J. Fish Dis.* **26**: 499–509.

Altinok, I., Kayis, S. & Capkin, E. 2006. *Pseudomonas putida* infection in rainbow trout. *Aquaculture* **261**: 850–855.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., *et al.* 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**: 3389–3402.

Amann, R. & Ludwig, W. 2000. Ribosomal RNA-targeted nucleic acid probes for studies in microbial ecology. *FEMS Microbiol. Rev.* **24**: 555–565.

Arumugam, M., Raes, J., Pelletier, E., Le Paslier, D., Yamada, T., Mende, D.R. *et al.* 2011. Enterotypes of the human gut microbiome. *Nature* **473**: 174–180.

Austin, B. 2006. The bacterial microflora of fish, revised. *Scientific World Journal* **6**: 931–945.

Austin, B.B. & Austin, D.A. 1999. *Bacterial Fish Pathogens: Disease of Farmed and Wild Fish*, 3rd edn. Springer-Verlag Berlin Heidelberg, New York.

Austin, B. & Austin, D.A. 2007. *Bacterial Fish Pathogens: Diseases of Farmed and Wild Fish*, 4th edn. Praxis Publishing Ltd, Chichester, UK.

Austin, B. & Stobie, M. 1992. Recovery of *Micrococcus luteus* and presumptive *Planococcus sp.* from moribund fish during an outbreak of rainbow trout, *Oncorhynchus mykiss* (Walbaum), fry syndrome in England. *J. Fish Dis.* **15**: 203–206.

Austin, B., Embley, T.M. & Goodfellow, M. 1983. Selective isolation of *Renibacterium salmoninarum*. *FEMS Microbiol. Lett.* **17**: 111–114.

Austin, B., Baudet, E. & Stobie, M. 1992a. Inhibition of bacterial fish pathogens by *Tetraselmis suecica*. *J. Fish Dis.* **15**: 55–61.

Austin, B., Gonzalez, C.J., Stobie, M., Curry, J.I. & McLoughlin, M.F. 1992b. Recovery of *Janthinobacterium lividum* from diseased rainbow trout, *Oncorhynchus mykiss* (Walbaum), in Northern Ireland and Scotland. *J. Fish Dis.* **15**: 357–359.

Bäckhed, F., Ley, R.E., Sonnenburg, J.L., Peterson, D.A. & Gordon, J.I. 2005. Host-bacterial mutualism in the human intestine. *Science* **307**: 1915–1920.

Baxter, M.J. 1995. Standardization and transformation in principal component analysis, with applications to archaeometry. *J. Roy. Statist. Soc. Ser. C* **44**: 513–527.

Berg, G., Eberl, L. & Hartmann, A. 2005. The rhizosphere as a reservoir for opportunistic human pathogenic bacteria. *Environ. Microbiol.* **7**: 1673–1685.

Bernadet, J.-F., Segers, P., Vancanneyt, M., Berthe, F., Kersters, K. & Vandamme, P. 1996. Cutting a Gordian Knot: Emended Classification and Description of the Genus *Flavobacterium*, Emended Description of the Family *Flavobacteriaceae*, and Proposal of *Flavobacterium hydatis* nom. nov. (Basonym, Cytophaga aquatilis Strohl and Tait 1978). *Int. J. Syst. Evol. Microbiol.* **46**: 128–148.

Bernatchez, L. 2004. Ecological theory of adaptive radiation: an empirical assessment from coregorine fishes (Salmoniformes). In: *Evolution Illuminated: Salmon and their Relatives A. P. S.* (S. Hendry, ed.), pp. 176–207. Oxford University Press, New York.

Bernatchez, L., Chouinard, A. & Lu, G. 1999. Integrating molecular genetics and ecology in studies of adaptive radiation: Whitefish, Coregonus sp., as a case study. *Biol. J. Linn. Soc.* **68**: 173–194.

Bernatchez, L., Renaut, S., Whiteley, A.R., Derome, N., Jeukens, J., Landry, L. *et al.* 2010. On the origin of species: insights from the ecological genomics of lake whitefish. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **365**: 1783–1800.

Bodaly, R. 1979. Morphological and ecological divergence within the lake whitefish (*Coregonus clupeaformis*) species complex in Yukon Territory. *J. Fish. Board Can.* **36**: 1214–1222.

Bontemps, C., Elliott, G.N., Simon, M.F., Dos Reis Junior, F.B., Gross, E., Lawton, R.C. *et al.* 2010. *Burkholderia* species are ancient symbionts of legumes. *Mol. Ecol.* **19**: 44–52.

Bosshard, P.P., Stettler, R. & Bachofen, R. 2000. Seasonal and spatial community dynamics in the meromictic Lake Cadagno. *Arch. Microbiol.* **174**: 168–174.

Boutin, S., Sevellec, M., Pavey, S.A., Bernatchez, L. & Derome, N. 2012. A fast, highly sensitive double-nested PCR-based method to screen fish immunobiomes. *Mol. Ecol. Resour.* **12**: 1027–1039.

Boutin, S., Bernatchez, L., Audet, C. & Derôme, N. 2013. Network Analysis Highlights Complex Interactions between Pathogen, Host and Commensal Microbiota. *PLoS ONE* **8**: e84772.

Bruno, D.W., Noguera, P.A. & Poppe, T.T. 2013. *A Colour Atlas of Salmonid Diseases*, 2nd edn. Springer, Dordrecht, Heidelberg, New York, London.

Brunskill, G.J. & Schindler, D.W. 1971. Geography and bathymetry of selected lake basins, experimental lakes area, Northwestern Ontario. *J. Fish. Res. Board Can.* **28**: 139–155.

Buffie, C.G., Jarchum, I., Equinda, M., Lipuma, L., Gobourne, A. & Viale, A. *et al.* 2011. Profound alterations of intestinal

microbiota following a single dose of Clindamycin results in sustained susceptibility to C. difficile-induced colitis. *Infect. Immun.* **80**: 62–71.

Buras, N., Duek, L. & Niv, S. 1985. Reactions of fish to microorganisms in wastewater. *Appl. Environ. Microbiol.* **50**: 989–995.

Burggraf, S., Stetter, K.O., Rouviere, P. & Woese, C.R. 1991. *Methanopyrus kandleri*: an archaeal methanogen unrelated to all other known methanogens. *Syst. Appl. Microbiol.* **14**: 346–351.

Cahill, M.M. 1990. Bacterial flora of fishes: a review. *Microb. Ecol.* **19**: 21–41.

Casadevall, A. & Pirofski, L.-A. 1999. Host-pathogen interactions: redefining the basic concepts of virulence and pathogenicity. *Infect. Immun.* **67**: 3703–3713.

Collin, H., Burri, R., Comtesse, F. & Fumagalli, L. 2013. Combining molecular evolution and environmental genomics to unravel adaptive processes of MHC class IIB diversity in European minnows (*Phoxinus phoxinus*). *Ecol. Evol.* **3**: 2568–2585.

Danguy, A., Genten, F. & Terwinghe, E. 2011. *Histologie Illustrée du Poisson*. Editions Quæ, Versailles.

Danilo, E. 2004. PCR-DGGE fingerprinting: novel strategies for detection of microbes in food. *J. Microbiol. Methods* **56**: 297–314.

Di Maiuta, N., Schwarzentruber, P., Schenker, M. & Schoelkopf, J. 2013. Microbial population dynamics in the faeces of wood-eating loricariid catfishes. *Lett. Appl. Microbiol.* **56**: 401–407.

Dionne, M., Miller, K.M., Dodson, J.J. & Bernatchez, L. 2009. MHC standing genetic variation and pathogen resistance in wild Atlantic salmon. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **364**: 1555–1565.

Dopazo, C.P., Lemos, M.L., Lodeiros, C., Bolinches, J., Barja, J.L. & Toranzo, A.E. 1988. Inhibitory activity of antibiotic-producing marine bacteria against fish pathogens. *J. Appl. Microbiol.* **65**: 97–101.

Ekman, E. 2003. Natural and experimental infections with Flavobacterium psychrophilum in salmonid fish.

Ekman, E., Börjeson, H. & Johansson, N. 1999. *Flavobacterium psychrophilum* in Baltic salmon Salmo salar brood fish and their offspring. *Dis. Aquat. Org.* **37**: 159–163.

Evans, M.L. & Neff, B.D. 2009. Major histocompatibility complex heterozygote advantage and widespread bacterial infections in populations of Chinook salmon (*Oncorhynchus tshawytscha*). *Mol. Ecol.* **18**: 4716–4729.

Evans, M.L., Praebel, K., Peruzzi, S. & Bernatchez, L. 2012. Parallelism in the oxygen transport system of the lake whitefish: the role of physiological divergence in ecological speciation. *Mol. Ecol.* **21**: 4038–4050.

Evans, M.L., Chapman, L.J., Mitrofanov, I. & Bernatchez, L. 2013. Variable extent of parallelism in respiratory, circulatory, and neurological traits across lake whitefish species pairs. *Ecol. Evol.* **3**: 546–557.

Falkow, S. 1997. What is a pathogen? *Am. Soc. Microbiol. News* **63**: 359–370.

Fenderson, O.C. 1964. Evidence of subpopulations of lake whitefish, *Coregonus clupeaformis*, involving a dwarfed form. *Trans. Am. Fish. Soc.* **93**: 77–94.

Frank, S.A. 2002. *Immunology and Evolution of Infectious Disease*. Princeton University Press, Princeton.

Gagnaire, P.-A., Pavey, S.A., Normandeau, E. & Bernatchez, L. 2013. The genetic architecture of reproductive isolation during speciation with gene flow in lake whitefish species pairs assessed by RAD sequencing. *Evolution* **67**: 2483–2497.

Gatesoupe, J. & Lésel, R. 1998. Flore digestive des poissons: approche environnementale. *Cah. Agric.* **7**: 29–35.

Ghiglione, J.F. & Murray, A.E. 2011. Pronounced summer to winter differences and higher wintertime richness in coastal Antarctic marine bacterioplankton. *Environ. Microbiol.* **14**: 617–629.

Gil, P., Vivas, J., Gallardo, C.S. & Rodrıguez, L.A. 2000. First isolation of *Staphylococcus warneri*, from diseased rainbow trout, *Oncorhynchus mykiss* (Walbaum), in Northwest Spain. *J. Fish Dis.* **23**: 295–298.

Goetz, F., Rosauer, D., Sitar, S., Goetz, G., Simchick, C., Roberts, S. *et al.* 2010. A genetic basis for the phenotypic differentiation between siscowet and lean lake trout (*Salvelinus namaycush*). *Mol. Ecol.* **19**: 176–196.

Goldschmidt-Clermont, E., Wahli, T., Frey, J. & Burr, S.E. 2008. Identification of bacteria from the normal flora of perch, *Perca fluviatilis* L., and evaluation of their inhibitory potential towards *Aeromonas* species. *J. Fish Dis.* **31**: 353–359.

Grice, E. & Segre, J. 2012. The Human Microbiome: our Second Genome. *Annu. Rev. Genomics Hum. Genet.* **13**: 151–170.

Hart, S., Wrathmell, A.B., Harris, J.E. & Grayson, T.H. 1988. Gut immunology in fish: a review. *Dev. Comp. Immunol.* **12**: 453–480.

Hatai, K., Egusa, S., Nakajima, M. & Chikahata, H. 1975. *Pseudomonas chlororaphis* as a fish pathogen. *Bull. Jpn. Soc. Sci. Fish* **41**: 1203.

Hau, H.H. & Gralnick, J.A. 2007. Ecology and biotechnology of the genus *Shewanella*. *Annu. Rev. Microbiol.* **61**: 237–258.

Hooper, L.V., Midtvedt, T. & Gordon, J.I. 2002. How host-microbial interactions shape the nutrient environment of the mammalian intestine. *Annu. Rev. Nutr.* **22**: 283–307.

Huber, J.A., Mark Welch, D.B., Morrison, H.G., Huse, S.M., Neal, P.R. & Butterfield, D.A. 2007. Microbial population structures in the deep marine biosphere. *Science* **318**: 97–100.

Huse, S.M., Dethlefsen, L., Huber, J.A., Welch, D.M., Relman, D.A. & Sogin, M.L. 2008. Exploring microbial diversity and taxonomy using SSU rRNA hypervariable tag sequencing. *PLoS Genet.* **4**: e1000255.

Janeway, C. 2001. *Immunobiology: The Immune System in Health and Disease*, 5th edn. Garland Science, New York.

Jeukens, J., Renaut, S., St-Cyr, J., Nolte, A.W. & Bernatchez, L. 2010. The transcriptomics of sympatric dwarf and normal lake whitefish (*Coregonus clupeaformis* spp., Salmonidae) divergence as revealed by next-generation sequencing. *Mol. Ecol.* **19**: 5389–5403.

Jutfelt, F., Olsen, R.E., Glette, J., Ringø, E. & Sundell, K. 2006. Translocation of viable *Aeromonas salmonicida* across the intestine of rainbow trout, *Oncorhynchus mykiss* (Walbaum). *J. Fish Dis.* **29**: 255–262.

Kaneko, T., Nakamura, Y., Sato, S., Asamizu, E., Kato, T., Sasamoto, S. *et al.* 2000. Complete genome structure of the nitrogen-fixing symbiotic bacterium *Mesorhizobium loti*. *DNA Res.* **7**: 331–338.

Kelly, D.P., Rainey, F.A. & Wood, A. 2006. The Genus Paracoccus. In: *The Prokaryotes Volume 5: Proteobacteria: Alpha and Beta Subclasses* (M. Dwarkin, ed.), 232–249. Springer Science+ Business Media, New York.

Knudsen, D., Jutfelt, F., Sundh, H., Sundell, K., Koppe, W. & Frøkiær, H. 2008. Dietary soya saponins increase gut

permeability and play a key role in the onset of soyabean-induced enteritis in Atlantic salmon (*Salmo salar* L.). *Br. J. Nutr.* **100**: 120–129.

Koizumi, Y., Kojima, H., Oguri, K., Kitazato, H. & Fukui, M. 2004. Vertical and temporal shifts in microbial communities in the water column and sediment of saline meromictic Lake Kaiike (Japan), as determined by a 16S rDNA-based analysis, and related to physicochemical gradients. *Environ. Microbiol.* **6**: 622–637.

Kuffner, M., Hai, B., Rattei, T., Melodelima, C., Schloter, M., Zechmeister-Boltenstern, S. *et al.* 2012. Effects of season and experimental warming on the bacterial community in a temperate mountain forest soil assessed by 16S rRNA gene pyrosequencing. *FEMS Microbiol. Ecol.* **82**: 551–562.

Kwon, K.K., Woo, J.-H., Yang, S.-H., Kang, J.-H., Kang, S.G., Kim, S.-J. *et al.* 2007. *Altererythrobacter epoxidivorans* gen. nov., sp. nov., an epoxide hydrolase-active, mesophilic marine bacterium isolated from cold-seep sediment, and reclassification of *Erythrobacter luteolus* Yoon et al. 2005 as *Altererythrobacter luteolus* comb. nov. *Int. J. Syst. Evol. Microbiol.* **57**: 2207–2211.

Landry, L. & Bernatchez, L. 2010. Role of epibenthic resource opportunities in the parallel evolution of lake whitefish species pairs (*Coregonus* sp.). *J. Evol. Biol.* **23**: 2602–2613.

Landry, L., Vincent, W.F. & Bernatchez, L. 2007. Parallel evolution of lake whitefish dwarf ecotypes in association with limnological features of their adaptive landscape. *J. Evol. Biol.* **20**: 971–984.

Lane, D.J., Pace, B., Olsen, G.J., Stahl, D.A., Sogin, M.L. & Pace, N.R. 1985. Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. *PNAS* **82**: 6955–6959.

Langefors, Å., Lohm, J., Grahn, M., Andersen, Ø. & von Schantz, T. 2001. Association between major histocompatibility complex class IIB alleles and resistance to *Aeromonas salmonicida* in Atlantic salmon. *Proc. R. Soc. Lond. B* **268**: 479–485.

Laplante, K., Sébastien, B. & Derome, N. 2013. Parallel changes of taxonomic interaction networks in lacustrine bacterial communities induced by a polymetallic perturbation. *Evol. Appl.* **6**: 643–659.

Larsen, J.L. & Mellergaard, S. 1981. Microbiological and hygienic problems in marine aquaculture: furunculosis and vibriosis in rainbow trout (*Salmo gairdneri*). *Bull. Eur. Assoc. Fish Pathol.* **1**: 29–32.

Larsen, M.H., Blackburn, N., Larsen, J.L. & Olsen, J.E. 2004. Influences of temperature, salinity and starvation on the motility and chemotactic response of *Vibrio anguillarum*. *Microbiology* **150**: 1283–1290.

Larsen, A., Tao, Z., Bullard, S.A. & Arias, C.R. 2013. Diversity of the skin microbiota of fishes: evidence for host species specificity. *FEMS Microbiol. Ecol.* **85**: 483–494.

Laskin, A.I. & White, D.C. 1999. Preface to special issue on Sphingomonas. *J. Ind. Microbiol. Biotechnol.* **23**: 231.

Lavery, T.J., Roudnew, B., Seymour, J., Mitchell, J.G. & Jeffries, T. 2012. High nutrient transport and cycling potential revealed in the microbial metagenome of Australian sea lion (*Neophoca cinerea*) faeces. *PLoS ONE* **7**: e36478.

Lebeis, S.L., Rott, M., Dangl, J.L. & Schulze-Lefert, P. 2012. Culturing a plant microbiome community at the cross-rhodes. *New Phytol.* **196**: 341–344.

Li, G., Guo, Y., Zhao, D., Qian, P., Sun, J., Xiao, C. *et al.* 2006. Effects of levamisole on the immune response and disease resistance of *Clarias fuscus*. *Aquaculture* **253**: 212–217.

Lidstrom, M.E. & Chistoserdova, L. 2002. Plants in the pink: cytokinin production by methylobacterium. *J. Bacteriol.* **184**: 1818.

Lindström, E.S., Kamst-Van Agterveld, M.P. & Zwart, G. 2005. Distribution of typical freshwater bacterial groups is associated with pH, temperature, and lake water retention time. *Appl. Environ. Microbiol.* **71**: 8201–8206.

Linnenbrink, M., Wang, J., Hardouin, E.A., Künzel, S., Metzler, D. & Baines, J.F. 2013. The role of biogeography in shaping diversity of the intestinal microbiota in house mice. *Mol. Ecol.* **22**: 1904–1916.

Lopez-Velasco, G., Welbaum, G.E., Boyer, R.R., Mane, S.P. & Ponder, M.A. 2011. Changes in spinach phylloepiphytic bacteria communities following minimal processing and refrigerated storage described using pyrosequencing of 16S rRNA amplicons. *J. Appl. Microbiol.* **110**: 1203–1214.

Lu, G. & Bernatchez, L. 1999. Correlated trophic specialization and genetic divergence in sympatric lake whitefish ecotypes (*Coregonus clupeaformis*): support for the ecological speciation hypothesis. *Evolution* **53**: 1491–1505.

MacColl, A.D.C. 2009. Parasite burdens differ between sympatric three-spined stickleback species. *Ecography* **32**: 153–160.

Mackay, R.J. & Kalff, J. 1969. Seasonal variation in standing crop and species diversity of insect communities in a small Quebec stream. *Ecology* **50**: 101–109.

Magurran, A.E. 2003. *Measuring Biological Diversity*. University of St Andrews, St Andrews, UK.

Maidak, B.L., Cole, J.R., Lilburn, T.G., Parker, C.T. Jr, Saxman, P.R., Farris, R.J. *et al.* 2001. The RDP-II (Ribosomal Database Project). *Nucleic Acids Res.* **29**: 173–174.

Marchesi, J.R., Sato, T., Weightman, A.J., Martin, T.A., Fry, J.C., Hiom, S.J. *et al.* 1998. Design and evaluation of useful bacterium-specific PCR primers that amplify genes coding for bacterial 16S rRNA. *Appl. Environ. Microbiol.* **64**: 795–799.

Matarese, G. & La Cava, A. 2004. The intricate interface between immune system and metabolism. *Trends Immunol.* **25**: 193–200.

Mc Cure, B., Grace, J.B. & Urban, D.L. 2002. *Analysis of Ecological Communities*. MjM Software, Gleneden Beach, OR.

McDowell, A., Gao, A., Barnard, E., Fink, C., Murray, P.I., Dowson, C.G. *et al.* 2011. A novel multilocus sequence typing scheme for the opportunistic pathogen Propionibacterium acnes and characterization of type I cell surface-associated antigens. *Microbiology* **157**: 1990–2003.

McKenzie, V.J., Bowers, R.M., Fierer, N., Knight, R. & Lauber, C.L. 2012. Co-habiting amphibian species harbor unique skin bacterial communities in wild populations. *ISME J.* **6**: 588–596.

McKnite, A.M., Perez-Munoz, M.E., Lu, L., Williams, E.G., Brewer, S., Andreux, P.A. *et al.* 2012. Murine Gut Microbiota Is Defined by Host Genetics and Modulates Variation of Metabolic Traits. *PLoS ONE* **7**: e39191.

Muyzer, G. & Smalla, K. 1998. Application of denaturing gradient gel electrophoresis (DGGE) and temperature gradient gel electrophoresis (TGGE) in microbial ecology. *Antonie Van Leeuwenhoek* **73**: 127–141.

Nakajima, T., Palchevsky, V., Perkins, D., Belperio, J. & Finn, P. 2011. Lung transplantation: infection, inflammation, and the microbiome. *Semin. Immunopathol.* **33**: 135–156.

Natsopoulou, M.E., Pálsson, S. & Ólafsdóttir, G.Á. 2012. Parasites and parallel divergence of the number of individual MHC alleles between sympatric three-spined stickleback *Gasterosteus aculeatus* morphs in Iceland. *J. Fish Dis.* **81**: 1696–1714.

Navarrete, P., Espejo, R.T. & Romero, J. 2009. Molecular analysis of microbiota along the digestive tract of juvenile atlantic salmon (*Salmo salar* L.). *Microb. Ecol.* **57**: 550–561.

Nematollahi, A., Decostere, A., Pasmans, F. & Haesebrouck, F. 2003. *Flavobacterium psychrophilum* infections in salmonid fish. *J. Fish Dis.* **26**: 563–574.

Nieto, T.P., Toranzo, A.E. & Barja, J.L. 1984. Comparison between the bacterial flora associated with fingerling rainbow trout cultured in two different hatcheries in the North-West of Spain. *Aquaculture* **42**: 193–206.

Olivier, G., Lallier, R. & Larivière, S. 1981. A toxigenic profile of *Aeromonas hydrophila* and *Aeromonas sobria* isolated from fish. *Can. J. Microbiol.* **27**: 330–333.

Ovreås, L., Forney, L., Daae, F.L. & Torsvik, V. 1997. Distribution of bacterioplankton in meromictic Lake Saelenvannet, as determined by denaturing gradient gel electrophoresis of PCR-amplified gene fragments coding for 16S rRNA. *Appl. Environ. Microbiol.* **63**: 3367–3373.

Pavey, S.A., Sevellec, M., Adam, W., Normandeau, E., Lamaze, F.C., Gagnaire, P.-A. *et al.* 2013. Nonparallelism in MHCIIβ diversity accompanies nonparallelism in pathogen infection of lake whitefish (*Coregonus clupeaformis*) species pairs as revealed by next-generation sequencing. *Mol. Ecol.* **22**: 3833–3849.

Peet, R.K. 1974. The measurement of species diversity. *Rev. Ecol. Evol. Syst.* **5**: 285–307.

Perru, O. 2006. Aux origines des recherches sur la symbiose vers 1868–1883. *Revue d'histoire des Sci.* **59**: 5–27.

Prescott, L., Harley, J. & Klein, D. 2002. *Microbiology*, 5th edn. McGraw-Hill, New York.

Press, C.M. & Evensen, Ø. 1999. The morphology of the immune system in teleost fishes. *Fish Shellfish Imm.* **9**: 309–318.

Reche, I., Pulido-Villena, E., Morales-Baquero, R. & Casamayor, E.O. 2005. Does ecosystem size determine aquatic bacterial richness? *Ecology* **86**: 1715–1722.

Ringø, E. & Olsen, R.E. 1999. The effect of diet on aerobic bacterial flora associated with intestine of Arctic charr (*Salvelinus alpinus* L.). *J. Appl. Microbiol.* **86**: 22–28.

Ringø, E., Sperstad, S., Myklebust, R., Refstie, S. & Krogdahl, Å. 2006. Characterisation of the microbiota associated with intestine of Atlantic cod (*Gadus morhua* L.): the effect of fish meal, standard soybean meal and a bioprocessed soybean meal. *Aquaculture* **261**: 829–841.

Roald, S. & Hastein, T. 1980. Infection with an acinetobacter-like bacterium in Atlantic salmon (*Salmo salar*) broodfish. In: *Fish Diseases* (W. Ahne, ed.), pp. 154–156. Springer, Berlin, Heidelberg.

Roesch, L.F., Fulthorpe, R.R., Riva, A., Casella, G., Hadwin, A.K. & Kent, A.D. 2007. Pyrosequencing enumerates and contrasts soil microbial diversity. *ISME J.* **1**: 283–290.

Roesch, L.F.W., Fulthorpe, R.R., Pereira, A.B., Pereira, C.K., Lemos, L.N., Barbosa, A.D. *et al.* 2012. Soil bacterial community abundance and diversity in ice-free areas of Keller Peninsula. *Antarctica Appl. Soil Ecol.* **61**: 7–15.

Roeselers, G., Mittge, E.K., Stephens, W.Z., Parichy, D.M., Cavanaugh, C.M., Guillemin, K. *et al.* 2011. Evidence for a core gut microbiota in the zebrafish. *ISME J.* **5**: 1595–1608.

Rogers, S. & Bernatchez, L. 2007. The genetic architecture of ecological speciation and the association with signatures of selection in natural lake whitefish (*Coregonus sp.* Salmonidae) species pairs. *Mol. Biol. Evol.* **24**: 1423–1438.

Ryan, R.P., Monchy, S., Cardinale, M., Taghavi, S., Crossman, L., Avison, M.B. *et al.* 2009. The versatility and adaptation of bacteria from the genus *Stenotrophomonas*. *Nat. Rev. Microbiol.* **7**: 514–525.

Sakai, M., Atsuta, S. & Kobayashi, M. 1989. *Pseudomonas fluorescens* isolated from the diseased rainbow trout, Oncorhynchus mykiss. *Kitasato Arch. Exp. Med.* **62**: 157.

Salgado-Miranda, C., Palomares, E., Jurado, M., Marín, A., Vega, F. & Soriano-Vargas, E. 2010. Isolation and distribution of bacterial flora in farmed rainbow trout from Mexico. *J. Aquat. Anim. Health* **22**: 244–247.

Savini, V., Catavitello, C., Masciarelli, G., Astolfi, D., Balbinot, A., Bianco, A. *et al.* 2010. Drug sensitivity and clinical impact of members of the genus *Kocuria*. *J. Med. Microbiol.* **59**: 1395–1402.

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B. *et al.* 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **75**: 7537–7541.

Sepkoski, J.J. Jr 1988. Alpha, beta, or gamma: where does all the diversity go? *Paleobiology* **14**: 221–234.

Siqueira, J.F. Jr, Alves, F.R.F. & Rôças, I.N. 2011. Pyrosequencing analysis of the apical root canal microbiota. *J. Endod.* **37**: 1499–1503.

Snedecor, G.W. & Cochran, W.G. 1980. *Statistical Methods*, 7th edn. Iowa State University Press, Ames, IA.

Sousa, J.A.d. & Silva-Souza, Â.T. 2001. Bacterial community associated with fish and water from Congonhas River, Sertaneja, Paraná, Brazil. *Braz. Arch. Biol. Technol.* **44**: 373–381.

St-Cyr, J., Derome, N. & Bernatchez, L. 2008. The transcriptomics of life-history trade-offs in whitefish species pairs (*Coregonus* sp.). *Mol. Ecol.* **17**: 1850–1870.

Toranzo, A.E., Novoa, B., Romalde, J.L., Núñez, S., Devesa, S., Mariño, E. *et al.* 1993. Microflora associated with healthy and diseased turbot (*Scophthalmus maximus*) from three farms in northwest Spain. *Aquaculture* **114**: 189–202.

Trust, T.J. & Sparrow, R.A.H. 1974. The bacterial flora in the alimentary tract of freshwater salmonid fishes. *Can. J. Microbiol.* **20**: 1219–1228.

Turnbaugh, P.J., Ley, R.E., Hamady, M., Fraser-Liggett, C.M., Knight, R. & Gordon, J.I. 2007. The human microbiome project. *Nature* **449**: 804–810.

Turner, S., Pryer, K.M., Miao, V.P.W. & Palmer, J.D. 1999. Investigating deep phylogenetic relationships among cyanobacteria and plastids by small subunit rRNA sequence analysis. *J. Eukaryot. Microbiol.* **46**: 327–338.

White, J.R., Nagarajan, N. & Pop, M. 2009. Statistical methods for detecting differentially abundant features in clinical metagenomic samples. *PLoS Comput. Biol.* **5**: e1000352.

Wu, S., Wang, G., Angert, E.R., Wang, W., Li, W. & Zou, H. 2012. Composition, diversity, and origin of the bacterial community in grass carp intestine. *PLoS ONE* **7**: e30440.

Xu, X.-W., Wu, Y.-H., Wang, C.-S., Wang, X.-G., Oren, A. & Wu, M. 2009. *Croceicoccus marinus* gen. nov., sp. nov., a yellow-pigmented bacterium from deep-sea sediment, and emended description of the family *Erythrobacteraceae*. *Int. J. Syst. Evol. Microbiol.* **59**: 2247–2253.

Yildirim, S., Yeoman, C.J., Sipos, M., Torralba, M., Wilson, B.A., Goldberg, T.L. *et al.* 2010. Characterization of the fecal microbiome from non-human wild primates reveals species specific microbial communities. *PLoS ONE* **5**: e13963.

Yoon, J.-H., Lee, S.T. & Park, Y.-H. 1998. Inter- and intraspecific phylogenetic analysis of the genus *Nocardioides* and related taxa based on 16S rDNA sequences. *Int. J. Syst. Evol. Bacteriol.* **48**: 187–194.

Youngblut, N.D., Shade, A., Read, J.S., McMahon, K.D. & Whitaker, R.J. 2013. Lineage-specific responses of microbial communities to environmental change. *Appl. Environ. Microbiol.* **79**: 39–47.

Yourno, J. 1992. A method for nested PCR with single closed reaction tubes. *Genome Res.* **2**: 60–65.

Zamora, L., Vela, A., Palacios, M.A., Dominguez, L. & Fernandez-Garayzabal, J. 2012. First isolation and characterization of *Chryseobacterium shigense* from rainbow trout. *BMC Vet. Res.* **8**: 77.

Zhou, Z., Karlsen, Ø., He, S., Olsen, R.E., Yao, B. & Ringø, E. 2012. The effect of dietary chitin on the autochthonous gut bacteria of Atlantic cod (*Gadus morhua* L.). *Aquacult. Res.* **44**: 1889–1900.

Zwart, G., Crump, B.C., Agterveld, M.P.K.-v., Hagen, F. & Han, S.-K. 2002. Typical freshwater bacteria: an analysis of available 16S rRNA gene sequences from plankton of lakes and rivers. *Aquat. Microb. Ecol.* **28**: 141–155.

Supporting information

Additional Supporting Information may be found in the online version of this article:

Figure S1 Box plot of the Chao index of bacterial richness for all ten study populations.

Figure S2 Dominance curve of frequency for the 50 most frequent genera.

Figure S3 Dominance curve of abundance for the 50 most abundant genera.

Figure S4 Box plot of only putative pathogen diversity by the Simpson index.

Table S1 Taxonomy of all bacteria found in the kidney of whitefish samples.

Table S2 The 50 most abundant OTUs in order of frequency.

Table S3 The first 50 OTUs in order of abundance.

Table S4 Bibliography study on species of selected putative pathogen genera which were highlighted using the BLAST algorithm.

Data deposited at Dryad: doi:10.5061/dryad.hv87d

Received 19 September 2013; accepted 16 March 2014