River Microbiome Composition Reflects Macroscale Climatic and Geomorphic Differences in Headwater Streams

Dawn R. Urycki, Stephen P. Good, Byron C. Crump, Jessica Chadwick and Gerrad D. Jones

Departments of Biological and Ecological Engineering, Oregon State University, Corvallis, OR, United States.

Maintaining the quality and quantity of water resources in light of complex changes in climate, human land use, and ecosystem composition requires detailed understanding of ecohydrologic function within catchments, yet monitoring relevant upstream characteristics can be challenging. In this study, we investigate how variability in riverine microbial communities can be used to monitor the climate, geomorphology, land-cover, and human development of watersheds. We collected streamwater DNA fragments and used 16S rRNA sequencing to profile microbiomes from headwaters to outlets of the Willamette and Deschutes basins, two large watersheds prototypical of the U.S. Pacific Northwest region. In the temperate, north-south oriented Willamette basin, microbial community composition correlated most strongly with geomorphic characteristics (mean Mantel test statistic $r = 0.19$). Percentage of forest and shrublands ($r = 0.34$) and latitude ($r = 0.41$) were among the strongest correlates with microbial community composition. In the arid Deschutes basin, however, climatic characteristics were the most strongly correlated to microbial community composition (e.g., $r = 0.11$). In headwater sub-catchments of both watersheds, microbial community assemblages correlated with catchment-scale climate, geomorphology, and land-cover ($r = 0.46$, $r = 0.38$, and $r = 0.28$, respectively), but these relationships were weaker downstream. Development-related characteristics were not correlated with microbial community composition in either watershed or in small or large sub-catchments. Our results build on previous work relating streamwater microbiomes to hydrologic regime and demonstrate that microbial DNA in headwater streams additionally reflects the structural configuration of landscapes as well as other natural and anthropogenic processes upstream. Our results offer an encouraging indication that streamwater microbiomes not only carry information about microbial ecology, but also can be useful tools for monitoring multiple upstream watershed characteristics.

Keywords: catchment hydrology, ungauged basins, watershed monitoring, 16S rRNA, microbial diversity, Pacific Northwest, ecohydrology
INTRODUCTION

Water quality and availability depends on the integrity of water resource systems, which are sensitive to changes in climate (Jiménez Cisneros et al., 2014), human land use (Foley et al., 2005), and ecosystem composition. Climate change is projected to increase drought frequency and reduce surface and groundwater availability in arid regions, and to negatively impact freshwater ecosystems and reduce surface water quality (Delpla et al., 2009; Jiménez Cisneros et al., 2014). It is well-understood that human development and land-use change results in increased surface runoff, eventually leading to larger flooding events and reduced groundwater recharge (e.g., Carter, 1961; Moscrop and Montgomery, 1997; Gregory et al., 2006; Wheater and Evans, 2009). Interactions among these factors exacerbate impacts to water resources (e.g., Feddema et al., 2005; IPCC, 2019), and better tools are needed to diagnose these effects on catchment ecohydrology at local scales.

In watershed catchments lacking sufficient hydrologic data, Seibert and McDonnell (2015) found that key pieces of data collected on short field campaigns (i.e., “soft data”) can prove useful for understanding catchment function. Leveraging a source of information-rich soft data could thus prove especially powerful in characterizing ungauged basins. DNA is gaining traction as a valuable source of data across research disciplines. One of the appeals of genetic data is the vast quantity of information (thousands of features or more) that can be extracted with relative ease from a single sample. In addition to falling costs, improved methods of DNA extraction and sequencing, resulting in higher-quality data (Li et al., 2015), have increased the appeal of genetic data for a wider range of applications, including hydrological studies. For instance, Mächler et al. (2019) used environmental DNA (eDNA) released from macro organisms to characterize hydrologic flow paths in an Alpine catchment in Switzerland. Analysis of aquatic DNA in boreal forests has been indicative of key gradients in catchment condition similar to morphologically derived stream macroinvertebrate metrics (Emilson et al., 2017), while similar DNA information has also been used to map landscape-level terrestrial biodiversity (Sales et al., 2020). Sugiyama et al. (2018) found that microbial DNA analysis, coupled with general information about optimal growth conditions of certain microbial groups, revealed information about groundwater flow paths that was not captured with chemical analyses. Microbial communities, or “microbiomes,” with their high biodiversity, are hypothesized to hold new clues that traditional hydrologic tracers have not been able to elucidate.

Streamwater microbiomes respond to conditions that are also likely related to biogeochemical cycling and streamflow including nutrient concentrations, pH, and temperature (Fortunato et al., 2013; Read et al., 2015; Savio et al., 2015; Doherty et al., 2017). Given that streamwater microbiomes are composed primarily of microbes that originate in upslope soil and groundwater environments (Crump et al., 2007, 2012; Sorensen et al., 2013; but see Hermans et al., 2019), it follows that these microbiomes could be rich sources of data about hydrologic function and upslope catchment conditions. Beyond potential source variation, the characteristics of how genetic material is transported, retained, and resuspended has been examined for environmental DNA fragments (Jerde et al., 2016; Shogren et al., 2017) as well as for microorganisms themselves (Droppo et al., 2009; Newby et al., 2009), wherein stream water microbial composition is influenced by both abiotic factors (e.g., flow rate, mixing with sediment waters) and biotic factors such as predation, intrinsic cell mobility, and reproduction. Read et al. (2015) found that stream microbiomes were related to upstream hydrology, and Good et al. (2018) employed microbial community composition to characterize flow regimes in a set of large Arctic rivers. Savio et al. (2015) found that stream microbiomes were more strongly correlated with macroscale properties related to catchment hydrology, such as stream length and catchment size, than with water temperature or pH along the length of the Danube River. It is unclear, however, whether and to what extent other macroscale catchment factors may shape the microbial community. Our overarching objective is to employ stream microbiomes to gain insight about watershed conditions and catchment function and how they shape downstream water quality and availability. The first step toward this goal is to understand how watersheds may influence streamwater microbial community composition. In this study, we explored how streamwater microbial community composition, characterized with aquatic DNA fragments, correlates with upstream catchment properties. 16S rRNA gene fragments have been used in microbiology since the 1980s to classify bacterial taxonomy (Kolbert and Persing, 1999), and the 16S rRNA gene is the most widely used phylogenetic marker gene for assessing prokaryotic microbiomes (Goodrich et al., 2014). We isolated and sequenced 16S rRNA collected from streamwater and examined how differences between microbial (bacteria and archaea) communities among catchments were related to differences in catchment characteristics across major watersheds and across spatial scales.

MATERIALS AND METHODS

Watershed and Sub-catchment Characteristics

The Willamette and Deschutes basins, two similarly sized large (Willamette = 29,000 km²; Deschutes = 27,700 km²), adjacent watersheds separated by the Cascade Mountains in Oregon, USA were surveyed for variability in streamwater microbial communities. Mean elevation is 560 m in the Willamette Basin and 1,230 m in the Deschutes Basin. The Willamette Basin, on the windward side of the Cascades, receives a mean annual precipitation of 1,640 mm, but the Deschutes Basin on the leeward side receives a mean of just 530 mm annually. Mean annual discharge of the Willamette River (933 m³/s) is thus much greater than the Deschutes River (165 m³/s; U.S. Geological Survey, 2016a) at their respective outlets draining into the Columbia River at Oregon’s northern border. Temperatures are comparable between the basins (mean annual minimum and maximum temperatures are 4 and 15°C in the Willamette Basin and 0 and 14°C in the Deschutes Basin, respectively). Both basins
exhibit a range of land use, including minimally disturbed upper
elevation headwater streams, crop and livestock agriculture,
and highly developed urban areas; however the Willamette
Basin is more developed overall, with greater percentages
of impervious area and low-, medium-, and high-intensity
development (Supplementary Table 1).

Sub-catchment characteristics for areas upstream of DNA
sample collection points throughout the Willamette and
Deschutes watersheds were obtained using the StreamStats
tool developed by the United States Geological Survey (Ries
et al., 2008; U.S. Geological Survey, 2016b). StreamStats is
a national map-based web application that allows users to
obtain basin boundaries, basin characteristics, and streamflow
statistic estimates for gauged or ungauged user-specified sites.
StreamStats employs a wide array of digital geospatial raster
data layers which are processed through an online geographic
information system to define drainage basin characteristics for
any location (although available information varies by state). A
script was developed using the Python programming language
(www.python.org) to obtain all available StreamStats basin
characteristics for each of our 61 sampling sites distributed
throughout the two watersheds (URycki and Good, 2020a). The
program queries the StreamStats Service API for a list of available
basin characteristics for a specified state (OR) and the values
for those characteristics given the latitude and longitude for
each sample location retrieved from an input spreadsheet. The
StreamStats Service API outputs a JSON file with the requested
data, which our Python script decodes and writes to a csv
file containing the list of sites and associated drainage basin
characteristics. For this analysis, StreamStats quantities reported
in English units were converted to metric.

Our analysis considered 42 macroscale sub-catchment
characteristics for a relationship with streamwater microbial
communities (Table 1). In addition to the StreamStats
characteristics, we calculated topographic index and added
latitudinal and longitudinal coordinates. From an initial suite
of 46 available characteristics for each sample point, we then
eliminated all but one of any redundant variables and summed
three StreamStats characteristics “MINOR ROADS,” “STATE
HIGHWAY ROADS,” and “MAJOR ROADS” into a new variable
“ALL ROADS” (Supplementary Table 1). We then grouped
remaining variables into four categories: climatic, geomorphic,
land-cover, and development characteristics.

DNA Collection and Sequencing
We collected streamwater DNA from a set of 61 sites in
the summer of 2017. We sampled 40 sites in the Willamette
watershed (25 July–8 August) and 21 sites in the Deschutes
watershed (2–4 August; Figure 1). Sample sites spanned
headwaters and tributaries to the main stem outlet at the bottom
of the watershed. Sites were selected to capture the range of land
cover, land use, and level of disturbance in each sub-catchment.

Prior to sample collection each day, the plastic collection
bucket was washed with antimicrobial soap, soaked for several
minutes in a 10% hydrochloric acid bath, and triple rinsed
with ultrapure water. Other reusable field equipment, such as
collection tubing, was sterilized in an autoclave. Equipment was
sample-rinsed prior to each collection and then triple-rinsed with
deionized water following each collection.

Streamwater was collected from the approximate center of the
waterway. Most samples were collected using a clean 2-gallon
bucket lowered with a rope from a bridge. Although it was
difficult to standardize collection depth, samples were collected
near the surface and care was taken not to disturb the streambed.
At very small streams, such as those in HJ Andrews Experimental
Forest, samples were collected into a bucket as water exited
flumes. At the few sites where either a flume or bridge were not
available, a rope and bucket were tossed from the riverbank to
the approximate center of the stream and pulled back to shore, again
taking care not to disturb the streambed.

DNA samples were filtered and extracted from collected
streamwater as described in Crump et al. (2003). Briefly,
streamwater was pumped through a 0.2 µm Sterivex-GP filter
(Millipore, Billerica, MA, USA) with a peristaltic pump (Geotech
Environmental Equipment, Denver, CO, USA) until the filter
clogged. DNA preservation/extraction buffer (100 mM Tris,
100 mM NaEDTA, 100 mM phosphate buffer, 1.5 M NaCl, 1% CTAB) was added to the filter with a syringe, and then filters
were sealed and stored on dry ice until transferred to a
−80°C freezer the same day, where samples were stored until
further processing. DNA was isolated using phenol-chloroform
extraction and isoamyl alcohol precipitation (Zhou et al., 1996;
Crump et al., 2003; Amaral-Zettler et al., 2009) and stored at
−20°C until amplification.

16S rRNA genes were PCR-amplified with dual-barcoded
primers targeting the V4 region (515f GTGCCAGCMGCCGCG
GTAA, 806r GGACTACHVGGGTWTCTAAT; Caporaso et al.,
2011) that were linked to barcodes and illumina adapters
following Kozich et al. (2013). PCRs of DNA samples and
no-template negative controls were run with HotStarTaq
DNA Polymerase Master Mix (Qiagen) and thermocycler
conditions: 3 min at 94°C followed by 30 cycles of 94°C 45 s,
50°C for 60 s, and 72°C for 90 s, followed by 10 min at 72°C
(Caporaso et al., 2012). PCR products were purified and
normalized with SequalPrep Normalization Plates (Thermo-
Fisher), and sequenced with Illumina Miseq V2 paired end
250 bp sequencing. Sequences were denoised using DADA2
(Callahan et al., 2016) implemented in QIME2 (Bolten et al.,
2019) using default settings to prepare an abundance table of
unique amplified sequences variants (ASVs). Sequences were
taxonomically classified with the SILVA 16S rRNA gene database
v.132 (Quast et al., 2013), and ASVs were removed if they were
classified as chloroplasts or mitochondria, or if they were not
classified to the domains Bacteria and Archaea. The dataset was
then rarefied to 1,164 sequences per sample prior to calculation
of biodiversity metrics. Rarefaction to 1,164 sequences resulted
in undersampling of some communities, which is not unusual
in microbial studies, but allowed for retention of the most
samples for the analysis (Supplementary Figure 1). Sequences
from no-template PCR controls that passed DADA2 quality
control represented 6 ASVs that did not appear in the rarefied
ASV dataset. In total, DNA was sequenced from 38 samples
within the Willamette watershed and 17 from the Deschutes
watershed. DNA sequence data is archived under BioProject
| Characteristic | Wil mean (SD) | Wil r | Des mean (SD) | Des r | All mean (SD) | All r | Sm mean (SD) | Sm r | Lg mean (SD) | Lg r | Description |
|----------------|---------------|-------|---------------|-------|---------------|-------|---------------|------|---------------|------|-------------|
| Geomorphic     |               |       |               |       |               |       |               |      |               |      |             |
| ORREG2         | 10,001 (0)    | –     | 363 (0)       | –     | 7,020 (4,490) | 0.31***| 7,590 (4,250) | 0.62***| 6,430 (4,740) | 0.04 | Oregon Region Number |
| ELEV (m²)      | 693 (326)     | 0.19  | 1,480 (214)   | 0.02  | 935 (469)     | 0.26***| 925 (491)     | 0.44***| 946 (454)     | 0.07 | Mean basin elevation |
| DISTANCE (km)  | 83.01 (50.27) | 0.36***| 72.20 (55.63) | 0.15  | 104.1 (57.64) | 0.24***| 81.60 (55.63) | 0.41***| 116.1 (61.58) | 0.02 | Great-circle distance between sample sites |
| MINBELEV (m)   | 258 (275)     | –0.10 | 1,030 (415)   | 0.21  | 498 (483)     | 0.26***| 621 (486)     | 0.39** | 370 (494)     | 0.11 | Minimum basin elevation |
| BSLOPD (degrees) | 16 (6.0)    | 0.15  | 6.5 (1.6)     | –0.03 | 13 (7)        | 0.14  | 16 (8)        | 0.38***| 11 (5)        | 0.03 | Mean basin slope measured in degrees |
| LATITUDE (decimal degrees) | 44.58 (0.560) | 0.41***| 44.15 (0.546) | 0.12  | 44.45 (0.586) | 0.21***| 44.28 (0.516) | 0.35***| 44.62 (0.614) | 0.02 | Latitudinal coordinate |
| ELEVMAX (m)    | 1,720 (930)   | 0     | 2,630 (523)   | 0.26  | 2,000 (923)   | –0.01 | 1,500 (704)   | 0.30** | 2,520 (845)   | 0.10 | Maximum basin elevation |
| OR_HIPERMA (percent) | 10.9 (13.3) | 0.35**| 6.89 (7.04)   | –0.02 | 9.64 (11.8)   | 0.13  | 5.44 (10.3)   | 0.17  | 14.0 (11.9) | 0.07 | Percent basin surface area containing high permeability aquifer units as defined in SIR 2008-5126 |
| Climatic       |               |       |               |       |               |       |               |      |               |      |             |
| JANMIN TEMP (degrees C) | –1.40 (1.26) | 0.08  | –6.89 (0.758) | 0.01  | –3.09 (2.80) | 0.27**| –2.75 (2.77) | 0.55***| –3.45 (2.83) | 0.04 | Mean minimum January temperature |
| MINTEMP (degrees C) | 3.65 (1.25)  | 0.09  | –1.01 (0.975) | 0.06  | 2.21 (2.47)  | 0.27***| 2.38 (2.56)  | 0.53***| 2.02 (2.39)  | 0.04 | Mean annual minimum air temperature over basin surface area as defined in SIR 2008-5126 |
| JANMAX TEMP (degrees C) | 6.15 (0.99)  | –0.10 | 2.45 (1.00)   | 0.00  | 5.01 (1.99)  | 0.25**| 5.23 (2.25) | 0.50***| 4.77 (1.68) | 0.04 | Mean maximum January temperature |
| JANMIN T2K (degrees C) | –0.83 (1.18) | –0.09 | –6.46 (0.66) | 0.11  | –2.57 (2.83) | 0.22***| –2.12 (2.70) | 0.49***| –3.04 (2.93) | 0.02 | Mean minimum January temperature from 2K resolution PRISM 1961–1990 data |
| MAXTEMP (degrees C) | 15.2 (1.1)   | –0.08 | 12.6 (1.8)    | 0.09  | 14.4 (1.8)   | 0.23**| 14.5 (2.3)   | 0.46** | 14.3 (1.3) | 0.10 | Mean annual maximum air temperature over basin area from PRISM 1971–2000 800-m grid data |
| JANAVPRES2K (mm) | 264 (41)     | 0.21  | 141 (66)      | 0.13  | 226 (76)     | 0.17  | 250 (57)     | 0.45***| 201 (85) | 0.05 | Mean January precipitation |
| PRECIP (mm)    | 1,860 (339)   | 0.14  | 979 (440)     | 0.16  | 1,590 (552)  | 0.13  | 1,800 (449)  | 0.39***| 1,370 (569) | 0.03 | Mean Annual precipitation |
| JANMAX T2K (degrees C) | 6.34 (1.07)  | –0.13 | 4.08 (1.40)   | 0.33  | 5.64 (1.57)  | 0.24***| 5.55 (1.84) | 0.39** | 5.74 (1.26) | 0.12 | Mean maximum January temperature from 2K resolution PRISM 1961–1990 data |
| Land-cover     |               |       |               |       |               |       |               |      |               |      |             |
| SOILPERM (mm per hour) | 47.7 (28.0) | –0.05 | 151 (79.3)    | –0.02 | 79.8 (68.8)  | 0.22**| 77.1 (69.1)  | 0.38** | 82.5 (69.8) | 0.07 | Average soil permeability |
| LC11FORSHB (percent) | 83 (19)     | 0.34**| 89 (6)        | 0.01  | 85 (16)      | 0.10  | 90 (14)      | 0.18  | 79 (17)     | 0.02 | Percentage of forests and shrub lands, classes 41 to 52, from NLCD 2011 |

Note: only basin characteristics that were significantly correlated with microbial community similarity (Bonferroni-adjusted p < 0.1) in at least one group of sub-catchments are presented here (see Supplementary Table 2 for results of all characteristics).
Microbial Metrics of Biodiversity

Biodiversity is a fundamental metric used to characterize a microbial community. Alpha diversity describes diversity within a site, whereas beta diversity describes diversity across sites (Whittaker, 1972). Alpha diversity is often described with Shannon’s index ($H$), which is the Shannon entropy (Shannon and Weaver, 1949) of measured ASVs within a site. Shannon’s index accounts for both the number of ASVs (richness) and the relative abundance of each ASV (evenness). Shannon’s index was calculated using the sci-kit bio package (v 0.5.1, http://scikit-bio.org/) developed for the Python programming language. In this implementation, Shannon’s index is calculated as

$$H = - \sum_{i=1}^{s} (p_i \log_2(p_i)),$$

where $s$ is the number of ASVs detected and $p_i$ is the proportion of $s$ represented by ASV $i$. Note that changing the base of the logarithm changes the units of $H$; when base 2 is used, as here, the resulting quantity is described in units of bits. Larger values of $H$ indicate greater diversity within a site. Diversity data were visually approximately normally distributed, and variance was similar between groups, so two-sample $t$-tests were used to check for differences in means between groups. Statistical tests were conducted using SciPy in Python 3.7.4 (Virtanen et al., 2020).

Microbial dissimilarity between sites was determined by the Bray-Curtis metric (Bray and Curtis, 1957). Bray-Curtis dissimilarity, $d_{BC}(u, v)$ [unitless], calculates dissimilarity in the number ($n$) of ASVs between sites $u$ and $v$ as

$$d_{BC}(u, v) = \frac{\sum_{i} |n_{u,i} - n_{v,i}|}{\sum_{i} |n_{u,i} + n_{v,i}|},$$

for all ASV $i$ (Bray and Curtis, 1957). Bray-Curtis dissimilarity is more robust than other distance measures like Euclidean distance, for example, where differences in abundance can overwhelm differences in ASVs between sites, leading to counterintuitive results, particularly with large, sparse matrices (Ricotta and Podani, 2017). The denominator of the Bray-Curtis dissimilarity index effectively weights the difference in abundance between sites by the overall abundance of each ASV, such that rare ASVs contribute less to differences between sites.
Bray-Curtis dissimilarity ranges \([0, 1]\), with a value of one indicating two sites have no ASVs in common and a value of zero indicating identical ASVs composition. The package sci-kit bio was again used to assemble a distance matrix of Bray-Curtis dissimilarity between all possible pairs of sites. Site-pairs were ranked by Bray-Curtis dissimilarity and the top 1% and top 5% most and least dissimilar site-pairs were analyzed to explore whether patterns emerged.

Additionally, we were interested in how relationships between microbial communities and macroscale characteristics might differ between the Willamette and Deschutes watersheds and between small and large sub-catchments. Therefore, distance matrices of Bray-Curtis dissimilarity between all pair of sites in: the Willamette Basin \((n = 38 \text{ sites}; 703 \text{ unique pairs})\), the Deschutes Basin \((n = 17 \text{ sites}; 136 \text{ unique pairs})\), small sub-catchments \((n = 28 \text{ sites}; 378 \text{ unique pairs})\), and large sub-catchments \((n = 27 \text{ sites}; 351 \text{ unique pairs})\) were assembled. Note for all 55 sites there were 1,485 unique pairs. Small sub-catchments were simply the smaller half of sample sites (drainage area \(\leq 520 \text{ km}^2\)) and large sub-catchments were the larger half. Dissimilarity values were skewed low, so to check for differences between groups, a Mann-Whitney \(U\)-test, which requires no distribution assumptions, was used. Mann-Whitney \(U\)-test assumes independent samples and tests the null hypothesis that the two distributions are equal.

**Macroscale Characteristics and Microbial Community Composition**

To investigate the relationship between microbial community and macroscale characteristics, differences in microbial community composition were examined in relation to differences in catchment characteristics between sites. Using the basin characteristics from StreamStats, a distance matrix of pairwise differences between sites for each characteristic was assembled. Distances between points \((u, v)\) for StreamStats characteristic \(c\) were calculated as absolute value arithmetic differences,

\[
d_{SS,c}(u, v) = |c_u - c_v|
\]

with the exception of the physical distance between points. The physical distance was the great-circle distance (km), which is the distance between the two points on the surface of a sphere containing the diameter of Earth. As with beta diversity, distances matrices were assembled for all characteristics and differences between all pairs of sites in: the Willamette Basin, the Deschutes Basin, small sub-catchments, large sub-catchments, and all sub-catchments. We used Mantel tests to evaluate the correlation between dissimilarly matrices of community composition and each of the macroscale properties (Mantel, 1967). Mantel correlations assume linear correlations between dissimilarity matrices and have been previously used to evaluate correspondence between microbial communities and other environmental factors (Fagervold et al., 2014; Repert et al., 2014; Read et al., 2015). When applied to raw data tables this approach may introduce bias due to spatial autocorrelation, though extensive testing has found this method is appropriate for evaluation of dissimilarly matrices constructed from autocorrelated data (Legendre et al., 2015). The Mantel statistic is the standardized Pearson correlation between the distance matrices and ranges \([-1, 1]\), with values close to 1 \((-1\) indicating strong positive (negative) correlation and a value of zero indicating no correlation between the distance matrices. To assess the relationship between microbial community composition and macroscale environment, we examined Mantel statistics between microbial distances \(d_{BC}(u, v)\) and individual characteristic distances \(d_{SS,c}(u, v)\), as well as mean Mantel statistics for the four groups of related properties: geomorphic, climate, land-cover, and development characteristics. We again used the package sci-kit bio for Python and calculated Mantel statistic \((r)\) and Bonferroni-corrected \(p\)-value for multiple comparisons over 10,000 permutations.

We might expect to detect stronger correlations between the microbial community and watershed characteristics when the microbial community samples represent a greater range of conditions. For example, consider the case where catchment characteristics are identical so their standard deviation is zero. Because dissimilarity among catchment characteristics would be zero in this case, correlation between microbial community dissimilarity and catchment dissimilarity (Mantel \(r\)) would also necessarily be zero. Thus, as variability in sub-catchment characteristics decreases near to geospatial measurement noise associated with the StreamStats outputs we potentially may see lower Mantel statistics. To explore whether the strength of correlations identified by Mantel tests could be related to variability in those catchment characteristics, we analyzed the relationship between Mantel statistics and the standard deviation of each characteristic. We calculated the relative sensitivity \(\varepsilon_{SS,c}\) [unitless] of the Mantel statistic \((r)\) to the standard deviation \((\sigma)\) of each StreamStats characteristic \((c)\) as:

\[
\varepsilon_{SS,c}(m, n) = \frac{r_m - r_n}{\sigma_m - \sigma_n} \times \frac{\sigma_m}{r_m}
\]

The relative sensitivity value quantifies the extent to which a change in variability in a StreamStats characteristics translates to a respective change in the correlation between microbial community composition and that watershed characteristic, where higher absolute values of \(\varepsilon\) indicate greater sensitivity and a stronger relationship between microbial community similarity and watershed characteristic variability. Values of \(\varepsilon\) that significantly differ from zero may suggest variability as a potential driver of the strength of correlations and offer insight into the conditions in which microbiomes may be useful monitoring tools. Sensitivity was calculated for each StreamStats variable for both Willamette vs. Deschutes watersheds and small vs. large sub-catchments, and then median sensitivity for each group of characteristics for both sets of sub-catchments was estimated. Sensitivity data contained several outliers and could not be assumed normal. Therefore, the non-parametric one-sample Wilcoxon signed-rank test was used to test the null hypothesis that the median sensitivity value for each group was equal to zero.

Finally, to assess whether and how different ways of grouping sequence data or applying different diversity metrics to characterize the microbial communities impacts the results of
the analysis, the data were reanalyzed several ways. First, rather than using ASVs (100% similar), stream microbial communities were instead characterized by grouping raw sequences into 95%, 97%, or 99% similar operational taxonomic units (OTUs). Sequence data were then rarefied to standardize sampling effort while retaining the greatest number of samples (95% similarity: 1,025 sequences per sample; 97% similarity: 1,023 sequences; 99% similarity: 1,014 sequences). Communities were also alternatively characterized using one of five major groups of ASVs: Actinobacteria (rarefied to 100 sequences per sample), Bacteroidetes (500 sequences), Cyanobacteria (50 sequences), Gammaproteobacteria (500 sequences), or Verrucomicrobia (100 sequences). Two alternative measures of alpha diversity, Chao-1 index of taxonomic diversity and abundance-based coverage estimators (ACE) index, were calculated (also using sci-kit bio). Finally, Weighted UniFrac distances (Lozupone and Knight, 2005) were calculated with QIIME2 as an alternate measure of beta diversity. Unlike the taxon-based Bray-Curtis dissimilarity, divergence-based UniFrac distances (Lozupone and Knight, 2008) consider the similarity of different taxa by incorporating information from a phylogenetic tree relating the genetic sequences from each sample.

All code developed for this analysis is available at www.zenodo.org (URycki and Good, 2020b).

RESULTS
Spatial Patterns in Microbial Community Similarity
We evaluated how 16S rRNA sequence data for 55 DNA samples varied across the Willamette and Deschutes watersheds. Among those samples, 3,530 unique ASVs were detected, including typical freshwater members of the classes Bacteroidetes, Actinobacteria, Verrucomicrobia, and Gammaproteobacteria (which includes Betaproteobacteria; Figure 2). Some samples also featured high abundances of Cyanobacteria (up to 28% of community).

Alpha diversity was similar in the Willamette and Deschutes watersheds (Wil: mean \( H = 5.81, SD = 1.08 \); Des: mean \( H = 6.58, SD = 0.93 \); two-sample \( t = 2.50, p = 0.016 \); Figure 1) and was not related to the sample volume filtered (\( R^2 = 0.04, p = 0.76 \), Supplementary Figure 2). Tributaries exhibited both the highest and lowest biodiversity values in both watersheds. Alpha diversity generally increased from south to north in the Willamette Basin (latitude \( R^2 = 0.55, p < 0.001 \)), but this pattern was not observed in the Deschutes Basin (latitude \( R^2 = -0.41, p > 0.1 \); Figure 3). Across both watersheds, small and large sub-catchments exhibited similar levels of biodiversity (small sub-catchments: mean \( H = 6.09, SD = 1.22 \), large sub-catchments: mean \( H = 6.00, SD = 0.95 \); two-sample \( t = 0.30, p = 0.762 \)), and alpha diversity was not related to sub-catchment drainage area (Willamette: \( R^2 = 0.04, p = 0.80 \); Deschutes: \( R^2 = -0.38, p = 0.13 \); Supplementary Figure 3).

The greatest beta diversity in microbial stream communities was observed between watersheds. Bray-Curtis dissimilarity (BC) was higher on average for site-pairs spanning watersheds (mean BC = 0.912) than for pairs within watersheds (mean BC = 0.832, Mann-Whitney \( U = 164584.5, p < 0.001 \), and the vast majority of the top 5% BC scores were for inter-watershed sample pairs (Figure 4). In fact, the most dissimilar pairs (top 1% BC scores) were between points extending from the headwaters of the Deschutes River to the mouth of the Willamette River. A few of the top 5% most dissimilar pairs were points in the same watershed that were geographically distant. Visual inspection reveals no clear patterns between small and large watersheds among the top 5% least similar pairs of samples, and beta diversity was similar within small and large sub-catchments (small: mean BC = 0.856, large: 0.875; Mann-Whitney \( U = 65169.5, p = 0.340 \)).

Beta diversity was higher within the Willamette Basin (mean BC = 0.897) than within the Deschutes Basin (mean BC = 0.820; Mann-Whitney \( U = 31305.5, p < 0.001 \)). The lowest beta diversity in microbial stream communities (5% lowest BC scores) was observed between samples within in the upper Willamette Basin (Figure 4). Points in the HJ Andrews Forest exhibited some of the lowest dissimilarity to other points within the HJ Andrews Forest and to several other points within the Willamette Basin. Also, a few of the lowest dissimilarity scores were between inter-watershed sample pairs, including pairs spanning headwaters of one watershed to the mouth of the other watershed. Almost all of the most similar inter-watershed pairs are samples from large sub-catchments (i.e., points along the mainstem of the rivers).

Relation of Macroscale Catchment Properties to Microbial Community Similarity
Among StreamStats characteristics determined to be statistically significant in at least one group of sub-catchments (Table 1), geomorphic-related characteristics were on average the most strongly correlated with microbial community composition in the Willamette Basin (mean \( r = 0.19 \); Figure 5). Land-cover (mean \( r = 0.15 \)) characteristics were more strongly correlated with microbial community composition than climatic (mean \( r = 0.01 \)) characteristics. Among the top five strongest correlates in the Willamette Basin were latitude (\( r = 0.41 \)), percentage of area containing high permeability aquifer units (\( r = 0.35 \)), and percentage of forest and shrublands (\( r = 0.34 \); Table 1). In the Deschutes Basin, climatic (mean \( r = 0.11 \)) characteristics were more strongly correlated with microbial community composition than geomorphic (mean \( r = 0.10 \)) characteristics. Land-cover characteristics were very weakly anticorrelated (mean \( r = -0.01 \); Figure 5). Among the top five strongest correlates with the microbial community in the Deschutes Basin were mean maximum January temperature (\( r = 0.33 \)), percentage of low-intensity development (\( r = 0.30 \)), topographic index (\( r = 0.33 \)), and topographic relief (\( r = 0.29 \)), although none of these were statistically significant (all \( p > 0.1 \); Table 1, Supplementary Table 2). No development-related characteristics were found to be correlated with microbial community similarity in the Willamette or Deschutes watersheds.
Small sub-catchments exhibited the strongest correlations between macroscale catchment characteristics and microbial community composition. Among StreamStats characteristics determined to be statistically significant in at least one group of sub-catchments (Table 1), microbial community composition in small sub-catchments correlated most strongly with climatic characteristics (mean \( r = 0.46 \)), followed by geomorphic (mean \( r = 0.38 \)), and land-cover (mean \( r = 0.28 \); Figure 5) characteristics. The most strongly correlated characteristics were watershed identifier (Willamette or Deschutes; \( r = 0.62 \)), January minimum temperature (\( r = 0.55 \)), and annual minimum temperature (\( r = 0.53 \); Table 1). Microbial community composition in large sub-catchments exhibited much weaker correlations. Mean maximum January temperature (\( r = 0.12 \)), minimum basin elevation (\( r = 0.10 \)), and mean annual maximum temperature (\( r = 0.10 \)) were among the top five strongest correlates with microbial community similarity, although none of these were statistically significant (all \( p > 0.1 \)). No development-related characteristics were found to be statistically correlated with microbial community similarity in small or large watersheds.

For the macroscale characteristics we analyzed, the strength of the correlation with the microbial community was not sensitive to the variability of those characteristics across catchments. Across all sub-catchment characteristics, an increase in the standard deviation of a characteristic did not translate to a statistically significant increase in Mantel statistic for either the Willamette versus Deschutes watersheds (median \( \epsilon = 0.14 \); Wilcoxon signed-rank \( W = 76.0, p = 0.981 \)) or for small versus large sub-catchments (median \( \epsilon = -3.40 \); Wilcoxon signed-rank \( W = 58.0, p = 0.381 \)). Similarly, median sensitivity was not statistically different from zero for any of the characteristic groups for either the Willamette vs. Deschutes
URycki et al. River Microbiome and Sub-catchment Characteristics

FIGURE 4 | Map of the 1% (dark lines) and 5% (light lines) most (left) and least (right) dissimilar microbial communities throughout the Willamette and Deschutes watersheds in Oregon, USA. Large (filled symbols) and small (unfilled symbols) sub-catchments are those with more than or less than median drainage area, respectively. Inset shows vicinity of HJ Andrews Experimental Forest.

Patterns were consistent when sequence data were grouped into OTUs or when different diversity metrics were applied. The strongest relationships between microbial communities and watershed characteristics were observed in small watersheds at all OTU sequence similarity levels (95%, 97%, 99%, or 100%) and for three major taxonomic groups: Bacteroidetes, Gammaproteobacteria, and Verrucomicrobia (Supplementary Figure 4). The microbial groups Actinobacteria and Cyanobacteria exhibited no significant correlations with watershed characteristics. The strongest relationships were observed between microbial communities and geomorphic and climatic related characteristics, although some microbial groups were also related to land cover characteristics (Supplementary Figure 4). The Chao-1 and ACE alpha diversity metrics were strongly correlated with Shannon index (Chao-1: Spearman $r = 0.91$, $p << 0.001$; ACE: Spearman $r = 0.90$, $p << 0.001$). Characterizing community similarity using Weighted UniFrac instead of Bray-Curtis dissimilarity also resulted in strong correlations with watershed characteristics in small watersheds. Also, some microbial groups exhibited comparably strong relationships to watershed characteristics in the Willamette and Deschutes watersheds (Supplementary Figure 5).

DISCUSSION

We explored the potential influence of the upstream macroscale environment in shaping streamwater microbiomes. Previous studies found that streamwater microbial community composition is more strongly correlated with catchment-scale hydrologic parameters such as stream distance and catchment area than with water sample physio-chemical water properties (Read et al., 2015; Savio et al., 2015). Our study builds on this previous work by expanding the suite of macroscale variables analyzed for relationships with the microbial community assemblage to more than 40 basin characteristics reflecting the geomorphology, climate, land cover, and level of human development in stream catchments.

Our results suggest that, in headwater catchments, microbial community assemblages are shaped by catchment-scale geomorphology and climate, but these influences weaken downstream. The streamwater microbial metacommunity is “seeded” from a diverse amalgamation of microbes dispersed into headwater streams from surrounding soil and groundwater (Crump et al., 2007, 2012), which develops and continually combines with local inputs as the community moves downstream with the flow of water (Ruiz-González et al., 2015). In headwater streams in particular, where the contributing area is large relative to the stream volume (Read et al., 2015; Savio et al., 2015), it follows that this immigrant microbial community would reflect the dispersal area and upslope environment and thus exhibit a stronger correlation with macroscale catchment characteristics. Downstream, however, biotic and abiotic factors increasingly drive ecological succession (i.e., species-sorting; Leibold et al., 2004), and the microbial assemblage shifts to a core riverine community, dampening signals from the upstream catchment (Savio et al., 2015).

The shift from diverse communities of upslope emigrants that are tightly coupled to the catchment to a core riverine community shaped by the local environment may explain the decreasing...
strength of the relationship between microbial community composition and macroscale watershed characteristics. This is supported by the low overall sensitivity of the correlations to variability in sub-catchment characteristics moving from upstream to downstream, in that moving to more (or less) variable sub-catchment characteristics did not result in higher (or lower) correlations uniformly. Since sensitivity values were distributed between both positive and negative values and not statistically different than zero, we conclude that changes in the homogeneity of the landscape across which samples are collected, as measured by the standard deviation of landscape characteristics, is not uniformly the driving processes determining differences in microbial composition.

Watershed location in either the Willamette or Deschutes basins was the strongest correlate with the microbial community across all basins and within small basins, indicating that there are additional factors shaping distinctive microbial communities in each basin. In the temperate, more developed Willamette Basin, latitude and the percentage of forest and shrub land and were most important characteristics. Due to the north-south orientation of these basins, latitude is likely a proxy for multiple other interacting factors (e.g., elevation and up-stream drainage area), and the Willamette Basin in particular has a strong human-driven gradient due to Portland at the outlet. On the other hand, in the Deschutes Basin, geomorphic characteristics such as topographic index and relief were more important. It is possible that the differences in important influences on streamwater microbial communities between basins is related to basin hydrology and the nature of the inputs to the stream. In the wetter, more crop-dominated landscape of the Willamette Basin, more constant, stable inputs of water—and thus microbes—may contribute to the development of a certain, less diverse community than in the arid, less populated Deschutes Basin, where flashier water inputs result in more variable contributing areas and a different, more diverse streamwater microbial community (Nippgen et al., 2015).

We note that, as with many scientific analyses, our results are a product of the decisions made throughout the analysis. Decisions about DNA extraction methods and sequencing depth may have impacted our results. Some catchment characteristics obtained from StreamStats that were considered redundant were eliminated from analysis (see Supplementary Table 1), but correlation among variables was not explored. Eliminating correlated variables may have yielded different results, as could have applying different significance criteria. On the other hand, results were robust to the similarity of ASV groups. When ASVs were grouped into operational taxonomic units (OTUs) based on 95%, 97%, or 99% DNA sequence similarity, the relative strength and importance of characteristic categories and of spatial scale were generally stable (Supplementary Figure 4). Although some major microbial groups (e.g., classes Bacteroidetes, Gammaproteobacteria, Verrucomicrobia) exhibited stronger relationships with watershed characteristics than other groups, we are not aware of any method that would allow for reliable targeted sampling or analysis effort in such a way as to appreciably conserve resources. Our results build upon those of Good et al. (2018), in which streamwater microbiomes were used in a machine learning algorithm to predict hydrologic regime of a set of large rivers in adjacent and detached watersheds in the Arctic. Here, we found that the summer microbial community within small headwater streams reflects both the structural configuration of the landscape as well as upstream processes. Coupled with the results of Good et al. (2018), our results offer an encouraging indication that streamwater microbial DNA may thus carry information about upstream macroscale conditions as well as hydrology and may therefore hold potential as a useful tool in watershed monitoring. More research is needed to determine whether these relationships hold in other seasons and how to optimally extract this information from microbiomes.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in the Sequence Read Archive at the National Center for Biotechnology Information (NCBI; https://www.ncbi.nlm.nih.gov/sra/PRJNA642636).

AUTHOR CONTRIBUTIONS

SG conceived the study, produced the original code, and reviewed the manuscript. DU collected the data, contributed to the data analysis, and prepared the manuscript and figures. JC contributed to sample and data analysis. BC contributed sample analysis and reviewed the manuscript. GJ contributed to data collection and reviewed the manuscript. SG and BC secured funding for this
ACKNOWLEDGMENTS

We conducted part of this research at HJ Andrews Experimental Forest, which was funded by the US Forest Service, Pacific Northwest Research Station. We gratefully acknowledge the efforts of the many graduate and undergraduate students who generously contributed many long hours of lab and field work. We additionally thank Associate Editor TS and two reviewers for their helpful, constructive reviews, which improved this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/frwa.2020.574728/full#supplementary-material
