The biogeography of the Atlantic salmon (*Salmo salar*) gut microbiome.

Running title: Biogeography of the Atlantic Salmon microbiome

Authors

Martin S. Llewellyn[1,2,3]
Philip McGinnity[4]
Melanie Dionne[5]
Justine Letourneau[2]
Florian Thonier[2]
Gary R Carvalho[1]
Simon Creer[1]
Nicolas Derome[2]

Affiliations

[1]Molecular Ecology and Fisheries Genetics Laboratory, Bangor University, Wales
[2]Insititut de Biologie Intergatif et des Systèmes, Université Laval, Quebec, Canada
[3]Institute of Biodiversity, Animal Health and Comparative Medicine, University of Glasgow
[4]School of Biological, Earth & Environmental Sciences, University College Cork, Cork, Ireland
[5]Ministère des Ressources Naturelles et de la Faune, Quebec, Canada

Corresponding Author

Martin S. Llewellyn
Institute of Biodiversity, Animal Health and Comparative Medicine,
University of Glasgow
Glasgow G12 8QQ
Email: martin.llewellyn@glasgow.ac.uk

Abstract

Although understood in many vertebrate systems, the natural diversity of host-associated microbiota has been little studied in teleosts. For migratory fishes, successful exploitation of multiple habitats may affect and be affected by the composition of the intestinal microbiome. We collected 96 *Salmo salar* from across the Atlantic encompassing both freshwater and marine phases. Dramatic differences between environmental and gut bacterial communities were observed. Furthermore, community composition was not significantly impacted by geography. Instead lifecycle stage strongly defined both the diversity and identity of microbial assemblages in the gut, with evidence for community destabilisation in migratory phases. *Mycoplasmataceae* phylotypes were abundantly recovered in all lifecycle stages. Patterns of *Mycoplasmataceae* phylotype recruitment to the intestinal microbial community among sites and lifecycles stages support a dual role for deterministic and stochastic processes in defining the composition of the *S. salar* gut microbiome.

Keywords: Atlantic Salmon, microbiome, biogeography, intestine, next generation sequencing

Introduction

Atlantic Salmon (*Salmo salar*) are non-obligate anadramous salmonids of significant commercial, cultural and recreational importance. Growth, development and migration in anadromous *Salmo salar* involves a radical shift across an ecological and trophic spectrum (Jacobsen & Hansen 1999; Orlov *et al.* 2006). Accompanying the physiological, behavioural and dietary adaptations necessary to cope with transition between freshwater and marine environments (McCormick *et al.* 2013), significant and potentially adaptive shifts in host-associated microbiota might be expected.

The advent of culture-free microbial meta-sequencing techniques means that bacterial communities can be profiled in unprecedented detail. As such, teleost-associated intestinal microbiota are increasingly subject to scrutiny (Llewellyn *et al.* 2014). Salmonids have received particular attention in view of their importance in aquaculture (e.g. (Zarkasi *et al.* 2014)) and the urgent need for innovation in feed sources (Green *et al.* 2013). Attempts to establish the natural identity and stability of commercially exploited teleost intestinal microbiomes have been limited to focal studies from single aquaculture facilities (e.g. (Zarkasi *et al.* 2014)) and single sites in the wild (Star *et al.* 2013). The ecological succession of gut bacterial phylotypes during wild teleost development and migration is an excellent system in which to explore the relative contribution of host and environmental factors to shaping microbiome recruitment, especially in euryhaline species (Schmidt *et al.* 2015). Furthermore, exploration of the biogeography of microbiome composition among species with wide geographic distributions is required to form a sound baseline for experimental study. In this study we set out to explore microbiome ontogeny and biogrography in wild *S. salar.*

Material and methods

Bacterial 16S SSU rRNA (V4 region) diversity was profiled from the intestinal contents of 96 wild *S. salar* on an Ilumina MiSeq platform and analysed using Mothur (Schloss *et al.* 2009) and QIIME (Kuczynski *et al.* 2012) (full methods in Supplementary Data 1). Juveniles and returning adults were sampled in eastern Canada and western Ireland. Marine adults were sampled in cooperation with the West Greenland subsistence fishery (www.nasco.int) (Supplementary Table 1).

Results

Study site (or country of origin) had no discernible effect on microbiome identity among samples form freshwater lifecycle stages, including returning adults (p = 0.7264, PERMANOVA UniFrac distances (PUd) (Oksanen *et al.* 2015) – a full list of statistical tests

applied and sample assizes is included in Supplementary Table 2). Nonetheless, study-site specific variation in intestinal operational taxonomic unit (OTU) richness (Chao1) was a consistent feature across juvenile lifecycles stages (Figure 1a, $p = 0.003$, Kruskal-Wallis test (K-WT)). Diversity estimates (Shannon Index) corroborate this finding (Supplementary Figure 1, $p = 0.009$).

In contrast to those between study sites, clear differences were observed in the microbial community identity of the *S. salar* intestine between freshwater and marine lifecycle stages ($p<0.019$ (PUd) (Oksanen *et al.* 2015)) Figure 1b), in which returning adults retain a large proportion of their microbial assemblage from the marine environment. Microbiome identity within fresh and saltwater ecotopes was not impacted by lifecycle stage (smolt *vs* parr: $p=0.268$; marine *vs* returning adult $p=0.522$), PUd). Over study sites, inter-individual microbial composition was most stable among parr and marine lifecycle stages and least stable among smolts and returning adults (open and green bars, (Figure 1c)). OTU richness (Chao1) and was dramatically affected by lifecycle stage across marine and freshwater environments ($p = 0.0001$) (K-WT)), but not diversity (Shannon, $p = 0.276$). By contrast both OTU richness and diversity was purged during the transition from parr to smolt in freshwater (Chao1, $p < 1.8\text{x}10^{-5}$, Shannon, $p = 0.0105$ (K-WT)) (Figure 1a, Supplementary Figure 1)).

Discussion

Phylum-level assignment of OTUs (Figure 2b) indicated the dominance of Proteobacteria among all samples. All lifecycle stages, especially marine adults, were enriched for Tenericutes (Genus *Mycoplasma* especially). In contrast, Firmicutes, Bacteriodetes and Actinobacteria - abundant in returning adults, smolt and parr - occur at negligible levels in marine adults. Both Tenericutes and Firmicutes OTUs occur in freshwater environmental samples, but at minimal levels with respect to those recovered from fish. By contrast Verrucomicrobia, common among freshwater samples, occurs at minimal levels in freshwater lifecycles stages. Although a common soil, freshwater and marine bacterial phylum (Freitas *et al.* 2012), as well as a frequent member of mammalian gut flora (Zhang *et al.* 2009), our study corroborates the low abundance Verrucomicrobia found in fish by others (Rawls *et al.* 2006). To buffer inter-sample variability among OTUs assigned to genus level, we first evaluated the diversity of core OTUs (defined here as those occurring in 85% of individuals) present among fish from each lifecycle stage (Figure 2a). OTUs assigned to genus *Yersinia* and other unclassified Enterobacteriacae dominate the core microbiota of freshwater parr. Some *Yersinia* species (*Y. ruckeri, Y. intermedia*) are important pathogens of salmon (Bruno *et al.* 2013). However, the healthy state of the parr we sampled, as well as several SNPs

between the principal *Yersinia* OTU in our dataset and the 16S V4 region of both *Y. ruckeri,* and *Y. intermedia* (data not shown), suggest that the *Yersinia* we sampled were likely to be commensals. Enterobacteriacae were also abundant among smolt, however, not among the core at 85% - unsurprising given raised beta-diversity within this group (Figure 1b). Genus *Mycoplasma* phylotypes were the most abundant and consistently recovered phylotypes from adult salmon. More typical members of marine teloest intestinal microflora – Genera *Allivibrio* and *Photobacterium* (e.g. (Sullam *et al.* 2012)), were also well represented. OTUs attributable to family Mycoplasmataceae were also recovered in large numbers from other freshwater lifecycles stages. Clear differences in the frequency distribution of Mycoplasmataceae OTUs exist between adult and juvenile salmon (Figure 2c). Biogeographic differences in Mycoplasmataceae OTU distributions between Canada and Ireland are apparent in juveniles, but not in returning adults (Figure 2c). With the exception of a single Mycoplasma OTU isolated from the Trinite river, Mycoplasmataceae OTUs abundant in parr and smolt were absent or rare (<5/11000) in local freshwater samples.

Both bacterial OTU richness and community stability declined over lifecycle stage in the intestine of *S. salar*, in stark contrast to mammals where community diversity increases after weaning and stabilises in early to late childhood (Yatsunenko *et al.* 2012). Returning adult salmon were characterised by low richness, highly variable microbial assemblages in comparison to parr. Although poor in absolute numbers of OTUs (i.e. richness), diversity estimates from returning adults were not significantly different from juveniles, suggesting a fairly even frequency distribution of those OTUs present. Dietary complexity in juvenile salmonids could explain rich associated microbial assemblages (Orlov *et al.* 2006). Meanwhile physiological disturbances and fasting in migratory phases (smolt and returning adults) could underlie reduced community stability with respect to corresponding non-migratory phases (i.e. parr and marine adults, Figure 1b). In particular, drinking rates increase during smoltification, as well as overall intestinal fluid re-absorption rates, perhaps affecting microbiome equilibrium (Stefansson *et al.* 2008). Additionally, documented variation in the response of different intestine regions (midgut vs hindgut) during smoltification could impact associated microbiota accordingly, an interesting future avenue for investigation (Stefansson *et al.* 2008). The microbial community of feeding marine adults was less rich and diverse than that of freshwater juveniles; perhaps attributable to the dominance of Mycoplasmataceae phylotypes among adult intestinal microbiota. Mycoplasmataceae, especially genera *Candidatus* and *Mycoplasma* frequently colonise vertebrate and invertebrate mucosae, both as pathogens and commensals (Frey & Herrmann 2002; Holben *et al.* 2002; Nechitaylo *et al.* 2009). Indeed *Mycoplasma* have been isolated from *S. salar* in the past (Holben *et al.* 2002;

Zarkasi *et al.* 2014). The abundance of Mycoplasmataceae (and individual OTUs, Figure 2c), among sites suggests an association with the salmon gut niche robust to developmental change and could point to some more complex level of interdependence with the host. In a recent laboratory study of microbiome ontology in euryhaline fishes, Schmidt et al. (2015), suggest a dominant role for deterministic forces (e.g. niche appropriation) over neutral ones (e.g. colonisation) (Schmidt *et al.* 2015). Mycoplasmataceae abundance and diversity in this study suggest a dual role in the wild: geography and environment influence colonisation source (and thus a proportion of the microbiome variation at the genus level); however the intra-host niche likely determines the abundance of Mycoplasmataceae in general across *S. salar*. More widely, a combination of deterministic host effects and stochastic environmental factors underpin diversity in the *S. salar* microbiome whereby the microbiota of freshwater juvenile and returning adults, while sharing many OTUs with local environmental samples, show radically different patterns of abundance and enrichment. Broad-scale shifts in the composition of key components of *S. salar* gut microbiomes pose fundamental questions in relation to functional significance of qualitative change. Such inferences demand an experimental approach to assess empirically the impact of microbiome diversity on fish health and survival in distinct environments, especially in the context of aquaculture.

Acknowledgements

We gratefully support the assistance of Julien April and Denise Deschamps, Ministère des Ressources Naturelles et de la Faune, Quebec Canada; Tim Sheehan, NOAA Fisheries Service; Woods Hole, USA; Ger Rogan, Russell Poole, Katie Thomas, Niall O'Maoileidigh, Deirdre Cotter and Elvira de Eyto, Marine Institute, Ireland; Paddy Gargan and Michael Hughes, Inland Fisheries Ireland; Brian Clarke and Tom Reed, University College Cork in sample collection. MSL was funded by a Marie Skłodowska-Curie International Outgoing Fellowship, grant number 302503. JL was funded by a NSERC-CREATE Fellowship. FT was funded by a NSERC discovery grant awarded to ND. PMcG was supported by the Beaufort Marine Fish Award in Fish Population Genetics funded by the Irish Government under the Sea Change Programme.

Conflict of Interest

The authors declare no conflict of interest.

Supplementary information is available at The ISME Journal's website

References

Bruno D, Noguera P, Poppe T (2013) A colour atlas of salmonid diseases. Springer Heidelberg, Germany.

Freitas S, Hatosy S, Fuhrman JA, Huse SM, Mark Welch DB, Sogin ML, et al. (2012) Global distribution and diversity of marine Verrucomicrobia. ISME J 6, 1499-1505.

Frey J, Herrmann R (2002) Mycoplasmas of animals. In: Molecular Biology and Pathogenicity of Mycoplasmas (eds. Razin S, Herrmann R). Kluwer. New York.

Green TJ, Smullen R, Barnes AC (2013) Dietary soybean protein concentrate-induced intestinal disorder in marine farmed Atlantic salmon, *Salmo salar* is associated with alterations in gut microbiota. Veterinary Microbiology 166, 286-292.

Holben WE, Williams P, Gilbert M, Saarinen M, Sarkilahti LK, Apajalahti JH (2002) Phylogenetic analysis of intestinal microflora indicates a novel Mycoplasma phylotype in farmed and wild salmon. Microbial Ecology 44, 175-185.

Jacobsen J, Hansen L (1999) Feeding Habits of Atlantic Salmon at Different Life Stages at Sea. In: The Ocean Life of Atlantic Salmon: Environmental and Biological Factors Influencing Survival (ed. Mills D). Wiley- Blackwell, London.

Kuczynski J, Stombaugh J, Walters WA, Gonzalez A, Caporaso JG, Knight R (2012) Using QIIME to analyze 16S rRNA gene sequences from microbial communities. Curr Protoc Microbiol Chapter 1, Unit 1E 5.

Llewellyn M, Boutin S, Hoseinifar SH, Derome N (2014) Teleost microbiomes: progress towards their characterisation, manipulation and applications in aquaculture and fisheries. Frontiers in Microbiology 5.

McCormick SD, Sheehan TF, Björnsson BT, Lipsky C, Kocik JF, Regish AM, et al. (2013) Physiological and endocrine changes in Atlantic salmon smolts during hatchery rearing, downstream migration, and ocean entry. Canadian Journal of Fisheries and Aquatic Sciences 70, 105-118.

Nechitaylo TY, Timmis KN, Golyshin PN (2009) 'Candidatus Lumbricincola', a novel lineage of uncultured Mollicutes from earthworms of family Lumbricidae. Environmental Microbiology 11, 1016-1026.

Oksanen J, Blanchet F, Kindt K, Legendre P, Minchin P, O'Hara R, et al. (2015) vegan: Community Ecology Package. R package version 2.2-1. http://CRAN.R-project.org/package=vegan.

231 Orlov AV, Gerasimov YV, Lapshin OM (2006) The feeding behaviour of cultured and wild
232 Atlantic salmon, *Salmo salar* L., in the Louvenga River, Kola Peninsula, Russia. ICES
233 Journal of Marine Science: Journal du Conseil 63, 1297-1303.

234 Rawls JF, Mahowald MA, Ley RE, Gordon JI (2006) Reciprocal Gut Microbiota Transplants
235 from Zebrafish and Mice to Germ-free Recipients Reveal Host Habitat Selection. Cell 127,
236 423-433.

237 Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et al. (2009)
238 Introducing mothur: open-source, platform-independent, community-supported software for
239 describing and comparing microbial communities. Appl Environ Microbiol 75, 7537-7541.

240 Schmidt VT, Smith KF, Melvin DW, Amaral-Zettler LA (2015) Community assembly of a
241 euryhaline fish microbiome during salinity acclimation. Molecular Ecology, 24:2537-50.

242 Star B, Haverkamp THA, Jentoft S, Jakobsen KS (2013) Next generation sequencing shows
243 high variation of the intestinal microbial species composition in Atlantic cod caught at a
244 single location. BMC Microbiology 13, 248-248.

245 Stefansson S, Bjornsson B, Ebbensson L, McCormick S (2008) Smoltification. In: Fish
246 Larval Physiology (eds. Kappor, BJ Finn, RN). Wiley- Blackwell, London.

247 Sullam KE, Essinger SD, Lozupone CA, O'Connor MP, Rosen GL, Knight R, et al. (2012)
248 Environmental and ecological factors that shape the gut bacterial communities of fish: a
249 meta-analysis. Mol Ecol 21, 3363-3378.

250 Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, et al.
251 (2012) Human gut microbiome viewed across age and geography. Nature 486, 222-227.

252 Zarkasi KZ, Abell GCJ, Taylor RS, Neuman C, Hatje E, Tamplin ML, et al. (2014)
253 Pyrosequencing-based characterization of gastrointestinal bacteria of Atlantic salmon (Salmo
254 salar L.) within a commercial mariculture system. Journal of Applied Microbiology 117, 18-
255 27.

256 Zhang H, DiBaise JK, Zuccolo A, Kudrna D, Braidotti M, Yu Y, et al. (2009) Human gut
257 microbiota in obesity and after gastric bypass. Proceedings of the National Academy of
258 Sciences 106, 2365-2370.

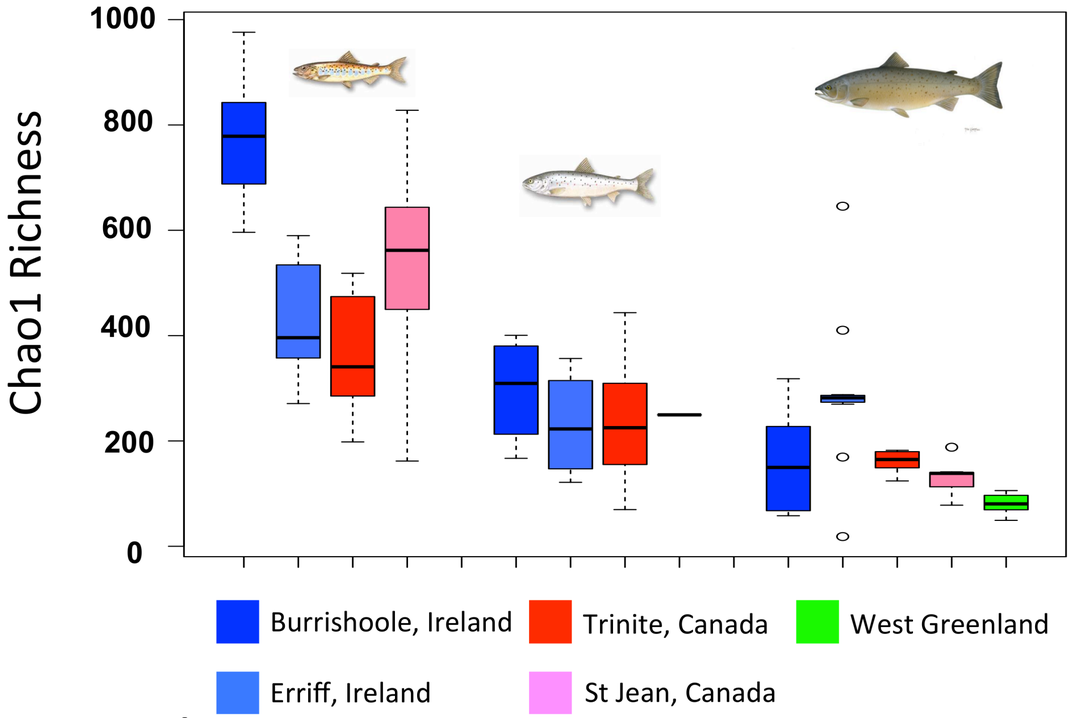
259 **Figure legends**

260 **Figure 1 – Alpha and beta diversity comparisons of Atlantic Salmon intestinal**
261 **microbiota**. **a.** Box plot showing alpha diversity (Chao1 richness estimator) variation across
262 sites and lifecycle stages: blue bars – Burrishoole, Ireland; pale blue bars – Erriff, Ireland; red

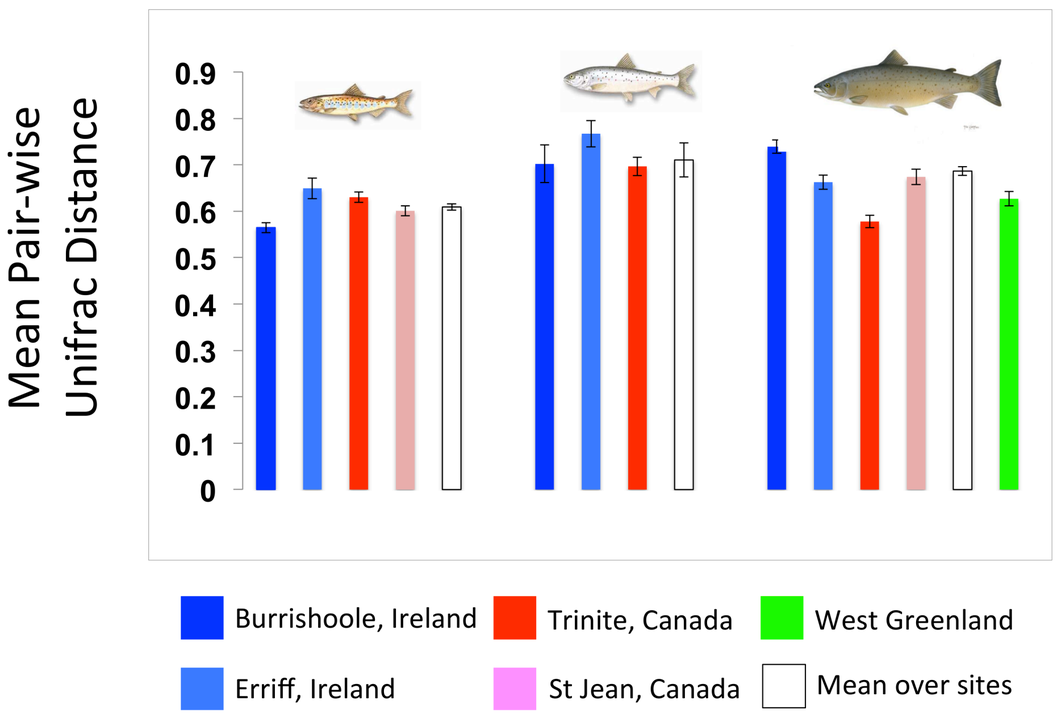
bars – St Jean, Canada; pink bars Trinite, Canada. Marine adults from West Greenland are represented in green. **b.** Mean beta diversity distances (unweighted unifrac) among individuals grouped by site and lifecycle stage. Error bars represent standard error about the mean. **c.** Principal coordinates analysis of pairwise unweighted unifrac distances between all salmon and environmental samples. Axes represent the two synthetic variables explaining the greatest proportion of variation in the dataset.

Figure 2 – Taxonomic composition of the Atlantic Salmon intestinal microbiome. a. Phlyum-level composition of total OTU abundances among distinct lifecycles stages and environmental samples. **b.** Core (present in ≥ 85% of individuals) 97% identity OTUs assigned to genus level from each lifecycle stage are represented by red-outlined ellipses. Black-outlined ellipses denote the presence of these 'core' OTUs among other lifecycle stages. Ellipse area is proportional to the mean abundance of OTUs assigned to each genus over all samples from each lifecycles stage. Core genera that occur at a mean frequency >1000 in each sample time are either bold (adults), underlined (parr) or italicised (freshwater). **c.** Heatmap displaying the frequency distribution of OTUs belonging to family Mycoplasmataceae across distinct lifecycle stages and countries of origin. Genera within the Mycoplasmataceae are indicated on the maximum likelihood phylogeny (left) on which values indicated % bootstrap support for the respective clades. Single asterisk indicates the *Candidatus* OTU also recovered from a sympatric environmental sample. Double asterisk indicates the most abundant core Mycoplasma OTU recovered from adult lifecycles stages.

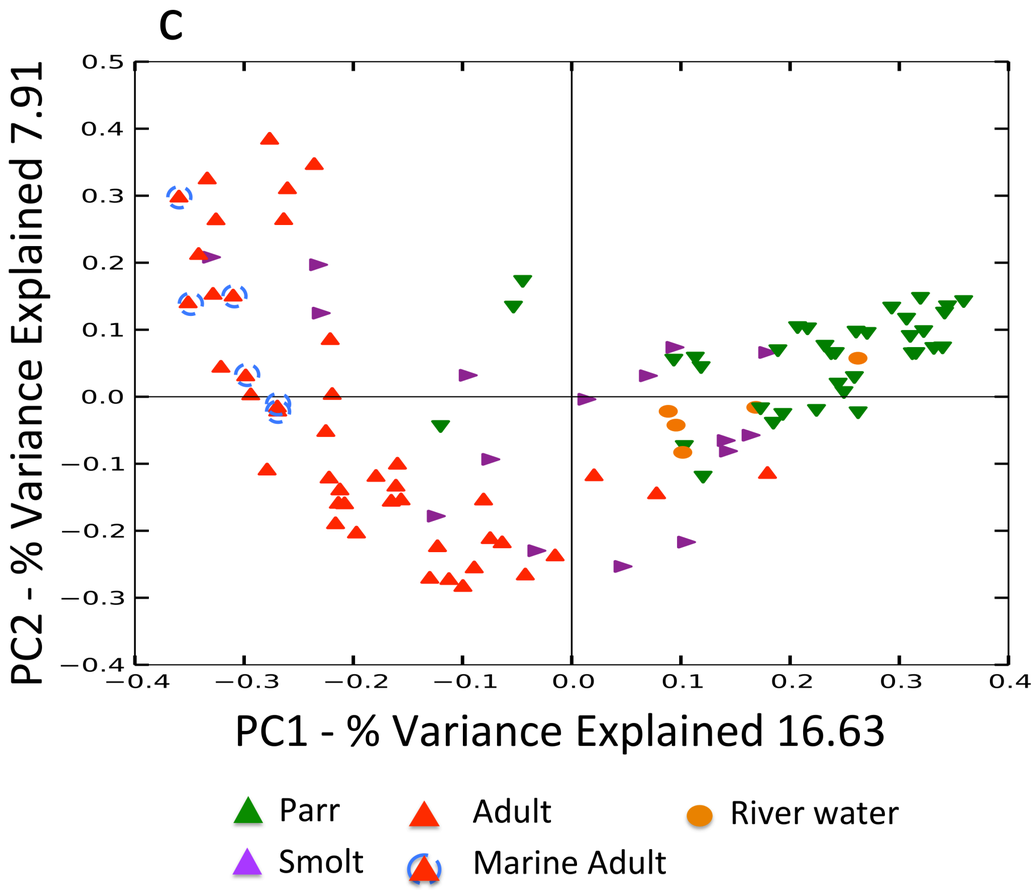
a


Chao1 Richness
1000
800
600
400
200
0
Burrishoole, Ireland
Trinite, Canada
West Greenland
Erriff, Ireland
St Jean, Canada

b


Mean Pair-wise
Unifrac Distance
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Burrishoole, Ireland
Trinite, Canada
West Greenland
Erriff, Ireland
St Jean, Canada
Mean over sites

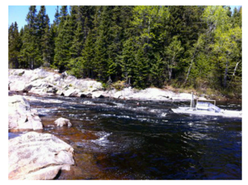
c


PC2 - % Variance Explained 7.91
0.5
0.4
0.3
0.2
0.1
0.0
−0.1
−0.2
−0.3
−0.4
−0.4
−0.3
−0.2
−0.1
0.0
0.1
0.2
0.3
0.4
PC1 - % Variance Explained 16.63
Parr
Adult
River water
Smolt
Marine Adult

a.







N = 7 N = 36 N = 16 N = 35 N = 5

Genus

Aliivibrio
Aquabacterium
Candidatus_Planktophila
Caulobacter
GKS98_freshwater_group
hgcI_clade
Hyphomicrobium
LD28_freshwater_group
Limnobacter
Limnohabitans
Lysinimonas
Methylobacterium
Moritella
Mucilaginibacter
Mycobacterium
Mycoplasma
Noviherbaspirillum
Opitutus
Pelomonas
Photobacterium
Polynucleobacter
possible_genus_04
Prosthecobacter
Reyranella
Rhizomicrobium
Sediminibacterium
Sphingomonas
Staphylococcus
Tsukamurella
Vibrio
Yersinia

Family

Comamonadaceae
Enterobacteriaceae
Gemmatimonadaceae
Hyphomicrobiaceae
KD3-93
LD12_freshwater_group
Microbacteriaceae
Mycoplasmataceae

Order

Sporichthyaceae
Burkholderiales
Rhizobiales
Sphingomonadales
TRA3-20
vadinHA64

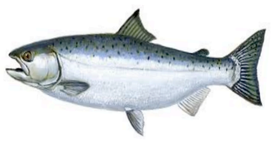
**

6000
1000
500

River Water

Parr

Smolt



Marine Adult

Returning Adult

b.

Nitrospirae
Fibrobacteres
Armatimonadetes
Elusimicrobia
Deferribacteres
Gemmatimonadetes
Acidobacteria
Synergistetes
Deinococcus-Thermus
Chloroflexi
Verrucomicrobia
Planctomycetes
Spirochaetae
Cyanobacteria
Fusobacteria
Actinobacteria
Bacteroidetes
Firmicutes
Tenericutes
Proteobacteria

*

c.

Mycoplasma
100
Mycoplasma/ Ureaplasma
100
Candidatus
100

**
*

5000
1000
100
1
0

Ireland
Greenland
Canada

Supplementary Material

Supplementary Data 1 – Methods Section

Sample collection and preparation

Parr stage Atlantic salmon were collected by electrofishing in four rivers in County Mayo, Ireland (Burrishoole catchment: 53°54''10N 9°33''26'W; Erriff river: 53°37''19N 9°39''32'W) and Quebec, Canada (Riviere St Jean: 48°49'' 31'N 64°34''15'W; Riviere Trinite: 49°24''55'N 67°19''23'W) under the supervision of the Marine Institute, Inland Fisheries Ireland and the Ministère des Ressources Naturelles et de la Faune respectively. Marine adults were sampled with the assistance of the North Atlantic Salmon Conservation Organization (www.nasco.int) in cooperation with the West Greenland subsistence fishery (Sisimiut: 66°56"33'N 53 42 28W; Maniitsoq: 65°21''11'N 53°13''50'W)

Freshwater smolt lifecycle stages were captured in rotary traps. Whole juvenile fish euthanized in MS222 were aseptically dissected to remove mid and distal intestinal contents. Mid & distal intestinal contents were collected from live returning adults and freshly landed marine adults via a pastette wash with temperature and UV sterilised phosphate buffered saline. In sampling both juveniles and adults our aim was to robustly sample (via wash and/or dissection) intestinal contents to collect a portion of adherent and allochthonous bacteria (Ringo & Birkbeck 1999). Bacterial community sampling from water was achieved from 6 litres of water sampled 1m below the surface, in the centre of the current). A two-step filtration was undertaken using peristaltic filtration equipment (Masterflex L/S Pump System with Easy-Load II Pump Head, Cole-Parmer) cleaned up with 5% HCl and rinsed with Milli-Q water before each filtration. The first stage filtration was through a 3.0 µm pore-width nitrocellulose membrane to exclude organic detritus, the second – to collect bacteria – through a sterile 0.22 µm pore-width membrane. DNA was extracted from all samples using the MoBio Powersoil DNA Isolation Kit, according to the manufacturer's protocol. Amplification of the V4 region was achieved with primers 519_f 5'-CAGCMGCCGCGGTAA-3' and 785_r 5'-TACNVGGGTATCTAATCC-3' using Takara *Taq* Polymerase (CloneTech, USA) and a final concentration of 1 pM of each primer. V4 was chosen in the light of its widespread use to profile vertebrates-associated microbiota as well as its suitability for Illumina paired end sequence read lengths at the time of sequencing (Werner *et al.* 2012). Each primer was 5' tagged with a common 22 base pair tag for barcode attachment (CS1-ACACTGACGACATGGTTCTACA; CS2-TACGGTAGCAGAGACTTGGTCT). Reaction conditions were 95°C for five minutes, followed by 30°C cycles and of 95°C for 30 seconds, 55°C for 30 seconds and 72°C for 30 seconds, followed by a final elongation step of 72°C for 10 minutes. Each amplification was

run in triplicate and pooled to minimise PCR bias, purified using an AxyPrep™ Mag PCR Clean-Up Kit (Corning, USA) and sequenced in a single run on an Illumina MiSeq platform.

Sequence analysis

A total of 3,059,424 sequences were recovered across the dataset after error checking, trimming (SICKLE (Joshi & Fass 2011)) and assembly of each paired end read into a single overlapping 290bp fragment (PANDASeq (Masella *et al.* 2012)). Sequences were kmer aligned against the *E. coli* 16S rRNA gene (95.5% reads aligned) and trimmed in Mothur (Schloss *et al.* 2009) prior to operational taxonomic unit clustering in UPARSE at 97% identity (Edgar 2013). Putatively chimeric OTUs were filtered out in reference to the genomes online database (GOLD) in UCHIME (Edgar *et al.* 2011). Subsequently, the following steps were undertaken in QIIME (Kuczynski *et al.* 2012): After exclusion of chimeric OTUs, samples containing <11,000 reads were discarded and all samples were rarefied to an even depth of 11,000 reads. 11,000 was chosen as a cut off to maximise inclusion of samples while ensuring sufficient read depth to endure saturation. Supplemental Figure 1 shows Chao1 rarefaction curves for each sample. A final dataset of 101 samples was selected (Supplementary Table 1). OTUs with fewer than 100 reads or that only occurred in a single sample were filtered out as a step to improve accuracy and diversity estimates (Bokulich *et al.* 2013). OTU richness and diversity was estimated using the Chao1 (Chao 1984) and Shannon index respectively. Alpha diversity comparisons between groups were made with non-parametric Kruskal-Wallis tests as a Shapiro-Wilks test clearly rejected normality in the dataset (Chao1, $p < 1.5 \times 10^{-5}$, Shannon $p = 0.0053$). Beta diversity estimations between samples were based on the unweighted unifrac metric (Lozupone *et al.* 2011). All OTU taxonomical assignments were made against the SILVA dataset and 'core' microbiota in each lifecycles stage were defined as being present in 85% of samples. On completion of QIIME analysis, the relationship between microbiome composition and different countries, sample sites and lifecycles stages was tested using a permutation based multivariate analyses of variance (PERMANOVA) in ADONIS in the Vegan package in R (Oksanen *et al.* 2015).

Supplementary Table 1 – Samples and provenance

Site	Parr	Smolt	Adult	Water samples
Burrishoole, Ireland	7	4	12	2
Erriff, Ireland	7	4	12	1
St Jean, Quebec, Canada	10	7	6	0
Trinite, Quebec, Canada	12	1	7	2
Maniitsoq, Greenland	0	0	3	0
Sisimiut, Greenland	0	0	4*	0

*Included two fish of North American and two fish of European origin.

Supplementary Table 2 – Summary of statistical comparisons made and sample sizes involved.

Test Summary	**Test groups (sample sizes compared)**	Kruskall-Wallis Test (Chao1)*	Kruskall-Wallis Test (Shannon)*	PERMANOVA (Unifrac)*
Microbiome Identity by study site	Freshwater St Jean (23) v Freshwater Trinite (20) v Freshwater Burrishoole (23) v Freshwater Erriff (23)	-	-	0.7264
Effect of study site on microbiome diversity (Juveniles only)	Freshwater Juvenile St Jean (17) v Freshwater Juvenile Trinite (13) v Freshwater Juvenile Burrishoole (11) v Freshwater Juvenile Erriff (11)	*0.003*	*0.009*	-
Microbiome identity between Adults and Juveniles	Marine (including returning adults) (52) vs Freshwater Juveniles (44)	-	-	*0.019*
Microbiome diversity by lifecycle stage	Parr (36) vs Smolt (16) vs Adult (37) vs Marine Adult (7)	*0.0001*	0.276	-
Microbiome identity between 'saltwater' lifecycles stages	Returning Adult (37) vs Marine Adult (7)	-	-	0.522
Microbiome diversity between freshwater lifecycles stages	Parr (36) vs Smolt (16)	$1.8x10^{-5}$	*0.0105*	
Microbiome identity between freshwater lifecycles stages	Parr (36) vs Smolt (16)	-	-	0.268

* p values

Supplementary Figure 1 - Box plot showing alpha diversity (Shannon diversity estimator) variation across sites and lifecycle stages: blue bars – Burrishoole, Ireland; pale blue bars – Erriff, Ireland; red bars – St Jean, Canada; pink bars Trinite, Canada. Marine adults from West Greenland are represented in green.

Supplementary Figure 2 – Alpha rarefaction curves of all 101 samples analysed (Chao1 richness estimator)

References

Bokulich NA, Subramanian S, Faith JJ, Gevers D, Gordon JI, Knight R, et al. (2013) Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. Nat Methods 10, 57-59.

Chao A (1984) Nonparametric estimation of the number of classes in a population. Scandinavian Journal of statistics, 265-270.

Edgar RC (2013) UPARSE: highly accurate OTU sequences from microbial amplicon reads. Nat Methods 10, 996-998.

Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R (2011) UCHIME improves sensitivity and speed of chimera detection. Bioinformatics 27, 2194-2200.

Joshi N, Fass J (2011) Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files

(Version 1.33).

Kuczynski J, Stombaugh J, Walters WA, Gonzalez A, Caporaso JG, Knight R (2012) Using QIIME to analyze 16S rRNA gene sequences from microbial communities. Curr Protoc Microbiol Chapter 1, Unit 1E 5.

Lozupone C, Lladser ME, Knights D, Stombaugh J, Knight R (2011) UniFrac: an effective distance metric for microbial community comparison. ISME J 5, 169-172.

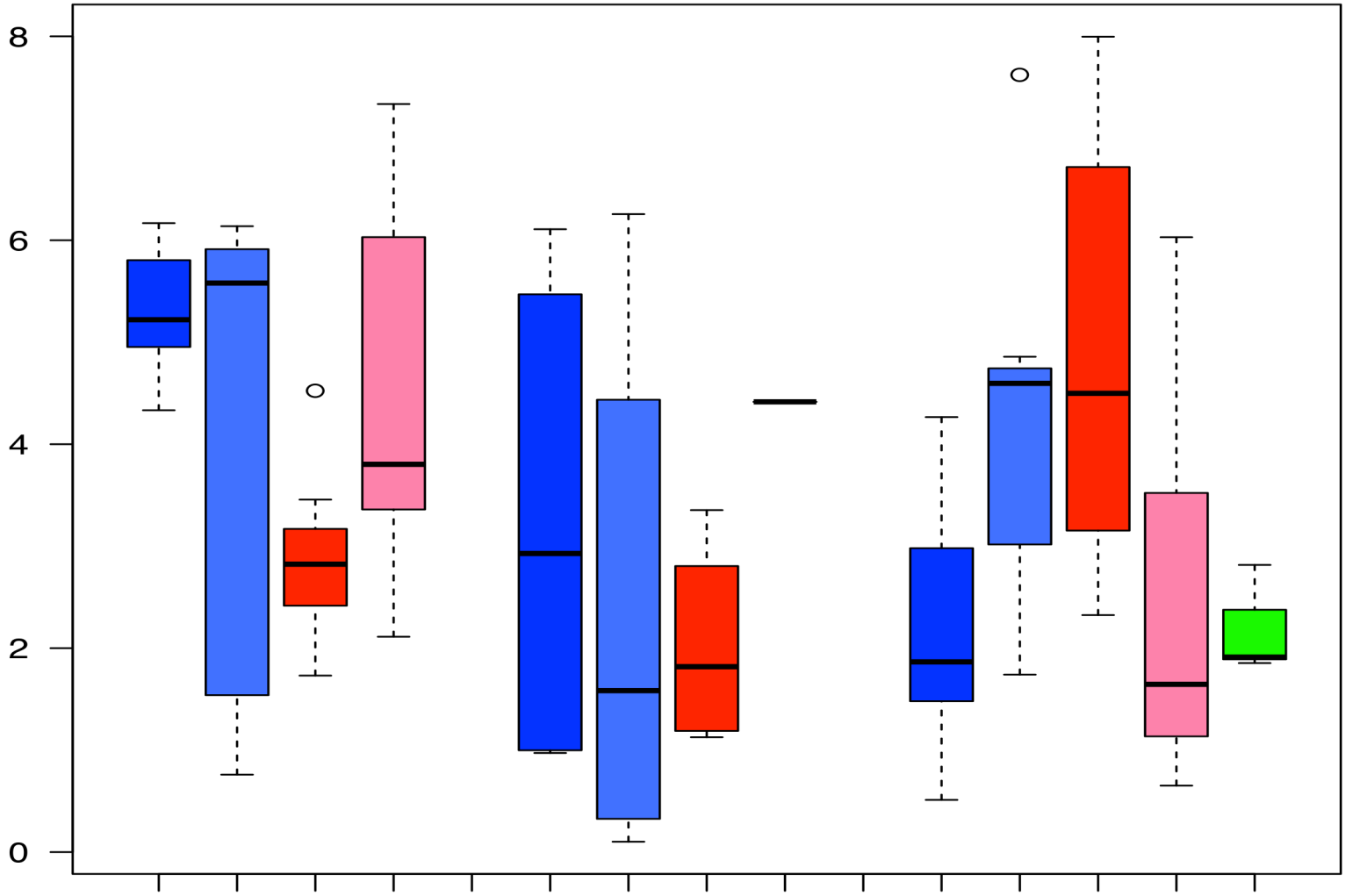
Masella AP, Bartram AK, Truszkowski JM, Brown DG, Neufeld JD (2012) PANDAseq: paired-end assembler for illumina sequences. BMC Bioinformatics 13, 31.

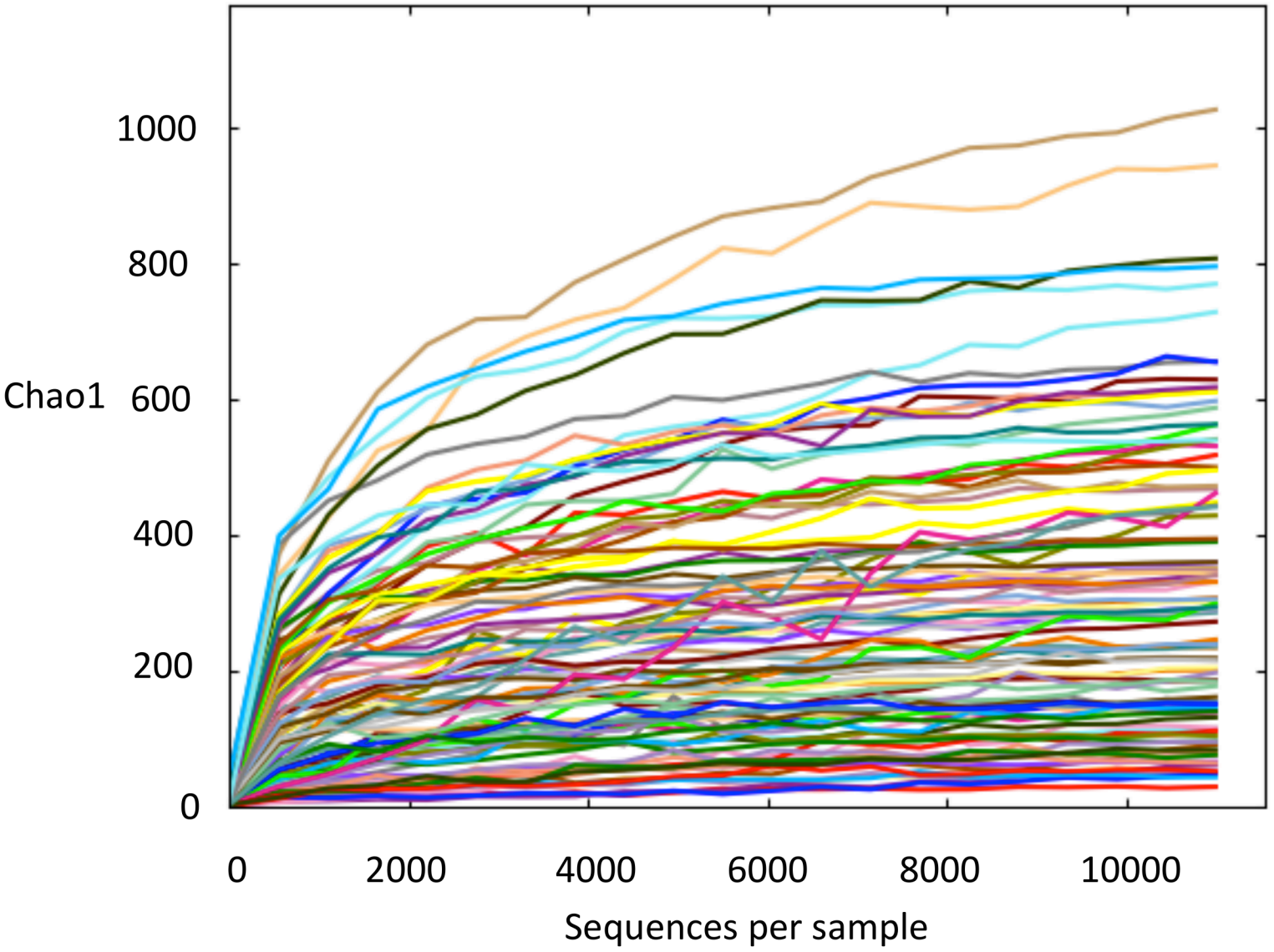
Oksanen J, Blanchet F, Kindt K, Legendre P, Minchin P, O'Hara R, et al. (2015) vegan: Community Ecology Package. R package version 2.2-1. http://CRAN.R-project.org/package=vegan.

Ringo E, Birkbeck T (1999) Intestinal microflora of fish larvae and fry. Aquaculture Rsearch 30, 73-93.

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et al. (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol 75, 7537-7541.

Werner JJ, Zhou D, Caporaso JG, Knight R, Angenent LT (2012) Comparison of Illumina paired-end and single-direction sequencing for microbial 16S rRNA gene amplicon surveys. ISME J 6, 1273-1276.


8
6
4
2
0


Chao1
1000
800
600
400
200
0
0
2000
4000
6000
8000
10000
Sequences per sample