

Microbial biogeography of streams in the Feniak Lake region, Noatak National Preserve, Alaska: Exploring influences of lithology and habitat

Julia R. Larouche¹
William B. Bowden¹
Rosanna Giordano²
and
Michael B. Flinn³

¹Rubenstein School of Environment and Natural Resources, University of Vermont, Burlington, VT. email: julia.larouche@uvm.edu / breck.bowden@uvm.edu

²Institute of Natural Resource Sustainability, University of Illinois at Urbana-Champaign, IL. email: giordano@life.uiuc.edu

³Biological Sciences, Murray State University, Murray, KY. email:

Final Report to the National Park Service
Arctic Network
Fairbanks, Alaska

31 December 2009

Abstract

Microbes are of critical importance but are a poorly understood component of arctic stream ecosystems. They are responsible for recycling organic matter and regenerating nutrients that are essential to the food webs of aquatic ecosystems. We tested the hypothesis that differences in highly contrasting parent lithologies (non-carbonate and ultramafic), stream habitat (sediments and rocks), and stream biogeochemistry influence the structure of bacterial biofilm communities in arctic streams. Terminal restriction fragment length polymorphism and 16S rRNA gene sequencing were used to explore the genetic diversity of bacterial communities in biofilms on sediments (epipssamon) and rocks (epilithon) in stream reaches that drain watersheds with contrasting lithologies in the Noatak National Preserve, Alaska. We observed different patterns in bacterial community composition primarily by stream habitat and secondarily by lithology. Non-metric multidimensional scaling ordination of terminal restriction fragment peaks showed highly significant separation between the two habitats and moderately significant separation between lithologies. Positive correlations were detected between particular biogeochemical variables (e.g. nutrients, base cations, and dissolved organic carbon) and bacterial community structure unique to each lithology. Although clone libraries indicated high diversity of bacterial operational units within and across stream sites, biogeographical patterns were observed that depended on lithology type and stream habitat. Our results show relationships at macro- and micro-scales at the landscape level and in ecological niches within a single stream.

Introduction

The interpretation of results from microbial biogeographical studies in natural environments has become increasingly controversial in recent years (Hughes-Martiny *et al.*, 2006). Typing microbial communities has been acknowledged as an important, yet neglected area of ecological research, especially in reference to the interaction between bacteria, the characteristics of their ecosystems, and their response to environmental changes (Green *et al.*, 2008). The historical view of microbial biogeography as formalized by Baas-Becking (1934), states that “everything is everywhere, but the environment selects”. Or as otherwise stated, global microbial diversity is low, consisting of a relatively small number of cosmopolitan species with high dispersal capabilities, and relatively high local diversity shaped by contemporary environmental conditions that act as filters to maintain distinctive microbial assemblages. Recent molecular studies indicate that biogeographical patterns in microbial distribution and diversity (Fierer *et al.*, 2007; Gray *et al.*, 2007) correlate with environmental variables (Fierer *et al.*, 2007; Hughes-Martiny *et al.*, 2006), as well as classic patterns of taxa-area relationships driven by environmental heterogeneity (Horner-Devine *et al.*, 2004).

Knowledge of microbial community composition in freshwater and in particular arctic stream ecosystems, is limited compared to what is known about microbial structure and function in terrestrial or marine environments (Logue *et al.*, 2008). The importance of sessile bacteria in aquatic environments has been recognized; an example is the early biofilm research that focused on alpine streams (Geesey *et al.*, 1978; McFeters *et al.*, 1978). However, despite the likely importance of their role in arctic stream ecosystems, basic characteristics of microbial communities in arctic streams remain poorly understood. Microbial activity in stream ecosystems is influenced by the degree of chemical and biological interactions between the terrestrial and aquatic interface (Palmer *et al.*, 2000; Stanford and Ward, 1993; Ward and Stanford, 1995). Various biotic and abiotic factors likely influence microbial community composition in aquatic ecosystems, including physical variables (i.e. temperature variations, climate, topography, and light availability) (Autio, 1998; Kaplan and Bott, 1989) and biogeochemical variables (i.e. underlying and surrounding lithology, terrestrial vegetation and solute composition including the quality and quantity of carbon sources, inorganic nutrients, and electron acceptors) (Crump *et al.*, 2003; Drever, 2002; Eiler *et al.*, 2003).

The few studies that have sought to characterize microbial communities from aquatic ecosystems in arctic Alaska have focused on environments such as lakes (Crump *et al.*, 2007; Crump *et al.*, 2003), sub-arctic sub-glacial streams (Skidmore *et al.*, 2005) and the coastal Arctic Ocean (Galand *et al.*, 2006; Garneau *et al.*, 2006). However, the current understanding of the relationship between taxonomic groups, their genetic diversity, and varying environments in arctic Alaska environments and in particular stream ecosystems remains unclear. If microbial communities and their biogeochemical environment influence each other, it is necessary to understand this interaction in an environment currently responding to a rapidly changing climate. Few studies have investigated the effect of lithological and biogeochemical differences on microbial community structure (Oline, 2006; Skidmore *et al.*, 2005; Takai *et al.*, 2003) or the effect of habitat differences within a stream that reflect differences in organic matter availability and degree of hydrologic stressors (Hullar *et al.*, 2006). Takai (Takai *et al.*) found that major archaeal organisms present in the groundwater and substratum associated with subsurface Cretaceous rock shifted over small scales (tens of centimeters) to changes in the lithology and geochemical gradients. Skidmore *et al.* (Skidmore *et al.*, 2005) suggests a reasonable assumption

correlating microbial community composition with chemical weathering products controlled by bedrock mineralogy of two geographically distant glaciers. The work of Hullar *et al.* (Hullar *et al.*, 2006) suggests the presence of stable seasonal oscillations in bacterial community structure of stream habitats, sediment and epilithon. Hullar's work also determined that these two habitats were composed of both terrestrial and aquatic derived microorganisms, suggesting a close association between headwater streams and their watersheds.

We hypothesized that these fundamentally different lithologies not only influence terrestrial productivity but also impart different biogeochemical characteristics to water, which in turn influences the structure and function of the biological communities in stream ecosystems, in particular, the bacterial community. Our objectives were to determine the degree to which: (1) stream bacterial community composition differ among streams selected from the three contrasting lithologies (non-carbonate, complex sedimentary, and ultramafic), (2) differences are correlated with the biogeochemical characteristics of the host stream ecosystem, and (3) bacterial community composition differs by stream habitat (sediment vs. epilithon) within each individual stream. Our conclusions are based on T-RFLP and sequencing of 16S rRNA bacterial clones. Results suggest that there are differences in bacterial community composition both small and large scales. Differences in community composition across lithologies and between stream habitats can be related to linkages between streams and the terrestrial environment in addition to varying degrees of resource availability.

Materials and methods

Study Area

The Noatak River is in the Noatak National Preserve in Alaska (USA). It is the longest continuous river in the U.S. National Wild and Scenic system and the largest mountain-ringed river basin, virtually unaltered by human activity (Milner *et al.*, 2005). Jorgenson *et al.* (Jorgenson *et al.*, 2002) studied the relationship between lithology and vegetation composition in the Noatak Basin. The lithology of this area is complex but includes three important and strongly contrasting types that were the focus of this study. Ultramafic rocks (basalt, gabbro, peridotite, pyroxenite, dunite) of the Siniktanneyak mountains tend to be high in iron and magnesium with sparse vegetation. Non-carbonate rocks (glaciolacustrine deposits, conglomerate, sandstone, shales) of the Avingyak Hills support acidic, organic-rich soils and host shrub birch, willow and ericaceous plants. Complex sedimentary rocks (shale, basalt, limestone and mafic rocks) of the Aniuk mountains support vegetation similar to the non-carbonate lithology (Jorgenson *et al.*, 2002). Jorgenson *et al.* found that vegetation composition differs with lithology type and that the relationship is a consequence of variations in soil pH and possible phytotoxic effects of soluble minerals. Samples in this study were collected from headwater stream tributaries arising from uniform and contrasting lithologies within the Noatak National Preserve. We sampled 30 streams located in the Noatak River basin (Figure 1) in the vicinity of Feniak Lake (68°14'56.55" N and 158°19'19.90" W, elevation 1411 feet). Fifteen of these streams were used in the bacterial community analyses described in this paper.

Sampling Design

A suite of first and second-order streams were sampled within each separate lithology: four in the non-carbonate, five in the ultramafic and six in the complex sedimentary. Sediment and rock samples were collected at each location. The former were obtained in triplicate along a

25-meter reach while the latter were a composite of six rock scrubs from each location (see sample collection details below). A series of analyses to determine stream biogeochemistry were conducted. See Flinn *et al.* (in preparation) for the details of the streamside chemistry sample collection and processing.

Sample Collection of Benthic Microbial Communities

Sediment samples were collected in sterile 15-ml plastic tubes from the surface of the bottom of the stream to a depth of approximately 3 to 5 cm. Samples were preserved immediately by adding a sucrose lysis buffer (SLB: 20 mM EDTA, 400 mM NaCl, 0.7 M sucrose, 50 mM Tris pH 9.0) in a 1:1 ratio and then frozen on dry ice while in the field and later transferred to a -80°C freezer for long-term storage.

Epilithic material from the tops of six submerged rocks was obtained from riffle sections of each streambed sampled. Rocks were scrubbed with a nylon brush and the liberated biofilm was collected by squirting filtered (0.22 μ) stream water over the rock surfaces into a sterile plastic container. Water and biofilm material was transferred into a sterile syringe and filtered through a filter capsule with a 0.22 μ membrane to trap most of the bacteria. Filter capsules were removed from syringes and 1 ml of DNA extraction buffer (100 mM Tris (pH 8), 100 mM NaEDTA (pH 8), 100mM phosphate buffer (pH 8), 1.5 M NaCl, 1% CTAB) was injected into the filter capsule using a sterile syringe. Filter capsules were kept on dry ice in the field until long-term storage at -80°C.

DNA Extractions

DNA extractions were conducted using the MoBio Power Soil DNA extraction kit (MoBio Laboratories, Inc. Carlsbad, CA, USA) following the manufacturer's protocol with the following modification: a FastPrep Homogenizer and Isolation System (Thermo Fisher Scientific Waltham, MA, USA) was used to shake tubes at 4.5m/sec for 30 seconds to ensure complete cell lysis of bacteria in sediment and rock biofilm samples. Using sterile conditions, 500- μ l subsamples of streambed sediment (1 : 1 sediment : SLB slurry) were placed in the DNA extraction tubes provided with the MoBio kit. Similarly, filters of the rock biofilm samples were removed from their capsules and transferred to DNA extraction tubes. DNA was extracted immediately prior to downstream applications to avoid degradation and bulk DNA was stored at 4°C.

T-RFLP Profiles

Terminal-Restriction Fragment Length Polymorphism (T-RFLP) analysis was conducted on all triplicate sediment samples (42 total) and epilithon samples (8 total). The 16S rRNA gene was amplified from both sediment and rock biofilm samples for T-RFLP analysis via polymerase chain reaction (PCR) using the following primers: Bac8f (5' AGAGTTTGATCCTGGCTCAG, HEX labeled) (Reysenbach *et al.*, 1994) and an unlabeled reverse primer, Univ1492r (5'-GGTTACCTTTGTTACGACTT) (Edwards *et al.*, 1989). Forward and reverse primers were obtained from Sigma-Genosys (St. Louis, MO, USA) and Invitrogen (Carlsbad, CA, USA), respectively. PCR reactions were run using Illustra PuReTaq Ready-To-Go PCR Beads (GE Healthcare Life Sciences, Piscataway, NJ, USA) using 2 μ l of genomic DNA. The PCR reaction

protocol was as follows: initial denaturation at 94°C for 4 minutes, followed by 40 cycles of 94°C for 45 seconds, 54°C for 20 seconds, 72°C for 2.5 minutes with a final 4 minutes at 72°C. Two separate PCR reactions were performed for each DNA sample. Presence of PCR products were confirmed by running 2- μ l on a 1% agarose gel at 90 volts for 30 minutes and stained with ethidium bromide. For each sample the two PCR products were pooled and digested separately with three different restriction enzymes: *MspI*, *AluI*, and *HinPII* (New England BioLabs, Beverly, MA, USA). The restriction digest mixture consisted of 10 μ l of PCR product, 1 unit of restriction enzyme, and 2 μ l of 10X reaction buffer 2 (New England BioLabs) brought up to a total volume of 25 μ l with Sigma water (Sigma-Aldrich). Reactions were digested overnight at 37 °C.

Fluorescently labeled terminal restriction fragments (T-RFs) were size separated on an ABI Avant Genetic Analyzer 3100 (Applied Biosystems, Foster City, CA, USA) using an internal size standard (BioVentures MapMarker 1000, BioVentures, Inc. Murfreesboro, TN, USA). To test for variation within each digested PCR product as well as instrument performance, we ran a trial set of ten restriction enzyme digests in duplicates (1- μ l aliquots). Even though results from a preliminary assay did not show any substantial differences across duplicate profiles, we chose to run three replicates for each restriction digest to ensure reliable results.

Analysis of T-RFLP Profiles

T-RFLP electropherograms were analyzed using GeneMapper software version 3.7 (Applied Biosystems, Foster City, CA, USA). GeneMapper software calculates fragment length to 1/100 of a base pair (bp). The error associated with the determination of fragment size can be up to 0.5 bp (Dunbar *et al.*, 2001), therefore T-RF peaks that differed by less than 0.5 bp were considered the same and grouped. Common observations in T-RFLP data include primer-dimer artifact formation in the low base-pair region of T-RFLP profiles and low peak area for T-RFs due to incomplete digestion or excess noise. Given these considerations, all peaks that were sized >80 bp with >50 relative fluorescent units were included in the analysis. The raw data for each profile was examined to ensure that each peak was solely a result of the HEX-green fluorescence. Occasionally false peaks arise due to electrochemical noise from dust or bubbles present in the capillaries of the detector, therefore, careful consideration was taken to report only true peaks in the resulting profiles of each sample. For comparisons between T-RFLP profiles, normalized relative T-RF peak height and allele presence/absence data were considered. Triplicate profiles were collapsed into one average profile by including peaks that occurred in two of the three replicate profiles. T-RFs of different lengths inferred representation of distinct operational taxonomic units (OTUs) but should not be interpreted as specific bacterial species because similar restriction fragment sizes can be produced from different organisms (Liu *et al.*, 1997).

T-RFLP Statistical Analyses

T-RFLP statistical analyses were performed in DECODA (Database for Ecological Community DATA) version 3 (Minchin, 1990). Non-metric multidimensional scaling (NMS) formalized by Clarke (1993) was used for ordination of the T-RFLP data using T-RF length and normalized peak height from all three restriction enzymes as input data. Similarities between samples were based on the Bray-Curtis dissimilarity matrix (Bray and Curtis, 1957), which has

been recommended (Rees *et al.*, 2004) and commonly used for T-RFLP data (Fierer *et al.*, 2007, Denaro *et al.*, 2005, Deslippe *et al.*, 2005). Analysis of Similarity (ANOSIM) (Clarke, 1993; Clarke and Green, 1988) was used to determine which samples were most closely related with patterns of similarity between bacterial communities. The NMS describes relationships within the microbial community data, yet further relationships can be examined by superimposing corresponding environmental data on the microbial community ordination by way of a vector analysis to determine which biogeochemical variables are positively correlated with particular groups of samples that represent given microbial community assemblages.

Building Clone Libraries for 16S rRNA Gene

The sediment and rock biofilm samples from stream sites in each lithology with the highest degree of variation using T-RFLP were chosen for more detailed phylogenetic analyses. Nine clone libraries of the 16S rRNA gene were built from sediment (n=5) and epilithon (n=4) from two representative streams within the non-carbonate and ultramafic lithologies. A single clone library was built from a sediment sample in the complex sedimentary lithology. The bacterial 16S rRNA gene was amplified using primers, Bac8f (unlabeled) and Univ1492r (Invitrogen). PCR protocol was as follows: 94°C for 2 minutes, followed by 40 cycles of 94°C for 30 seconds, 54°C for 20 seconds, and 72°C for 1.5 minutes with a final extension of 15 minutes at 72°C. To minimize PCR drift in individual reactions, PCRs were run in triplicate and pooled for each of the DNA extracts from an individual stream site. PCR products were run on 0.75% agarose gels and visualized with ethidium bromide, excised with a sterile razor blade and purified with Zymoclean Gel DNA Recovery kit (Zymo Research, Orange, CA, USA), cloned into pCR[®]2.1 vector using the TA cloning kit (Invitrogen) and transformed into OneShot[®] Competent Cells (Invitrogen). Transformants were plated on Luria broth (LB) agar medium containing ampicillin, X-gal and isopropyl-β-d-thiogalactopyranoside (IPTG). Ampicillin-resistant and β-galactosidase-negative clones were randomly selected and grown overnight at 37°C in LB with ampicillin. Clones were tested for the presence of inserts by PCR amplification and ethidium bromide gel visualization. Clones with inserts were sequenced directly and grown in three mls of LB ampicillin broth for long-term storage at -80 °C. An excess of 100 clones for each sample were sequenced using primers designed to the pCR[®]2.1 vector: M13Long Forward (5'-CAGGAAACAGCTATGACCATGATTAC-3') and M13Long Reverse (5'-GTAAAACGACGGCCAGTGAATTGT-3') as well as internal 16S primers custom designed for specific clone groups to ensure complete overlap of sequence reads in both directions: 16S-A1F (5'-GTGCCAGCAGCCGCGTAATAC-3'); 16S-A1R (5'-GTATTACCGCGGCTGCTGGCAC-3'); 16S-B1F (5'-GGTGCTGCATGGCTGTCGTCAGC-3'); 16S-B1R (5'-GCTGACGACAGCCATGCAGCACC-3'); 16S-B2F (5'-GGTGGTGCATGGTTGTCGTCAGC-3'); and 16S-B2R (5'-GCTGACGACAACCATGCACCACC-3'). Clones from SedSNC03 were sequenced using the following protocol: 96°C for 1 minute, followed by 25 cycles of 96°C for 10 seconds, 50°C for 5 seconds, and 60°C for 4 minutes. Ready to load sequence reactions were run at Vermont Cancer Center, University of Vermont, Burlington, Vermont on an ABI Avant Genetic Analyzer 3100 (Applied Biosystems). PCR products of the subsequent eight clone libraries were cleaned and sent to Agencourt Bioscience Corporation (Beverly, MA, USA) for sequencing. A total of 9 clone libraries were constructed from five representative stream sites within each of the three lithologies (streams SNC03 and SNC08 from the non-carbonate lithology, SUM01 and SUM11

from the ultramafic lithology and site SCS07 from the complex sedimentary). Five sediment clone libraries were constructed at each of the five stream sites and four epilithon libraries were constructed from the two representative stream sites in the non-carbonate and ultramafic lithologies.

Processing of Sequence Data

Clone sequences were assembled and edited using Sequencher version 4.6 (Gene Codes, Ann Arbor, MI, USA). MacClade version 4.08 (Maddison and Maddison, 2002) was used to visualize aligned sequences. Basepair positions that varied between clones within each library were verified in the original contiguous Sequencher files. Sequences were aligned using the Ribosomal Database Project II (RDP) release 9.58 web resource (<http://rdp.cme.msu.edu/>) by Cole et al. (2007). The RDP makes available a general bacterial rRNA alignment model that uses a modified version of the RNACAD program that takes into consideration rRNA secondary structure in its internal model and is a Stochastic Context Free Grammar (SCFG) based rRNA aligner (Brown, 2000; Gutell *et al.*, 2002). We screened for potential chimeric sequences using the RDP's CHIMERA_CHECK program based on the Pintail algorithm by Ashelford (2005) and found X number of sequences to be chimeric. Sequences identified as chimeric were removed from the data. The final data set included a range of 70-95 non-chimeric sequences, approximately 1500 bp long, per clone library. Edited sequences were imported into PAUP (Swofford, 2001) where pairwise distances between sequences were computed and Jukes and Cantor (1969) distance matrices were exported for use in subsequent analyses.

Estimations of Richness and Diversity

Estimates of richness and diversity of 16S rRNA genes were determined for all clone libraries. Highly similar sequences were considered as part of a single operational taxonomic unit (OTU) determined using the farthest-neighbor criterion of Schloss and Handelsman's (2005) as implemented in DOTUR (version 1.53; Department of Plant Pathology, University of Wisconsin–Madison [<http://www.plantpath.wisc.edu/fac/joh/DOTUR.html>]), with a matrix of Jukes-Cantor distances used as input. An OTU was defined as having a 16S rRNA gene sequence similarity of $\geq 98\%$. The “species” level as per convention by Rossello-Mora and Amann (2001) suggests using $\geq 97\%$ sequence similarity; however we chose a slightly higher cutoff (98%) to increase the stringency of our results and subsequent conclusions. DOTUR calculated rarefaction curves based on the 98% OTU cutoff and bootstrapping procedures assessed the confidence limits of the rarefaction curves.

Clone Library Statistical Analyses

We examined variation of 16S gene sequences in clone libraries that represent bacterial communities from sediment and epilithon in non-carbonate and ultramafic lithologies. Analysis of molecular variance (AMOVA) and F_{ST} tests (Arlequin version 3.1; Genetics and Biometry Laboratory, University of Geneva: <http://lgb.unige.ch/arlequin>) were used to compare the samples. AMOVA uses a hierarchically partitioned matrix of genetic distances to assess, by permutation, the significance of variance components associated with each level of partitioning (Excoffier *et al.*, 1992). Input matrices consisted of distances computed in DNAsp (Rozas and

Rozas, 1999) and tested as follows: i) all clone libraries considered as distinct groups ($n = 9$); ii) sediment ($n = 5$) vs. epilithon ($n = 4$); and iii) ultramafic ($n = 4$) vs. non-carbonate ($n = 4$).

F_{ST} tests (Martin, 2002) were performed as tests of genetic differentiation among all pairs of samples. F_{ST} can also be considered as a measure of distance between pairs of samples that takes into account both the frequency of identical or closely related sequences and the amount of diversity within the sample pairs. A matrix of pairwise F_{ST} distances was used as the basis of the NMS plot in the program PRIMER 5 for Windows (version 5.2.7; PRIMER-E, Ltd: <http://www.primers-e.com>) to visualize samples that are similar in genetic composition. ANOSIM was also used to determine which samples were most closely associated with patterns of similarity between bacterial communities (Clarke, 1993; Clarke and Green, 1988). Additionally, Mantel tests were used to examine the relationship between geographic (straight-line distance between stream sites) and genetic distance

Taxonomic Associations of 16S rRNA Bacterial Clones

Groups of sequences were formed based on their similarities and by identifying their closest genus and species matches using the SeqMatch tool in the RDP, which has its own database as well as being linked to GenBank (Altschul *et al.*, 1990). Identification of unknown clones using GenBank data followed the guidelines established by Goebel and Stackebrandt (1994), which considers a 97-100% match an approximate identification to species level, 93-96% similarity as genus level identification, and 86-92% match a distant yet related organism. The $\geq 98\%$ sequence similarity criteria concurred with results obtained using DOTUR in that all sequences within an established OTU generally matched to at least the same genus as determined using GenBank. Unique sequences in each clone library have been submitted to GenBank under accession numbers FJ849067 to FJ849648.

Results

Biogeochemistry of stream study sites

Stream study sites with unique global positioning system location and elevation and a summary are found in Supplementary Material Table 1. Biogeochemical characteristics of stream sites by lithology are found in Table 1.

Lithology and Habitat Comparisons using T-RFLP

The NMS ordination of T-RFLP patterns, the 42 sediment samples obtained from the 14 stream sites arising in the three different lithologies form three clusters with varying degrees of overlap. A first analysis of these three groups using ANOSIM (DECODA; (Clarke and Warwick, 2001) showed differences (Global $R = 0.40$; $P < 0.001$) (Fig. 2). Pairwise comparisons indicate that ultramafic UM streams were significantly different from non-carbonate (NC) and complex sedimentary (CS) streams (UM vs. CS: $R = 0.50$, $P < 0.0001$; UM vs. NC: $R = 0.55$, $P < 0.001$; and NC vs. CS: $R = 0.06$, $P = 0.2$).

We also observed significant differences when comparing the sediment and rock habitats (Global $R = 0.98$; $P < 0.0001$) (Fig. 2). The overall number of phylotypes, or restriction fragments (T-RFs), for sediment samples ranged from 19 to 69 (mean: 51) with an average of 51

for CS, 54 for NC, and 50 for UM. T-RFs for rock biofilm samples ranged from 66 to 99 (mean: 79) with an average of 72 for CS, 86 for NC, and 75 for UM. To detect differences with the T-RFLP data we chose NMS with the ANOSIM multivariate method rather than diversity indices since it has been suggested that the use of diversity indices on T-RFLP data provides inaccurate estimates of true diversity in microbial communities (Blackwood *et al.*, 2007) and that multivariate methods (ordination and clustering) have greater sensitivity for detecting microbial community differences (Dunbar *et al.*, 2000; Hartmann and Widmer, 2006).

We examined the relationship between bacterial community structure and environmental variables in each of the three lithologies by using a vector analysis. A positive correlation was observed between bacterial communities and measured inorganic chemical variables (Bonferroni-corrected $P > 0.5$ in all cases) (Fig. 2). The vectors associated with cation variables (Ca^{2+} , Mg^{2+} , K^+ , and Na^+) were correlated with the complex sedimentary community, while the vectors associated with dissolved organic carbon (DOC) and nutrients (Total dissolved nitrogen – nitrate, (TDN- NO_3^-); Total dissolved nitrogen/Total dissolved phosphorus, (TDN/TDP)) were correlated with the non-carbonate community, and nitrate/total dissolved nitrogen, (NO_3^-/TDN) was correlated with the ultramafic community.

Lithology and Habitat Comparisons using Clone Libraries

Based on the analysis of T-RFLP data, samples that were furthest away in NMS genotype space stream sites were selected for 16S rRNA gene cloning and sequencing from each of the three lithologies differentiated in Figure 2. Two streams from NC and UM and 1 from CS were selected to further characterize the diversity of the bacterial communities in sediment and epilithon of each lithology as well as to identify specific phylogenetic groups that could be associated with the bacterial community composition pattern shown by the T-RFLP analysis. Subsequent to the removal of chimeric sequences, the number of clones for each library ranged from 77 to 95.

Analysis by Rarefaction

Using DOTUR a number of rarefaction curves were generated for each clone library. In general, sediment and rock biofilm curves were at or near plateau, indicating that we were successful in sampling nearly the full extent of bacterial species richness within each of the samples. The 95% confidence intervals of the rarefaction curves did not overlap at the highest common clone count when clone libraries were combined by lithology (UM and NC) and habitat (epilithon and sediment) (Fig 3a and 3b; 95% CIs not shown), indicating differences in species richness at both levels of partitioning. Figure 3a shows higher diversity within the UM samples and Figure 4 indicates higher diversity among NC streams sampled.

Comparisons of Clone Libraries using AMOVA

All nine communities had significantly different genetic composition ($P < 0.0001$) relative to the pool of total species in all samples. The AMOVA (Fig. 5) indicated that sediment and rock biofilm samples from UM and NC lithologies were composed of genetically different microbial communities ($P = 0.01$). It was not possible to show that microbial communities in contrasting lithologies were genetically different ($P = 1.00$; $P = 0.300$). However, pairwise

comparisons of clone library F_{ST} tests (Martin, 2002) showed that all were significantly different ($P < 0.00001$), with the exception of EpiSUM01 and EpiSNC03 ($P = 0.11$). A NMS ordination (Fig. 6) using pairwise F_{ST} comparisons, which represents distance in genetic and community composition, segregates samples by sediment and rock biofilm communities. A test of isolation of stream site bacterial community composition by geographic distance using the Mantel Test, did not show any structure ($P > 0.3$).

Taxonomic Associations of 16S rRNA Bacterial Clones

Using sequence matches obtained from GenBank and the RDP, we determined the presence of 32 families of bacteria belonging to 14 classes. The most commonly represented bacterial families in sediment samples were as follows: *Enterobacteriaceae* (25.1%), *Paenibacillaceae* (28.3%), *Pseudomonadaceae* (9.4%), and *Xanthomonadaceae* (11.6%) (Supplementary Material Table 2). Rock biofilm samples were dominated by species of *Cyanobacteria* (43.5%) as well as the following bacterial families: *Flexibacteraceae* (17.7%), *Comamonadaceae* (12.9%), and *Deinococcaceae* (5.5%) (Supplementary Material Table 2). Sediment clones included 20 families belonging to 11 classes, whereas rock biofilm clones included 16 families representing 11 classes. Sediment samples had four bacterial families that overlapped across all three lithologies. The ultramafic sediment, with 16 bacterial families, had the highest diversity at the family level, while 12 were detected in NC and five in CS. Eight bacterial families were found in both NC and UM sediment samples. Families unique to sediment NC and absent in sediment UM were *Sporolactobacillaceae* and *Aeromonadaceae*, while several families were unique to sediment UM: *Microbacteriaceae*, *Hyphomicrobiaceae*, *Carnobacteriaceae*, *Alcaligenaceae*, *Geobacteraceae*, *Flavobacteriaceae*, *Planctomycetaceae*, and *Crenotrichaceae*. Clones in the single CS library consisted of six families represented by three classes. Fifty-seven of the 82 total clones matched at the genus level (98%) to the organism *Paenibacillus borealis*.

For rock biofilm clones, nine of 16 families were shared between the NC and UM samples. In rock biofilm samples, 12 and 13 bacterial families were found in the UM and NC, respectively. Unique families in rock biofilm from NC sites were as follows: *Oxalobacteraceae*, *Cryomorphaceae*, *Planctomycetaceae*, and *Verrucomicrobiaceae*, while those unique to rock biofilm from UM sites were *Bdellovibrionaceae*, *Enterobacteriaceae*, and *Sphingobacteriales*.

Discussion

Results from this study indicate that there are high levels of OTU diversity among and within bacterial communities in the streams of the Noatak National Preserve. We used both T-RFLP and 16S rRNA gene sequencing, two methods that differ in their resolution, to determine how the methods would resolve samples collected from our study sites. Differing patterns in bacterial community composition at both the small-scale (stream habitat) and large-scale (lithology) were observed. Using a T-RFLP approach we detected distinct fingerprints of bacterial communities by lithology and habitat, whereas 16S rRNA gene clone libraries differentiated habitats and marginally differentiated lithologies.

Comparisons using T-RFLP

NMS ordination of T-RFLP peaks show a high degree of separation between UM and NC communities and only partial separation between NC and CS. These trends in bacterial community composition mirror stream biogeochemistry across the three lithologies in that UM and NC have significantly different biogeochemical characteristics, while the NC and CS are similar. Our results are similar to previously reported differences in microbial community composition as influenced by lithology and parent material in soils (Dunbar *et al.*, 2000; Oline, 2006), glaciers (Skidmore *et al.*, 2005), groundwater and substratum (Takai *et al.*, 2003). NMS results of T-RFLP data also showed clear separation by habitat, between sediment and epilithon samples. The average number of phylotypes (T-RFs) detected in sediment samples were similar, suggesting low variation in bacterial diversity across lithologies. However, differences in the number of T-RFs were observed between sediment and epilithon within a lithology, indicating differences in bacterial diversity at the habitat scale.

Relationships between Bacterial Communities and Biogeochemistry

Vector analysis indicates that certain biogeochemical variables explain NMS ordination of T-RFLP data (Fig. 2), although causal relationships can only be inferred. Specifically, we observed a positive correlation between base cations and CS lithology, suggesting bacterial community structure may be influenced by the CS streams' high base cation concentration, whereas the scarcity of cations in NC and UM streams may constrain the composition of resident bacterial communities. Notably calcium and magnesium are well known to enhance bacterial adhesion to substrates within the exopolysaccharide matrix of biofilms (Geesey *et al.*, 2000), which may facilitate a niche for a more stable bacterial community in the CS lithology. Other constituents such as DOC, TDN-NO₃⁻, and TDN/TDP were correlated with NC community composition while NO₃⁻/TDN was correlated with the UM community. DOC and nutrient concentrations were lowest in streams of the CS and UM lithologies and highest in streams of the NC lithology. These trends suggest that bacterial community composition may be influenced by the abundance or scarcity of resources. Similar relationships have been observed in other studies that investigated the influence of microbial activity on redox chemistry and mineral processes in natural environments (Ehrlich, 1998; Neilson and Stahl, 1997). Furthermore, other studies have determined that microbial community composition can be correlated with observed aqueous geochemistry in subglacial chemical weathering (Skidmore *et al.*, 2005), streamwater pH, quality of fine benthic organic matter, and quantity of DOC and nitrogen in stream water (Fierer *et al.*, 2007), and seasonal changes in temperature, nutrient availability and light in estuarine biofilms (Moss *et al.*, 2006).

Comparisons using 16S rRNA Gene Sequencing

The T-RFLP approach used in this study revealed general patterns in community composition indicative of differences unique to varying types lithology and habitat. Construction of 16S rRNA gene libraries offered greater resolution by allowing the identification of dominant taxa present in each sample. Although analysis of clone libraries allowed the discrimination of community composition by habitat and only marginally by lithology, it permitted the identification of bacterial taxonomic composition at both the stream reach and landscape scale.

While we are aware that 77-95 clones per library is not a large number compared to the potential microbial diversity of these communities, we nonetheless detected high diversity among samples using 16S rRNA gene sequencing. Similar levels of microbial diversity were detected in studies in which comparable or fewer number of clones were sequenced (Fierer *et al.*, 2007; Oline, 2006). In this study, twenty-nine out of thirty possible F_{ST} pairwise comparisons of the 9 samples were significant, suggesting distinct genetic differentiation among all communities with the exception of EpiSUM01 vs. EpiSNC03. Furthermore, there was significant variation in the genetic structure of communities residing in different habitats (AMOVA). This heterogeneity in community composition may be a result of differences in hydrologic stressors and substrate availability in the sediment versus the rock biofilm habitats.

Sequencing of the 16S rRNA gene revealed the presence of genera including gram-negative bacteria that have previously been isolated from aquatic and terrestrial environments. There were distinct differences in taxonomic identities (at the family level) of clones from sediment versus rock biofilm habitats. Only four families are shared between the 20 and 16 families found in sediment and rock biofilm samples, respectively. At the class level, eight of 11 classes were shared between sediment and epilithon. Our results at the class level are similar to those reported by Hullar (2006), who sampled headwater streams in southeast Pennsylvania and found a high degree of similarity in taxa type overlap between sediment and rock biofilm samples. Furthermore, all representatives of the rock biofilm in our work, including 1 division and 3 classes, were also present in the 13 sediment classes detected in Hullar's study. Our results differ from those of Hullar (2006) in that there are few similar taxa between the sediment and epilithon at the family-level identification in our study and we detected a high abundance of cyanobacteria exclusively in rock biofilm samples (44%), whereas Hullar (2006) found that cyanobacteria clones comprised the majority (40%) of their sediment-derived sequences and a smaller proportion (25%) of the epilithic-derived sequences.

GenBank was used to identify matches to the sediment clones at the genus level. Groups of clones from the sediment samples: *Pseudomonas*, *Flavobacterium*, *Alcaligenes*, *Aeromonas*, *Enterobacter*, *Xanthomonas*, and *Sporolactobacillus*, known to be heterotrophic bacteria, were also previously isolated from similarly classified pristine stream bed sediments in forested watersheds (Halda-Alija and Johnston 1999). *Aeromonadaceae*, *Pseudomonadaceae*, and *Xanthomonadaceae* species are obligately aerobic, whereas *Enterobacteriaceae* and *Sporolactobacilliaceae* members are facultatively anaerobic with *Enterobacteriaceae* species having the ability to reduce nitrate to nitrite. *Paenibacillus borealis*, a nitrogen-fixing species, also isolated from Norway and Finnish spruce forest humus, was present in sediment samples from all three lithologies, but found to be dominant in the CS clone library.

Unclassifiable *Cyanobacteria* species, the most dominant of the epilithon clones, metabolize via oxygenic photosynthesis. Other dominant epilithic members include the following families: *Flexibacteriaceae*, *Comamonadaceae*, and *Deinococcaceae*, which are all chemoorganotrophic as well as obligately aerobic. Thirteen clones identified as belonging to the genus *Spingomonas* were found exclusively associated with the epilithic community (3.4%). *Spingomonas* have been isolated from a range of environments, including ultraoligotrophic waters, in which species such as *S. alaskensis* has been shown to possess physiological characteristics adapted to very low carbon substrate concentrations (Eiler *et al.*, 2003). That the detection of *Spingomonas* in our study was restricted to epilithon samples, may indicate lower availability of carbon sources for bacterial metabolism in this habitat. In contrast to the high

loads of particulate and dissolved organic matter associated with upwelling areas from the hyporheic zone, associated with stream sediment habitats (Sobczak and Findlay, 2002).

Figure 7 shows the percent of dominant clones from each category: sediment NC (n=2); sediment UM (n=2); epilithon NC (n=2) and epilithon UM (n=2) and reveals that clone identifications to family are particularly unique to each habitat type, whereas unique families span the two contrasting lithologies. This suggests that differences in community composition are more pronounced at the habitat scale than at the lithology scale, which is supported by the lack of distinction between communities by lithology in the AMOVA.

Conclusions and Relevance to Microbial Biogeography

In the past decade, studies on the taxonomic, phylogenetic, and physiological diversity of prokaryotes have begun to provide more comprehensive information about microbial communities and their natural environments, and in particular, whether microbes exhibit biogeographical patterns. Structural geographic patterns as detected in microbial communities within stream ecosystems have been attributed to the following factors: geographic distance (<10 km) and connectivity between lakes and streams (Crump *et al.*, 2007); biome-level control in low-order streams (Findlay *et al.*, 2008); variation of chemical characteristics in streams across the southeastern and midwestern U.S. (Gao *et al.*, 2005); and landscape-level controls on streams due to biogeochemical factors (Fierer *et al.*, 2007). In general, very few studies have focused on low-order streams (Findlay *et al.*, 2008; Hullar *et al.*, 2006), as we have done in this study. However, these types of streams are important because they function as links between the terrestrial environment and the aquatic ecosystem network.

Results obtained by comparing the bacterial OTUs from our study to the GenBank database showed matches with similar bacterial taxa isolated from various locations around the globe, suggesting (as Baas-Becking hypothesized) that bacteria have a cosmopolitan distribution. However, these results should be interpreted with caution given that the 16S rRNA gene sequencing method used in this study can be used as a measure of phylogenetic relatedness which does not necessarily reflect levels of similarity at the physiological level. It is possible that even slight differences in bacterial physiology can be related to biogeochemical processing. Future work that focuses on the metagenomics of environmental samples could be an alternative means to clarify the physiological characteristics of bacteria in a given environment allowing greater discriminatory power in determining bacterial contribution to ecosystem function.

While we observed similarities in bacterial clone types between our study sites and other geographically distant locations, we also detected biogeographic structure of species richness and taxon type. This is shown in the rarefaction curves (Fig. 4) by habitat type within NC and UM lithologies, which reveal distinct variation in species richness in the NC landscape, while UM streams track each other. These results indicate that NC streams may be more diverse than UM streams at the landscape scale. Furthermore, non-carbonate and ultramafic rarefaction curves (Fig. 3a) as well as sediment and rock biofilm rarefaction curves (Fig. 3b) are significantly different from one another (Figure 2.4 a). With regards to the controversy related to the Baas-Becking hypothesis (Whitfield, 2005), the biogeographical patterns observed in the stream bacterial communities of the Noatak River Basin lead us to conclude that whether a given bacterial taxon is cosmopolitan or biogeographically restricted depends on the environment and its biogeochemical requirements. Differences in bacterial community composition between sediment and epilithic habitats could result from different hydrologic stressors. For example,

varying flow regimes alter sediment structure via erosion and redistribute bacteria, exposing them to different environmental conditions (Hullar *et al.*, 2006). While the epilithic community is not as likely to experience the same degree of disturbance as that found in the sediment, differences in hydrodynamic conditions are known to influence the structure and activity of epilithic biofilms (Battin, 2000; Battin *et al.*, 2003).

Biogeographical bacterial patterns as influenced by lithology may be the result of differences in resource availability across different lithologies. The NC lithology is a richer environment, hosting streams with greater nutrient and DOC availability, higher chlorophyll *a* productivity, and concentrations of benthic organic matter, whereas streams of the UM lithology lack the resources abundant in the NC in addition to experiencing a scarcity of base cations. Given these biogeochemical conditions, we infer that the NC landscape may allow for the survival of more varied metabolic types given the abundance of resources, creating opportunities for bacterial diversity to persist across streams. The lack of biogeochemical resources in the UM streams may provide an environment where bacteria that are specialists and thriving on low nutrient and substrate concentrations may competitively exclude others. Thus, the UM lithology displays a predictable species richness perhaps because it is a highly selective environment where a limited type of organisms can persist.

Our results suggest that there are differences in bacterial community composition across contrasting lithologies that can be related to large-scale linkages between streams and the terrestrial environment and parent material in which they are embedded. In turn, this relationship is reflected in differences in resource availability. Furthermore, the resident microorganisms of sediment and epilithon habitats are composed of significantly different bacterial taxa, indicating the presence of contrasting ecological niches at the small-scale within stream ecosystems. Thus, our study of arctic streams using T-RFLP and 16S rRNA gene sequencing indicates that bacterial community composition is dependent on the physical characteristics of the habitat within a stream as well as the stream location on the lithological landscape.

ACKNOWLEDGEMENTS

We gratefully acknowledge D Sanzone, K Rattenbury, and T Whitsell of the National Park Service for logistical support. We thank A Allen and B Peterson for field assistance and S Herman for helicopter support. B Crump for their assistance with molecular methods and to G Druschel for lab space and equipment. The comments from C Nelson and B Crump greatly improved this manuscript. This study was supported by the National Park Service. The opinions and conclusions in this report are those of the authors and not the National Park Service.

REFERENCES

- Altschul SF, Warren G, Miller W, Myers EW, Lipman DJ (1990). Basic Local Alignment Search Tool. *J Mol Biol* **215**: 403-410.
- Ashelford KE, Chuzhanova NA, Fry JC, Jones AJ, Weightman AJ (2005). At least 1 in 20 16S rRNA sequence records currently held in public repositories is estimated to contain substantial anomalies. *Appl Environ Microbiol* **71**: 7724-7736.
- Autio R (1998). Response of seasonally cold-water bacterioplankton to temperature and substrate treatments. *Estuar Coast Shelf Sci* **46**: 465-474.
- Baas-Becking LGM (1934). *Geologie of Inleiding Tot de Milieukunde*. Van Stockum & Zoon: The Netherlands.
- Battin TJ (2000). Hydrodynamics is a major determinant of streambed biofilm activity: From the sediment to the reach scale. *Limnol Oceanogr* **45**: 1308-1319.
- Battin TJ, Kaplan A, Newbold D, Cheng X, Hansen C (2003). Effects of current velocity on the nascent architecture of stream microbial biofilms. *Appl Environ Microbiol* **69**: 5443-5452.
- Blackwood CB, Hudleston D, Zak DR, Buyer JS (2007). Interpreting Ecological Diversity Indices Applied to Terminal Restriction Fragment Length Polymorphism Data: Insights from Simulated Microbial Communities. *Appl Environ Microbiol* **73**: 5276-5283.
- Bray JR, Curtis JT (1957). An ordination of the Upland Forest Communities of Southern Wisconsin. *Ecol Monogr* **27**: 325-349.
- Brown MPS. (2000). *Proceedings of the Eighth International Conference on Intelligent Systems for Molecular Biology* San Diego, California, USA, pp 57-66.
- Chao A (1984). Nonparametric-estimation of the number of classes in a population. *Scand J Stat* **11**: 265-270.
- Chao A, Lee SM (1992). Estimating the number of classes via sampling coverage. *J Am Stat Assoc* **87**: 210-264.
- Chao A, Ma MC, Yang MCK (1993). Stopping rules and estimation for recapture debugging with unequal failure rates. *Biometrika* **80**: 193-201.
- Clarke KR (1993). Non-parametric multivariate analyses of changes in community structure. *Aust J Ecol* **18**: 117-143.
- Clarke KR, Green RH (1988). Statistical design and analysis for a 'biological effects' study. *Mar Ecol Prog Ser* **46**: 213-226.

Microbial biogeography of arctic streams

- Clarke KR, Warwick RM (2001). A further biodiversity index applicable to species lists: variation in taxonomic distinctness. *Mar Ecol Prog Ser* **216**: 265-278.
- Cole JR, Chai B, Farris J, Wang Q, Kulam-Syed-Mohideen AS, Bandela AM *et al* (2007). The ribosomal database project (RDP-II): introducing myRDP space and quality controlled public data. *Nucleic Acids Res* **35** D169-D172.
- Crump BC, Adams HE, Hobbie JE, Kling GW (2007). Biogeography of Bacterioplankton in Lakes and Streams of an Arctic Tundra Catchment. *Ecology* **88**: 1365-1378.
- Crump BC, Kling GW, Bahr M, Hobbie JE (2003). Bacterioplankton Community Shifts in an Arctic Lake Correlate with Seasonal Changes in Organic Matter Source. *Appl Environ Microbiol* **69**: 2253-2268.
- Denaro R, D'Auria G, Di Marco G, Genovese M, Troussellier M, Yakimov MM, *et al* (2005). Assessing terminal restriction fragment length polymorphism suitability for the description of bacterial community structure and dynamics in hydrocarbon-polluted marine environments. *Environ Microbiol* **7**: 78-87.
- Deslippe JR, Egger KN, Henry GHR (2005). Impacts of warming and fertilization on nitrogen-fixing microbial communities in the Canadian High Arctic. *FEMS Micro Ecology* **53**: 41-50.
- Drever JI (2002). *The Geochemistry of Natural Waters: Surface and Groundwater Environments*, Third edn. Prentice-Hall, Inc.: Englewood Cliffs, NJ.
- Dunbar J, Lawrence TO, Kuske CR (2000). Assessment of Microbial Diversity in Four Southwestern United States Soils by 16S rRNA Gene Terminal Restriction Fragment Analysis. *Appl Environ Microbiol* **66**: 2943-2950.
- Dunbar J, Ticknor LO, Kuske CR (2001). Phylogenetic Specificity and Reproducibility and New Method for Analysis of Terminal Restriction Fragment Profiles of 16S rRNA Genes from Bacterial Communities. *Appl Environ Microbiol* **67**: 190-197.
- Edwards U, Rogall T, Bloeker H, Ende MD, Boettge EC (1989). Isolation and Direct Complete Nucleotide Determination of Entire Genes, Characterization of Gene Coding for 16S Ribosomal RNA. *Nucleic Acids Res* **17**: 7843-7853.
- Ehrlich HI (1998). Geomicrobiology: its significance for geology. *Earth Sci Rev* **45**: 45-60.
- Eiler A, Laggenheder S, Bertilsson S, Tranvik LJ (2003). Heterotrophic bacterial growth efficiency and community structure at different natural organic carbon concentrations. *Appl Environ Microbiol* **69**: 3701-3709.
- Excoffier L, Smouse PE, Quattro JM (1992). Analysis of Molecular Variance Inferred From Metric Distances Among DNA Haplotypes: Application to Human Mitochondrial DNA Restriction Data. *Genetics* **131**: 479-491.

Microbial biogeography of arctic streams

Fierer N, Morse JL, Berthrong ST, Bernhardt ES, Jackson RB (2007). Environmental Controls on the Landscape-Scale Biogeography of Stream Bacterial Communities. *Ecology* **88**: 2162-2173.

Findlay RH, Yeates C, Hullar MAJ, Stahl DS, Kaplan LA (2008). Biome-Level Biogeography of Streambed Microbiota. *Appl Environ Microbiol* **74**: 3014-3021.

Galand PE, Lovejoy C, Vincent WF (2006). Remarkably diverse and contrasting archaeal communities in a large arctic river and the coastal Arctic Ocean. *Aquat Microb Ecol* **44**: 115-126.

Gao X, Olapade OA, Leff LG (2005). Comparison of benthic bacterial community composition in nine streams. *Aquat Microb Ecol* **40**: 51-60.

Garneau ME, Vincent WF, Alonso-Saez L, Gratton Y, Lovejoy C (2006). Prokaryotic community structure and heterotrophic production in a river-influenced coastal arctic ecosystem. *Aquat Microb Ecol* **42**: 27-40.

Geesey GG, Mutch R, Costerton JW, Green RB (1978). Sessile Bacteria: An Important Component of the Microbial Population in Small Mountain Streams. *Limnol Oceanogr* **23**: 1214-1223.

Geesey GG, Wigglesworth-Cooksey B, Cooksey KE (2000). Influence of calcium and other cations on surface adhesion of bacteria and diatoms: A review. *Biofouling* **15**: 195-205.

Goebel BM, Stackebrandt E (1994). Cultural and phylogenetic analysis of mixed microbial populations found in natural and commercial bioleaching environments. *Appl Environ Microbiol* **60**: 1614-1621.

Gray ND, Brown A, Nelson DR, Pickup RW, Rowan AK, Head IM (2007). The biogeographical distribution of closely related freshwater sediment bacteria is determined by environmental selection. *ISME J* **1**: 596-605.

Green JL, Bohannon BJM, Whitaker RJ (2008). Microbial Biogeography: From Taxonomy to Traits. *Science* **320**: 1039-1043.

Gutell RR, Lee JC, Cannone JJ (2002). The accuracy of ribosomal RNA comparative structure models. *Curr Opin Struct Biol* **12**: 301-310.

Hartmann M, Widmer F (2006). Community structure analyses are more sensitive to differences in soil bacterial communities than anonymous diversity indices. *Appl Environ Microbiol* **72**: 7804-7812.

Horner-Devine MC, Lage M, Hughes J, Bohannon BJM (2004). A taxa-area relationship for bacteria. *Nature* **432**: 750-753.

Microbial biogeography of arctic streams

- Hughes-Martiny JB, Bohannon BJM, Brown JH, Colwell RK, Fuhrman JA, Green JL *et al* (2006). Microbial biogeography: putting microorganisms on the map. *Nat Rev Microbiol* **4**: 102-112.
- Hullar M, Kaplan LA, Stahl DA (2006). Recurring Seasonal Dynamics of Microbial Communities in Stream Habitats. *Appl Environ Microbiol* **72**: 713-722.
- Jorgenson MT, Swanson DK, Macander M. (2002). Fairbanks, AK, pp 1-54.
- Jukes TH, Cantor CR (1969). Evolution of protein molecules. In: Munro HN (ed). *Mammalian Protein Metabolism*. Academic Press: New York. pp 21-132.
- Kaplan L, Bott T (1989). Diel fluctuations in bacterial activity on streambed substrata during vernal algal blooms: effects of temperature, water chemistry, and habitat. *Limnol Oceanogr* **34**: 718-733.
- Liu W, Marsh TL, Cheng H, Forney LJ (1997). Characterization of Microbial Diversity by Determining Terminal Restriction Fragment Length Polymorphisms of Gene Encoding 16S rRNA. *Appl Environ Microbiol* **63**: 4516-4522.
- Logue JB, Burgmann H, Robinson CT (2008). Progress in the Ecological Genetics and Biodiversity of Freshwater Bacteria. *Bioscience* **58**: 103-113.
- Maddison DR, Maddison WP. (2002). Sinauer Associates, Inc.: Sunderland, MA.
- Martin AP (2002). Phylogenetic Approaches for Describing and Comparing the Diversity of Microbial Communities. *Appl Environ Microbiol* **68**: 3673-3682.
- McCune B, Grace JB (2002). *Analysis of Ecological Communities*. MjM Software Design: Glenden Beach, Oregon USA.
- McFeters GA, Stuart SA, Olson SB (1978). Growth of Heterotrophic Bacteria and Algal Extracellular Products in Oligotrophic Waters. *Appl Environ Microbiol* **35**: 383-391.
- Milner AM, Oswood MW, Munkittrick KR (2005). Rivers of Arctic North America. In: Benke A and Cushing C (eds). *Rivers of North America*. Academic Press.
- Minchin P. (1990). Australian National University: Canberra.
- Moss JA, Nocker A, Lepo JE, Snyder RA (2006). Stability and Change in Estuarine Biofilm Bacterial Community Diversity. *Appl Environ Microbiol* **72**: 5679-5688.
- Nealson KH, Stahl DA (1997). Microorganisms and biogeochemical cycles: what can we learn from layered microbial communities? In: Nealson KH and Banfield JF (eds). *Geomicrobiology: interactions between microbes and minerals*. Mineralogical Society of America: Washington, DC. pp 5-34.

Microbial biogeography of arctic streams

Oline DK (2006). Phylogenetic Comparisons of Bacterial Communities from Serpentine and Nonserpentine Soils. *Appl Environ Microbiol* **72**: 6965-6971.

Palmer MA, Covich AP, Lake S, Biro P, Brooks JJ, Cole J *et al* (2000). Linkages between Aquatic Sediment Biota and Life above Sediments as Potential Drivers of Biodiversity and Ecological Processes. *Bioscience* **50**: 1062-1075.

Rees GN, Baldwin DS, Watson GO, Perryman S, Nielsen DL (2004). Ordination and significance testing of microbial community composition derived from terminal restriction fragment length polymorphisms: application of multivariate statistics. *Antonie Van Leeuwenhoek* **86**: 333-347.

Reysenbach AL, Wickham GS, Pace NR (1994). Phylogenetic analysis of the hyperthermophilic pink filament community in Octopus Spring, Yellowstone National Park. *Appl Environ Microbiol* **60**: 2113-2119.

Rosello-Mora R, Amann R (2001). The Species Concept for Prokaryotes. *FEMS Microbiol Rev* **25**: 39-67.

Rozas J, Rozas R (1999). DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* **15**: 174-175.

Schloss PD, Handelsman J (2005). Introducing Dotur, a computer program for defining operational taxonomic units and estimating species richness. *Appl Environ Microbiol* **71**: 1501-1506.

Skidmore M, Anderson SP, Sharp M, Foght J, Lanoil BD (2005). Comparison of Microbial Community Compositions of Two Subglacial Environments Reveals a Possible Role for Microbes in Chemical Weathering Processes. *Appl Environ Microbiol* **71**: 6986-6997.

Sobczak WV, Findlay S (2002). Variation in bioavailability of dissolved organic carbon among stream hyporheic flowpaths. *Ecology* **83**: 3194-3209.

Stanford JA, Ward JV (1993). An ecosystem perspective of alluvial rivers: connectivity and the hyporheic corridor. *Journal N Am Benthol Soc* **12**: 48-60.

Swofford DL. (2001). Sinauer Associates, Inc.: Sunderland, MA.

Takai K, Mormile MR, McKinley JP, Brockman FJ, Holben WE, Kovacik WP *et al* (2003). Shifts in archaeal communities associated with lithological and geochemical variations in subsurface Cretaceous rock. *Environ Microbiol* **5**: 309-320.

Ward JV, Stanford JA (1995). Ecological connectivity in alluvial river ecosystems and its disruption by flow regulation. *Regul Rivers: Res Manage* **11**: 105-119.

Whitfield J (2005). Biogeography: Is Everything Everywhere? *Science* **310**: 960-961.

Microbial biogeography of arctic streams

Table 1. Values of biological, chemical and physical parameters for study streams by lithology (CS = Complex Sedimentary; NC = Non-carbonate; and UM = Ultramafic)

| Lithology | n | | Ca | Mg | K | Na | NO ₃ ⁻ | TDN | TDP | Cu | Al | Fe | Si | E _c | pH | Temp | DO | DOC | Chl a |
|-----------|---|-----------|------|------|------|------|------------------------------|------|------|------|-------|------|------|----------------|-------|------|------|------|--------------------|
| | | | mg/l | mg/l | mg/l | mg/l | μM | μM | μM | μg/l | μg/l | μg/l | μg/l | μS/cm | units | °C | mg/l | mg/l | μg/cm ² |
| CS | 6 | mean | 38.1 | 27.4 | 1.1 | 5.1 | 2.6 | 7.7 | 0.1 | 0.5 | 92.3 | 12.3 | 1703 | 502.2 | 8.8 | 8.8 | 10.9 | 2.8 | 0.3 |
| | | std error | 9.0 | 4.9 | 0.1 | 1.5 | 0.9 | 0.7 | 0.01 | 0.04 | 1.4 | 1.1 | 73 | 140.2 | 0.6 | 0.6 | 0.3 | 0.2 | 0.1 |
| NC | 5 | mean | 14.0 | 9.1 | 0.9 | 1.6 | 1.1 | 19.9 | 0.1 | 1.5 | 102.1 | 42.3 | 3176 | 168.7 | 7.5 | 8.5 | 10.9 | 8.3 | 0.3 |
| | | std error | 2.4 | 1.8 | 0.02 | 0.0 | 0.1 | 2.6 | 0.01 | 0.1 | 1.8 | 6.8 | 608 | 26.1 | 0.1 | 1.2 | 0.7 | 0.9 | 0.1 |
| UM | 5 | mean | 3.8 | 1.9 | 0.9 | 1.3 | 4.3 | 7.3 | 0.1 | 0.5 | 95.7 | 33.0 | 2278 | 38.8 | 7.3 | 7.9 | 10.9 | 2.1 | 0.3 |
| | | std error | 0.8 | 0.6 | 0.00 | 0.0 | 1.4 | 1.2 | 0.00 | 0.0 | 2.7 | 10.4 | 134 | 4.4 | 0.1 | 0.3 | 0.2 | 0.2 | 0.0 |

Figure 1. Study area of Feniak Lake region with stream site locations across contrasting lithologies in the Noatak National Preserve, Alaska (Map credit: Andrew Balsler).

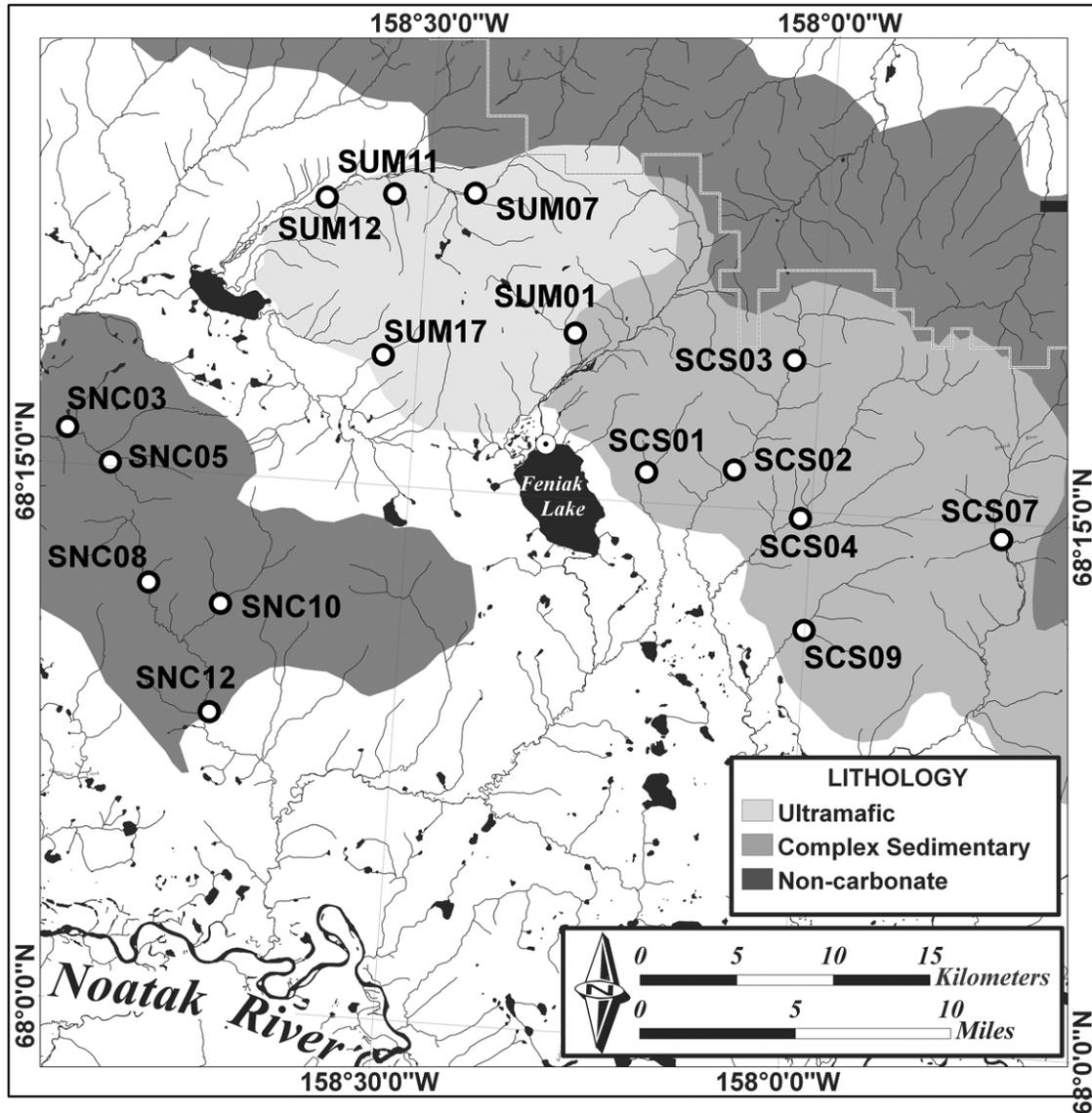


Figure 2. Non-metric multidimensional scaling (NMS) ordination (1st & 2nd of the 3-dimensional solution) of stream sediment bacterial communities (left side, n=42) and epilithon bacterial communities (right side, n=8) based on pairwise similarity estimates (Bray-Curtis). Points that are close together represent communities with similar bacterial community composition based on the T-RFLP method. The associated normal stress value of the ordination is 0.11, indicating a good approximation of the overall structure of the data in multivariable space. Significant ($p < 0.005$), Bonferonni adjusted biogeochemical variables were overlaid (arrows) showing the degree of correlation with sediment data. Abbreviations: Nutrients = TDN-NO₃⁻, nitrate and TDN/TDP, total dissolved nitrogen/total dissolved phosphorus; DOC, dissolved organic carbon; Cations = Ca²⁺, Mg²⁺, K⁺, and Na⁺; NO₃⁻/TDN, nitrate/total dissolved nitrogen.

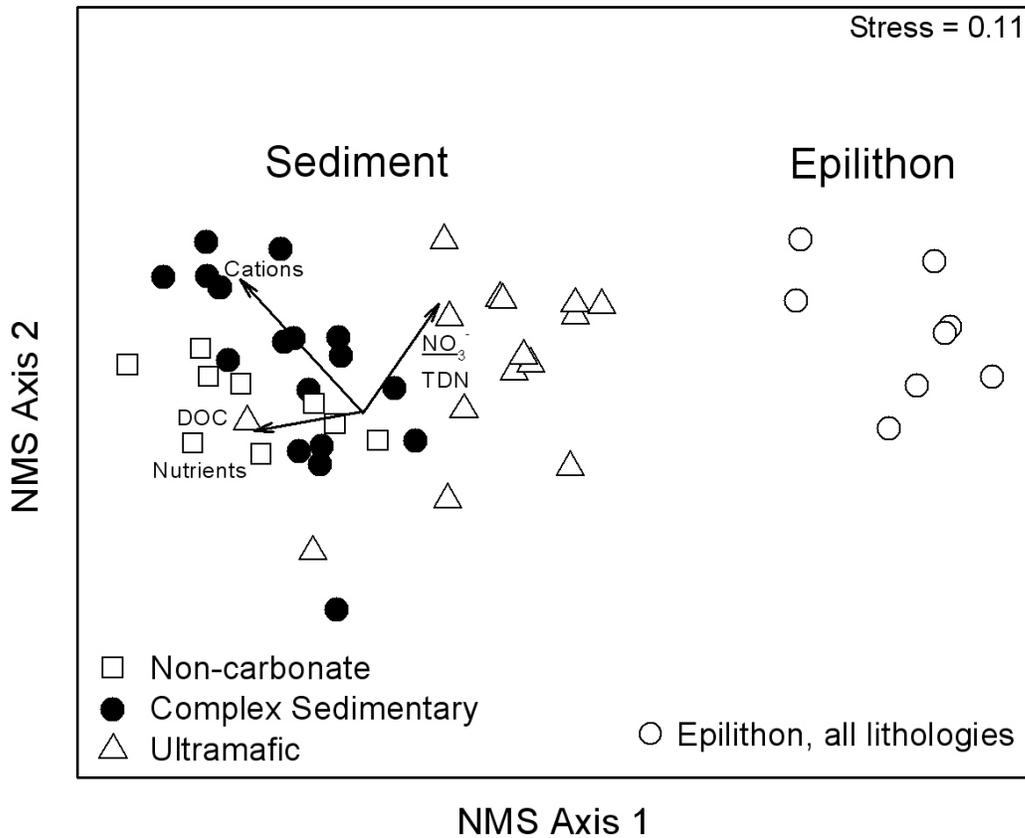


Figure 3. a) Rarefaction curves of observed OTU richness in sediment and epilithon samples within each lithology (non-carbonate and ultramafic). b) Rarefaction curves of observed OTU richness in sediment and epilithon samples regardless of lithology. The variance of the number of OTUs drawn in 100 randomizations at each sample size was calculated as 95% CIs (not shown).

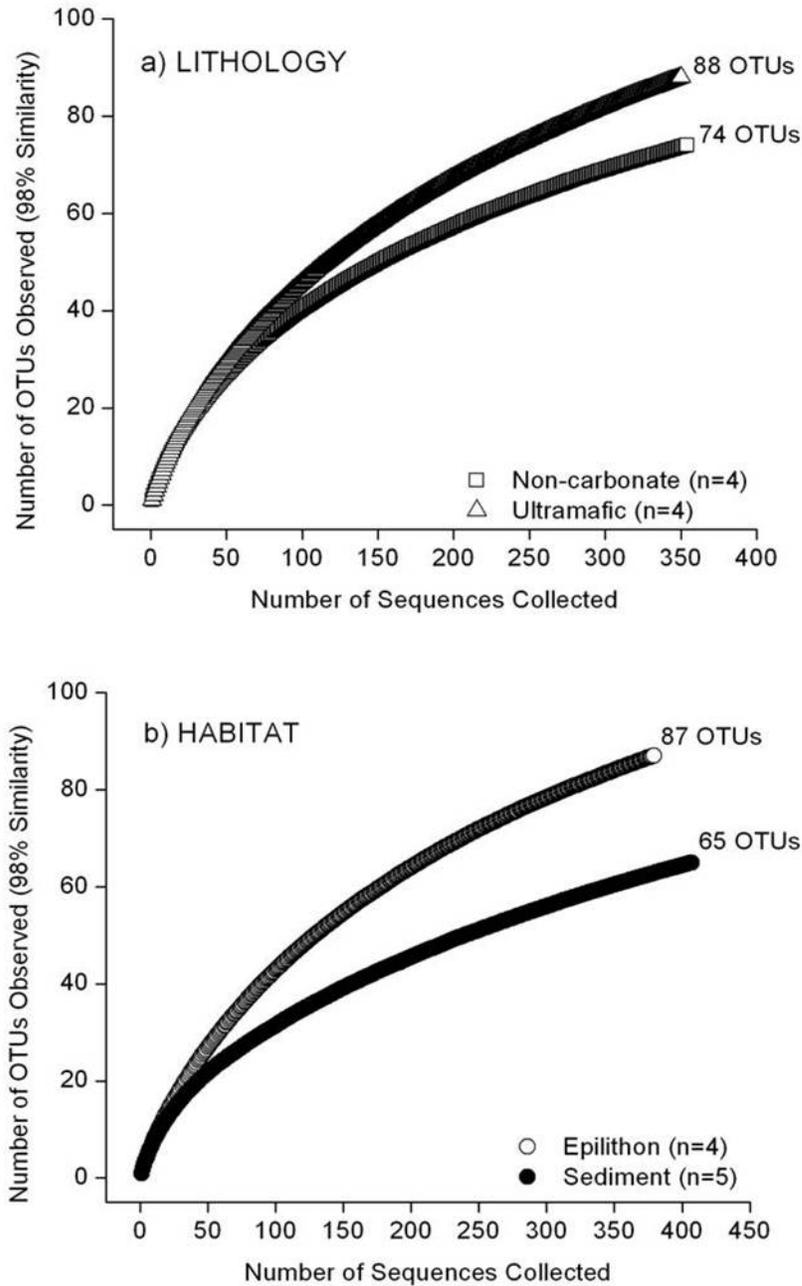
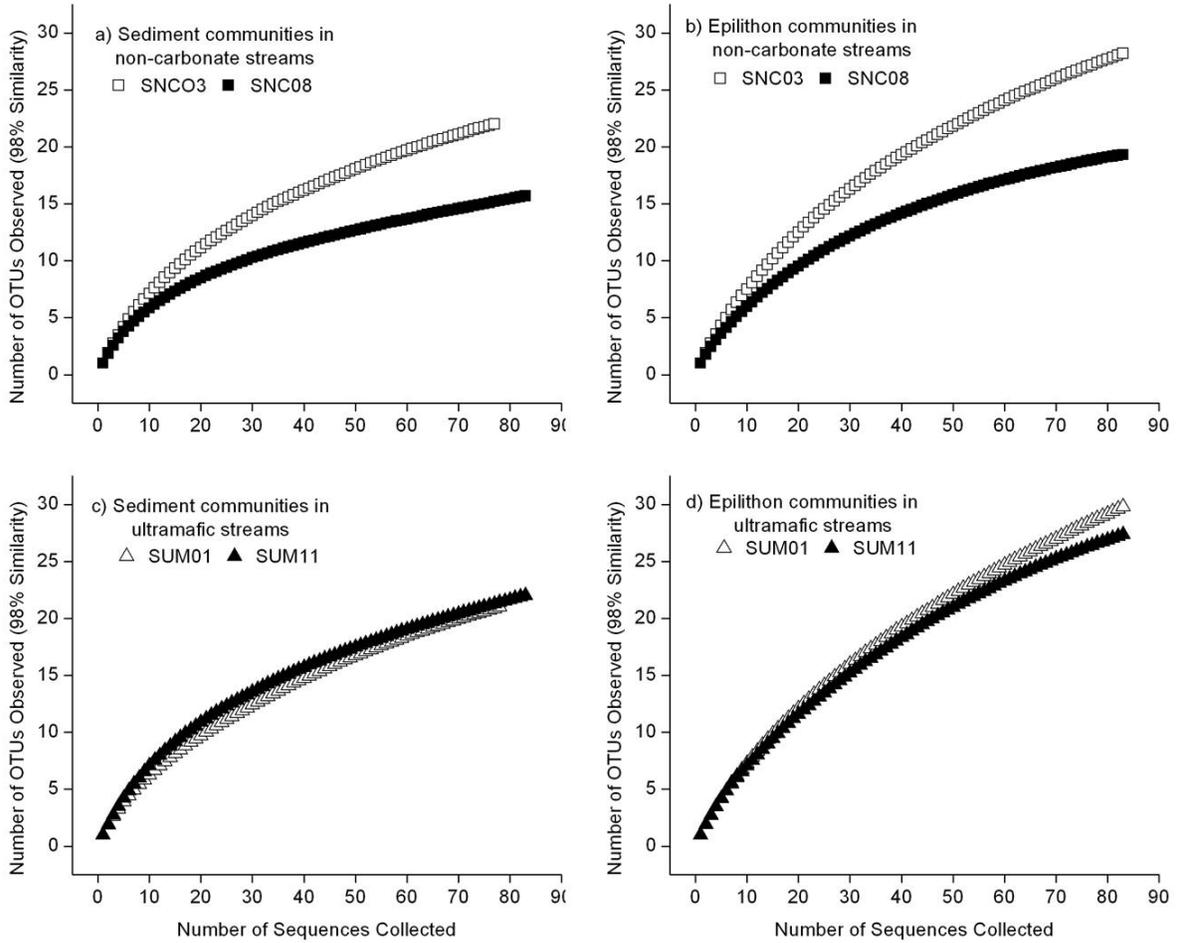
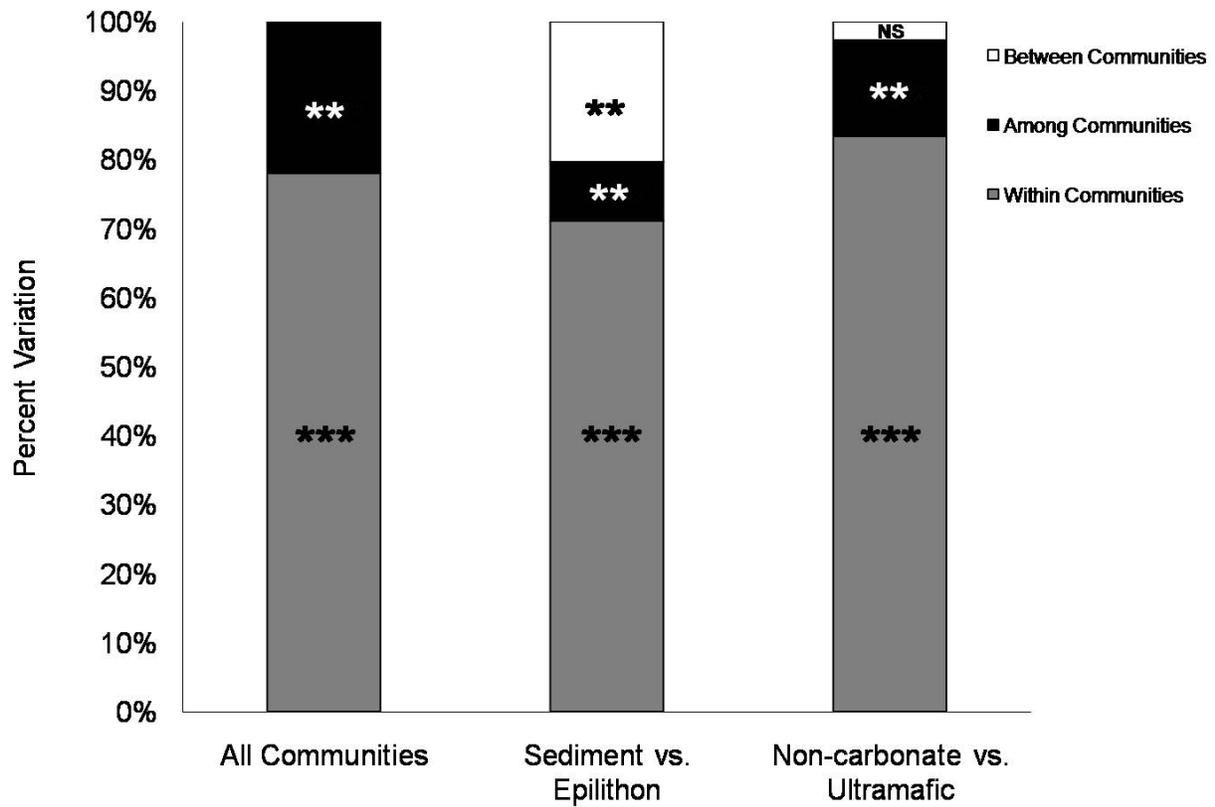


Figure 4. Rarefaction curves of observed OTU richness in individual stream and stream habitat comparisons (a) Sediment non-carbonate streams; (b) Epilithon non-carbonate streams; (c) Sediment ultramafic streams; (d) Epilithon ultramafic streams.



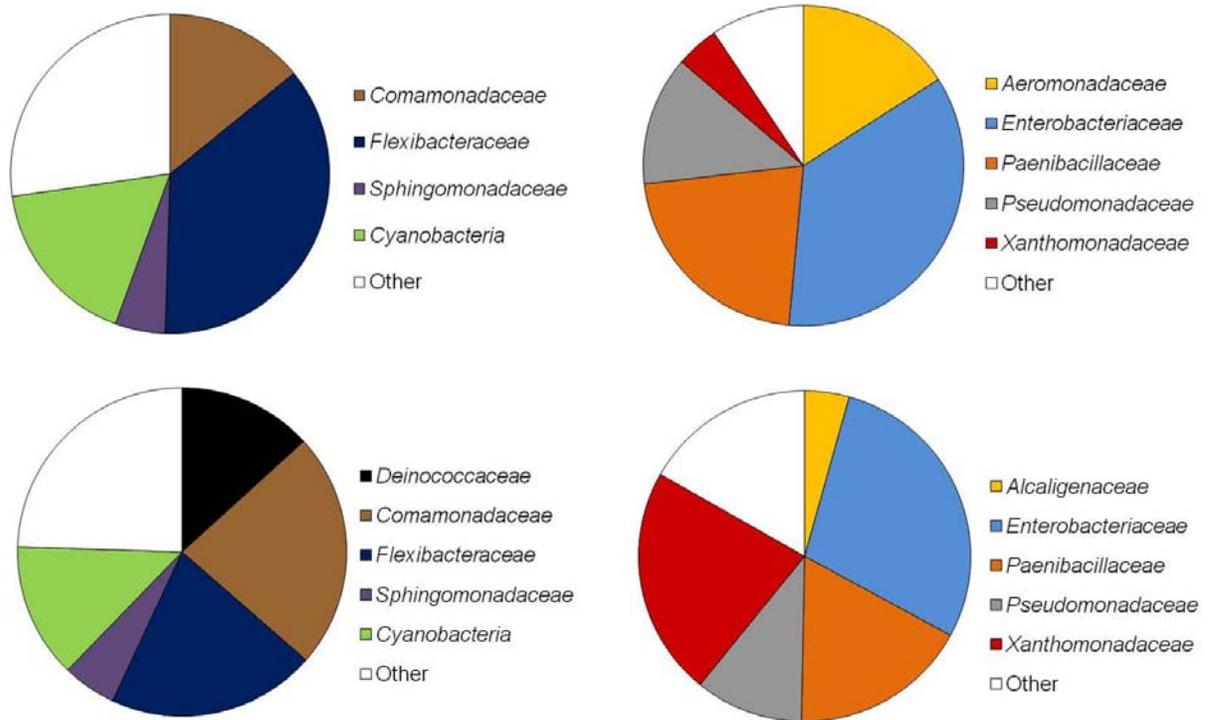
Microbial biogeography of arctic streams

Figure 5. Percent variation from analysis of molecular variance (AMOVA) analysis at each level of partitioning: a) all communities includes the 9 clone libraries; b) sediment clone libraries (n = 5) versus epilithon clone libraries (n = 4); and non-carbonate clone libraries (n = 4) versus ultramafic clone libraries (n = 4). NS = no significance.



Microbial biogeography of arctic streams

Figure 6. Pie charts representing the *dominant* clone frequency distribution among the sediment and epilithon stream samples in the non-carbonate and ultramafic lithologies (n=2 clone libraries for each category for a total of 8 clone libraries; the SCS07 clone library is not included because only a sediment sample was analyzed from the CS lithology).



Microbial biogeography of arctic streams

Supplementary Table 1. Summary of individual study streams with GPS locations.

| Lithology | Stream Site ID | Sample Date | Latitude (DD) | Longitude (DD) | Elevation (m) |
|---------------------|-----------------------|--------------------|----------------------|-----------------------|----------------------|
| Non-carbonate | SNC03 | July 11, 2006 | 68.266972 | 158.927500 | 342 |
| Non-carbonate | SNC05* | July 8, 2006 | 68.251750 | 158.872860 | 418 |
| Non-carbonate | SNC08 | July 8, 2006 | 68.198617 | 158.816683 | 468 |
| Non-carbonate | SNC10* | July 8, 2006 | 68.178950 | 158.741817 | 426 |
| Non-carbonate | SNC12 | July 11, 2006 | 68.139444 | 158.729750 | 358 |
| Complex Sedimentary | SCS01 | July 12, 2006 | 68.265361 | 158.204972 | 570 |
| Complex Sedimentary | SCS02 | July 12, 2006 | 68.267917 | 158.098778 | 451 |
| Complex Sedimentary | SCS03 | July 13, 2006 | 68.320300 | 158.035400 | 535 |
| Complex Sedimentary | SCS04 | July 12, 2006 | 68.252083 | 158.004833 | 490 |
| Complex Sedimentary | SCS07 | July 13, 2006 | 68.245472 | 157.785278 | 556 |
| Complex Sedimentary | SCS09 | July 13, 2006 | 68.195222 | 158.001806 | 474 |
| Ultramafic | SUM01 | July 9, 2006 | 68.327139 | 158.306750 | 550 |
| Ultramafic | SUM07 | July 10, 2006 | 68.392306 | 158.425778 | 660 |
| Ultramafic | SUM11 | July 10, 2006 | 68.386528 | 158.542361 | 560 |
| Ultramafic | SUM12 | July 10, 2006 | 68.382972 | 158.627056 | 447 |
| Ultramafic | SUM17 | July 9, 2006 | 68.303361 | 158.574611 | 501 |

*Sample not used from site due to questionable sample preservation.

Microbial biogeography of arctic streams

Supplementary Table 2. Phylogenetic affiliation of clones amplified from sediment and epilithon samples.

| Habitat | GenBank Identifications | | No. of clones | | | % of clone library | | | % of total clone library | |
|-------------------------|----------------------------|---|---------------|----|-----|--------------------|------|------|--------------------------|------|
| | Class (Domain*) | Family | NC | UM | CS | NC | UM | CS | | |
| Sediment n = 5 | <i>Acidobacteria</i> | <i>Acidobacteriaceae</i> | 1 | 3 | 0 | 0.6 | 1.9 | 0.0 | 1.0 | |
| | <i>Actinobacteria</i> | <i>Microbacteriaceae</i> | 0 | 1 | 0 | 0.0 | 0.6 | 0.0 | 0.2 | |
| | <i>Alphaproteobacteria</i> | <i>Unclassifiable Alphaproteobacteria</i> | | 2 | 0 | 0 | 1.2 | 0.0 | 0.0 | 0.5 |
| | | <i>Hyphomicrobiaceae</i> | | 0 | 3 | 0 | 0.0 | 1.9 | 0.0 | 0.7 |
| | <i>Bacilli</i> | <i>Rhizobiaceae</i> | | 1 | 5 | 0 | 0.6 | 3.1 | 0.0 | 1.5 |
| | | <i>Carnobacteriaceae</i> | | 0 | 1 | 0 | 0.0 | 0.6 | 0.0 | 0.2 |
| | | <i>Paenibacillaceae</i> | | 30 | 28 | 57 | 18.3 | 17.4 | 69.5 | 28.3 |
| | <i>Bacteria*</i> | <i>Sporolactobacillaceae</i> | | 25 | 0 | 0 | 15.2 | 0.0 | 0.0 | 6.2 |
| | | <i>Unclassifiable Bacteria</i> | | 1 | 0 | 0 | 0.6 | 0.0 | 0.0 | 0.2 |
| | <i>Betaproteobacteria</i> | <i>Alcaligenaceae</i> | | 0 | 7 | 0 | 0.0 | 4.3 | 0.0 | 1.7 |
| | | <i>Comamonadaceae</i> | | 4 | 5 | 0 | 2.4 | 3.1 | 0.0 | 2.2 |
| | <i>Deltaproteobacteria</i> | <i>Aeromonadaceae</i> | | 22 | 0 | 6 | 13.4 | 0.0 | 7.3 | 6.9 |
| | | <i>Geobacteraceae</i> | | 0 | 1 | 0 | 0.0 | 0.6 | 0.0 | 0.2 |
| | <i>Flavobacteria</i> | <i>Flavobacteriaceae</i> | | 0 | 2 | 0 | 0.0 | 1.2 | 0.0 | 0.5 |
| | <i>Gammaproteobacteria</i> | <i>Enterobacteriaceae</i> | | 49 | 46 | 7 | 29.9 | 28.6 | 8.5 | 25.1 |
| | | <i>Pseudomonadaceae</i> | | 18 | 17 | 3 | 11.0 | 10.6 | 3.7 | 9.4 |
| | | <i>Shewanellaceae</i> | | 4 | 2 | 4 | 2.4 | 1.2 | 4.9 | 2.5 |
| <i>Xanthomonadaceae</i> | | | 6 | 36 | 5 | 3.7 | 22.4 | 6.1 | 11.6 | |
| <i>Planctomycetacia</i> | <i>Planctomycetaceae</i> | | 0 | 1 | 0 | 0.0 | 0.6 | 0.0 | 0.2 | |
| <i>Sphingobacteria</i> | <i>Crenotrichaceae</i> | | 0 | 3 | 0 | 0.0 | 1.9 | 0.0 | 0.7 | |
| Epilithon n = 4 | <i>Alphaproteobacteria</i> | <i>Sphingomonadaceae</i> | 5 | 8 | n/a | 2.6 | 4.2 | n/a | 3.4 | |
| | <i>Bacteria*</i> | <i>Unclassifiable Bacteria</i> | 7 | 7 | n/a | 3.7 | 3.7 | n/a | 3.7 | |
| | <i>Betaproteobacteria</i> | <i>Betaproteobacteria</i> | | 4 | 8 | n/a | 2.1 | 4.2 | n/a | 3.2 |
| | | <i>Burkholderiales</i> | | 4 | 1 | n/a | 2.1 | 0.5 | n/a | 1.3 |
| | <i>Cyanobacteria*</i> | <i>Comamonadaceae</i> | | 14 | 35 | n/a | 7.4 | 18.5 | n/a | 12.9 |
| | | <i>Oxalobacteraceae</i> | | 6 | 0 | n/a | 3.2 | 0.0 | n/a | 1.6 |
| | <i>Deinococci</i> | <i>Unclassifiable Cyanobacteria</i> | 107 | 58 | n/a | 56.3 | 30.7 | n/a | 43.5 | |
| | <i>Deinococci</i> | <i>Deinococcaceae</i> | 1 | 20 | n/a | 0.5 | 10.6 | n/a | 5.5 | |
| | <i>Deltaproteobacteria</i> | <i>Bdellovibrionaceae</i> | 0 | 1 | n/a | 0.0 | 0.5 | n/a | 0.3 | |
| | <i>Flavobacteria</i> | <i>Cryomorphaceae</i> | | 1 | 0 | n/a | 0.5 | 0.0 | n/a | 0.3 |
| | | <i>Flavobacteriaceae</i> | | 3 | 2 | n/a | 1.6 | 1.1 | n/a | 1.3 |
| | <i>Gammaproteobacteria</i> | <i>Enterobacteriaceae</i> | | 0 | 2 | n/a | 0.0 | 1.1 | n/a | 0.5 |
| | <i>Planctomycetacia</i> | <i>Planctomycetaceae</i> | | 1 | 0 | n/a | 0.5 | 0.0 | n/a | 0.3 |
| | <i>Spingobacteria</i> | <i>Flexibacteraceae</i> | | 36 | 31 | n/a | 18.9 | 16.4 | n/a | 17.7 |
| <i>Verrucomicrobiae</i> | <i>Sphingobacteriales</i> | | 0 | 16 | n/a | 0.0 | 8.5 | n/a | 4.2 | |
| | <i>Verrucomicrobiaceae</i> | | 1 | 0 | n/a | 0.5 | 0.0 | n/a | 0.3 | |