Supporting Information

S.1 Fish sampling, processing, water measurements and DNA amplification procedures

Licenses to fish for scientific purposes were obtained in accordance with section 52 of the general fishery regulations of the Fisheries Act, Department of Fisheries and Oceans Canada (DFO). Animal care permits were issued by the Freshwater Institute Animal Care Committee of DFO (S-18/19-1045-NU and FWI-ACC AUP-2018-63). Fish sampling was performed over two years using fishing nets, either commercial (140 mm mesh) or multi-mesh (5 or 8 panels of 38–140 mm), during the August-September char migration preceding ice-up in rivers and estuaries. Both net types were also used in December-January, in freshwater lakes and rivers under ice, as well as in May-June when these were under ~2 m thick ice. Net soak times ranged from 4-24 h. Fish were photographed, weighed, and measured, followed by dissection in an on-site mobile lab. Entire intestines were removed using aseptic technique, then placed in sterile sample bags and frozen at −20°C for subsequent shipping and storage. For age interpretation, otoliths were collected and dried overnight at room temperature and subsequently prepared and transversely thin-sectioned through the origin, then ground and polished (Casselman, 2015) in the Calcified Structure Age and Growth Interpretation Laboratory at Queen's University using validated age-assessment techniques for slow-growing Arctic fishes (Campana *et al.*, 2008). Specific conductance of surface waters at fishing sites was measured using a Model 30 conductivity meter (YSI, Yellow Springs, OH).

Deleted: Inc.

Dissected intestines were partially thawed to excise three slices within the distal large intestine (~2 cm from the vent), comprising a total of 5–100 mg of epithelial tissue, avoiding any faeces and connective tissue. The slices were pooled and DNA extracted using the UltraClean Tissue and Cells DNA Isolation Kit (QIAGEN, Toronto, ON) following the manufacturer's instructions with the exception that pure water was used to elute DNA. Intestinal DNA extracts were quantified with the QuantiFluor ONE dsDNA System (Promega Corporation, Madison, WI, USA) and a Qubit 4 Fluorometer (Invitrogen). If required, samples were diluted to 30 ng μL^{-1} prior to polymerase chain reaction (PCR) amplification so that all extracts ranged between 1 and 30 ng μL^{-1}. Pre-amplification of the variable V1–V9 region of the bacterial 16S ribosomal RNA (rRNA) gene used primers 8F and 1406R (Lane, 1991; Coolen *et al.*, 2005). Each reaction mix (50 µL total volume) contained 2.5 U DreamTaq DNA polymerase (ThermoFisher) and the corresponding DreamTaq Buffer with 2 mM $MgCl_2$, 0.4 µM forward and reverse primers, 200 µM dNTPs, 400 ng bovine serum albumin (BSA), and 2 µL of template, with the reaction conducted as follows: 95°C for 5 min, 25 cycles of 95°C for 1 min, 52°C for 1 min, 72°C for 1 min, with a final extension at 72°C for 7 min.

Preparation for high-throughput sequencing involved a nested PCR conducted by amplification of the V4-V5 region of the 16S rRNA gene using primers 515F-Y (Parada *et al.*, 2016) and 926R (Quince *et*

al., 2011). Each primer contained a 6-base index sequence for sample multiplexing as well as Illumina flow cell binding and sequencing sites (Bartram *et al.*, 2011). The PCR mix (25 µL total volume) contained 1X ThermoPol Buffer, 0.2 µM forward and reverse primers, 200 µM dNTPs, 15 µg BSA, 0.625 U Hot Start *Taq* DNA polymerase, and 1 µL of template from the previous PCR described above. Amplification was performed as follows: 95°C for 3 min, 35 cycles of 95°C for 30 s, 50°C for 30 s, 68°C for 1 min, and a final extension at 68°C for 7 min. Each sample was amplified in triplicate and pooled subsequently. Although samples that did not yield a visible PCR product were not processed further, all no-template controls (NTCs) were added to the sequencing pool (5 µL), even when amplicons were not detected. Pooled 16S rRNA gene amplicons were subsequently excised from an agarose gel and purified using the Wizard SV Gel and PCR Clean-Up System (Promega, WI, USA). A 5 pM library containing 15% PhiX Control v3 (Illumina, NB, Canada) was sequenced on a MiSeq (Illumina, CA, USA) using a 2 × 250 cycle MiSeq Reagent Kit v2 (Illumina Canada). Procedures for DNA sequencing and analysis of water samples have been described previously (Hamilton *et al.*, 2019).

S.2 Experimental procedures: Controls for contamination and reproducibility

The collection of fish and subsequent DNA extractions took place over a two-year period, with PCR amplification in multiple 96-well PCR plates and sequencing in two separate MiSeq runs. To verify PCR and sequencing consistency, we used replicate samples. For these, DNA extracts from some samples that were PCR amplified and sequenced in the first year were re-amplified and sequenced again in the second year. Additionally, randomly selected samples from the different sample sites were included in multiple PCR plates to test for differences in amplification between plates. No-template-controls (NTCs) were used to assess the contribution of contaminant DNA sequences. For the first year of sequencing, NTCs were placed within 96-well plates (hereafter referred to as “in-plate” controls) and consisted of water used to elute the DNA extracts. Because DNA used in the first sequencing year had been extracted using double-distilled water (ddH_2O), whereas samples in the second year used MilliQ filtered water, controls in the second year included a ddH_2O well to detect any differences resulting from the type of pure water used. Second year controls were PCR amplified at the same time as samples, but were contained in separate tubes (hereafter referred to as “out-plate” controls) and used MilliQ filtered water that had been run through the entire extraction protocol to identify contaminants potentially originating from the extraction kit (Salter *et al.*, 2014).

S.3 Results: Sample yields, seasonal habitats, and intestinal content

Arctic char intestines were dissected from 538 specimens from fishing sites on, and within 180 km of, King William Island, Nunavut. Of these, 500 DNA extractions were performed (including 9 replicate samples), sequenced and yielded 399 intestinal communities from 9 distinct fishing sites. Generally, fish could be assigned to one of four types of seasonal habitat, based on the site geography, salinity, and time of year (Table 1). The first phase of the autumn upriver migration where Arctic char

were caught near the sea shore before entering river systems was designated as the autumn brackish run (ABR), with those sampled further up river defined as the autumn freshwater run (AFR). Collections made from a frozen lake (site 1) represented winter freshwater overwintering (WFO) habitat, and those from this lake as well as others under thick ice were designated as spring freshwater overwintering (SFO) habitat. Of the autumn samples, 13 were caught from a site that was not analyzed for water conductivity and thus could not be confidently assigned, but were nonetheless retained for analysis.

Of the 399 samples, 29 were rejected due to low sequence read counts implementing controls for contamination (S.4). Of the remaining 370 (including 8 replicate samples), 10 of the source intestines contained evidence of piscivory, with 5 each from both autumn habitats (5 intestinal samples were too small to assess). Evidence of feeding and parasitism were more frequently seen in autumn samples than in overwintering fish (Figure S7), with 79% (48/61) of ABR samples containing faeces and 25% (15/61) containing parasitic worms. As well, 32% (31/98) of AFR samples contained faeces and 23% (23/98) contained parasitic worms. Of the autumn samples for which salinity was not classified, 46% (6/13) contained faeces and 54% (7/13) contained parasitic worms. For the WFO samples, only 5% (2/44) contained faeces and none contained parasitic worms, and these proportions were 25% (37/146) and 3% (5/146) for SFO samples respectively.

Deleted: , respectively

OTU tables comprised bacterial sequences, with no archaeal OTUs identified. There were 28,983 unique OTUs in a total of 2,111,944 sequence reads that averaged 5,306 reads per sample with a standard deviation (SD) of 7,892 reads. This OTU table was later rarefied to 2,016 reads (S.4). A separate rarefied OTU table (1000 reads) contained both OTUs from the water and intestinal samples for comparison. Overall, less than 8% of the OTUs from any fishing site were shared between the water and intestinal samples (Figure S1). The 2,016 OTU table was used for all subsequent analyses. PERMDISP testing of Bray-Curtis community dissimilarity showed that male and female intestinal communities did not have significantly different dispersions ($p > 0.05$; with gender explaining only about 1% of the variation at R^2 = 0.01, pseudo-F = 3.3, $p < 0.001$), and thus in all subsequent analyses, males and females were grouped together.

Deleted: =

Deleted: =

S.4 Results: Contamination and reproducibility

In order to increase confidence in sample-specific microbiome profiles, NTCs were assessed for contaminant OTUs and their relation to samples from the entire multi-salmonid-species dataset. Ordinations indicated that these NTC profiles were associated with two distinct clusters, corresponding to in-plate and out-plate controls used in the first and second year of sequencing respectively (Figure S8). The two extraction controls from 2018 that contained ddH_2O and MilliQ water (NTCs 11 and 16 respectively) grouped tightly with the other controls from that year. However, several NTCs grouped closely with intestinal communities, and some NTC OTUs were associated with genera known to occur in fish intestinal samples, including *Mycoplasma* and *Pseudomonas* (Figure S9). Therefore, we conservatively filtered from the entire data set, any NTC OTUs that were observed more than 27 times across all NTCs (a rate of more than 3 reads per NTC). In addition, a rarefaction plot showed that samples could be rarefied to as low as 2000 reads while retaining most of the available Chao1 diversity

Deleted: , respectively

Deleted: , respectively

(Figure S10), and thus 29 samples with fewer than 2000 sequence reads after filtering were discarded. The corresponding OTU table was rarefied to a depth of 2016 reads per sample for subsequent analyses.

Ordination of *S. alpinus* microbial communities using PCR plates with randomly selected samples as a factor showed that any amplification bias, if present, was minor (Figure S11 A). The SFO samples from a single PCR plate (plate "Three2017") grouped separately within an ordination from the other plates. However, all samples in this plate originated from a single sampling season and it was not possible to determine if the grouping was due to the samples or PCR effects. Therefore, these 53 samples were omitted from subsequent statistical analysis.

Analysis of replicate samples revealed some differences in sequencing results depending on the year sequences were obtained, although in some cases replicate samples grouped well and contained very similar taxonomic compositions (Figure S11 B and Figure S12). All replicate samples from the second year of sequencing were removed in subsequent analyses.

The OTU filtering method used to remove contaminant OTUs in this study has not been previously documented and thus to further examine the analysis, the hypothesis that seasonal habitat is a driver of community composition was also tested using all the samples with an unfiltered OTU table. The OTU table was rarefied to the lowest number of reads in a sample (6098), then PERMANOVA was performed. The results, showing dissimilarity between seasonal habitats, were highly similar to those derived from the filtered table (R^2 = 0.12, pseudo-*F* = 14.9, *p* value <0.001), suggesting that removal of contaminant OTUs did not impact the overall conclusions.

Deleted: =

Deleted: =

S.5 Results: SIMPER analysis

To identify bacteria that defined each seasonal habitat, a similarity percentages (SIMPER) test at the genus level was performed (Table S1). *Photobacterium* and other OTUs associated with *Vibrionaceae* together contributed >50% of similarity within the ABR habitat, whereas *Mycoplasma* and sequences associated with *Mycoplasmataceae* distinguished the AFR habitat with ~64% of within group similarity. Given that the ratios of similarity/SD for these taxa were less than one, this suggested some varying abundances between fish. In contrast to autumn samples, 11 – 21% of similarity between overwintering fish was contributed by *Pseudomonas*, but again there was variation (similarity/SD = 0.3). Much of the remaining similarity between samples from the overwintering habitats was contributed by unclassified OTU sequences (~17 – 23%), with these appearing to be the most consistent in abundance (similarity/SD = 0.8 – 1.1). Unclassified bacterial sequences contributed 6.7% of the similarity within the ABR habitat, with similar abundance consistency to those in the overwintering environments.

Deleted: =

Deleted: =

S.6 Microbial community diversity

Variation in microbial community diversity was analyzed with respect to sample site, seasonal habitat, sex, presence of faeces and parasitic worms, and evidence of piscivory. Both Chao1 and Shannon indices appeared to have lognormal distributions and so were $\log_{10}$ transformed to

approximate normal distributions. Only sample site and seasonal habitat were significantly correlated with both diversity measurements ($p < 0.05$), whereas the presence of faeces was positively correlated with Chao1 only ($R^2 = 0.04$, $p = 0.008$). For both metrics, stepwise model selection revealed that all factors other than seasonal habitat could be removed from the linear regression model without a significant loss of explanatory power. For this minimal acceptable model, there were significant differences between seasonal habitats for Chao1 ($F = 4.0$, $p = 0.008$) and Shannon index ($F = 5.8$, $p = 7.2e\text{-}4$). Tukey's HSD tests revealed that the significant differences in mean Chao1 were between the autumn habitats ($p = 0.024$), as well as between ABR and SFO ($p = 0.006$), with ABR slightly greater than the other habitats. For Shannon indices, significant differences in means were between autumn habitats ($p = 0.004$) with ABR slightly greater, AFR and WFO ($p = 0.028$) with WFO slightly greater, and AFR and SFO ($p = 0.006$) with SFO slightly greater.

To further account for the effects of geography on community diversity, a single migratory system, represented by sites along the Back River migration route (Figure 1) including site 6 (ABR), sites 18 and 19 (AFR), and site 13 (SFO), was interrogated for differences in mean diversity. Here there appeared to be less structure in the distribution of residuals for the Shannon model when this variable was not $\log_{10}$ transformed. Within the Back River system, there were again differences between seasonal habitats for the Chao1 ($F = 4.3$, $p = 0.016$) and Shannon index ($F = 3.4$, $p = 0.037$) (Figure S3). For Chao1, the significant difference in mean diversity was between the ABR and SFO sites ($p = 0.014$) and for the Shannon index, the significant difference was between the autumn habitats ($p = 0.030$).

These differences between habitats prompted the separate testing of continuous factor models within each habitat. Age and Fulton's condition factor ($K = 100\,000*(\text{weight}/\text{length}^3)$) were considered, but for all four seasonal habitats, the null hypotheses that K and age explained none of the variation seen in Chao1 and Shannon indices could not be rejected ($p > 0.05$), either when testing these variables separately or in a multiple model.

Deleted: =
Deleted: =
Deleted: =
Deleted: =
Deleted: =
Deleted: =
Deleted: =
Deleted: =
Deleted: =
Deleted: =
Deleted: =
Deleted: =
Deleted: =
Deleted: =
Deleted: =
Deleted: =
Deleted: =