AN ABSTRACT OF THE DISSERTATION OF

Dawn URycki for the degree of Doctor of Philosophy in Water Resources Engineering and Biological and Ecological Engineering presented on March 10, 2022.

Title: Streamwater Microbial Communities as Hydrologic Observation: Insights Gained from Investigation of a Novel Dataset

Abstract approved:

_____

Stephen P. Good

Interactions and feedbacks among climate change effects and continued human impacts will exacerbate impacts to water resources in complex ways. An urgent imperative of the hydrologic community is to understand the response of hydrologic systems to these perturbations, thus contributing to long-term sustainability of water resources in an uncertain future. Critical to anticipating and mitigating changes to water resources is a thorough characterization of hydrologic processes across a variety of systems and spatiotemporal scales. However, major gaps in our understanding persist, particularly regarding the movement of water through the catchment and streamflow generation, despite decades of research supported by large amounts of highly complex hydrological observation data. We suggest that a new type of information-rich data, which can be easily collected and analyzed, might be the key to new insights that propel the field toward a deeper, more fundamental understanding of hydrologic processes. Genetic material (i.e., DNA) is a naturally-occurring, high-dimensional digitally-encoded dataset, and DNA analysis has become much cheaper and more widely available in recent years. Microbial communities, characterized taxonomically by sequencing the 16S rRNA gene in DNA, are highly diverse and respond dynamically to environmental conditions. Here, we investigate the streamwater microbial community as a novel hydrologic dataset. We collected DNA samples over three years, from 2017-2020, from more than 60 streams across the

Willamette, Deschutes, and John Day watersheds, three large and characteristically divergent watersheds in Oregon, USA. We found that differences in microbial community composition among streams, particularly in headwater streams, was statistically related to differences in geomorphic and climatic characteristics of the drainage catchment. We furthermore found through an information-theoretic approach, that specific summer community constituents (i.e., microbial taxa) were related to stream discharge metrics at multiple temporal and flow scales. We also observed that streamwater microbial diversity exhibited a rich and dynamic response to event hydrograph dynamics on the Marys River in the Willamette Valley of Oregon. In that analysis, we furthermore classified microbial taxa (and broader phylogenetic groups) according to whether they are mobilized or diluted with streamflow, potentially contributing new insights regarding the sources of streamflow, as well as a new way of characterizing taxa in microbial ecology studies. Results of this research support further investigation of the hydrologic information encoded within microbial communities, as well as ways in which the field might best extract and apply this new information to further our understanding of hydrologic systems and contribute fresh insight to unsolved questions in hydrology.

Streamwater Microbial Communities as Hydrologic Observation: Insights Gained from Investigation of a Novel Dataset

by
Dawn URycki

A DISSERTATION

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Presented March 10, 2022
Commencement June 2022

Doctor of Philosophy dissertation of Dawn URycki presented on March 10, 2022

APPROVED:

Major Professor, representing Water Resources Engineering

Major Professor, representing Biological and Ecological Engineering

Director of the Water Resources Graduate Program

Head of the Department of Biological and Ecological Engineering

Dean of the Graduate School

I understand that my dissertation will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my dissertation to any reader upon request.

Dawn URycki, Author

ACKNOWLEDGEMENTS

This work was supported the National Science Foundation grant EAR 1836768. I gratefully acknowledge STEM Scholarship support from NSF grant 1153490. Data and facilities for a portion of this research were provided by the HJ Andrews Experimental Forest and Long Term Ecological Research (LTER) program, administered cooperatively by the USDA Forest Service Pacific Northwest Research Station, Oregon State University, and the Willamette National Forest. This material is based upon work supported by the National Science Foundation under the LTER Grants: LTER8 DEB-2025755 (2020-2026) and LTER7 DEB-1440409 (2012-2020).

I solemnly acknowledge that Oregon State University in Corvallis, Oregon is located within the traditional homelands of the Ampinefu Band of Kalapuya. Following the Willamette Valley Treaty of 1855, Kalapuya people were forcibly removed to reservations in western Oregon. Today, living descendants of these people are a part of the Confederated Tribes of Grand Ronde Community of Oregon and the Confederated Tribes of the Siletz Indians.

It would be impossible to overstate the depth of my gratitude to my major advisor, Dr. Stephen P. Good, for his unwavering enthusiasm and support of me in this research. From his example, I learned that our most important and impactful work is to employ the sum of our resources, both tangible and intangible, to support and empower others, and to relentlessly seek new opportunities to that end. He is a credit to the scientific community far beyond his innumerable intellectual contributions. Thank you to Dr. Byron C. Crump for sharing his domain knowledge, as expansive, it seems, as microbial communities are diverse; I am particularly grateful for his kind generosity of time and attention, even as both were in such short supply. Thank you to committee members Dr. Kevin Bladon and Dr. Hong Liu for their support and helpful review of my research program and dissertation. Thank you to Dr. Andrew Meigs, who committed many hours to ensuring our programmatic path was as smooth as possible. My sincere gratitude to Dr. Mary Santelmann, Dr. Chad Higgins, and the faculty and staff in Biological and Ecological Engineering, particularly Jennifer Cohen, who have always supported and valued me as an individual first and a student second. Thank you

to Maoya Bassiouni for the bits of advice, snacks, code, and encouragement; I regret that I can only offer these lame words of gratitude rather than a hilarious cat/pizza gif. I thank my dear friend Carly Lettero for encouraging me to apply to this program and for all the professional and personal support she has offered me over more than two decades of friendship.

Thank you to my Dad, the most brilliant engineer I know, for teaching me to think like one, even though he would never call it that. To my Mom, for always insisting that I look up in the Dictionary words I didn't know how to spell ("*How can I look it up if I don't know how to spell it?*"), among many other lessons in resourcefulness; I have you to thank first for my grit and the ability to write. Thank you to my mother- and father-in-law, especially Paulette, for your consistent praise of my character and ambition, even as your son devoted every ounce of his energy to our family during this endeavor. Thank you to Weston and Gilia, who unwittingly forced me to take much-needed breaks and whose simple joy offered a grander perspective when it was sometimes otherwise difficult to see. Finally, thank you to my husband, Joseph Zumpano URycki, whose faith in me, and whose delight at all my fancy figures, sustained me through countless unexpected challenges. Your sacrifices to support me on this journey both humble and inspire me beyond measure.

CONTRIBUTION OF AUTHORS

This dissertation is the product of a collaborative effort. Individual contributions are as follows. Stephen P. Good conceptualized this research, acquired funding, and contributed resources, analysis, writing, and review of chapters 2-4. Byron C. Crump acquired funding and contributed resources, analysis, writing and review of chapters 2-4. Jessica Chadwick contributed sample collection and analysis for Chapter 2. Gerrad D. Jones contributed data collection and manuscript review for Chapter 2. Maoya Bassiouni contributed analysis, writing, and review for Chapter 3. Bonan Li contributed methodology for Chapter 3. J. Renée Brooks contributed data and analysis to Chapter 4. Natalie Ceperley contributed analysis to Chapter 4.

TABLE OF CONTENTS

Page

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

Page

LIST OF FIGURES

Figure Page

Figure Page

LIST OF FIGURES (Continued)

Figure Page

LIST OF FIGURES (Continued)

Figure Page

LIST OF TABLES

DEDICATION

For Zump,

whose kindness is skewed left in the range $(0, \infty]$,

and to Weston and Gilia:

I love you both $\infty \times \infty + n$, for n = all the days of your life.

1 Introduction

In recent years, earth systems have been reflecting more conspicuously the ramifications of a rapidly changing climate, in the form of more dramatic weather events and increasing fire frequency and severity, for example. As the climate continues to warm, drought frequency is projected to continue to increase, reducing surface and groundwater availability in arid regions and negatively impacting freshwater ecosystems and reducing surface water quality (Jiménez Cisneros et al., 2014; Delpla et al., 2009). Interactions and feedbacks among these effects and continued human impacts will exacerbate impacts to water resources in complex ways (e.g., IPCC, 2019; Feddema et al., 2005). A central imperative of contemporary hydrology is to understand and predict the response of hydrologic systems to these perturbations, thus contributing to long-term sustainability of water resources; yet major gaps in our understanding of these systems persist, particularly regarding subsurface processes (Blöschl et al., 2019).

Streamflow volume is arguably the most fundamental hydrologic measurement. An expression of landscape scale hydrology, stream discharge integrates physical aspects of the catchment such as topographic organization , lithology, and climatology, as well as biological influences, such as vegetation cover and soil type (Dingman, 2015). Thus, much can be learned about hydrologic function by analyzing stream gage records; yet many parts of the world, and indeed parts of the American arid southwest, Alaska, and Hawai'i (Kiang et al., 2013) lack sufficient spatial coverage of stream gages for even basic water resource monitoring. Seibert & McDonnell (2015) found, however, that key pieces of data collected on short, intensive field campaigns (i.e., 'soft data') could be powerful tools in characterizing catchment function in ungaged basins.

Yet even in areas covered by expansive networks of sophisticated observation data, including meteorological, eddy covariance, stable isotopes, and remotely sensed earth observation data, the specific mechanisms and drivers of many aspects of hydrologic function are still not well understood. For example, the field still lacks a thorough process-based understanding of the dynamics of streamflow generation, including specific sources of streamwater, flow paths of precipitation inputs, and transit time distributions, for example, which remain areas of active research (Kirchner, 2016;

Stockinger et al., 2016; Blöschl et al., 2019). It may be that new hydrologic knowledge may be limited not by the amount of data, but by the type of data traditionally employed in hydrologic studies. Thus, there is opportunity in hydrologic research, both in heavily instrumented and in ungaged basins, for a novel, information-dense dataset to promote transformative hydrologic insight.

DNA is a large, digitally-encoded dataset that has been applied across an increasingly wide array of research domains in recent years. Microbial communities native to an environment (i.e., microbiomes), commonly characterized by sequencing the 16S rRNA gene in DNA, are commonly employed in studies from oceanography to human health. Microbial communities are composed of many thousands of taxonomically and functionally diverse groups and are sensitive to perturbations, shifting species composition in response to changes in environmental conditions (Thompson et al., 2017). The composition of streamwater microbial communities has been found to be strongly related to hydrologic properties, including stream discharge (Crump and Hobbie, 2005; Doherty et al., 2017), and characteristics related to in-stream residence time, such as cumulative stream length upstream of the community (Read et al., 2015), river kilometer, dendritic stream length, and catchment size (Savio et al., 2015). These highly diverse communities may therefore comprise a rich dataset of information valuable to hydrologic studies.

Here, we investigate the potential utility and application of microbial DNA as a novel, informative hydrologic observation. We collected and analyzed DNA samples from more than 60 streams and rivers across three major, ecoclimatically divergent watersheds in Oregon, USA between 2017 and 2020. First, we relate streamwater microbial community composition to the physical characteristics of the catchment that also shape hydrologic function (Chapter 2). We next identify and quantify relationships between microbial taxa and hydrologic discharge at multiple temporal and flow scales (Chapter 3). We then capture and analyze the response of the microbial community to stream discharge dynamics over the course of an isolated precipitation event on the Marys River, and we compare this response to that of a traditional geochemical tracer (stable water isotopes; Chapter 4). Finally, we highlight our most

significant findings, discuss the broader implications of this research, and propose opportunities for further investigation (Chapter 5).

1.1 References

Blöschl, G., Bierkens, M. F. P., Chambel, A., Cudennec, C., Destouni, G., Fiori, A., … Zhang, Y. (2019). Twenty-three unsolved problems in hydrology (UPH) – a community perspective. *Hydrological Sciences Journal*, *64*(10), 1141–1158. https://doi.org/10.1080/02626667.2019.1620507

Crump, B. C., & Hobbie, J. E. (2005). Synchrony and seasonality in bacterioplankton communities of two temperate rivers. *Limnology and Oceanography*, *50*(6), 1718–1729. https://doi.org/10.4319/lo.2005.50.6.1718

Delpla, I., Jung, A. V., Baures, E., Clement, M., & Thomas, O. (2009, November 1). Impacts of climate change on surface water quality in relation to drinking water production. *Environment International*. Elsevier Ltd. https://doi.org/10.1016/j.envint.2009.07.001

Dingman, S. L. (2015). *Physical Hydrology* (3rd ed.). Prentice Hall.

Doherty, M., Yager, P. L., Moran, M. A., Coles, V. J., Fortunato, C. S., Krusche, A. V, … Crump, B. C. (2017). Bacterial Biogeography across the Amazon River-Ocean Continuum. *Frontiers in Microbiology*, *8*, 882. https://doi.org/10.3389/fmicb.2017.00882

Feddema, J. J., Oleson, K. W., Bonan, G. B., Mearns, L. O., Buja, L. E., Meehl, G. A., & Washington, W. M. (2005). The Importance of Land-Cover Change in Simulating Future Climates. *Science*, *310*(5754).

Kiang, J. E., Stewart, D. W., Archfield, S. A., Osborne, E. B., & Eng, K. (2013). A national streamflow network gap analysis. *Scientific Investigations Report*. https://doi.org/10.3133/SIR20135013

Kirchner, J. W. (2016). Aggregation in environmental systems-Part 2: Catchment mean transit times and young water fractions under hydrologic nonstationarity. *Hydrology and Earth System Sciences*, *20*(1), 299–328. https://doi.org/10.5194/HESS-20-299-2016

Read, D. S., Gweon, H. S., Bowes, M. J., Newbold, L. K., Field, D., Bailey, M. J., & Griffiths, R. I. (2015). Catchment-scale biogeography of riverine bacterioplankton. *The ISME Journal*, *9*(2), 516–526. https://doi.org/10.1038/ismej.2014.166

Savio, D., Sinclair, L., Ijaz, U. Z., Parajka, J., Reischer, G. H., Stadler, P., … Eiler, A. (2015). Bacterial diversity along a 2600?km river continuum. *Environmental Microbiology*, *17*(12), 4994–5007. https://doi.org/10.1111/1462-2920.12886

Seibert, J., & McDonnell, J. J. (2015). Gauging the Ungauged Basin: Relative Value of Soft and Hard Data. *Journal of Hydrologic Engineering*, *20*(1), A4014004. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000861

Stockinger, M. P., Bogena, H. R., Lücke, A., Diekkrüger, B., Cornelissen, T., & Vereecken, H. (2016). Tracer sampling frequency influences estimates of young water fraction and streamwater transit time distribution. *Journal of Hydrology*, *541*, 952–964. https://doi.org/10.1016/J.JHYDROL.2016.08.007

Thompson, L. R., Sanders, J. G., McDonald, D., Amir, A., Ladau, J., Locey, K. J., … Knight, R. (2017). A communal catalogue reveals Earth's multiscale microbial diversity. *Nature*, *551*(7681), 457–463. https://doi.org/10.1038/nature24621

RIVER MICROBIOME COMPOSITION REFLECTS MACROSCALE CLIMATIC AND GEOMORPHIC DIFFERENCES IN HEADWATER STREAMS

D. R. URycki, S. P. Good, B. C. Crump, J. Chadwick, and G. D. Jones

Frontiers in Water
Holyoke Building, 107 Spring Street, Seattle, WA 98104
Volume 2

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

2.1 Abstract

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, 0.38, and 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.

2.2 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 (Jiménez Cisneros et al., 2014; Delpla et al., 2009). 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; Gregory et al., 2006; Moscrip & Montgomery, 1997; Wheater and Evans, 2009). Interactions among these factors exacerbate impacts to water resources (e.g., IPCC, 2019; Feddema et al., 2005), and better tools are needed to diagnose these effects on catchment ecohydrology at local scales.

In watershed catchments lacking sufficient hydrologic data, Seibert & McDonnell (2015) found that key pieces of data collected on short field campaigns (i.e., 'soft data') can prove useful for understanding catchment functions. 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, temperature (Read et al., 2015; Savio et al., 2015; Fortunato et al., 2013; Doherty et al., 2017). Given that streamwater microbiomes are composed primarily of microbes that originate in upslope soil and groundwater environments; 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 (Shogren et al., 2017; Jerde et al., 2016) 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

& 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.

2.3 Materials and Methods

2.3.1 Watershed and Sub-Catchment Characteristics

The Willamette and Deschutes basins, two similarly sized large (Willamette = 29,000 km^2; Deschutes = 27,700 km^2), 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 m^3/s) is thus much greater than the Deschutes River (165 m^3/s; U.S. Geological Survey, 2016) 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 °C and 15 °C in the Willamette Basin and 0 °C 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 (Table 2.S1).

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 (https://streamstats.usgs.gov/ss/; Ries et al., 2017). 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 python script was developed to obtain all available StreamStats basin characteristics for each of our 61sampling sites distributed throughout the two watersheds (URycki and Good, 2020*a*). 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 2.1). 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’ (Table 2.S1). We then grouped remaining variables into four categories: climatic, geomorphic, land-cover, and development characteristics.

2.4 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; Fig. 2.1). 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.

Table 2.1. Mean, standard deviation, and correlation with microbial community similarity (Mantel statistic [*r*]) for statistically significant StreamStats macroscale basin characteristics by watershed and in small and large sub-catchments across the Willamette and Deschutes watersheds, Oregon, USA. Note that 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 Table 2.S2 for results of all characteristics).

Characteristic		Willamette		Deschutes		All		Small		Large		Description
		mean (SD)	*r*	mean (SD)	*r*	mean (SD)	*r*	mean (SD)	*r*	mean (SD)	*r*	
Geomorphic	ORREG2 (dimensionless)	10001 (0)	-	363 (0)	-	7020 (4490)	0.31***[a]	7590 (4250)	0.62***	6430 (4740)	0.04	Oregon Region Number
	ELEV (m[b])	693 (326)	0.19	1480 (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	1030 (415)	0.21	498 (483)	0.26***	621 (486)	0.39**	370 (454)	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)	1720 (930)	0	2630 (523)	0.26	2000 (923)	-0.01	1500 (704)	0.30*	2520 (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 perm-eability aquifer units as defined in SIR 2008-5126
Climatic	JANMINTMP (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
	JANMAXTMP (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
	JANMINT2K (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 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
	JANAVPRE2K (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)	1860 (339)	0.14	979 (440)	0.16	1590 (552)	0.13	1800 (449)	0.39***	1370 (569)	0.03	Mean Annual precipitation
	JANMAXT2K (degrees C)	6.34 (1.07)	-0.13	4.08 (1.40)	0.33	5.64 (1.57)	0.24***	5.55 (1.84)	0.33**	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

[a] * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ (Bonferroni-adjusted for multiple comparisons)

[b] StreamStats quantities obtained in English units were converted to metric.

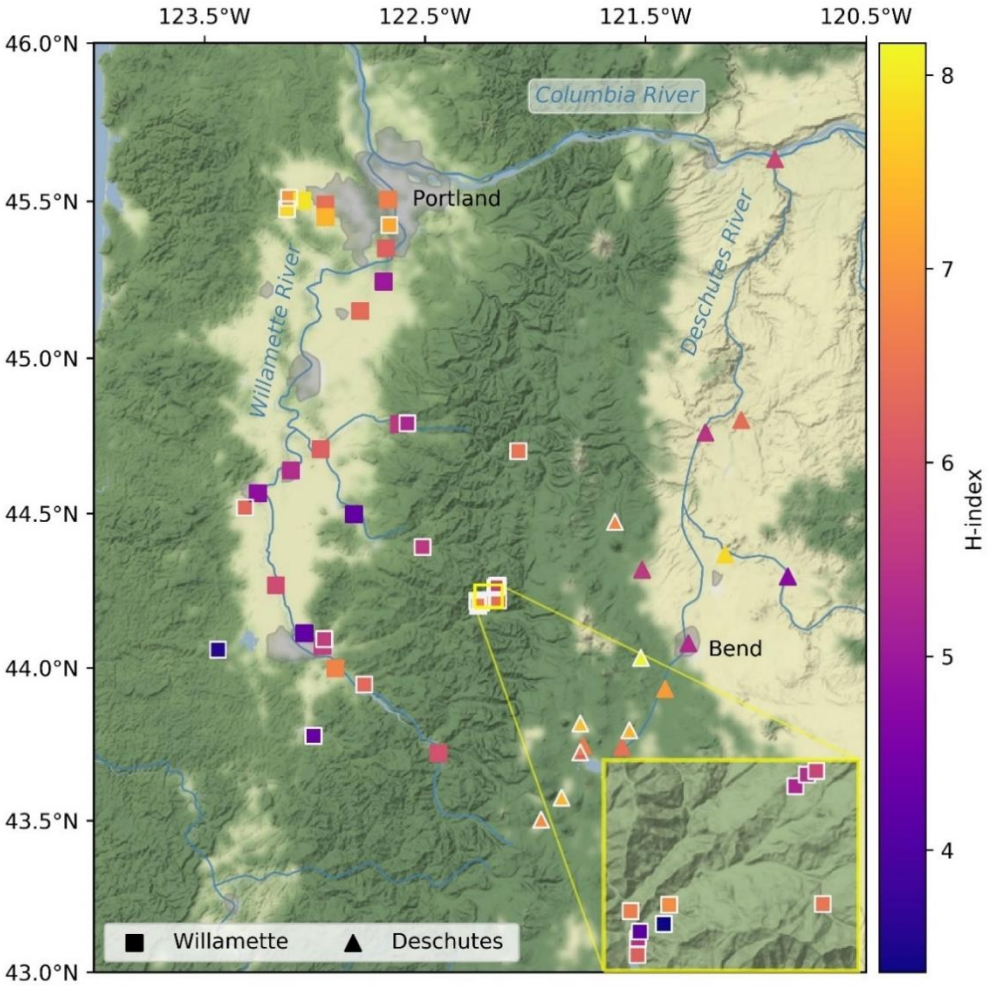

Figure 2.1. Alpha diversity (Shannon's index [*H*]) of streamwater microbiomes in the Willamette (squares) and Deschutes (triangles) watersheds in Oregon, USA. Outlined symbols indicate small sub-catchments (i.e., those with less than median drainage area). Inset shows vicinity of H.J. Andrews Experimental Forest.

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 H.J. 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 isopropanol precipitation (Crump et al., 2003; Zhou et al., 1996; 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 GTGCCAGCMGCCGCGGTAA, 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 sec, 50 °C for 60 sec, and 72 °C for 90 sec, 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 V.2 paired end 250bp sequencing. Sequences were denoised using DADA2 (Callahan et al., 2016) implemented in QIIME2 (Bolyen 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 samples resulted in undersampling of some communities, which is not unusual in microbial studies, but allowed for retention of the most samples for the analysis (Fig 2.S1). 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 PRJNA642636 at the National Center for Biotechnology Information (NCBI).

2.4.1 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 ASV within a site. Shannon's index accounts for both the number of ASV (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

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 ASV 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 ASV contribute less to differences between sites. Bray-Curtis dissimilarity ranges [0, 1], with a value of one indicating two sites have no ASV in common and a value of zero indicating identical ASV 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 sites; 703 unique pairs), the Deschutes Basin (n = 17 sites; 136 unique pairs), small sub-catchments (n = 28 sites; 378 unique pairs), and large sub-catchments (n = 27 sites; 351 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 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.

2.4.2 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 for 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 (Read et al., 2015; Fagervold et al., 2014; Repert et al., 2014). 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 (Z) 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 (σ) 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 ε indicate greater sensitivity and a stronger relationship between microbial community similarity and watershed characteristic variability. Values of ε 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 versus Deschutes watersheds and small versus 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, 2020*b*).

2.5 Results

2.5.1 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; Fig. 2.2). Some samples also featured high abundances of Cyanobacteria (up to 28% of community).

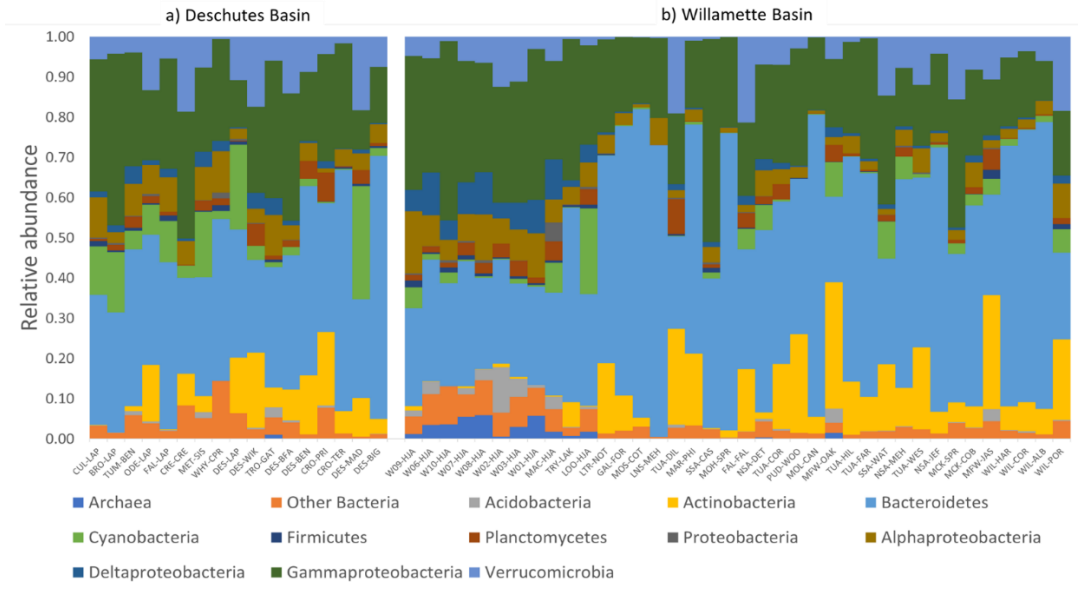

Figure 2.2. Phylogenetic biodiversity (relative abundance of unique amplified sequences variants [ASVs]) of tributary and main-stem streamwater microbial DNA samples throughout the a) Deschutes and b) Willamette watersheds in Oregon, USA. Samples for each watershed are presented in order of increasing sub-catchment drainage area (Table 2.S1).

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; Fig. 2.1) and was not related to the sample volume filtered (R^2 = 0.04, p = 0.76, Fig. 2.S2). 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, $p > 0.1$; Fig. 2.3). 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; Fig. 2.S3).

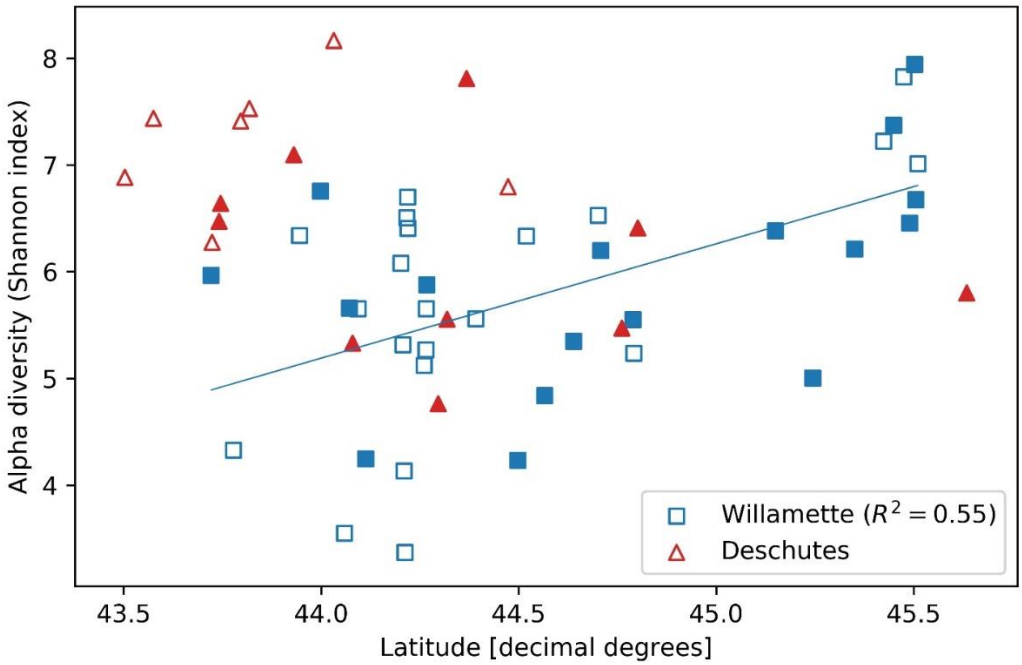

Figure 2.3. Alpha diversity (Shannon index) vs. latitudinal coordinate of streamwater microbial DNA samples collected from small (unfilled symbols) and large (filled symbols) sub-catchments throughout the Willamette (squares) and Deschutes (triangles) watersheds in Oregon, USA.

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 (Fig. 2.4). 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).

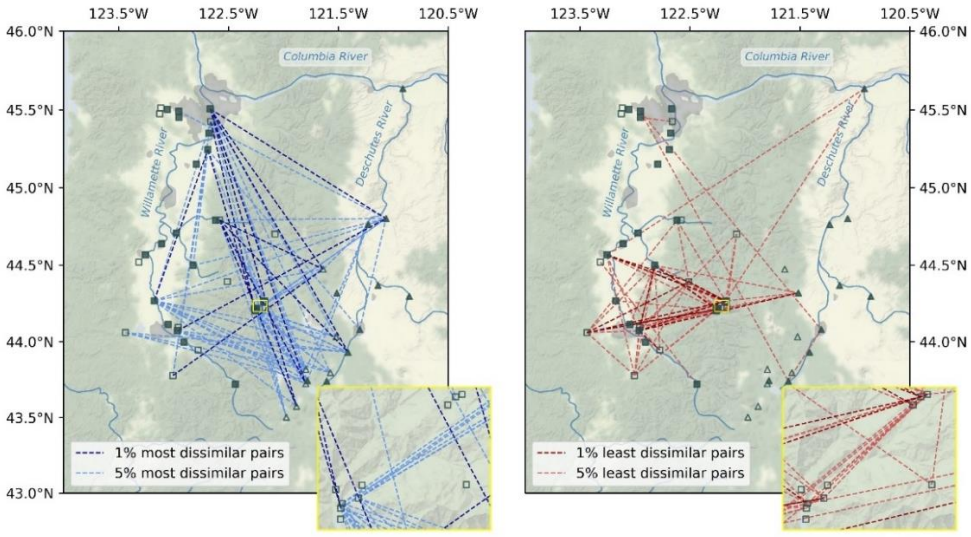

Figure 2.4. Map of the 1% (dark lines) and 5% (light lines) most (left) and least (right) dissimilar microbial communities throughout the Willamette and Deschutes watersheds in Oregon, USA. Large (filled symbols) and small (unfilled symbols) sub-catchments are those with more than or less than median drainage area, respectively. Inset shows vicinity of H.J. Andrews Experimental Forest.

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 (Fig. 2.4). Points in the H.J. Andrews Forest exhibited some of the lowest dissimilarity to other points within the H.J. 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).

2.5.2 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 2.1), geomorphic-related characteristics were on average the most strongly correlated with microbial community composition in the Willamette Basin (mean r = 0.19; Fig 2.5). 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 2.1). In the Deschutes Basin, climatic (mean r = 0.11) characteristics were more strongly correlated with microbial community composition than were geomorphic (mean r = 0.10) characteristics. Land-cover characteristics were very weakly anticorrelated (mean r = -0.01; Fig. 2.5). 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 2.1, Table 2.S2). No development-related characteristics were found to be statistically 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 2.1), 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; Fig. 2.5) 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 2.1). 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 significantly correlated with microbial community similarity in small or large watersheds.

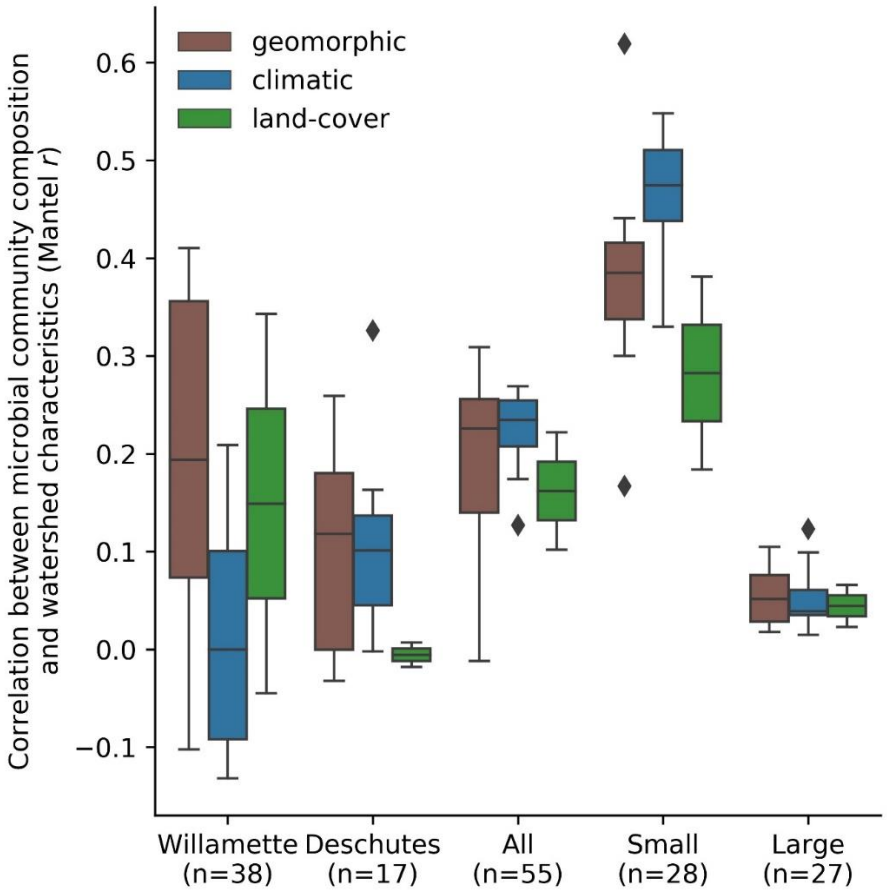

Figure 2.5. Mean correlation between microbial community composition (Mantel test statistic [r]) for land-cover, geomorphic, and climatic related StreamStats basin characteristics by watershed and in small and large sub-catchments across the Willamette and Deschutes watersheds, Oregon, USA.

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 $\varepsilon = 0.11$; Wilcoxon signed-rank $W = 324.0$, $p = 0.678$) or for small versus large sub-catchments (median $\varepsilon = -0.01$; Wilcoxon signed-rank $W = 357.0$, $p = 0.476$). Similarly, median sensitivity was not statistically different from zero for any of the characteristic groups for either the Willamette versus Deschutes watersheds or for small versus large sub-catchments (all Wilcoxon signed-rank $p > 0.1$).

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 (Fig 2.S4). 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 (Fig 2.S4). 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 (Fig 2.S5).

2.6 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 analysed 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 (Savio et al. 2015; Read 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 a likely 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 Table 2.S1), 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 (Fig. 2.S4). 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.

2.7 Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

2.8 Author Contributions

SG conceived of the study, produced the original code, and reviewed the manuscript. DU collected data, contributed to 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 research.

2.9 Funding

This work was supported the National Science Foundation grant EAR 1836768. The first author would like to acknowledge STEM Scholarship support from NSF grant #1153490.

2.10 Acknowledgments

We conducted part of this research at HJ Andrews Experimental Forest, which is 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 T. Scheibe and two anonymous reviewers for their helpful, constructive reviews, which improved this manuscript.

2.11 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).

2.12 References

Amaral-Zettler, L. A., McCliment, E. A., Ducklow, H. W., and Huse, S. M. (2009). A Method for Studying Protistan Diversity Using Massively Parallel Sequencing of V9 Hypervariable Regions of Small-Subunit Ribosomal RNA Genes. *PLoS One* 4, e6372. doi:10.1371/journal.pone.0006372.

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., et al. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* 37, 852–857. doi:10.1038/s41587-019-0209-9.

Bray, J. R., and Curtis, J. T. (1957). An Ordination of the Upland Forest Communities of Southern Wisconsin. *Ecol. Monogr.* 27, 325–349. doi:10.2307/1942268.

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., and Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* 13, 581–583. doi:10.1038/nmeth.3869.

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Huntley, J., Fierer, N., et al. (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* 6, 1621–1624. doi:10.1038/ismej.2012.8.

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Lozupone, C. A., Turnbaugh, P. J., et al. (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc. Natl. Acad. Sci. U. S. A.* 108, 4516–4522. doi:10.1073/pnas.1000080107.

Carter, R. W. (1961). Magnitude and frequency of floods in suburban areas. US Geological Survey Professional Paper, 424, 9-11.

Crump, B. C., Adams, H. E., Hobbie, J. E., and Kling, G. W. (2007). BIOGEOGRAPHY OF BACTERIOPLANKTON IN LAKES AND STREAMS OF AN ARCTIC TUNDRA CATCHMENT. *Ecology* 88, 1365–1378. doi:10.1890/06-0387.

Crump, B. C., Amaral-Zettler, L. A., and Kling, G. W. (2012). Microbial diversity in arctic freshwaters is structured by inoculation of microbes from soils. *ISME J.* 6, 1629–1639. doi:10.1038/ismej.2012.9.

Crump, B. C., Kling, G. W., Bahr, M., and Hobbie, J. E. (2003). Bacterioplankton community shifts in an Arctic lake correlate with seasonal changes in organic matter source. *Appl. Environ. Microbiol.* 69, 2253–2268. doi:10.1128/AEM.69.4.2253-2268.2003.

Delpla, I., Jung, A. V., Baures, E., Clement, M., and Thomas, O. (2009). Impacts of climate change on surface water quality in relation to drinking water production. *Environ. Int.* 35, 1225–1233. doi:10.1016/j.envint.2009.07.001.

Doherty, M., Yager, P. L., Moran, M. A., Coles, V. J., Fortunato, C. S., Krusche, A. V, et al. (2017). Bacterial Biogeography across the Amazon River-Ocean Continuum. *Front. Microbiol.* 8, 882. doi:10.3389/fmicb.2017.00882.

Droppo, I. G., Liss, S. N., Williams, D., Nelson, T., Jaskot, C., and Trapp, B. (2009). Dynamic existence of waterborne pathogens within river sediment compartments. Implications for water quality regulatory affairs. *Environ. Sci. Technol.* 43, 1737–1743. doi:10.1021/es802321w.

Emilson, C. E., Thompson, D. G., Venier, L. A., Porter, T. M., Swystun, T., Chartrand, D., et al. (2017). DNA metabarcoding and morphological macroinvertebrate metrics reveal the same changes in boreal watersheds across an environmental gradient. *Sci. Rep.* 7, 1–11. doi:10.1038/s41598-017-13157-x.

Fagervold, S. K., Bourgeois, S., Pruski, A. M., Charles, F., Kerhervé, P., Vétion, G., et al. (2014). River organic matter shapes microbial communities in the sediment of the Rhône prodelta. *ISME J.* 8, 2327–2338. doi:10.1038/ismej.2014.86.

Feddema, J. J., Oleson, K. W., Bonan, G. B., Mearns, L. O., Buja, L. E., Meehl, G. A., et al. (2005). The Importance of Land-Cover Change in Simulating Future Climates. *Science (80-.).* 310.

Foley, J. A., DeFries, R., Asner, G. P., Barford, C., Bonan, G., Carpenter, S. R., et al. (2005). Global consequences of land use. *Science (80-.).* 309, 570–574. doi:10.1126/science.1111772.

Fortunato, C. S., Eiler, A., Herfort, L., Needoba, J. A., Peterson, T. D., and Crump, B. C. (2013). Determining indicator taxa across spatial and seasonal gradients in the Columbia River coastal margin. *ISME J.* 7, 1899–1911. doi:10.1038/ismej.2013.79.

Good, S. P., URycki, D. R., and Crump, B. C. (2018). Predicting Hydrologic Function With Aquatic Gene Fragments. *Water Resour. Res.* 54, 2424–2435. doi:10.1002/2017WR021974.

Goodrich, J. K., Di Rienzi, S. C., Poole, A. C., Koren, O., Walters, W. A., Caporaso, J. G., et al. (2014). Conducting a microbiome study. *Cell* 158, 250–262. doi:10.1016/j.cell.2014.06.037.

Gregory, J. H., Dukes, M. D., Jones, P. H., and Miller, G. L. (2006). Effect of urban soil compaction on infiltration rate. *J. Soil Water Conserv.* 61.

Hermans, S. M., Buckley, H. L., Case, B. S., and Lear, G. (2019). Connecting through space and time: catchment-scale distributions of bacteria in soil, stream

water and sediment. *Environ. Microbiol.*, 1462-2920.14792. doi:10.1111/1462-2920.14792.

IPCC, 2019: Climate Change and Land: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems [P.R. Shukla, J. Skea, E. Calvo Buendia, V. Masson-Delmotte, H.-O. Pörtner, D. C. Roberts, P. Zhai, R. Slade, S. Connors, R. van Diemen, M. Ferrat, E. Haughey, S. Luz, S. Neogi, M. Pathak, J. Petzold, J. Portugal Pereira, P. Vyas, E. Huntley, K. Kissick, M. Belkacemi, J. Malley, (eds.)]. In press.

Jerde, C. L., Olds, B. P., Shogren, A. J., Andruszkiewicz, E. A., Mahon, A. R., Bolster, D., et al. (2016). Influence of Stream Bottom Substrate on Retention and Transport of Vertebrate Environmental DNA. *Environ. Sci. Technol.* 50, 8770–8779. doi:10.1021/acs.est.6b01761.

Jiménez Cisneros, B.E. et al. 2014: Freshwater resources. In: Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Field, C.B. et al. (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 229-269.

Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K., and Schloss, P. D. (2013). Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the miseq illumina sequencing platform. *Appl. Environ. Microbiol.* 79, 5112–5120. doi:10.1128/AEM.01043-13.

Legendre, P., Fortin, M., and Borcard, D. (2015). Should the Mantel test be used in spatial analysis? *Methods Ecol. Evol.* 6, 1239–1247. doi:10.1111/2041-210X.12425.

Leibold, M. A., Holyoak, M., Mouquet, N., Amarasekare, P., Chase, J. M., Hoopes, M. F., et al. (2004). The metacommunity concept: a framework for multi-scale community ecology. *Ecol. Lett.* 7, 601–613. doi:10.1111/j.1461-0248.2004.00608.x.

Li, P., Yang, S. F., Lv, B. B., Zhao, K., Lin, M. F., Zhou, S., et al. (2015). Comparison of extraction methods of total microbial DNA from freshwater. *Genet. Mol. Res.* 14, 730–738. doi:10.4238/2015.January.30.16.

Lozupone, C. A., and Knight, R. (2008). Species divergence and the measurement of microbial diversity. *FEMS Microbiol. Rev.* 32, 557–578. doi:10.1111/j.1574-6976.2008.00111.x.

Lozupone, C., and Knight, R. (2005). UniFrac: A new phylogenetic method for comparing microbial communities. *Appl. Environ. Microbiol.* 71, 8228–8235. doi:10.1128/AEM.71.12.8228-8235.2005.

Mächler, E., Walser, J.-C., Zurich, E., Schaefli, B., Altermatt, F., Salyani, A., et al. (2019). Water tracing with environmental DNA in a high-Alpine catchment. doi:10.5194/hess-2019-551.

Mantel, N. (1967). The Detection of Disease Clustering and a Generalized Regression Approach. *Cancer Res.* 27, 209–220.

Moscrip, A. L., and Montgomery, D. R. (1997). URBANIZATION, FLOOD FREQUENCY, AND SALMON ABUNDANCE IN PUGET LOWLAND STREAMS. *J. Am. Water Resour. Assoc.* 33, 1289–1297. doi:10.1111/j.1752-1688.1997.tb03553.x.

Newby, D. T., Pepper, I. L., and Maier, R. M. (2009). "Microbial Transport," in *Environmental Microbiology* (Elsevier Inc.), 365–383. doi:10.1016/B978-0-12-370519-8.00019-5.

Nippgen, F., Mcglynn, B. L., and Emanuel, R. E. (2015). The spatial and temporal evolution of contributing areas. *Water Resour. Res.* doi:10.1002/2014WR016719.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2013). The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Res.* 41, D590. doi:10.1093/nar/gks1219.

Read, D. S., Gweon, H. S., Bowes, M. J., Newbold, L. K., Field, D., Bailey, M. J., et al. (2015). Catchment-scale biogeography of riverine bacterioplankton. *ISME J.* 9, 516–26. doi:10.1038/ismej.2014.166.

Repert, D. A., Underwood, J. C., Smith, R. L., and Song, B. (2014). Nitrogen cycling processes and microbial community composition in bed sediments in the Yukon River at Pilot Station. *J. Geophys. Res. Biogeosciences* 119, 2328–2344. doi:10.1002/2014JG002707.

Ricotta, C., and Podani, J. (2017). On some properties of the Bray-Curtis dissimilarity and their ecological meaning. *Ecol. Complex.* 31, 201–205. doi:10.1016/j.ecocom.2017.07.003.

Ries, K.G., III, Guthrie, J.D., Rea, A.H., Steeves, P.A., Stewart, D.W., 2008, StreamStats: A water resources web application: U.S. Geological Survey Fact Sheet 2008-3067, 6 p.

Ruiz-González, C., Niño-García, J. P., and del Giorgio, P. A. (2015). Terrestrial origin of bacterial communities in complex boreal freshwater networks. *Ecol. Lett.* 18, 1198–1206. doi:10.1111/ele.12499.

Sales, N. G., McKenzie, M. B., Drake, J., Harper, L. R., Browett, S. S., Coscia, I., et al. (2020). Fishing for mammals: Landscape-level monitoring of terrestrial and semi-aquatic communities using eDNA from riverine systems. *J. Appl. Ecol.* 57, 707–716. doi:10.1111/1365-2664.13592.

Savio, D., Sinclair, L., Ijaz, U. Z., Parajka, J., Reischer, G. H., Stadler, P., et al. (2015). Bacterial diversity along a 2600?km river continuum. *Environ. Microbiol.* 17, 4994–5007. doi:10.1111/1462-2920.12886.

Seibert, J., and McDonnell, J. J. (2015). Gauging the Ungauged Basin: Relative Value of Soft and Hard Data. *J. Hydrol. Eng.* 20, A4014004. doi:10.1061/(ASCE)HE.1943-5584.0000861.

Shogren, A. J., Tank, J. L., Andruszkiewicz, E., Olds, B., Mahon, A. R., Jerde, C. L., et al. (2017). Controls on eDNA movement in streams: Transport, Retention, and Resuspension /704/158/2464 /704/242 /45/77 article. *Sci. Rep.* 7, 1–11. doi:10.1038/s41598-017-05223-1.

Sorensen, J. P. R., Maurice, L., Edwards, F. K., Lapworth, D. J., Read, D. S., Allen, D., et al. (2013). Using Boreholes as Windows into Groundwater Ecosystems. *PLoS One* 8, e70264. doi:10.1371/journal.pone.0070264.

Sugiyama, A., Masuda, S., Nagaosa, K., Tsujimura, M., and Kato, K. (2018). Tracking the direct impact of rainfall on groundwater at Mt. Fuji by multiple analyses including microbial DNA. *Biogeosciences* 15, 721–732. doi:10.5194/bg-15-721-2018.

URycki, D. R., and Good, S.P. (2020*a*) StreamStats Data Query. https://doi.org/10.5281/zenodo.3902476

URycki, D.R., and Good, S. P. (2020*b*) River Microbiome and Watershed Characteristics Analysis. https://doi.org/10.5281/zenodo.3902478

U.S. Geological Survey, 2016, National Water Information System data available on the World Wide Web (USGS Water Data for the Nation), accessed 9 May 2020. http://dx.doi.org/10.5066/F7P55KJN

Wheater, H., & Evans, E. (2009). Land use, water management and future flood risk. Land use policy, 26, S251-S264.

Whittaker, R. H. (1972). EVOLUTION AND MEASUREMENT OF SPECIES DIVERSITY. *Taxon* 21, 213–251. doi:10.2307/1218190.

Zhou, J., Bruns, M. A., and Tiedje, J. M. (1996). DNA recovery from soils of diverse composition. *Appl. Environ. Microbiol.* 62, 316–22. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8593035 [Accessed April 29, 2017].

2.13 Supplementary Material

Table 2.S1. Watershed basin characteristics derived from StreamStats (https://streamstats.usgs.gov/ss/; Ries et al., 2017) for sample locations in the Willamette and Deschutes watersheds, Oregon, USA.

Basin Characteristic	Study Site											
	LTR-NOT	MAR-PHI	WIL-COR	WIL-HAR	TUA-DIL	GAL-FOR	WIL-ALB	TUA-COR	MCK-COB	MOS-COT	NSA-JEF	MCK-SPR
LONGITUDE (decimal degrees)	-123.438	-123.316	-123.257	-123.175	-123.125	-123.116	-123.107	-123.056	-123.047	-123.005	-122.973	-122.964
LATITUDE (decimal degrees)	44.05909	44.51943	44.56549	44.2677	45.475	45.51067	44.63873	45.50206	44.11254	43.77813	44.7079	44.07167
ASPECT (degrees)	184	177	183	185	167	165	183	165	190	183	189	192
BSLOPD (degrees)	12.1	11.8	13.7	15.7	11.3	13.6	13.2	11.7	15.4	20	16.9	16
DRNAREA (square km)	116.03	396.27	11447.75	8883.66	323.75	193.47	12587.34	562.03	3444.68	246.83	1895.87	2952.59
DRNDENSITY (dimensionless)	0.54	0.77	0.64	0.62	0.69	0.67	0.64	0.7	0.59	0.7	0.72	0.58
ELEV (m)	250.85	289.86	731.52	886.97	310.9	310.9	688.85	294.44	963.17	603.5	920.5	1048.51
ELEVMAX (m)	640.08	1246.63	3139.44	3139.44	1057.66	957.07	3139.44	1057.66	3139.44	1444.75	3200.4	3139.44
FOREST (percent)	80.3	82.2	75.6	84.4	73.5	81.6	72.3	72.4	86.4	92.5	82.4	87.3
I24H2Y (mm)	67.06	74.17	67.31	69.09	62.99	65.53	66.29	62.99	75.44	64.26	80.77	77.72
IMPERV (percent)	0.29	0.65	1.2	0.86	1	0.79	1.32	1.19	0.37	0.0489	0.58	0.23
JANAVPRE2K (mm)	284.48	337.82	240.28	240.79	244.09	287.02	236.98	254	264.16	246.89	322.58	271.78
JANMAXT2K (degrees C)	8.17	7.78	6.94	6.61	6.22	5.89	7	6.22	5.89	8.39	5.17	5.5
JANMAXTMP (degrees C)	7.89	7.67	6.39	5.94	5.83	5.72	6.5	5.89	5.44	7.22	4.72	5.06
JANMINT2K (degrees C)	1.22	0.5	-0.56	-0.94	-0.17	-0.39	-0.44	-0.22	-1.89	0.83	-1.89	-2.44
JANMINTMP (degrees C)	0.5	0.33	-1.33	-1.89	-0.11	-0.39	-1.17	-0.17	-2.56	-0.22	-2.89	-3
JULAVPRE2K (mm)	9.4	13.97	18.8	21.59	17.27	16.26	18.54	16.51	25.4	12.7	26.92	26.42
LC11BARE (percent)	2	0	1	2	1	1	1	1	3	0	1	4
LC11CRPHAY (percent)	3	8	12	4	16	10	15	17	2	1	7	1
LC11DEVHI (percent)	0	0	0	0	0	0	0	0	0	0	0	0
LC11DVLO (percent)	1	1	1	1	1	1	2	2	0	0	0	0
LC11DVMD (percent)	0	0	1	0	0	0	1	0	0	0	0	0

LC11DVOPN (percent)	6	8	2	1	3	3	2	3	1	0	1	1
LC11FORSHB (percent)	77	75	76	87	66	79	73	68	89	92	84	90
LC11HERB (percent)	7	5	4	4	10	5	4	8	4	5	5	3
LC11IMP (percent)	1.03	1.4	1.34	0.8	0.82	0.77	1.49	1.09	0.26	0.0661	0.33	0.18
LC11WATER (percent)	0	0	1	1	1	0	1	1	0	0	1	0
LC11WETLND (percent)	3	2	1	1	1	1	1	1	0	0	1	0
MAJ_ROADS (km)	26.39	44.9	2269.07	1348.57	54.55	31.7	2671.39	111.36	251.05	19.63	204.38	167.36
MAXBSLOPD (degrees)	45.5	47.6	70	70	47.6	48.2	70	48.2	70	50.8	66	70
MAXTEMP (degrees C)	17.11	16.22	15.33	14.89	14.94	14.78	15.5	15	14.33	16.11	13.39	13.94
MINBELEV (m)	128.93	68.28	60.96	89.92	49.68	52.12	54.25	43.28	121.31	207.87	75.59	137.46
MINBSLOPD (degrees)	0	0	0	0	0	0	0	0	0	0	0	0
MINTEMP (degrees C)	5.39	5.06	3.61	3.11	4.83	4.56	3.78	4.78	2.56	4.61	2.44	2.06
MIN_ROADS (km)	70	535.89	12021.24	9253.3	423.24	218.86	13389.12	703.25	2751.85	302.54	1754.1	2156.42
ORREG2 (dimensionless)	10001	10001	10001	10001	10001	10001	10001	10001	10001	10001	10001	10001
OR_HIPERMA (percent)	16.2	13.7	17	7.57	12.9	8.02	19.9	14	5.51	1.21	11.3	4.47
OR_HIPERMG (percent)	0	0	23.7	30.5	0	0	21.6	0	42.2	0	37	49.3
PRECIP (mm)	1480.82	1831.34	1620.52	1694.18	1617.98	1653.54	1597.66	1597.66	1945.64	1503.68	2108.2	2009.14
RELIEF (m)	512.06	1176.53	3078.48	3048	1008.89	905.26	3078.48	1014.98	3023.62	1237.49	3108.96	3008.38
SOILPERM (mm per hour)	18.54	36.07	48.51	54.1	23.11	19.3	47.24	21.34	70.1	22.35	90.42	76.45
STATE_HWY (km)	16.09	60.35	836.82	494.05	11.2	32.19	1012.23	51.5	191.5	0	108.3	185.07
STATSGODEP (mm)	1150.62	1107.44	1163.32	1145.54	1290.32	1292.86	1176.02	1300.48	1130.3	1206.5	1165.86	1117.6
STRMTOT (km)	62.28	302.54	7273.9	5551.98	223.69	129.71	8014.16	395.88	2027.68	172.19	1358.22	1721.92
WATCAPORC (mm)	3.81	3.81	3.3	3.3	4.06	4.06	3.56	4.06	3.3	3.3	3.3	3.3
WATCAPORR (mm per mm)	0.15	0.14	0.14	0.13	0.16	0.16	0.14	0.16	0.13	0.13	0.13	0.13
TI_index ()	6.293847	7.547943	10.75706	10.36106	7.390311	6.684248	10.89054	7.906157	9.43395	6.519383	8.738742	9.239599
ALL ROADS (km)	112.48	641.14	15127.13	11095.92	488.99	282.75	17072.74	866.11	3194.4	322.17	2066.78	2508.85

Table 2.S1. (CONTINUED)

Basin Characteristic	Study Site											
	TUA-HIL	MFW-JAS	SSA-WAT	PUD-WOO	FAL-FAL	TUA-WES	WIL-POR	NSA-MEH	MFW-OAK	W09-HJA	W01-HJA	W02-HJA
LONGITUDE (decimal degrees)	-122.952	-122.906	-122.821	-122.793	-122.775	-122.677	-122.667	-122.618	-122.438	-122.258	-122.258	-122.245
LATITUDE (decimal degrees)	45.48999	43.998	44.49765	45.15036	43.94499	45.35047	45.505	44.78884	43.721	44.20148	44.20723	44.21244
ASPECT (degrees)	174	182	181	192	185	177	183	185	183	264	197	215
BSLOPD (degrees)	9.12	17	18.2	5.52	18.9	7.74	11.7	18.4	19.1	30	27.9	26.7
DRNAREA (square km)	1225.06	3496.48	1644.64	821.03	481.74	1833.71	29007.87	1696.44	1017.87	0.03	1.01	0.62
DRNDENSITY (dimensionless)	0.72	0.63	0.65	0.77	0.61	0.77	0.7	0.7	0.65	0	0	0
ELEV (m)	236.83	1008.89	755.9	265.48	704.09	194.77	557.78	1002.79	1203.96	612.65	722.38	804.67
ELEVMAX (m)	1057.66	2654.81	1767.84	1338.07	1517.9	1057.66	3200.4	3200.4	2654.81	713.23	1018.03	1078.99
FOREST (percent)	58.4	89.1	86.2	30.7	93.5	47.4	65.3	88.8	87.8	100	99.8	100
I24H2Y (mm)	57.66	66.55	74.42	58.42	65.53	53.85	66.04	83.31	68.33	64.26	67.31	76.45
IMPERV (percent)	2.27	0.23	0.58	3.44	0.0688	6.95	2.36	0.36	0	0	0	0
JANAVPRE2K (mm)	233.17	225.55	274.32	209.8	232.66	211.33	246.13	335.28	224.28	250.95	250.44	256.54
JANMAXT2K (degrees C)	6.5	6.5	6.5	7.61	7.83	6.83	6.78	4.83	5.67	6.72	6.67	6.5
JANMAXTMP (degrees C)	6.17	5.72	6.17	7.39	7.39	6.61	6.44	4.39	5	5.67	5.89	6.67
JANMINT2K (degrees C)	-0.22	-1	-0.72	0.22	0.94	0	-0.39	-2.11	-2.39	-0.78	-0.83	-1
JANMINTMP (degrees C)	-0.06	-2.28	-1.83	0.22	-0.78	0.22	-0.94	-3.28	-3.11	-1.67	-1.83	-1.33
JULAVPRE2K (mm)	15.24	22.86	29.21	20.32	20.07	15.49	19.81	27.69	23.88	19.05	18.54	18.03
LC11BARE (percent)	1	1	0	0	0	0	1	1	1	0	0	0
LC11CRPHAY (percent)	27	1	2	52	0	27	21	1	0	0	0	0
LC11DEVHI (percent)	0	0	0	1	0	2	1	0	0	0	0	0
LC11DVLO (percent)	3	0	0	4	0	9	3	0	0	0	0	0
LC11DVMD (percent)	1	0	0	2	0	6	1	0	0	0	0	0
LC11DVOPN (percent)	3	1	1	3	0	5	3	1	0	0	0	0
LC11FORSHB (percent)	57	92	89	33	96	44	65	91	95	100	100	100

LC11HERB (percent)	5	3	6	4	2	4	4	5	4	0	0	0
LC11IMP (percent)	2.57	0.16	0.3	3.26	0.11	9.14	2.56	0.18	0.0607	0	0	0
LC11WATER (percent)	0	2	1	0	1	0	1	1	1	0	0	0
LC11WETLND (percent)	2	0	0	2	0	2	2	0	0	0	0	0
MAJ_ROADS (km)	416.8	518.18	199.55	284.84	68.55	1208.56	8078.53	141.29	154.01	0	0	0
MAXBSLOPD (degrees)	48.2	67.2	66.4	57.4	53.9	48.2	70	66	64.5	34.8	42.2	44.2
MAXTEMP (degrees C)	15.44	14.56	14.94	16.06	16.33	15.89	15.39	13	14.06	16.5	15.94	15.83
MINBELEV (m)	34.44	159.72	119.18	31.7	207.87	16.58	1.46	185.62	371.86	481.58	463.3	560.83
MINBSLOPD (degrees)	0	0	0	0	0	0	0	0	0	17.4	1.15	2.34
MINTEMP (degrees C)	5	2.78	3.56	5.22	4.39	5.28	4.11	2.11	1.89	2.67	2.94	3.72
MIN_ROADS (km)	1599.61	3701.32	1593.18	1131.32	579.34	4393.31	38944.32	1503.06	1110.4	0	0	0
ORREG2 (dimensionless)	10001	10001	10001	10001	10001	10001	10001	10001	10001	10001	10001	10001
OR_HIPERMA (percent)	29.2	3.06	3.81	41.6	0	38.8	26.2	6.2	1.46	0	0	0
OR_HIPERMG (percent)	0	35.9	9.29	6.72	0	0	20.1	41	53.6	0	0	0
PRECIP (mm)	1450.34	1584.96	2082.8	1445.26	1691.64	1336.04	1638.3	2179.32	1582.42	2199.64	2250.44	2303.78
RELIEF (m)	1024.13	2496.31	1648.97	1304.54	1310.64	1042.42	3169.92	2999.23	2282.95	232.87	557.78	518.16
SOILPERM (mm per hour)	19.81	48.77	44.45	27.94	29.21	19.56	46.23	94.49	61.21	17.27	35.05	35.05
STATE_HWY (km)	165.75	95.27	73.06	124.56	0	394.27	2928.87	88.99	0	0	0	0
STATSGODEP (mm)	1371.6	1130.3	1132.84	1297.94	1163.32	1412.24	1211.58	1150.62	1120.14	1247.14	1201.42	1201.42
STRMTOT (km)	885.1	2204.7	1073.38	629.22	296.11	1414.55	20276.79	1187.64	661.41	0	0	0
WATCAPORC (mm)	4.06	3.3	3.3	3.81	3.3	4.06	3.56	3.3	3.05	3.3	2.79	2.79
WATCAPORR (mm per mm)	0.16	0.13	0.14	0.15	0.13	0.16	0.14	0.13	0.13	0.13	0.11	0.11
TI_index ()	8.940007	9.344563	8.517633	9.047308	7.249249	9.509813	11.84992	8.536934	7.985972	-2.95725	0.645824	0.209233
ALL ROADS (km)	2182.16	4314.77	1865.79	1540.72	647.89	5996.14	49951.72	1733.34	1264.41	0	0	0
H_index ()	4.297152	3.41858	3.101057	3.314169	2.760258	2.768154	2.914974	3.761218	3.068951	3.356252	2.607017	1.936592

Table 2.S1. (CONTINUED)

Basin Characteristic	Study Site											
	W03-HJA	W06-HJA	W07-HJA	W08-HJA	MAC-HJA	NSA-DET	CRE-CRE	ODE-LAP	CUL-LAP	DES-LAP	MET-SIS	DES-WIK
LONGITUDE (decimal degrees)	-122.243	-122.181	-122.175	-122.171	-122.167	-122.076	-121.973	-121.881	-121.796	-121.782	-121.639	-121.607
LATITUDE (decimal degrees)	44.21915	44.26072	44.26487	44.26616	44.21944	44.701	43.5029	43.57535	43.81817	43.74484	44.4731	43.7411
ASPECT (degrees)	210	148	171	183	180	197	149	147	172	178	134	161
BSLOPD (degrees)	25.5	13.9	17.3	14.6	24.8	14.9	6.63	8.02	4.87	6.66	7.7	6.15
DRNAREA (square km)	0.98	0.1	0.14	0.24	5.78	520.59	147.63	123.8	39.37	670.81	242.42	1261.32
DRNDENSITY (dimensionless)	0	0	0	0	0.6	0.76	0.4	0.56	0.36	0.4	0.42	0.37
ELEV (m)	783.34	941.83	1014.98	1078.99	1203.96	1280.16	1731.26	1682.5	1551.43	1636.78	1274.06	1581.91
ELEVMAX (m)	1078.99	996.7	1094.23	1179.58	1621.54	3200.4	2648.71	2563.37	1914.14	3108.96	2368.3	3108.96
FOREST (percent)	100	98.1	48.1	98.5	98.5	86.7	82.4	82.2	97.6	83.7	90.9	81.3
I24H2Y (mm)	84.07	87.88	87.88	88.14	86.87	84.84	74.93	62.74	71.37	70.1	45.72	59.94
IMPERV (percent)	0	0	0	0	0	0	0	0.0275	0	0	0.032	0.0027
JANAVPRE2K (mm)	256.54	292.1	292.1	292.1	287.02	297.18	214.88	193.8	226.57	226.57	188.98	190.5
JANMAXT2K (degrees C)	6.5	5	5	5	5.33	3.56	3.39	3.11	3.11	3	3.89	3.28
JANMAXTMP (degrees C)	6.56	6.61	6.61	7.17	5.78	3.06	1.39	1.22	2.06	1.61	3	1.94
JANMINT2K (degrees C)	-1	-2.11	-2.11	-2.11	-2.17	-3.72	-7.17	-5.83	-5.72	-6.06	-5.72	-6.33
JANMINTMP (degrees C)	-1.83	-1.5	-1.44	-2.22	-2.67	-4.61	-7.33	-7.11	-6.83	-7.11	-5.61	-7.28
JULAVPRE2K (mm)	18.03	27.94	27.94	28.19	34.04	31.5	35.31	24.89	29.46	29.21	19.56	24.89
LC11BARE (percent)	0	0	0	0	0	3	2	1	1	6	3	4
LC11CRPHAY (percent)	0	0	0	0	0	0	0	0	0	0	0	0
LC11DEVHI (percent)	0	0	0	0	0	0	0	0	0	0	0	0
LC11DVLO (percent)	0	0	0	0	0	0	0	0	0	0	0	0
LC11DVMD (percent)	0	0	0	0	0	0	0	0	0	0	0	0
LC11DVOPN (percent)	0	0	0	0	0	1	0	1	0	1	2	1
LC11FORSHB (percent)	100	100	100	100	100	86	85	86	97	86	84	83

LC11HERB (percent)	0	0	0	0	0	9	0	0	1	2	10	4
LC11IMP (percent)	0.041	0	0	0	0.0147	0.0496	0.0326	0.12	0.0696	0.0946	0.23	0.18
LC11WATER (percent)	0	0	0	0	0	1	12	12	0	4	1	6
LC11WETLND (percent)	0	0	0	0	0	0	0	0	0	1	0	2
MAJ_ROADS (km)	0	0	0	0	0	23.17	18.18	3.4	3.06	70.81	29.45	145
MAXBSLOPD (degrees)	51.8	24.4	29.5	23.4	48.4	65.9	65.8	55.6	28	61.3	71.5	61.3
MAXTEMP (degrees C)	15.89	16.28	16.17	15.67	13.89	11.94	10.56	10.44	11.39	10.89	12.78	11.5
MINBELEV (m)	481.58	868.68	935.74	972.31	762	518.16	1475.23	1356.36	1356.36	1322.83	886.97	1295.4
MINBSLOPD (degrees)	0.65	1.6	1.23	4.11	0.93	0	0	0	0	0	0	0
MINTEMP (degrees C)	3.22	3.56	3.56	2.94	2.06	0.61	-2.17	-1.83	-1.22	-1.5	-0.17	-1.5
MIN_ROADS (km)	0	0.04	0.47	0.24	7.43	318.64	37.01	71.93	21.4	574.51	342.77	1581.91
ORREG2 (dimensionless)	10001	10001	10001	10001	10001	10001	363	363	363	363	363	363
OR_HIPERMA (percent)	0	0	0	0	0	7.47	0	0	0	0	9.88	3.89
OR_HIPERMG (percent)	0	0	91.7	100	89.6	89.8	100	100	100	100	98.1	99.2
PRECIP (mm)	2334.26	2125.98	2123.44	2171.7	2291.08	2197.1	1526.54	1501.14	1389.38	1518.92	1353.82	1287.78
RELIEF (m)	600.46	129.84	158.19	204.52	859.54	2667	1173.48	1207.01	557.78	1795.27	1481.33	1825.75
SOILPERM (mm per hour)	38.86	71.12	73.66	73.66	76.71	152.4	91.19	122.43	126.49	165.61	205.23	182.37
STATE_HWY (km)	0	0	0	0	0	34.92	0	10.86	0	0	18.02	10.86
STATSGODEP (mm)	1181.1	1018.54	1008.38	1008.38	1041.4	1209.04	1089.66	1122.68	1242.06	1193.8	1338.58	1219.2
STRMTOT (km)	0	0	0	0	3.46	397.49	58.42	69.68	14.23	265.53	102.03	471.52
WATCAPORC (mm)	2.79	4.06	4.06	4.06	3.81	3.05	2.15	2.51	2.79	3.05	2.31	3.3
WATCAPORR (mm per mm)	0.11	0.16	0.16	0.16	0.15	0.12	0.0935	0.11	0.11	0.13	0.0991	0.13
TI_index ()	0.720087	-0.90614	-0.79965	-0.08189	2.526454	7.578923	7.146854	6.778395	6.135725	8.656075	7.491633	9.367838
ALL ROADS (km)	0	0.04	0.47	0.24	7.43	376.73	55.19	86.19	24.46	645.32	390.24	1737.77

Table 2.S1. (CONTINUED)

Basin Characteristic	Study Site									Description (category[a])
	FAL-LAP	TUM-BEN	DES-BFA	DES-BEN	DES-MAD	CRO-TER	TRO-GAT	DES-BIG	CRO-PRI	
LONGITUDE (decimal degrees)	-121.573	-121.521	-121.412	-121.306	-121.229	-121.139	-121.066	-120.914	-120.855	Longitudinal coordinate (G)
LATITUDE (decimal degrees)	43.79595	44.03193	43.93051	44.0792	44.7609	44.36792	44.80123	45.63391	44.29618	Latitudinal coordinate (G)
ASPECT (degrees)	147	144	177	176	173	179	190	173	180	basin average of topographic slope compass directions from elevation grid (G)
BSLOPD (degrees)	3.7	9.66	4.96	4.9	5.92	6.08	9.81	6.84	6.26	Mean basin slope measured in degrees (G)
DRNAREA (square km)	124.32	77.44	4506.58	4817.38	20901.2	11654.95	1724.93	27712.87	7174.27	Area that drains to a point on a stream (G)
DRNDENSITY (dimensionless)	0.18	0.55	0.33	0.33	0.48	0.54	0.76	0.56	0.64	Basin drainage density defined as total stream length divided by drainage area. (G)
ELEV (m)	1392.94	1911.1	1524	1517.9	1341.12	1359.41	1002.79	1225.3	1392.94	Mean Basin Elevation (G)
ELEVMAX (m)	1895.86	2535.94	3108.96	3108.96	3200.4	2331.72	1807.46	3413.76	2176.27	Maximum basin elevation (G)
FOREST (percent)	89.6	69.6	81.5	81.2	47.1	29.4	23.3	43.5	24.5	Percentage of area covered by forest (L)
I24H2Y (mm)	37.59	60.45	45.72	44.96	30.73	23.88	24.38	30.48	23.88	Maximum 24-hour precipitation that occurs on average once in 2 years - Equivalent to precipitation intensity index (C)
IMPERV (percent)	0	0	0.41	0.58	0.45	0.15	0.0729	0.36	0.0417	Percentage of impervious area
JANAVPRE2K (mm)	108.46	185.42	132.08	129.03	73.66	42.93	40.64	75.18	44.2	Mean January Precipitation (C)
JANMAXT2K (degrees C)	3.5	1.78	2.89	5.22	7.22	5.22	6.22	4.72	5.28	Mean Maximum January Temperature from 2K resolution PRISM 1961-1990 data (C)
JANMAXTMP (degrees C)	3.33	0.06	2.61	2.67	3.28	3.22	4.11	3.33	2.89	Mean Maximum January Temperature (C)
JANMINT2K (degrees C)	-6.28	-6.61	-6.94	-6.89	-6.89	-7.44	-5.28	-6.28	-7.72	Mean Minimum January Temperature from 2K resolution PRISM PRISM 1961-1990 data (C)
JANMINTMP (degrees C)	-7.44	-7.72	-7.44	-7.39	-6.78	-7.11	-4.78	-6.22	-7.44	Mean Minimum January Temperature (C)
JULAVPRE2K (mm)	17.78	33.53	19.3	19.3	15.49	14.22	11.18	14.48	14.99	Mean July Average Precipitation (C)
LC11BARE (percent)	0	12	2	3	1	0	0	1	0	Percentage of barren from NLCD 2011 class 31 (L)
LC11CRPHAY (percent)	0	0	0	0	3	2	4	3	1	Percentage of cultivated crops and hay, classes 81 and 82, from NLCD 2011 (L)
LC11DEVHI (percent)	0	0	0	0	0	0	0	0	0	Percentage of area developed, high intensity, NLCD 2011 class 24 (D)
LC11DVLO (percent)	0	0	0	1	1	0	0	1	0	Percentage of developed area, low intensity, from NLCD 2011 class 22 (D)
LC11DVMD (percent)	0	0	0	0	0	0	0	0	0	Percentage of area developed, medium intensity, NLCD 2011 class 23 (D)
LC11DVOPN (percent)	1	0	1	2	1	1	1	1	0	Percentage of developed open area from NLCD 2011 class 21 (D)

LC11FORSHB (percent)	98	86	87	86	89	94	94	89	96	Percentage of forests and shrub lands, classes 41 to 52, from NLCD 2011 (L)
LC11HERB (percent)	1	2	5	5	3	1	2	3	0	Percentage of herbaceous from NLCD 2011 classes 71-74 (L)
LC11IMP (percent)	0.26	0.0165	0.46	0.6	0.6	0.33	0.3	0.51	0.19	Average percentage of impervious area determined from NLCD 2011 impervious dataset (D)
LC11WATER (percent)	0	0	2	2	1	0	0	1	0	Percent of open water, class 11, from NLCD 2011 (L)
LC11WETLND (percent)	0	0	2	1	1	1	0	1	1	Percentage of wetlands, classes 90 and 95, from NLCD 2011 (L)
MAJ_ROADS (km)	25.43	3.91	455.42	537.5	2542.65	1194.08	197.94	3299	577.73	Length of non-state major roads in basin (D)
MAXBSLOPD (degrees)	40	56.8	65.8	65.8	74.1	58.8	68.4	74.1	57.2	Maximum basin slope, in degrees, using ArcInfo Grid with NHDPlus 30-m resolution elevation data. (G)
MAXTEMP (degrees C)	13.61	9.22	12.78	12.89	14.17	14.5	15.56	14.33	14.33	Mean annual maximum air temperature over basin area from PRISM 1971-2000 800-m grid (C)
MINBELEV (m)	1283.21	1438.66	1264.92	1072.9	411.48	813.82	429.77	50.6	868.68	Minimum basin elevation (G)
MINBSLOPD (degrees)	0	0	0	0	0	0	0	0	0	Minimum basin slope, in degrees, using ArcInfo Grid with NHDPlus 30-m resolution elevation data. (G)
MINTEMP (degrees C)	-1.22	-2.28	-1.56	-1.5	-0.56	-0.56	1.78	0.11	-0.67	Mean annual minimum air temperature over basin surface area as defined in SIR 2008-5126 (C)
MIN_ROADS (km)	271.97	51.5	7032.51	7692.31	23978.11	9880.91	1298.68	29771.48	5342.77	Length of non-state minor roads in basin (D)
ORREG2 (dimensionless)	363	363	363	363	363	363	363	363	363	Oregon Region Number (G)
OR_HIPERMA (percent)	22.1	0	17.1	16	9.68	8.46	2.47	7.51	6.68	Percent basin surface area containing high permeability aquifer units as defined in SIR 2008-5126 (G)
OR_HIPERMG (percent)	95.4	100	65.4	67.4	36.3	20.5	0	35	10.6	Percent basin surface area containing high permeability geologic units as defined in SIR 2008-5126 (G)
PRECIP (mm)	698.5	1264.92	916.94	896.62	533.4	350.52	342.9	530.86	365.76	Mean Annual Precipitation (C)
RELIEF (m)	609.6	1097.28	1856.23	2045.21	2773.68	1517.9	1377.7	3352.8	1307.59	Maximum - minimum elevation (G)
SOILPERM (mm per hour)	299.72	146.05	264.16	261.62	121.41	71.88	13.97	100.58	34.8	Average Soil Permeability (L)
STATE_HWY (km)	0	0	160.44	183.46	872.22	411.97	84.81	1221.44	202.77	Length of state highways in basin (D)
STATSGODEP (mm)	1524	1054.1	1397	1379.22	1038.86	924.56	1244.6	1026.16	838.2	Area-weighted average soil depth from NRCS STATSGO database (G)
STRMTOT (km)	22.37	42.32	1499.84	1604.44	10090.12	6324.43	1311.55	15690.38	4570.33	total length of all mapped streams (1:24,000-scale) in the basin (G)
WATCAPORC (mm)	4.83	1.79	4.06	4.06	3.05	2.79	3.3	3.05	2.54	Available water capacity from STATSGO data using methods from SIR 2005-5116 (G)
WATCAPORR (mm per mm)	0.19	0.0782	0.17	0.16	0.12	0.11	0.13	0.12	0.1	Available water capacity from STATSGO data using methods from SIR 2008-5126 (G)
TI_index ()	7.561362	6.120198	10.85761	10.93654	12.21388	11.60295	9.207927	12.35033	11.08831	Topographic index (G)
ALL ROADS (km)	297.4	55.41	7648.37	8413.27	27392.98	11486.96	1581.43	34291.92	6123.27	Length of state highways and non-state major and minor roads in basin (D)

[a]Categories: C – climatic; D – development; G – geomorphic; L – land cover

Table 2.S2. Mean, standard deviation (SD), correlation with microbial community similarity (Mantel statistic [*r*]) and associated Bonferroni-adjusted *p*-value for all StreamStats macroscale basin characteristics by watershed and in small and large sub-catchments across the Willamette and Deschutes watersheds, Oregon, USA.

	Willamette				**Deschutes**			
Basin characteristic	Mean	SD	*r*	*p*	Mean	SD	*r*	*p*
MINBELEV (m)	294.123	304.842	0.035	1	1021.813	435.058	0.277	0.73
JANMINTMP (degrees C)	-1.552	1.296	0.083	1	-6.906	0.797	-0.015	1
MINTEMP (degrees C)	3.468	1.243	0.074	1	-0.99	1.035	0.059	1
JANMINT2K (degrees C)	-1.029	1.168	0.05	1	-6.477	0.697	0.085	1
ORREG2 (dimensionless)	10001	0			363	0		
JANMAXTMP (degrees C)	6.083	1.059	0.067	1	2.448	1.056	-0.001	1
ELEV (m)	741.759	337.075	0.089	1	1475.029	226.315	0.027	1
JANMAXT2K (degrees C)	6.218	1.062	0.043	1	4.122	1.466	0.211	1
MAXTEMP (degrees C)	15.152	1.201	-0.015	1	12.597	1.849	0.097	1
PRECIP (mm)	1884.116	320.239	0.016	1	965.2	463.794	0.159	1
DISTANCE (km)	80.89	49.559	0.202	0.298	75.228	57.446	0.219	1
BSLOPD (degrees)	16.496	6.135	-0.074	1	6.544	1.71	-0.157	1
LATITUDE (decimal degrees)	44.532	0.54	0.201	0.61	44.17	0.57	0.186	1
OR_HIPERMG (percent)	26.604	33.251	0.042	1	68.527	37.817	0.227	1
JANAVPRE2K (mm)	263.858	34.971	0.083	1	138.193	70.054	0.158	1
LONGITUDE (decimal degrees)	-122.69	0.411	0.048	1	-121.446	0.351	0.242	1
TI_index ()	6.138	4.341	-0.063	1	9.168	2.22	0.325	0.389
ELEVMAX (m)	1858.942	979.223	-0.042	1	2619.451	531.879	0.19	1
MAJ_ROADS (km)	591.609	1605.174	-0.021	1	606.904	1005.218	0.306	0.413
DRNDENSITY (dimensionless)	0.497	0.305	-0.031	1	0.459	0.145	-0.033	1
LC11FORSHB (percent)	83.593	18.232	0.138	1	89.333	5.066	0.009	1
SOILPERM (mm per hour)	51.956	30.824	-0.004	1	147.167	83.689	0.002	1

DRNAREA (square km)	2700.613	5985.725	-0.023	1	5411.953	8457.813	0.326	0.278
ALL ROADS (km)	3997.38	9943.83	-0.024	1	6682.012	10561.01	0.319	0.235
MIN_ROADS (km)	3182.217	7765.083	-0.025	1	5863.318	9200.694	0.322	0.269
LC11WATER (percent)	0.481	0.58	0.011	1	2.733	4.131	-0.087	1
LC11CRPHAY (percent)	7.556	12.373	0.185	1	0.867	1.407	0.176	1
STRMTOT (km)	1803.267	4108.677	-0.023	1	2809.146	4608.79	0.304	0.442
LC11DVLO (percent)	1	1.961	0.041	1	0.2	0.414	0.348	0.418
OR_HIPERMA (percent)	9.754	12.043	0.121	1	6.918	7.092	0.08	1
LC11IMP (percent)	0.957	1.868	0.012	1	0.266	0.197	0.272	0.912
LC11HERB (percent)	3.556	2.926	0.066	1	2.6	2.64	0.201	1
STATE_HWY (km)	223.555	582.069	-0.02	1	211.79	364.285	0.288	0.638
LC11DVOPN (percent)	1.63	2.06	0.071	1	0.867	0.64	0.093	1
JULAVPRE2K (mm)	21.994	5.973	0.006	1	21.573	7.539	0.094	1
LC11WETLND (percent)	0.667	0.92	0.066	1	0.667	0.724	0.246	1
LC11BARE (percent)	0.852	1.099	-0.007	1	2.4	3.158	-0.127	1
WATCAPORC (mm)	3.517	0.432	-0.003	1	3.039	0.8	-0.116	1
RELIEF (m)	1562.495	1111.023	-0.026	1	1598.574	737.501	0.257	0.797
WATCAPORR (mm per mm)	0.139	0.016	-0.012	1	0.123	0.03	-0.114	1
FOREST (percent)	81.411	18.096	0.048	1	67.187	26.003	0.165	1
I24H2Y (mm)	72.531	10.225	0.027	1	47.125	18.357	0.174	1
ASPECT (degrees)	186.63	20.311	-0.212	1	165.333	16.855	-0.016	1
MINBSLOPD (degrees)	1.089	3.396	-0.118	1	0	0		
STATSGODEP (mm)	1173.762	100.949	0.083	1	1175.512	187.592	0.111	1
MAXBSLOPD (degrees)	53.381	14.488	0	1	60.3	12.439	-0.072	1
LC11DEVHI (percent)	0.148	0.456	-0.015	1	0	0		
LC11DVMD (percent)	0.407	1.217	0.001	1	0	0		

Table 2.S2. (Continued)

	Small				Large				All			
Basin characteristic	Mean	SD	*r*	*p*	Mean	SD	*r*	*p*	Mean	SD	*r*	*p*
MINBELEV (m)	745.802	495.714	0.411	0.005	362.222	430.851	-0.004	1	546.337	501.766	0.219	0.01
JANMINTMP (degrees C)	-3.261	2.954	0.398	0.014	-3.667	2.764	-0.006	1	-3.367	2.796	0.156	0.082
MINTEMP (degrees C)	1.873	2.679	0.377	0.029	1.878	2.272	-0.017	1	1.938	2.45	0.161	0.067
JANMINT2K (degrees C)	-2.712	2.776	0.356	0.053	-3.237	2.928	0	1	-2.859	2.763	0.146	0.091
ORREG2 (dimensionless)	6788.333	4655.596	0.344	0.058	6329.381	4795.997	0.027	1	6709.976	4627.121	0.168	0.072
JANMAXTMP (degrees C)	4.827	2.464	0.325	0.062	4.743	1.592	-0.032	1	4.831	2.052	0.18	0.024
ELEV (m)	1022.532	516.09	0.31	0.077	984.751	419.168	-0.01	1	994.146	466.389	0.134	0.13
JANMAXT2K (degrees C)	5.156	1.848	0.3	0.125	5.783	1.213	-0.047	1	5.474	1.595	0.195	0.014
MAXTEMP (degrees C)	14.124	2.459	0.292	0.106	14.355	1.158	-0.074	1	14.237	1.926	0.141	0.221
PRECIP (mm)	1784.29	444.05	0.271	0.134	1327.573	619.941	0.068	1	1584.96	556.086	0.071	1
DISTANCE (km)	80.665	58.836	0.26	0.043	115.581	61.142	0.046	1	104.244	57.873	0.14	0.13
BSLOPD (degrees)	14.697	8.057	0.257	0.134	11.186	5.228	0.067	1	13.104	6.944	0.047	1
LATITUDE (decimal degrees)	44.238	0.512	0.233	0.149	44.567	0.592	0.019	1	44.405	0.578	0.121	0.485
OR_HIPERMG (percent)	50.695	49.626	0.229	0.504	32.458	25.464	0.035	1	42.332	40.228	0.136	0.125
JANAVPRE2K (mm)	246.319	50.688	0.195	0.758	191.637	92.195	0.048	1	223.241	74.413	0.07	1
LONGITUDE (decimal degrees)	-122.269	0.561	0.171	1	-122.222	0.858	0.014	1	-122.279	0.69	0.081	1
TI_index ()	4.439	3.794	0.151	1	10.001	1.318	0.032	1	7.126	3.975	-0.017	1
ELEVMAX (m)	1644.614	793.192	0.147	1	2616.49	773.555	0.099	1	2129.437	928.256	-0.006	1
MAJ_ROADS (km)	23.008	30.906	0.125	1	1171.135	1834.118	0.03	1	597.543	1424.251	0.014	1
DRNDENSITY (dimensionless)	0.359	0.294	0.124	1	0.607	0.134	0.039	1	0.479	0.26	-0.009	1
LC11FORSHB (percent)	90	11.243	0.119	1	81.286	17.312	0.033	1	85.39	15.174	0.054	1
SOILPERM (mm per hour)	88.513	73.573	0.093	1	83.408	71.287	0.01	1	87.208	72.022	0.029	1
DRNAREA (square km)	182.624	219.409	0.092	1	7155.273	8638.501	-0.025	1	3583.453	7055.737	0.005	1
ALL ROADS (km)	223.606	282.492	0.086	1	9688.748	12766.33	-0.001	1	4927.711	10247.52	0.007	1

MIN_ROADS (km)	190.935	241.722	0.085	1	8088.571	10406.93	-0.009	1	4110.411	8397.442	0.005	1
LC11WATER (percent)	1.571	3.586	0.082	1	1	1.342	-0.063	1	1.317	2.715	0.072	1
LC11CRPHAY (percent)	2.095	5.127	0.081	1	8.238	13.262	0.056	1	5.268	10.514	0.082	1
STRMTOT (km)	107.43	140.921	0.077	1	4217.589	5330.402	-0.013	1	2103.782	4301.461	0.007	1
LC11DVLO (percent)	0.238	0.539	0.056	1	1.19	2.159	-0.08	1	0.732	1.644	0.005	1
OR_HIPERMA (percent)	4.583	7.118	0.054	1	12.899	11.86	0.03	1	8.791	10.663	0.066	1
LC11IMP (percent)	0.256	0.428	0.052	1	1.165	2.043	-0.088	1	0.723	1.546	-0.012	1
LC11HERB (percent)	2.714	3.73	0.048	1	3.714	1.419	0.123	1	3.293	2.822	0.027	1
STATE_HWY (km)	9.664	17.86	0.043	1	429.042	663.89	0.021	1	219.758	516.426	0.012	1
LC11DVOPN (percent)	1.238	2.166	0.039	1	1.476	1.167	-0.03	1	1.39	1.73	0.017	1
JULAVPRE2K (mm)	23.344	7.375	0.038	1	20.344	5.209	0.08	1	22.011	6.475	0.068	1
LC11WETLND (percent)	0.381	0.805	0.018	1	0.952	0.805	0.184	0.662	0.659	0.855	0.073	1
LC11BARE (percent)	1.524	2.839	0.011	1	1.286	1.271	0.024	1	1.439	2.191	0.028	1
WATCAPORC (mm)	3.294	0.787	0.006	1	3.398	0.42	0.024	1	3.366	0.62	0.021	1
RELIEF (m)	898.724	613.117	-0.013	1	2252.037	805.37	0.103	1	1581.912	995.584	0.027	1
WATCAPORR (mm per mm)	0.131	0.028	-0.018	1	0.135	0.017	0.069	1	0.134	0.023	0.015	1
FOREST (percent)	87.029	13.36	-0.027	1	65.633	24.037	0.107	1	77.595	20.724	0.041	1
I24H2Y (mm)	70.635	13.599	-0.029	1	56.279	19.745	0.058	1	64.422	17.361	0.023	1
ASPECT (degrees)	176.857	29.773	-0.032	1	181.19	7.587	0.007	1	179	21.838	-0.051	1
MINBSLOPD (degrees)	1.4	3.812	-0.041	1	0	0			0.717	2.787	-0.058	1
STATSGODEP (mm)	1175.899	126.099	-0.058	1	1172.875	148.247	0.067	1	1182.587	126.677	0.011	1
MAXBSLOPD (degrees)	46.971	13.649	-0.099	1	64.733	7.233	0.02	1	55.82	14.215	-0.016	1
LC11DEVHI (percent)	0	0			0.19	0.512	-0.049	1	0.098	0.374	-0.009	1
LC11DVMD (percent)	0	0			0.524	1.365	-0.107	1	0.268	1.001	-0.023	1

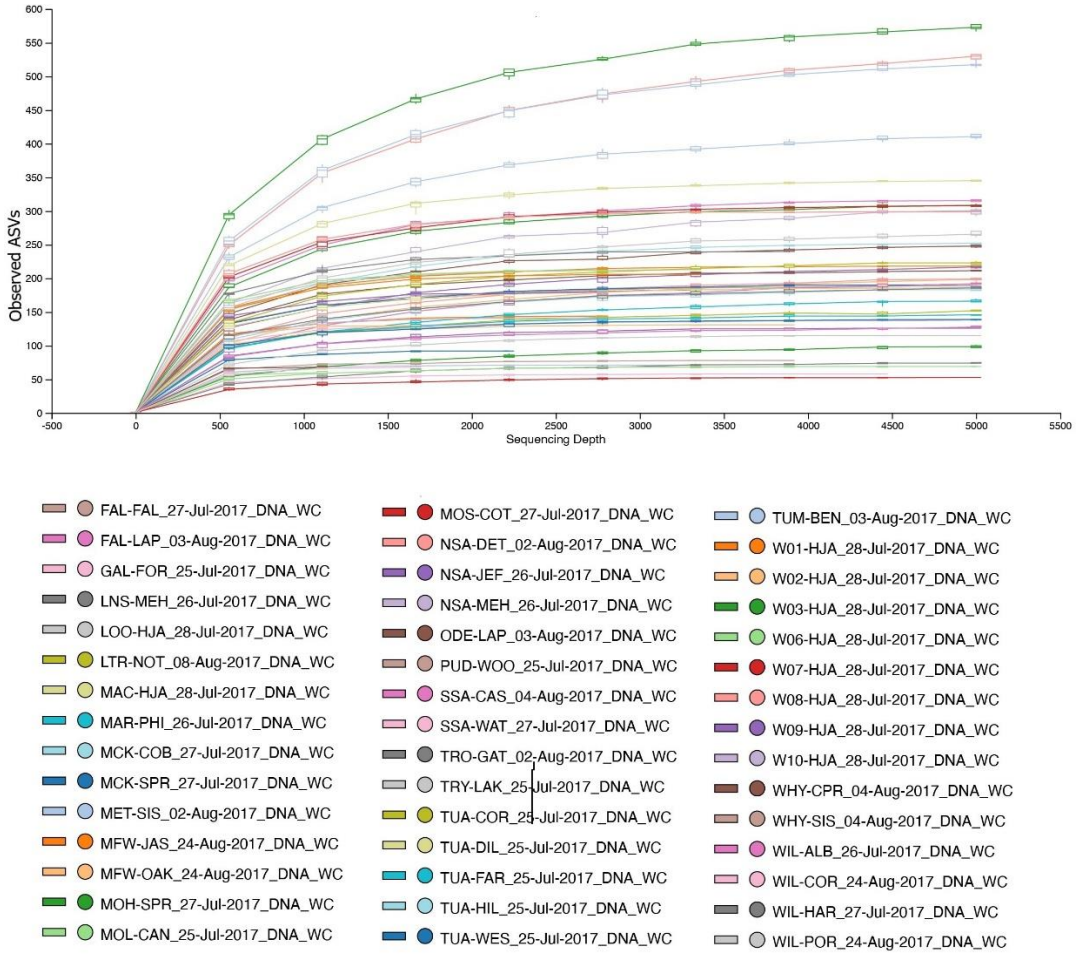

Figure 2.S1. Amplified sequence variant (ASV) rarefaction curves for each sample in the Willamette and Deschutes watersheds, Oregon, USA.

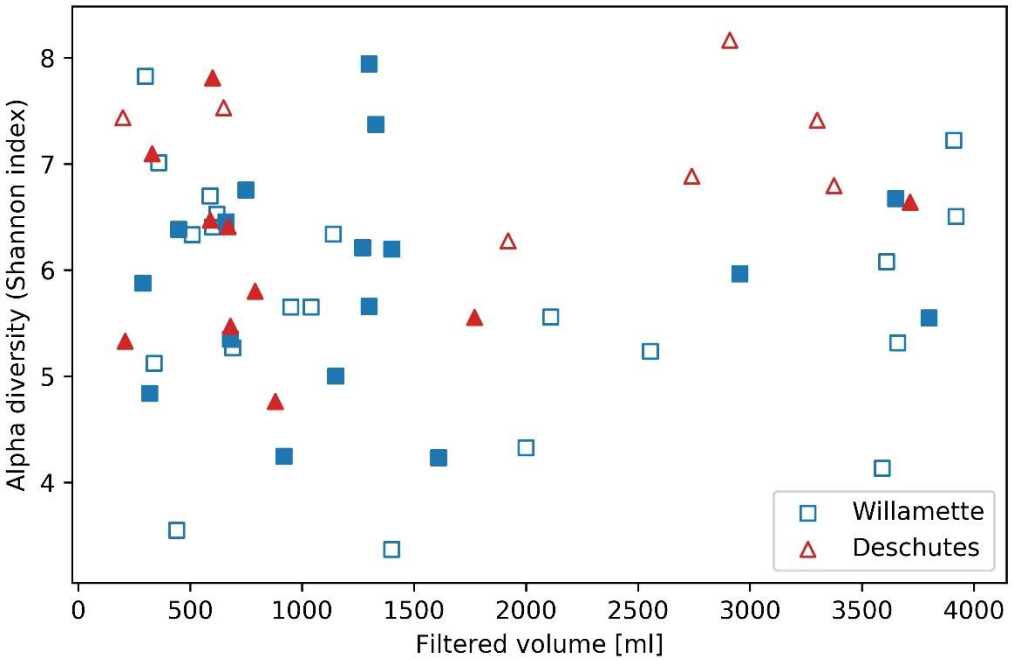

Figure 2.S2: Alpha diversity (Shannon index) of microbial DNA vs. filtered volume of streamwater collected from small (unfilled symbols) and large (filled symbols) sub-catchments throughout the Willamette (squares) and Deschutes (triangles) watersheds in Oregon, USA.

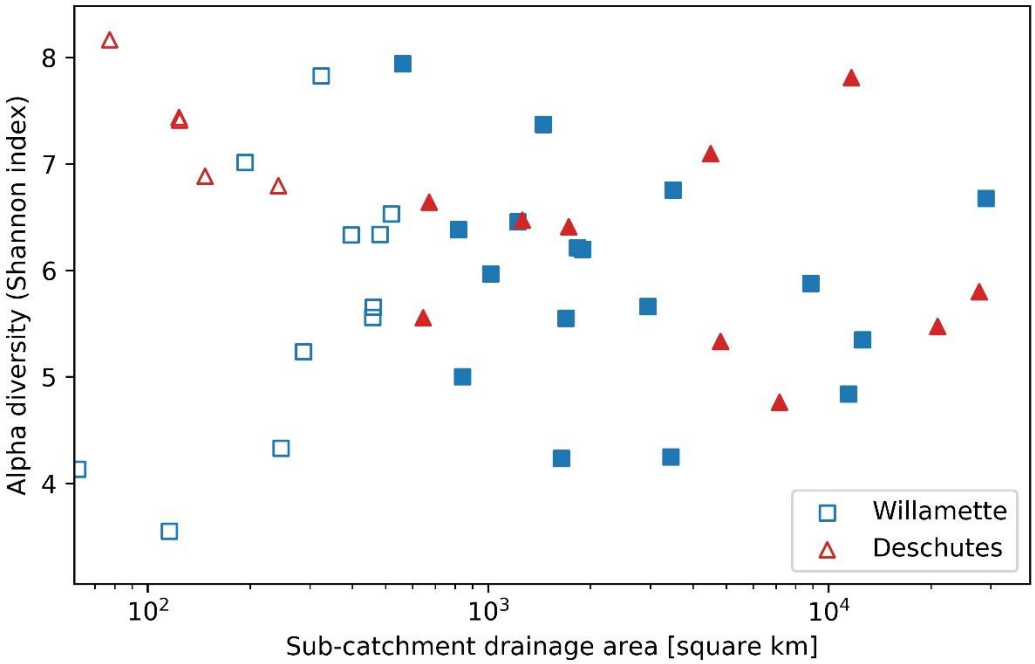

Figure 2.S3. Alpha diversity (Shannon index) of microbial DNA collected from streamwater vs. sub-catchment drainage area in small (unfilled symbols) and large (filled symbols) sub-catchments throughout the Willamette (squares) and Deschutes (triangles) watersheds in Oregon, USA.

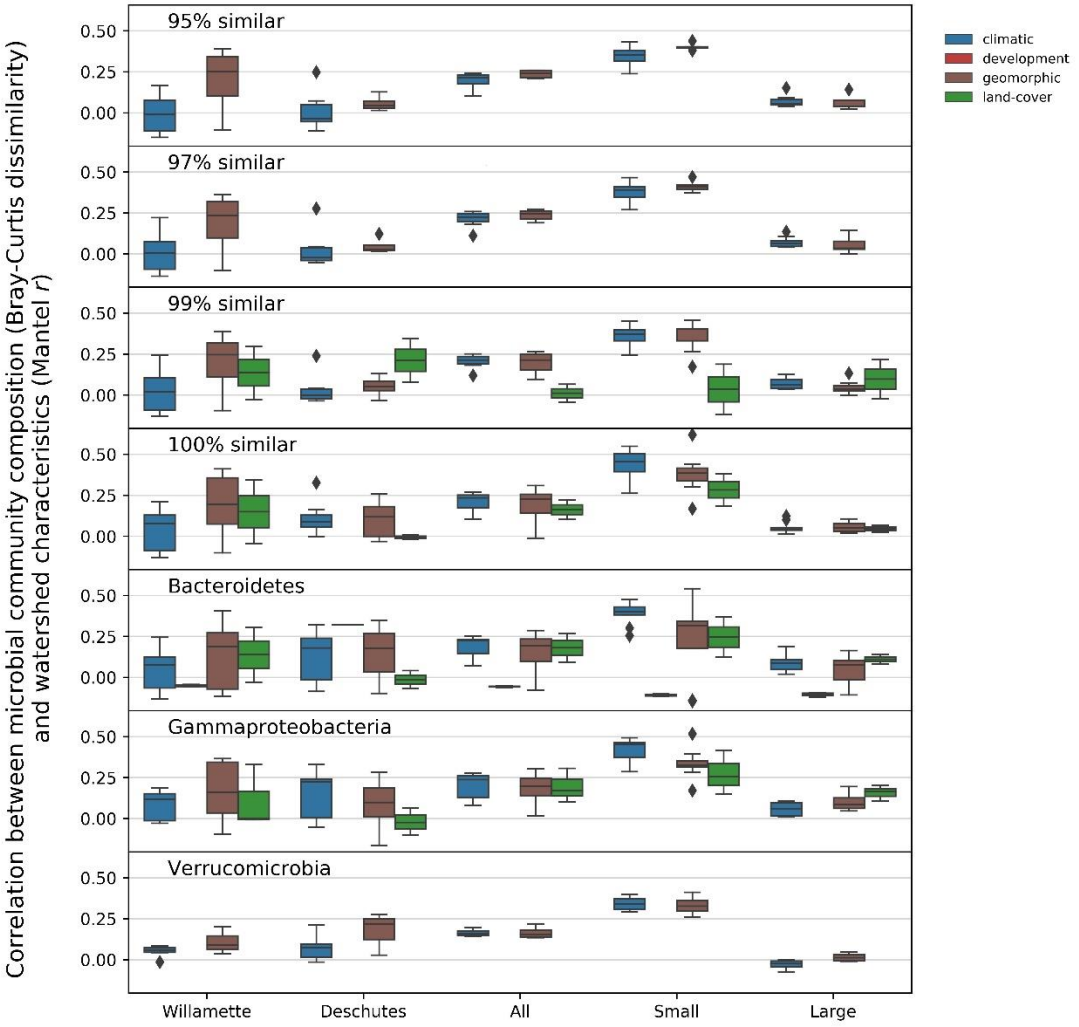

Figure 2.S4. Mean correlation between microbial community composition (Mantel test statistic [*r*]) for land-cover, geomorphic, climatic and development related StreamStats basin characteristics by watershed and in small and large sub-catchments across the Willamette and Deschutes watersheds, Oregon, USA. Microbial communities were characterized based on: 1) raw sequence data grouped into 95%, 97%, 99% operational taxonomic units (OTUs) or 100% similar amplified sequence variants (ASVs); or 2) one of three subsets of ASVs (classes Bacteroidetes, Gammaproteobacteria, Verrucomicrobia). Community similarity was based on Bray-Curtis dissimilarity metric.

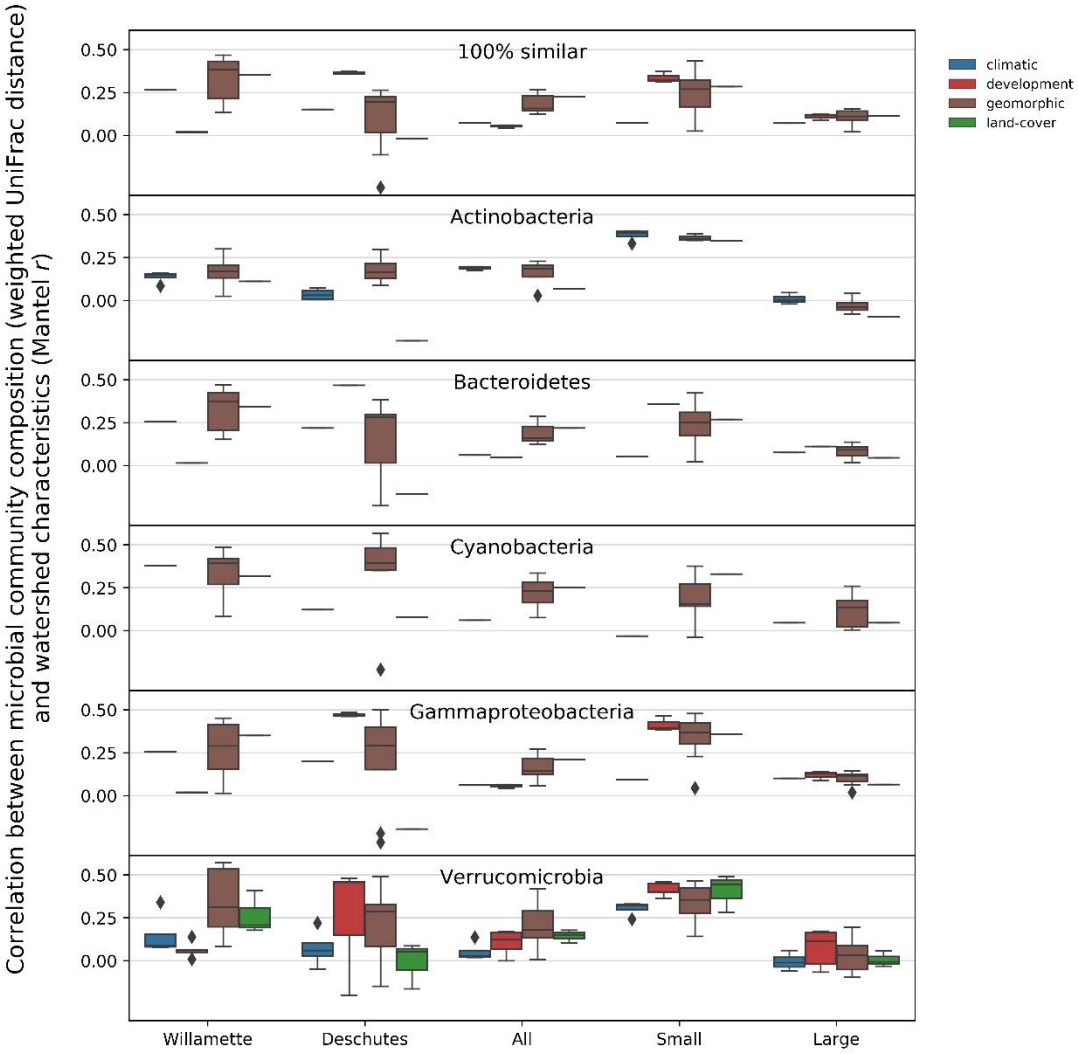

Figure 2.S5. Mean correlation between microbial community composition (Mantel test statistic [r]) for land-cover, geomorphic, climatic and development related StreamStats basin characteristics by watershed and in small and large sub-catchments across the Willamette and Deschutes watersheds, Oregon, USA. Microbial communities were characterized based on: 1) raw sequence data grouped into 95%, 97%, 99% operational taxonomic units (OTUs) or 100% similar amplified sequence variants (ASVs); or 2) one of five subsets of ASVs (classes Bacteroidetes, Gammaproteobacteria, Verrucomicrobia, Actinobacteria and Cyanobacteria). Community similarity was based on Weighted UniFrac distance metric.

THE STREAMWATER MICROBIOME ENCODES HYDROLOGIC DATA ACROSS SCALES

Dawn R URycki, Maoya Bassiouni, Stephen P Good, Byron C Crump, and Bonan Li

(In Review)

3 The Streamwater Microbiome Encodes Hydrologic Data across Scales

3.1 Abstract

Many fundamental questions in hydrology remain unanswered due to the limited information that can be extracted from existing data sources. Microbial communities constitute a novel type of environmental data, as they are comprised of many thousands of taxonomically and functionally diverse groups known to respond to both biotic and abiotic environmental factors. As such, these microscale communities reflect a range of macroscale conditions and characteristics, some of which also drive hydrologic regimes. Here, we assess the extent to which streamwater microbial communities (as characterized by 16S gene amplicon sequence abundance) encode information about catchment hydrology across scales. We analyzed 64 summer streamwater DNA samples collected from subcatchments within the Willamette, Deschutes, and John Day river basins in Oregon, USA, which range 0.03-29,000 km 2 in area and 343-2,334 mm/year of precipitation. We applied information theory to quantify the breadth and depth of information about common hydrologic metrics encoded within microbial taxa. Of the 256 microbial taxa that spanned all three watersheds, we found 9.6% (24.5/256) of taxa, on average, shared information with a given hydrologic metric, with a median 15.6% (range = 12.4 – 49.2%) reduction in uncertainty of that metric based on knowledge of the microbial biogeography. All of the hydrologic metrics we assessed, including daily discharge at different time lags, mean monthly discharge, and seasonal high and low flow durations were encoded within the microbial community. Summer microbial taxa shared the most information with winter mean flows. Our study demonstrates quantifiable relationships between streamwater microbial taxa and hydrologic metrics at different scales, likely resulting from the integration of multiple overlapping drivers of each. Streamwater microbial communities are rich sources of information that may contribute fresh insight to unresolved hydrologic questions.

3.2 Introduction

Hydrology spans earth and life sciences. The living environment is shaped by water and in turn shapes the hydrological cycle (e.g.,Dingman, 2015). A deeper knowledge of these interactions requires data that integrates processes at multiple

spatial and temporal scales, beyond our direct human observations. Meteorological, eddy covariance, stable isotopes, and remotely sensed earth observation data have made high-resolution hydrologic observations more widely available than ever before, through observation networks such as FLUXNET (Baldocchi et al., 2001), the National Ecological Observatory Network (NEON; Schimel et al., 2007), and the Critical Zone Collaborative Network (CZNet; https://criticalzone.org/). Despite this growing amount of data, the specific mechanisms and interactions of many drivers of hydrologic function are still not well understood. A fundamental process-based understanding of streamflow generation, for example, is necessary to predict water availability and adapt to changing climate and landcover conditions, yet the dynamics of streamflow sources and transit time distribution remain areas of active research (Blöschl et al., 2019). The ability to develop new hydrologic insight may not be limited by the amount of data, but instead by the type of data available for hydrologic study. Patterns that are not always apparent from data traditionally employed in hydrology may emerge when analyzed with biotic data that integrates information about the spatiotemporal dynamics of the hydrological cycle and its drivers (Seibert and McDonnell, 2002).

Microbial communities native to an environment (i.e., microbiomes) constitute a novel type of environmental data that comprise many thousands of taxonomically and functionally diverse groups. These communities are most often assessed taxonomically using DNA sequences of the phylogenetically informative 16S rRNA gene. This approach to assessing microbiomes is common in a range of research fields from oceanography to human health. The taxonomic composition of microbial communities is diagnostic of environment types (e.g., freshwater, soil, seawater, etc.) and is sensitive to perturbations, shifting species composition in response to changes in environment (Thompson et al., 2017). Moreover, these communities are often highly diverse and include hundreds of abundant taxa and many thousands of rare taxa, thus providing a rich dataset of information about biological responses to environmental conditions.

Streamwater microbiomes originate primarily from upslope soil, groundwater, and sediment (Crump et al., 2007, 2012; Sorensen et al., 2013; Hermans et al., 2019; Miller et al., 2021) and subsequently develop, through species-sorting and dispersal, in response to biotic factors such as predation and reproduction, but also to abiotic

subsurface and environmental factors, including soil saturation and streamflow rate (Newby et al., 2009), water residence time and network connectivity (Hrachowitz et al., 2016), and interactions with sediment (Droppo et al., 2009). These microscale communities thus reflect a range of macroscale characteristics and processes that also influence and interact with catchment scale hydrology.

Spatial characteristics related to water residence time have been identified as the strongest correlates with streamwater microbial community composition, even among physicochemical factors, along the River Thames (Read et al., 2015) and the Danube River (Savio et al., 2015). Microbial community composition has also been linked to a broader range of landscape scale climatological and geomorphological characteristics (URycki et al., 2020), which are important drivers of hydrologic function. Recent hydrologic studies have employed microbial communities as tracers to elucidate groundwater recharge and flow paths (Sugiyama et al., 2018; Miller et al., 2021), and it has been suggested that microbial information may be useful for hydrologic prediction (Good et al., 2018). However, the breadth of hydrologic data encoded within microbiomes, and the timescales over which microbial communities may be informative, has yet to be quantified and explored.

One major challenge of employing microbial communities as hydrologic observations is identifying informative constituents among thousands or even millions of taxa subject to complex ecosystem interactions that remain poorly understood. The tools of information theory, based on Shannon's entropy (Shannon, 1948), capture linear and nonlinear relationships; analysis of information flows are thus a powerful framework to understand complex patterns in Earth systems science (Goodwell et al., 2020). Here, we sought to leverage information theory to explore the extent to which microbial communities inform the hydrology of the ecosystems in which they occur. We analyze microbial community samples from 64 gauged streams across three major watersheds in the state of Oregon, USA. Our objective is to quantify the information shared between microbial taxa and a set of hydrologic metrics that represent watershed function and characteristic water balance dynamics. Results of this analysis can contribute to a broader understanding of the relationships between hydrology and

streamwater microbial communities and offer insights about the value of microbial communities as a novel source of information to answer open hydrological questions.

3.3 Data and Methods

3.3.1 Study Area

The Willamette, Deschutes, and John Day basins, three similarly sized adjacent watersheds, are exceptional in that together they span a wide range of ecoclimatic conditions yet all lie within the state of Oregon, USA; although all three basins experience wetter winters and drier summers characteristic of the Pacific Northwest, the effect diminishes eastward. Westernmost of the three, the Willamette Basin, encompassing 29,000 km^2 with a mean elevation of 560 m, is bounded by the Coast Range to the west and the Cascade Range to the east (United States Geological Survey, 2017). The Willamette Basin is the wettest and most temperate of the three watersheds, receiving 1,640 mm precipitation annually (246 mm in January and 20 mm in July, on average) and with mean annual minimum and maximum temperatures of 4 °C and 15 °C, respectively (Fig. 3.1). The Deschutes Basin, encompassing 27,700 km^2 to the east of the Cascades, lies at a mean elevation of 1,230 m and is considerably drier and cooler than the Willamette Basin, receiving just 530 mm precipitation annually (75 mm [Jan] and 14 mm [Jul]) and with mean annual minimum and maximum temperatures of 0 °C and 14 °C, respectively. To the east of the Deschutes Basin, the John Day Basin encompasses 20,500 km^2 and rises to a mean elevation of 1,170 m. Climatically more similar to the Deschutes Basin, the John Day Basin receives 460 mm precipitation annually (53 mm [Jan] and 15 mm [Jul]) with mean annual minimum and maximum temperatures of 1 °C and 14 °C, respectively.

All three basins are generally oriented with water flowing south to north and ultimately drain to the Columbia River at the northern border of Oregon. Median winter (Jan, Feb, Mar [*JFM*]) discharge (50% flow duration) at the outlet is 1,260 m^3/s on the Willamette River, 178 m^3/s on the Deschutes River, and 52 m^3/s on the John Day River; whereas median summer (Jul, Aug, Sep [*JAS*]) discharge is 289 m^3/s on the Willamette River, 132 m^3/s on the Deschutes River, and 3 m^3/s on the John Day River (Risley et al. 2008). The Willamette Basin is the most developed of the three study basins (6% developed area), containing the city of Portland at the mouth of the

Willamette River, as well as some additional urban development along the Willamette Valley; the Deschutes Basin (1% developed area), including the city of Bend, and the John Day Basin (<1% developed area) are far less developed (United States Geological Survey, 2017).

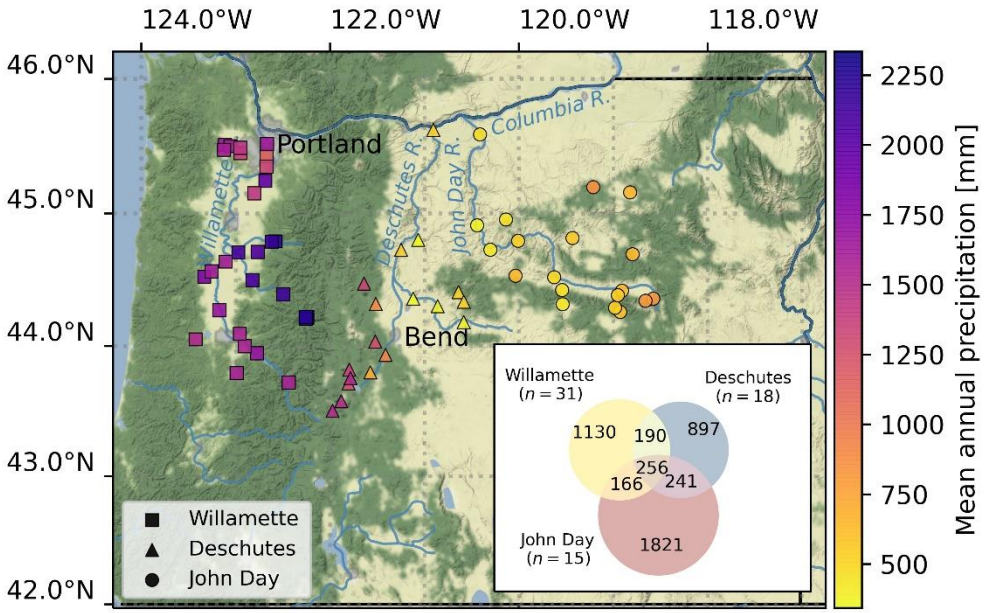

Figure 3.1. Map of co-located stream gages and streamwater DNA sample collections in streams across Willamette (2017), Deschutes (2017), and John Day (2018) basins in Oregon, USA. Marker colors indicate mean annual precipitation in sample catchments (United States Geological Survey, 2017). Inset indicates number of unique and common microbial amplified sequence variants detected in *n* samples across each basin.

3.3.2 DNA Collection and Sequencing

We collected DNA samples near active stream gages from 40 sites in the Willamette Basin and 21 sites in the Deschutes Basin between 21 July and 8 August 2017 and from 20 sites in the John Day Basin from 6-8 August 2018 (Fig. 3.1). We collected most samples from the approximate center of the waterway by lowering from a bridge a two-gallon bucket that had been sanitized at the beginning of the day and sample-rinsed between sites. We filtered and extracted DNA samples from collected streamwater as described in Crump, Kling, Bahr, & Hobbie (2003). We sequenced 16S rRNA with Illumina Miseq V.2 paired end 250bp sequencing after isolating the DNA

and preparing the library following common accepted protocols. A detailed description of collection and processing of this dataset is found in URycki et al. (2020). To approximate even sampling depth, we rarefied the sequence dataset to 1,450 sequences per sample. We selected a rarefaction threshold of 1,450 sequences because it represents the largest tolerable loss of data to retain as many samples as possible (to avoid excluding samples with fewer sequences). Sequences were taxonomically classified with the SILVA 16S rRNA gene database v.132 (Quast et al., 2013). We analyzed these data as a matrix of the relative abundance (i.e., sequence counts) of each unique amplified sequence variant (ASV) detected at each site. In microbiome research, individual ASVs are considered distinct taxonomic groups, usually at the genus level; as such, we use the terms *ASV* and *taxon* interchangeably throughout this paper.

After rarefying the raw sequences, our dataset consisted of 4,701 unique ASVs from 64 sample sites across the three watersheds. We detected 1,742 unique ASVs across 31 samples in the Willamette Basin, 1,584 unique ASVs across 18 samples in the Deschutes Basin, and 2,484 unique ASVs across 15 samples in the John Day Basin (Fig. 3.1). To reduce computational and analytical complexity, we selected a subset of the microbial taxa detected across our study sites. Here, we analyzed the 256 ASVs detected at least once in all three watersheds (Fig. 3.1 and within this subset we identify these unique ASVs by their abundance rank, with the most abundant ASV across all sites identified as *ASV 1* and the least abundant as *ASV 256* (Table 3.S2).

3.3.3 Hydrologic Metrics

We calculated hydrologic metrics from records at 64 stream gages (Fig. 3.1). Study site stream gages spanned headwaters, tributaries, and outlets of the three major rivers and with the intention to sample the range of land use, landcover, and disturbance in each watershed. Stream gages were managed by United States Geological Survey (USGS; U.S. Geological Survey, 2016), Oregon Water Resources Department (Oregon Water Resources Department, 2021), or H.J. Andrews Experimental Forest (Johnson, Rothacher, & Wondzell, 2020). We obtained available records of daily mean discharge (ft^3/s; converted to m^3/s) for each stream gage for the 10-year period preceding DNA sample collection (see *3.3.2 DNA Collection and Sequencing*). We obtained 10 years

of data for 54 sites; 6.9-9.9 years of data for six sites, and <5 years of data for four sites (Table 3.S1).

We sought to analyze a set of metrics that would characterize the hydrology in our study area at different time and flow scales. We therefore calculated 74 hydrologic metrics across three categories: daily discharge, mean monthly discharge, and seasonal flow durations. To describe current hydrologic conditions, we calculated daily absolute discharge [m^3/s] on to the date (*t*) of DNA sample collections and at lags of even numbers of days, up to 30 days, prior to sampling ($Q_t, Q_{(t-2\ days)}, Q_{(t-4\ days)}, \ldots, Q_{(t-30\ days)}$) at all 64 stream gages. To characterize typical catchment conditions, we calculated mean monthly absolute discharge ($\bar{Q}_{mon}$) over the period of study for each month of the water year from October to September for the 60 sites for which we had >6.9 years of data. To capture more extreme hydrologic responses, we calculated seasonal high and low flow durations over the period of study for each of four seasons and annually beginning at the start of the water year ($Q_{P,s}$ for *P = 5*- and *95*-percent exceedance probability [95th and 5th percentile, respectively] for seasons *s = OND*, *JFM*, *AMJ*, *JAS*, and *Ann* [annual]), again for the 60 sites for which we had >6.9 years of data. We normalized absolute discharge values by the sub-catchment area, derived from the StreamStats web application developed by the USGS (https://streamstats.usgs.gov/ss/; USGS, 2017), to obtain daily specific discharge ($q_{(t-n\ days)}$ [$m^3km^{-2}s^{-1}$]), mean monthly specific discharge ($\bar{q}_{mon}$), and seasonal specific discharge flow durations ($q_{P,s}$).

Across sites, daily discharge generally decreased in the 30 days prior to sampling, ranging $Q_{(t-2\ days)} = 0 - 297$ m^3/s, $Q_{(t-10\ days)} = 0 - 340$ m^3/s, and $Q_{(t-30\ days)} = 0 - 521$ m^3/s (Table 3.S1). Mean monthly discharge across sites in January and July ranges $\bar{Q}_{Jan} = 0 - 1{,}755$ m^3/s and $\bar{Q}_{Jul} = 0 - 343$ m^3/s. Summer and winter low flows across sites range $Q_{95,JAS} = 0 - 208$ m^3/s and $Q_{95,JFM} = 0 - 479$ m^3/s. Summer and winter high flows range $Q_{5,JAS} = 0 - 456$ m^3/s and $Q_{5,JFM} = 0 - 3512$ m^3/s.

Median catchment drainage area ranges $0.03 - 29{,}007.89$ km^2 (Table 3.S1). Daily specific discharge across sites ranged from zero to 0.05 m^{3}km^{-2}s^{-1}, 0.06 m^{3}km^{-2}s^{-1}, and 0.09 m^{3}km^{-2}s^{-1} for $q_{(t-2\ days)}$, $q_{(t-10\ days)}$, and $q_{(t-30\ days)}$ respectively. Mean monthly specific discharge in January and July ranges $\bar{q}_{Jan} = 0.0 - 0.22$ m^{3}km^{-2}s^{-1} and $\bar{q}_{Jul} = 0.0 - 0.05$ m^{3}km^{-2}s^{-1}. Summer and winter low flows across sites range $q_{95,JAS} = 0.0 - 0.04$ m^{3}km^{-2}s^{-1} and $q_{95,JFM} = 0.0 - 0.03$ m^{3}km^{-2}s^{-1}. Summer and winter high flows range $q_{5,JAS} = 0.0 - 0.06$ m^{3}km^{-2}s^{-1} and $q_{5,JFM}$ $0.0 - 0.66$ m^{3}km^{-2}s^{-1}.

3.3.4 Information Metrics

We leveraged information theory to analyze the relationships between microbial communities and catchment hydrology. Information theory is useful for understanding complex systems in a range of research domains, including in hydrology and in the geosciences more broadly (Ruddell and Kumar, 2009; Ehret et al., 2014; Olds et al., 2016; Franzen et al., 2020; Goodwell et al., 2020; Li et al., 2021). Information theory is based on Shannon's entropy (*H(X)* [*bits*]), which quantifies the amount of uncertainty in a discrete random variable *X* (Shannon, 1948) and is defined as:

$$H(X) = -\sum p(x) log_2 p(x), \quad (1)$$

where $p(x)$ is the probability distribution function of random variable *X* and summation is over all possible states of *X*=*x* Likewise, the joint entropy ($H(X,Y)[bits]$), is the total amount of information necessary to describe two random variables *X* and *Y*, is defined as:

$$H(X,Y) = -\sum p(x,y) log_2 p(x,y), \quad (2)$$

where ($p(x,y)$) is the joint probability distribution of *X* and *Y* and summation is over all possible states of *X*=*x* and *Y*=*y*. Mutual information ($I(X;\ Y)$ $[bits]$) is the reduction in uncertainty in one random variable *X* as a result of knowledge of a second random variable *Y* and is calculated from the joint and marginal probability distributions of *X* and *Y* as:

$$I(X;\ Y) = \sum p(x) log_2 \frac{p(x,y)}{p(x)p(y)}. \quad (3)$$

Mutual information can be equivalently expressed (Cover and Thomas, 2005) in terms of the entropy of X and Y as:

$$I(X;Y) = H(X) + H(Y) - H(X,Y). \quad (4)$$

We calculated the mutual information shared between each target hydrologic metric Y and abundance of microbial taxon X in terms of their marginal and joint entropies across study sites. To compute the information metrics, we (i) log-transformed hydrologic metrics and ASV abundances, adding 0.001 to each hydrologic metric and 1 to each ASV abundance value to avoid taking the log of zero; (ii) standardized the log-transformed data between zero and unity using the range of each variable, (iii) discretized the standardized data for each X ASV abundance and Y hydrologic metric into five evenly spaced bins. We estimated the marginal and joint probability density functions (*pdfs*) with a fixed binning method to ensure that information metrics for each variable are comparable and transparent. We selected this fixed number of bins based on sample size and standard deviation according to a commonly used rule of thumb (Scott, 1979). We computed the optimal bin size for all variables and found that it ranged between 4 and 9 bins and was ≥5 bins for 95% of the variables. For comparison, we computed information metrics with up to 20 bins and using gaussian kernel densities and verified that prescribing a fixed bin size of 5 provided the most robust results for this dataset. We also note that although the bin size and *pdf* estimation methods do influence entropy values, this choice does not typically affect comparison patterns (Loritz et al., 2019) which is the focus of this analysis. We then employed the marginal and joint fixed-bin *pdf*s in equations 1 and 2, respectively, to obtain $H(X)$, $H(Y)$, and $H(X,Y)$. Finally, we applied equation (4) to compute the mutual information between each microbial taxon and each hydrologic metric.

To identify statistically significant relationships between microbial taxa and hydrologic metrics, we used a shuffled surrogates method (Ruddell & Kumar, 2009). For 1,000 iterations of each pair of X and Y, we randomly shuffled X to destroy any structure while maintaining the distribution of the data and then computed mutual information of the shuffled data. We considered $I(X,Y)$ statistically significant if the value was greater than the 99th percentile of shuffled iterations.

Finally, we were interested in comparing mutual information values across microbial taxa and hydrologic metrics, so we normalized mutual information scores by the entropy of the target hydrologic metrics as:

$$I(X;Y)_{norm} = \frac{I(X;Y)}{H(Y)}. \quad (5)$$

The normalized mutual information value is thus the fraction reduction in uncertainty of the hydrologic metric that comes from observing the abundance of a particular microbial taxon.

To analyze patterns in hydrologic information shared with the microbial community, we summarized mutual information in two ways: 1) median value of normalized information across all informative microbial taxa for each hydrologic metric, and 2) the number of informative microbial taxa for each hydrologic metric. We define informative taxa as those with $I(X;Y)_{norm} > 0$.

We first sought to compare the value of information shared between the streamwater microbial community and absolute versus specific discharge and among the three major categories of hydrologic metrics: daily discharge, mean monthly discharge, and seasonal high and low flow durations. To compare information shared with absolute versus specific discharge, we applied a non-parametric Mann-Whitney *U*-test (Mann and Whitney, 1947) to test for differences in median value of information shared between microbial taxa and absolute versus specific discharge metrics. We then applied a paired sample *t*-test to compare the number of informative taxa for absolute versus specific discharge for each hydrologic metric. To satisfy the parameters of a paired sample test, we assigned a value of zero to all ASVs that did not share significant information with hydrologic metrics. (Throughout the rest of the analysis, non-significant values of mutual information were excluded from calculations). To compare the strength of the relationships among the three hydrologic categories, we applied Mann-Whitney *U*-tests to test for differences in median value of shared information, and one-way analysis of variance (ANOVA) *F*-test to test for differences in the mean number of informative ASVs, across all metrics for each category (e.g., across all time lags for daily discharge).

We performed regression analyses to analyze patterns in mutual information between the microbial community and discharge through time for daily discharge and

mean monthly discharge. For daily discharge, we fit regression functions to the median value of shared information and to the number of informative ASVs across time lags. For mean monthly discharge, we fit regression functions to the median value of shared information and the number of informative ASVs across months. We assessed model fit with Pearson's correlation (ρ).

To assess how the strength of relationships with streamwater microbial taxa may be related to discharge magnitude, we used non-parametric Kruskal-Wallis *H*-tests (i.e., one-way ANOVA on ranks; Kruskal and Wallis, 1952) to test for differences in median value of mutual information between high and low flow durations. We conducted a one-way ANOVA *F*-test to test for differences in the mean number of informative ASVs between high and low flow durations.

Finally, we sought to identify patterns in mutual information related to abundance of microbial taxa across sites. We applied a non-parametric Spearman's rank correlation (r_s) to test for relationships between the value of mutual information for each hydrologic metric and 1) total abundance of an ASV across all sites and 2) the number of sites at which an ASV was detected. We conducted all statistical tests at significance level $\alpha = 0.05$ using *SciPy* (v 1.6.2) for Python.

3.4 Results

3.4.1 Stream Microbial Community Composition

Overall, gammaproteobacteria was the most abundant taxonomic group, comprising 20-40% of most stream communities we sampled (Fig. 2). Phylum Bacteroidota was also common across all sites. The proportion of unclassified Other microbes, often consisting of more rare taxa, was generally greater in sites with lower daily stream discharge on the date of DNA sample collection (Q_t). Phyla Verrucomicrobiota and Actinobacteriota were, very broadly, more common at sites with higher discharge. Alphaproteobacteria, Cyanobacteria, and Planctomycetota were also detected in smaller numbers across most sites.

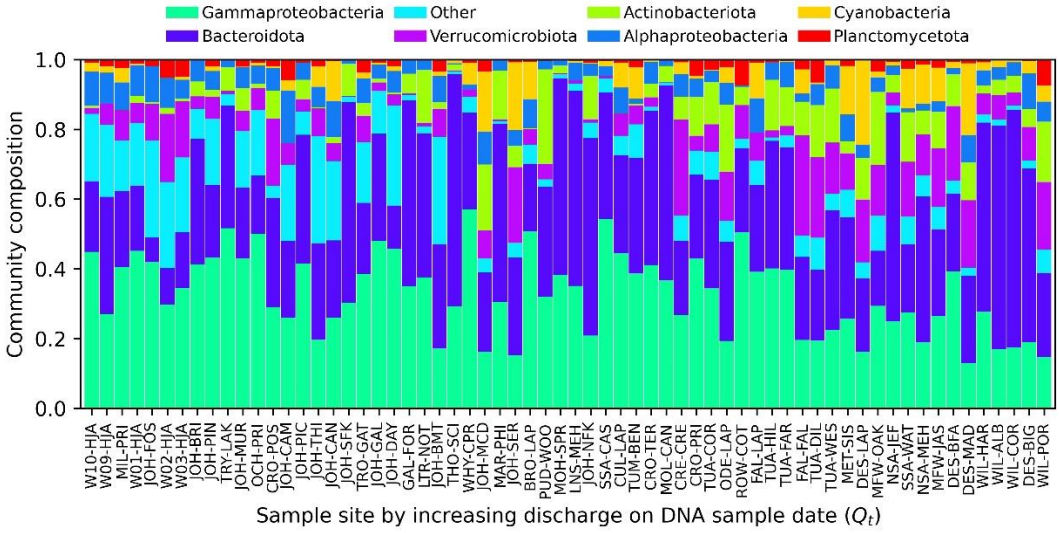

Figure 3.2. Microbial community composition by phylogenetic group for streamwater DNA samples collected in summer throughout the Deschutes (2017), Willamette (2017), and John Day (2018) basins in Oregon, USA. Samples are presented in order of increasing discharge on the date of DNA sample collection (Q_t; Table 3.S1).

3.4.2 Mutual Information between Microbial Communities and Hydrologic Metrics

We calculated normalized mutual information scores for each ASV common to all three watersheds for each hydrologic metric, resulting in a matrix of 256 ASVs × 76 hydrologic metrics (and area; Fig. 3.3). Of the 256 common ASVs, 102 had statistically significant mutual information with at least one hydrologic metric, and each hydrologic metric shared information with at least one ASV. Mutual information is generally concentrated among more abundant taxa with a notable exception. *ASV 69*, a Bacteroidota classified to the lake-inhabiting genus *Lacihabitans* (Joung et al., 2014) shares information with all of the absolute and nearly all of the specific hydrologic metrics. *ASV 69* demonstrates an especially strong relationship with daily discharge (Fig 3.3; Table S2). Across all absolute and specific hydrologic metrics, 9.7% of taxa, on average, share information with a given hydrologic metric, reducing uncertainty of that metric by a median of 15.6%. The maximum value of shared information is $I(ASV\ 4; \bar{Q}_{Feb})_{norm} = 0.447\ bits/bit$, corresponding to a 44.7% reduction in uncertainty of mean February absolute discharge by observing the abundance of *ASV 4*, a Gammaproteobacteris classified to the planktonic genus

Limnohabitans, the fourth most abundant microbial taxon detected across all sites; the minimum nonzero value is $I(ASV\ 3; q_{(t)})_{norm} = 0.124\ bits/bit$.

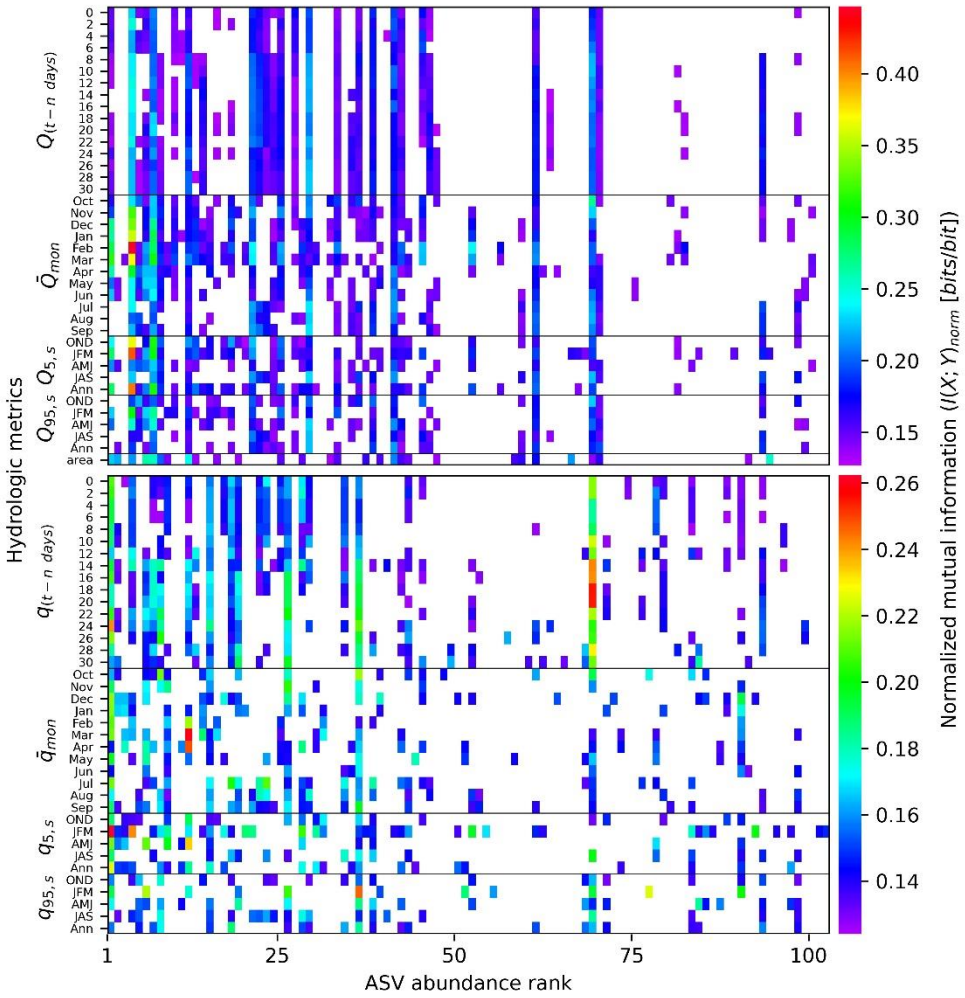

Figure 3.3. Heatmap illustrating the normalized mutual information ($I(X;Y)_{norm}$ [bits/bit]) between streamwater microbial amplified sequence variants (X=ASVs) and absolute discharge (Y=Q; top), specific discharge (Y=q; bottom) hydrologic metrics, as well as basin drainage area (top, bottom line), for study streams in Oregon, USA. We collected microbial DNA samples in summer in the Willamette (2017), Deschutes (2017), and John Day (2018) basins. Hydrologic metrics include daily discharge at time lags n days prior to DNA sample day t ($Q_{(t-n\ days)}$, $q_{(t-n\ days)}$), mean monthly discharge ($\bar{Q}_{mon}, \bar{q}_{mon}$) for months October to September, and seasonal high and low flow durations ($Q_{P,s}$, $q_{P,s}$ for P = 5- and 95-percent exceedance probability for seasons s = fall [*OND*], winter [*JFM*], spring [*AMJ*], summer [*JAS*], and annually [*Ann*]).

Microbial taxa share more information with absolute than with specific discharge metrics (Table 1; Fig. 3.3). Median value of mutual information is significantly higher for absolute versus specific discharge ($I(X;Q)_{norm} = 0.163\ bits/$

bit vs $I(X;q)_{norm} = 0.150\ bits/bit;\ U = 126{,}985,\ p < 0.001$). Furthermore, a greater number of taxa share information with absolute (28.2 ±4.1 SD) than with specific (21.4 ±5.5 SD) hydrologic metrics (two-sample one-sided $t = 6.13,\ p < 0.001$). Because relationships with the microbial community were so much stronger, in terms of both number of informative taxa and median value of shared information, we focus the remainder of the analysis on absolute discharge metrics only. (Results of the complementary analyses of specific discharge metrics are presented in supplementary figures 3.S1-3.S6.)

Table 3.1. Medians and ranges of entropy ($H(Y)$ [bits]) and normalized mutual information ($I(X;Y)/H(Y)$ [bits/bit]) between streamwater microbial amplified sequence variants (ASVs) and hydrologic metrics: absolute discharge (Q [m^3s^{-1}]) and specific (per unit area) discharge (q [$m^3km^{-2}s^{-}$]). We collected microbial DNA samples in summer in the Willamette (2017), Deschutes (2017), and John Day (2018) basins in Oregon, USA. Hydrologic metrics include daily discharge at time lags up to *n* = 30 days prior to DNA sample collection, mean monthly discharge for *mon* = all months October to September, and seasonal high and low flow durations for *P* = 5- and 95-perecent exceedance probability and seasons *s* = fall [*OND*], winter [*JFM*], spring [*AMJ*], summer [*JAS*], and annually [*Ann*]. Catchment area is derived from the StreamStats web application developed by the USGS (USGS, 2017).

		Information metric					
		(Absolute discharge Q)			(Specific Discharge q)		
		H(Y)	$\frac{I(X;Y)}{H(Y)}$		***H(Y)***	$\frac{I(X;Y)}{H(Y)}$	
Hydrologic metric			Median	(Range)		Median	(Range)
Daily discharge ($Q_{(t-n\ days)}$, $q_{(t-n\ days)}$)	0	2.107	0.151	(0.100)	2.181	0.145	(0.092)
	2	2.098	0.161	(0.120)	2.179	0.147	(0.088)
	4	2.082	0.155	(0.092)	2.185	0.148	(0.068)
	6	2.078	0.158	(0.082)	2.185	0.147	(0.069)
	8	2.064	0.157	(0.108)	2.150	0.147	(0.066)
	10	2.052	0.163	(0.105)	2.148	0.146	(0.097)
	12	2.070	0.169	(0.116)	2.136	0.143	(0.085)
	14	2.093	0.166	(0.115)	2.116	0.146	(0.117)
	16	2.079	0.166	(0.114)	2.073	0.148	(0.115)
	18	2.091	0.157	(0.105)	2.069	0.147	(0.126)
	20	2.100	0.162	(0.107)	2.062	0.158	(0.130)
	22	2.065	0.164	(0.118)	2.032	0.162	(0.089)
	24	2.060	0.164	(0.087)	1.960	0.155	(0.114)
	26	2.079	0.163	(0.091)	2.047	0.151	(0.083)
	28	2.055	0.157	(0.096)	2.041	0.151	(0.094)
	30	2.055	0.157	(0.095)	2.054	0.147	(0.082)
	Oct	2.100	0.167	(0.129)	1.995	0.155	(0.078)

Mean monthly discharge ($\bar{Q}_{mon}$, $\bar{q}_{mon}$)	Nov	2.208	0.164	(0.186)	1.972	0.172	(0.064)
	Dec	2.174	0.161	(0.197)	1.992	0.152	(0.077)
	Jan	2.150	0.158	(0.223)	2.115	0.153	(0.080)
	Feb	1.925	0.185	(0.304)	2.186	0.160	(0.081)
	Mar	2.003	0.164	(0.236)	2.121	0.159	(0.126)
	Apr	2.024	0.164	(0.160)	2.054	0.149	(0.114)
	May	2.000	0.165	(0.100)	2.103	0.145	(0.062)
	Jun	2.015	0.155	(0.103)	2.100	0.144	(0.041)
	Jul	1.964	0.174	(0.101)	2.142	0.151	(0.079)
	Aug	2.061	0.167	(0.091)	2.122	0.149	(0.051)
	Sep	2.092	0.172	(0.097)	2.114	0.146	(0.064)
Seasonal flow durations ($Q_{P,s}$, $q_{P,s}$)	5, OND	2.160	0.163	(0.236)	2.186	0.145	(0.060)
	5, JFM	2.041	0.165	(0.276)	2.100	0.150	(0.127)
	5, AMJ	2.056	0.154	(0.093)	2.074	0.165	(0.093)
	5, JAS	1.949	0.179	(0.086)	2.076	0.150	(0.066)
	5, Ann	2.012	0.171	(0.267)	2.168	0.152	(0.094)
	95, OND	2.006	0.161	(0.088)	2.205	0.142	(0.047)
	95, JFM	2.038	0.167	(0.166)	1.703	0.193	(0.095)
	95, AMJ	2.015	0.163	(0.118)	1.946	0.155	(0.049)
	95, JAS	2.067	0.165	(0.087)	2.236	0.149	(0.051)
	95, Ann	2.024	0.154	(0.088)	2.169	0.151	(0.055)
Catchment area		1.81	0.157	(0.131)			

In comparing mutual information between the streamwater microbial community and the three hydrologic categories, we find that the microbial community is, on average, more strongly related to mean monthly discharge and typical seasonal extreme flows, in terms of median value of shared information but not necessarily in terms of the number of informative taxa. Median value of shared information between microbial taxa and daily discharge ($I(X; Q_{(t-n\ days)})_{norm} = 0.159\ bits/bit$) was significantly lower than median value of shared information for mean monthly discharge ($I(X; \bar{Q}_m)_{norm} = 0.165\ bits/bit;\ U = 473{,}779, p < 0.001$) and seasonal flow durations (median $I(X; Q_{P,s})_{norm} = 0.165\ bits/bit; U = 619{,}418, p < 0.001$). Mean number of informative taxa for each category was similar: $27.4\ (\pm 2.6\ \text{SD})$ taxa, $29.3\ (\pm 4.7\ \text{SD})$ taxa, and $28.1\ (\pm 5.2\ \text{SD})$ taxa for daily discharge, monthly mean discharge, and seasonal flow durations, respectively.

For daily absolute discharge, information shared with the microbial community across all time lags up to 30 days before sampling ranges $I(ASV\ 47; Q_{(t-20\ days)})_{norm} = 0.127\ bits/bit$ to $I(ASV\ 4; Q_{(t-2\ days)})_{norm} = 0.250\ bits/bit$ (Fig. 3.4). Median value of mutual information over time follows a second-order polynomial function with a peak in shared information with discharge at 16 days prior to sampling (Pearson's $\rho = 0.73, p = 0.001$; Fig. 3.4). The number of informative ASVs through time also fits a second-order polynomial with a maximum number of informative taxa for discharge at 22 days prior to sampling ($\rho = 0.68, p = 0.004$; Fig. 3.4).

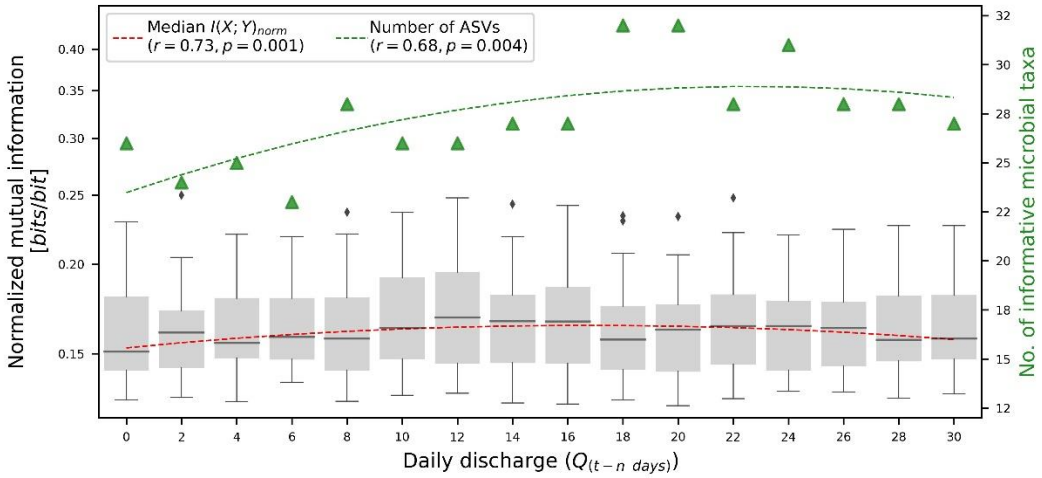

Figure 3.4. Median value of normalized mutual information ($I(ASV; Q_{(t-n\ days)})_{norm}$ [bits/bit]) between informative streamwater microbial amplified sequence variants (ASVs) and daily discharge at different time lags ($Q_{(t-n\ days)}$ for time lags up to *n* = 30 days before sample date) for study streams in Oregon, USA. Boxes show medians and interquartile ranges; whiskers show values within 1.5 times the interquartile range. Green triangles indicate the number of ASVs with statistically significant mutual information (99% confidence). Dashed lines show best fit curves between time lags and median value of normalized information (red) and number of informative ASVs (green), with Pearson correlation (*r*) and *p*-value indicated in legend. We collected microbial DNA samples in summer in the Willamette (2017), Deschutes (2017), and John Day (2018) basins.

For mean monthly discharge, shared information across all months ranges $I(ASV\ 4; \bar{Q}_{Feb})_{norm} = 0.447\ bits/bit$ to $I(ASV\ 24; \bar{Q}_{Jan})_{norm} = 0.133\ bits/bit$ (Fig 3.5). Unlike for daily discharge, we observed no trend in mutual information by month (Fig 3.5). Number of informative ASVs fits a second-order polynomial with the greatest number of taxa sharing information with January mean discharge ($\rho = 0.70, p = 0.01$; Fig. 3.5). Across seasons and high and low flow durations, mutual

information shared with microbial taxa ranges $I(ASV\ 20; Q_{5,OND})_{norm} = 0.130\ bits/bit$ to $I(ASV\ 4; Q_{5,JFM})_{norm} = 0.414\ bits/bit$ (Fig. 3.6). We did not detect a difference in shared information between high versus low flows, neither in terms of median value of shared information nor in the number of informative taxa.

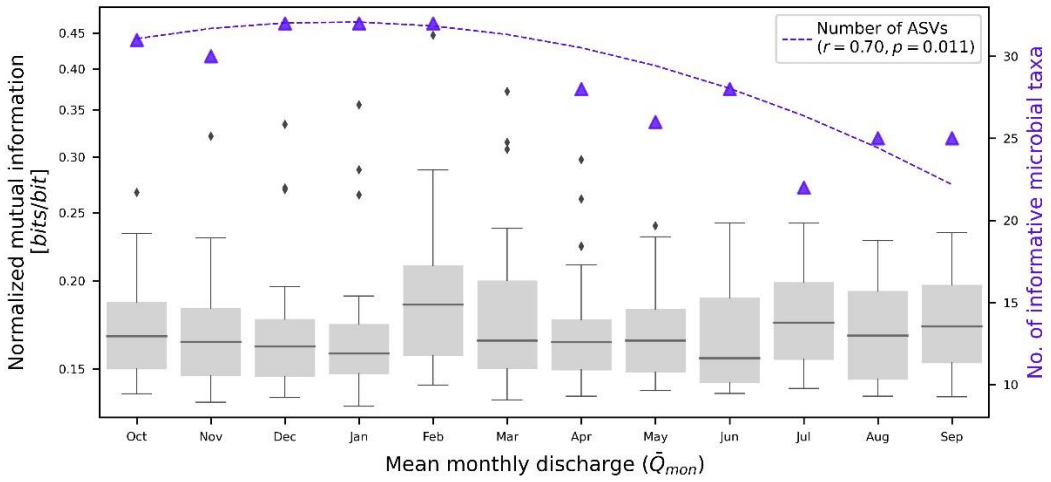

Figure 3.5. Median value of normalized mutual information ($I(ASV; \bar{Q}_{mon})_{norm}$ [bits/bit]) between informative streamwater microbial amplified sequence variants (ASVs) and mean monthly discharge ($\bar{Q}_{mon}$) of study streams in Oregon, USA. Boxes show medians and interquartile ranges; whiskers show values within 1.5 times the interquartile range. Purple triangles indicate the number of ASVs with statistically significant mutual information (99% confidence). Dashed purple line shows best fit curve between months and number of informative ASVs, with Pearson correlation (*r*) and *p*-value indicated in legend. We observed no significant relationship between month and median value of mutual information. We collected microbial DNA samples in July and August in the Willamette (2017), Deschutes (2017), and John Day (2018) basins.

Mutual information increases with increasing detection of microbial taxa for all three categories of hydrologic metrics, although the relationship is not as strong for daily discharge. For all three categories, mutual information increases linearly with the log of abundance of a microbial taxon across all sites for mean monthly discharge ($\rho = 0.42, p < 0.001$), seasonal flow durations ($\rho = 0.37, p < 0.001$), and daily discharge ($\rho = 0.12, p = 0.013$; Fig 3.7). Mutual information is most strongly correlated with the number of sites at which a microbial taxon is detected ($\rho = 0.49, 0.48, 0.26$ for mean monthly, seasonal flow durations, and daily discharge, respectively; all $p < 0.001$; Fig 3.S4).

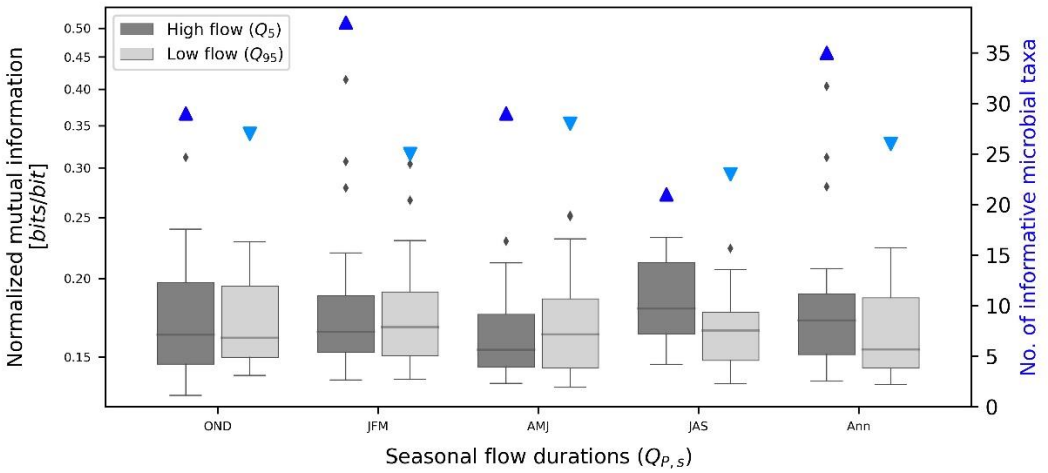

Figure 3.6. Normalized mutual information ($I(ASV; Q_{P,s})_{norm}$ [bits/bit]) between unique streamwater microbial amplified sequence variants (ASVs) and seasonal high (dark) and low (light) flow durations ($Q_{P,s}$ for P = 5- and 95-percent exceedance probability for seasons s = fall [*OND*], winter [*JFM*], spring [*AMJ*], summer [*JAS*], and annually [*Ann*]) at study streams in Oregon, USA. Boxes show medians and interquartile ranges; whiskers show values within 1.5 times the interquartile range. Blue triangles indicate the number of ASVs with statistically significant mutual information (99% confidence) for high flows (dark) and low flows (light). We collected microbial DNA samples in summer in the Willamette (2017), Deschutes (2017), and John Day (2018) basin.

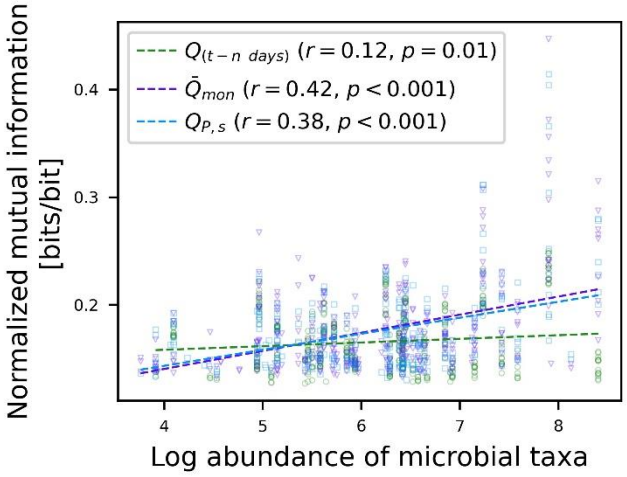

Figure 3.7. Normalized mutual information ($I(ASV; Y)_{norm}$ [bits/bit]) between hydrologic metrics and streamwater microbial taxa versus the log of abundance of taxa in streams across Oregon, USA. We collected microbial DNA samples in summer in the Willamette (2017), Deschutes (2017), and John Day (2018) basins. Hydrologic metrics include daily discharge at time lags up to n = 30 days prior to DNA sample collection ($Q_{(t-n\ days)}$; green circles), mean monthly discharge ($\bar{Q}_{mon}$; purple triangles), and seasonal high and low flow durations ($Q_{P,s}$; 5- and 95-percent exceedence probability for all seasons and annually; blue squares). Legend shows Pearson's correlation (r) and p-value of the linear regression.

3.5 Discussion

We detected significant relationships between summer streamwater microbial communities and both absolute and specific (per unit area) hydrologic metrics. Drainage area, the upslope area that drains to a point in a stream, is one of the strongest drivers of stream discharge across timescales, and as such, discharge normalized by area reflects a different set of controls than absolute discharge. Streams draining large watersheds generally respond more slowly to precipitation events because of greater infiltration rates, longer transit times, and diminishing peak event flows, usually resulting in lower rates of discharge per unit area than might be observed in an otherwise similar smaller watershed (Dingman, 2015). The strength of relationships with microbial taxa was greater for absolute than for specific discharge, as indicated by both a higher median value of mutual information and a greater number of informative taxa across absolute discharge metrics. However, patterns in the information shared between microbial taxa and absolute discharge appear different than patterns in information shared with specific discharge (Fig. 3.3); that is, information shared with discharge per unit area across the microbial community is not simply a characteristically lesser value of the information shared with absolute discharge. Furthermore, patterns in information shared between both absolute discharge and discharge per unit area appear independent of the patterns in information shared with catchment drainage area (Fig 3.3). The different strengths and patterns of mutual information between microbial taxa and absolute versus specific discharge, and that these relationships do not appear to be driven by catchment area, suggests that distinct aspects of the streamwater microbial communities integrate information from a broad range of complex processes and interactions.

We furthermore found that summer microbial taxa were, on average, more informative of long-term mean monthly discharge and seasonal flow durations than of daily discharge concurrent or up to 30-days antecedent to the microbial sampling dates. The hydrologic regime of a catchment, including typical extremes as well as average monthly or seasonal stream discharge, is shaped by larger scale time and space characteristics, such as climate and topographic organization of the catchment (McGuire et al., 2005). On the other hand, daily discharge is influenced to varying

degrees by short- and moderate-timescale localized variable conditions, including duration and intensity of recent precipitation events and antecedent soil moisture and storage, which impact infiltration rates and water transit times (Dingman, 2015). That the summer communities we sampled were more informative of metrics of general hydrologic regime suggests that these communities are shaped by broader catchment characteristics, further supporting earlier findings connecting streamwater microbial diversity to geomorphic and climatic characteristics of the catchment (URycki et al., 2020). However, the summer microbial community was also informative of recent daily discharge, suggesting that, although seasonality and properties of the catchment contribute to the development of characteristic microbial communities (Crump et al., 2009), these communities also respond to conditions at shorter timescales. Additional research on microbial community dynamics at higher temporal resolution would be useful to further test these responses.

The number of informative microbial taxa is greater for discharge metrics in the days or months prior to DNA sampling than the day or season of sampling. For daily discharge, this pattern suggests that informative taxa may accumulate over time (~2 weeks in this case) as microbial communities develop in response to environmental conditions that also control discharge. For monthly discharge, an explanation for this lag in microbial community response is less straightforward. Given that freshwater microbial communities experience seasonal shifts but return to characteristic core summer and winter communities (Crump and Hobbie, 2005; Crump et al., 2009), we might expect the strongest relationships between microbial taxa and hydrologic measures to be observed during periods when conditions are similar to those when microbial communities are sampled. Our results suggest the opposite pattern may be true. For monthly discharge, the number of informative ASVs is greater in months more distant (e.g., March and November) than those in which we sampled (i.e., July and August; Fig 3.5). Furthermore, although we did not observe a clear trend in the median value of mutual information across months, the highest values of mutual information shared between microbial taxa and monthly discharge occur with a time lag of 4-6 months, as observed by the outlying high values of mutual information in fall, winter, and spring months, but not in summer (Fig 3.5). One explanation for this

phenomenon is that patterns in shared information might be more related to discharge dynamics (e.g., flow volume) than to temporal lags. The peak in shared information we observed between summer microbes and winter flows might indicate that the microbial community is more informative of higher flow conditions, which typically occur in winter in the Mediterranean climate of the Pacific Northwest where our study is located, than the lower flows observed in summer when we collected DNA samples. However, the patterns in shared information we identified between seasonal high and low flows (i.e., Q_5, Q_{95}) would appear to contradict this explanation; if summer microbes are more informative of high flows (typically winter) than low flows (typically summer), we would expect to see more information shared between microbial taxa and seasonal high flow durations (Q_5), but we observed no difference in shared information between high and low seasonal flow durations (Fig. 3.6). Future research might examine whether this pattern of greater shared information with trends in the opposite season holds when winter microbial communities are sampled. Furthermore, considering that our results indicate that microbial communities also respond to local environmental conditions at shorter timescales (days and weeks), it may be that the relationships between microbial taxa and typical flow regime are confounded by short-term perturbations to the community. Additional research on microbial community diversity at higher temporal resolution is necessary to determine the timescale and magnitude at which the streamwater microbial community responds to variable local conditions.

Microbial community composition likely reflects different components of hydrologic and ecosystem function, even within the relatively small suite of discharge metrics we analyzed. Future work might also analyze more closely the sets of taxa that share information with one metric versus another, for instance daily discharge versus mean monthly discharge, and whether and how those sets of informative taxa overlap. Investigation of the taxonomy and phylogeny of informative taxa, coupled with analysis of the functional roles of these taxa within the microbial community and within the larger ecosystem, will improve our understanding of the dynamics and interactions between the streamwater microbial community and hydrology, as well as contribute new knowledge to the fields of microbiology and ecology, among others.

While we identified significant shared information between a number of microbial taxa and catchment hydrology, it is impossible from this analysis to quantify the sum total informational value of the community overall. At least some of the hydrologic information encoded within individual taxa likely overlaps with that of other taxa (i.e., is redundant); on the other hand it is also plausible that information encoded within different taxa is more valuable when considered jointly (i.e., is synergistic). Future work might enlist additional information theory metrics, such as mutual information conditioned on a third microbial or hydrologic variable and information decomposition (Goodwell et al., 2020; Gutknecht et al., 2021). Such techniques can quantify joint relations and parse information types (redundant, synergistic and unique) to better explore causality and broaden our understanding of the dynamics of microbial community development and catchment hydrology.

We evaluated possible sources of uncertainty in our results and used our best judgment in the analysis and parameter choices required for investigating this type novel type of hydrologic data. For instance, we selected only those microbial taxa that appeared in all three major study watersheds, which reduced our dataset from >4,000 taxa down to 265 taxa. Although we reasoned that greater detection over a wider range of conditions would likely result in identifying stronger relationships between microbial taxa and hydrologic metrics, we acknowledge that this course selection criteria may have precluded identification of some potentially meaningful relationships. However, the positive relationship between the value of information encoded within a particular taxon and the number of sites at which it was detected suggests that the strongest relationships were captured in the common taxa we analyzed (Fig. 3.S4). We also applied a strict significance threshold ($\alpha = 0.01$) for mutual information values, which resulted in more robust results. Although this conservative threshold likely resulted in the loss of some data, it also contributes to stronger confidence in the patterns we identify. On the other hand, we opted to include some study sites with less than a 10-year period of hydrologic record (minimum = 6.9 years; Table 3.S1) in calculations of mutual information for monthly discharge and seasonal flow durations. We determined that small differences in hydrologic statistics for a fraction of sites would not bias our overall results and that including a greater number

of co-located hydrological and microbial observations would strengthen the overarching insights gained from the analysis.

3.6 Conclusions

Microbial diversity and hydrologic metrics are driven by common processes, and we found that microbial taxonomic abundance is statistically related to all 74 hydrologic metrics we analyzed. An average of 9.6% of the summer streamwater microbial taxa we analyzed shared information with a given hydrologic metric, including mean monthly discharge, seasonal high and low flow durations, and daily discharge, in some cases reducing the uncertainty of a hydrologic metric by >40%. The summer microbial community shared the most information with winter mean flows, which also coincide with high flow periods in our study area. The microbial community also shared information with daily discharge, most strongly at an approximately two-week lag from sampling dates. Considering that a single streamwater DNA sample can yield thousands of data points, many more than traditional hydrologic observations, our study lends further support to the value of microbial community composition as a hydrologic observation at multiple timescales. Microbiome analyses therefore have the potential to contribute new and complex insights to outstanding questions in the field of hydrology, and the value and application of these data should continue to be explored.

3.7 Author Contributions

Author contributions are as follows: DRU, conceptualization, data curation, formal analysis, writing – original draft; MB: formal analysis, methodology, writing – review and editing; SPG, conceptualization, funding acquisition, resources, writing – review and editing; BCC, funding acquisition, data curation, resources, writing – review and editing; BL, methodology, writing – review and editing.

3.8 Funding

This work was supported by the National Science Foundation grant EAR 1836768. The first author would like to acknowledge STEM Scholarship support from NSF grant 1153490. Data and facilities for a portion of this research were provided by the HJ Andrews Experimental Forest and Long Term Ecological Research (LTER) program, administered cooperatively by the USDA Forest Service Pacific Northwest

Research Station, Oregon State University, and the Willamette National Forest. This material is based upon work supported by the National Science Foundation under the LTER Grants: LTER8 DEB-2025755 (2020-2026) and LTER7 DEB-1440409 (2012-2020).

3.9 Acknowledgments

We gratefully acknowledge the graduate and undergraduate students who contributed many long hours of lab and field work. The authors further acknowledge that Oregon State University in Corvallis, Oregon is located within the traditional homelands of the Ampinefu Band of Kalapuya. Following the Willamette Valley Treaty of 1855, Kalapuya people were forcibly removed to reservations in western Oregon. Today, living descendants of these people are a part of the Confederated Tribes of Grand Ronde Community of Oregon and the Confederated Tribes of the Siletz Indians.

3.10 Data Availability

DNA sequence data is archived under under run accessions SRX8627687-SRX8627763 at the National Center for Biotechnology Information (NCBI).

3.11 References

Baldocchi, D., Falge, E., Gu, L., Olson, R., Hollinger, D., Running, S., et al. (2001). FLUXNET: A New Tool to Study the Temporal and Spatial Variability of Ecosystem-Scale Carbon Dioxide, Water Vapor, and Energy Flux Densities. Bull. Am. Meteorol. Soc. 82, 2415–2434. doi:10.1175/1520-0477(2001)082<2415:FANTTS>2.3.CO;2.

Blöschl, G., Bierkens, M. F. P., Chambel, A., Cudennec, C., Destouni, G., Fiori, A., et al. (2019). Twenty-three unsolved problems in hydrology (UPH) – a community perspective. Hydrol. Sci. J. 64, 1141–1158. doi:10.1080/02626667.2019.1620507.

Cover, T. M., and Thomas, J. A. (2005). Elements of Information Theory. Elem. Inf. Theory, 1–748. doi:10.1002/047174882X.

Crump, B. C., Adams, H. E., Hobbie, J. E., and Kling, G. W. (2007). Biogeography of Bacterioplankton in Lakes and Streams of an Arctic Tundra Catchment. Ecology 88, 1365–1378. doi:10.1890/06-0387.

Crump, B. C., Amaral-Zettler, L. A., and Kling, G. W. (2012). Microbial diversity in arctic freshwaters is structured by inoculation of microbes from soils. ISME J. 6, 1629–1639. doi:10.1038/ismej.2012.9.

Crump, B. C., and Hobbie, J. E. (2005). Synchrony and seasonality in bacterioplankton communities of two temperate rivers. Limnol. Oceanogr. 50, 1718–1729. doi:10.4319/lo.2005.50.6.1718.

Crump, B. C., Kling, G. W., Bahr, M., and Hobbie, J. E. (2003). Bacterioplankton community shifts in an Arctic lake correlate with seasonal changes in organic matter source. Appl. Environ. Microbiol. 69, 2253–2268. doi:10.1128/AEM.69.4.2253-2268.2003.

Crump, B. C., Peterson, B. J., Raymond, P. A., Amon, R. M. W., Rinehart, A., McClelland, J. W., et al. (2009). Circumpolar synchrony in big river bacterioplankton. Proc. Natl. Acad. Sci. U. S. A. 106, 21208–12. doi:10.1073/pnas.0906149106.

Dingman, S. L. (2015). Physical Hydrology. 3rd ed. Prentice Hall.

Droppo, I. G., Liss, S. N., Williams, D., Nelson, T., Jaskot, C., and Trapp, B. (2009). Dynamic existence of waterborne pathogens within river sediment compartments. Implications for water quality regulatory affairs. Environ. Sci. Technol. 43, 1737–1743. doi:10.1021/es802321w.

Ehret, U., Gupta, H. V., Sivapalan, M., Weijs, S. V., Schymanski, S. J., Blöschl, G., et al. (2014). Advancing catchment hydrology to deal with predictions under change. Hydrol. Earth Syst. Sci. 18, 649–671. doi:10.5194/HESS-18-649-2014.

Franzen, S. E., Farahani, M. A., and Goodwell, A. E. (2020). Information Flows: Characterizing Precipitation-Streamflow Dependencies in the Colorado Headwaters With an Information Theory Approach. Water Resour. Res. 56, e2019WR026133. doi:10.1029/2019WR026133.

Good, S. P., Urycki, D. R., and Crump, B. C. (2018). Predicting Hydrologic Function With Aquatic Gene Fragments. Water Resour. Res. 54, 2424–2435. doi:10.1002/2017WR021974.

Goodwell, A. E., Jiang, P., Ruddell, B. L., and Kumar, P. (2020). Debates—Does Information Theory Provide a New Paradigm for Earth Science? Causality, Interaction, and Feedback. Water Resour. Res. 56, e2019WR024940. doi:10.1029/2019WR024940.

Gutknecht, A. J., Wibral, M., and Makkeh, A. (2021). Bits and pieces: understanding information decomposition from part-whole relationships and formal logic. Proc. R. Soc. A 477. doi:10.1098/RSPA.2021.0110.

Hermans, S. M., Buckley, H. L., Case, B. S., and Lear, G. (2019). Connecting through space and time: catchment-scale distributions of bacteria in soil, stream water and sediment. Environ. Microbiol., 1462-2920.14792. doi:10.1111/1462-2920.14792.

Hrachowitz, M., Benettin, P., van Breukelen, B. M., Fovet, O., Howden, N. J. K., Ruiz, L., et al. (2016). Transit times-the link between hydrology and water quality at the catchment scale. Wiley Interdiscip. Rev. Water 3, 629–657. doi:10.1002/wat2.1155.

Johnson, S.L., J.S. Rothacher, and S.M. Wondzell. 2020. Stream discharge in gaged watersheds at the HJ Andrews Experimental Forest, 1949 to present ver 33. Environmental Data Initiative. https://doi.org/10.6073/pasta/0066d6b04e736af5f234d95d97ee84f3.

Joung, Y., Kim, H., Kang, H., Lee, B. Il, Ahn, T. S., and Joh, K. (2014). Lacihabitans soyangensis gen. nov., Sp. nov., A new member of the family Cytophagaceae, isolated from a freshwater reservoir. Int. J. Syst. Evol. Microbiol. 64, 3188–3194. doi:10.1099/IJS.0.058511-0/CITE/REFWORKS.

Kruskal, W. H., and Wallis, W. A. (1952). Use of Ranks in One-Criterion Variance Analysis. J. Am. Stat. Assoc. 47, 583. doi:10.2307/2280779.

Li, B., Good, S. P., and URycki, D. R. (2021). The value of L-band soil moisture and vegetation optical depth estimates in the prediction of vegetation phenology. Remote Sens. 13. doi:10.3390/rs13071343.

Loritz, R., Kleidon, A., Jackisch, C., Westhoff, M., Ehret, U., Gupta, H., et al. (2019). A topographic index explaining hydrological similarity by accounting for the joint controls of runoff formation. Hydrol. Earth Syst. Sci. 23, 3807–3821. doi:10.5194/HESS-23-3807-2019.

Mann, H. B., and Whitney, D. R. (1947). On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. https://doi.org/10.1214/aoms/1177730491 18, 50–60. doi:10.1214/AOMS/1177730491.

McGuire, K. J., McDonnell, J. J., Weiler, M., Kendall, C., McGlynn, B. L., Welker, J. M., et al. (2005). The role of topography on catchment-scale water residence time. Water Resour. Res. 41, 1–14. doi:10.1029/2004WR003657.

Miller, J. B., Frisbee, M. D., Hamilton, T. L., and Murugapiran, S. K. (2021). Recharge from glacial meltwater is critical for alpine springs and their microbiomes. Environ. Res. Lett 16, 64012. doi:10.1088/1748-9326/abf06b.

Newby, D. T., Pepper, I. L., and Maier, R. M. (2009). “Microbial Transport,” in Environmental Microbiology (Elsevier Inc.), 365–383. doi:10.1016/B978-0-12-370519-8.00019-5.

Novick, K. A., Biederman, J. A., Desai, A. R., Litvak, M. E., Moore, D. J. P., Scott, R. L., et al. (2018). The AmeriFlux network: A coalition of the willing. Agric. For. Meteorol. 249, 444–456. doi:10.1016/J.AGRFORMET.2017.10.009.

Olds, B. P., Jerde, C. L., Renshaw, M. A., Li, Y., Evans, N. T., Turner, C. R., et al. (2016). Estimating species richness using environmental DNA. Ecol. Evol. 6, 4214–4226. doi:10.1002/ece3.2186.

Oregon Water Resources Department (2021). Near Real Time Hydrographics Data, accessed 3 September 2021.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2013). The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. Nucleic Acids Res. 41, D590. doi:10.1093/nar/gks1219.

Read, D. S., Gweon, H. S., Bowes, M. J., Newbold, L. K., Field, D., Bailey, M. J., et al. (2015). Catchment-scale biogeography of riverine bacterioplankton. ISME J. 9, 516–26. doi:10.1038/ismej.2014.166.

Risley, J., Stonewall, A., and Haluska, T. (2008). Estimating flow-duration and low-flow frequency statistics for unregulated streams in Oregon: U.S. Geological Survey Scientific Investigations Report 2008-5126, 22 p.

Ruddell, B. L., and Kumar, P. (2009). Ecohydrologic process networks: 1. Identification. Water Resour. Res. 45. doi:10.1029/2008WR007279.

Savio, D., Sinclair, L., Ijaz, U. Z., Parajka, J., Reischer, G. H., Stadler, P., et al. (2015). Bacterial diversity along a 2600?km river continuum. Environ. Microbiol. 17, 4994–5007. doi:10.1111/1462-2920.12886.

Schimel, D., Hargrove, W., Hoffman, F., and MacMahon, J. (2007). NEON: A hierarchically designed national ecological network. Front. Ecol. Environ. 5, 59. doi:10.1890/1540-9295(2007)5[59:NAHDNE]2.0.CO;2.

Scott, D. W. (1979). On optimal and data-based histograms. Biometrika 66, 605–610. doi:10.1093/BIOMET/66.3.605.

Seibert, J., and McDonnell, J. J. (2002). On the dialog between experimentalist and modeler in catchment hydrology: Use of soft data for multicriteria model calibration. Water Resour. Res. 38, 23–1. doi:10.1029/2001WR000978.

Shannon, C. (1948). A Mathematical Theory of Communication. Bell Syst. Tech. J. 27, 379–423. doi:10.1002/j.1538-7305.1948.tb01338.x.

Sorensen, J. P. R., Maurice, L., Edwards, F. K., Lapworth, D. J., Read, D. S., Allen, D., et al. (2013). Using Boreholes as Windows into Groundwater Ecosystems. PLoS One 8, e70264. doi:10.1371/journal.pone.0070264.

Sugiyama, A., Masuda, S., Nagaosa, K., Tsujimura, M., and Kato, K. (2018). Tracking the direct impact of rainfall on groundwater at Mt. Fuji by multiple analyses including microbial DNA. Biogeosciences 15, 721–732. doi:10.5194/bg-15-721-2018.

Thompson, L. R., Sanders, J. G., McDonald, D., Amir, A., Ladau, J., Locey, K. J., et al. (2017). A communal catalogue reveals Earth's multiscale microbial diversity. Nature 551, 457–463. doi:10.1038/nature24621.

URycki, D. R., Good, S. P., Crump, B. C., Chadwick, J., and Jones, G. D. (2020). River Microbiome Composition Reflects Macroscale Climatic and Geomorphic Differences in Headwater Streams. Front. Water 2, 574728. doi:10.3389/frwa.2020.574728.

U.S. Geological Survey, 2016. National Water Information System data available on the World Wide Web (USGS Water Data for the Nation), accessed 9 May 2020. http://dx.doi.org/10.5066/F7P55KJN.

U.S. Geological Survey, 2017. StreamStats, Version 4. doi:10.3133/fs20173046.

3.12 Supporting Information

Table 3.S1. Abundance rank, log abundance, number of sites detected, and taxonomy of streamwater microbial amplified sequence variants (ASVs) detected in summer across 64 sites in all three of the Willamette (2017), Deschutes (2017), and John Day (2018) basins of Oregon, USA.

Abundance rank	Log abundance	Number of sites	Taxonomy[a]
1	8.4	40	d__Bacteria, p__Bacteroidota, c__Bacteroidia, o__Cytophagales, f__Spirosomaceae, g__Pseudarcicella, s__uncultured_bacterium
2	8.13	38	d__Bacteria, p__Proteobacteria, c__Gammaproteobacteria, o__Burkholderiales, f__Comamonadaceae, g__unclassified_f__Comamonadaceae, s__unclassified_f__Comamonadaceae
3	7.93	32	d__Bacteria, p__Bacteroidota, c__Bacteroidia, o__Flavobacteriales, f__Flavobacteriaceae, g__Flavobacterium, s__unclassified_g__Flavobacterium
4	7.9	40	d__Bacteria, p__Proteobacteria, c__Gammaproteobacteria, o__Burkholderiales, f__Comamonadaceae, g__Limnohabitans, s__unclassified_g__Limnohabitans
5	7.59	28	d__Bacteria, p__Proteobacteria, c__Gammaproteobacteria, o__Burkholderiales, f__Comamonadaceae, g__unclassified_f__Comamonadaceae, s__unclassified_f__Comamonadaceae
6	7.42	40	d__Bacteria, p__Proteobacteria, c__Gammaproteobacteria, o__Burkholderiales, f__Comamonadaceae, g__unclassified_f__Comamonadaceae, s__unclassified_f__Comamonadaceae
7	7.23	37	d__Bacteria, p__Bacteroidota, c__Bacteroidia, o__Chitinophagales, f__Chitinophagaceae, g__Sediminibacterium, s__unclassified_g__Sediminibacterium
8	7.15	24	d__Bacteria, p__Bacteroidota, c__Bacteroidia, o__Flavobacteriales, f__Crocinitomicaceae, g__Fluviicola, s__uncultured_Flexibacter
9	7.12	39	d__Bacteria, p__Proteobacteria, c__Gammaproteobacteria, o__Burkholderiales, f__Comamonadaceae, g__Rhodoferax, s__unclassified_g__Rhodoferax
10	6.92	18	d__Bacteria, p__Bacteroidota, c__Bacteroidia, o__Flavobacteriales, f__Flavobacteriaceae, g__Flavobacterium, s__Antarctic_bacterium
11	6.9	23	d__Bacteria, p__Proteobacteria, c__Gammaproteobacteria, o__Pseudomonadales, f__Moraxellaceae, g__Enhydrobacter, s__unclassified_g__Enhydrobacter
12	6.85	34	d__Bacteria, p__Actinobacteriota, c__Actinobacteria, o__Frankiales, f__Sporichthyaceae, g__Sporichthyaceae, s__metagenome
13	6.67	23	d__Bacteria, p__Actinobacteriota, c__Actinobacteria, o__Frankiales, f__Sporichthyaceae, g__hgcI_clade, s__unclassified_g__hgcI_clade
14	6.64	16	d__Bacteria, p__Proteobacteria, c__Gammaproteobacteria, o__Burkholderiales, f__Comamonadaceae, g__unclassified_f__Comamonadaceae, s__unclassified_f__Comamonadaceae
15	6.6	21	d__Bacteria, p__Proteobacteria, c__Gammaproteobacteria, o__Burkholderiales, f__Burkholderiaceae, g__Polynucleobacter, s__unclassified_g__Polynucleobacter
16	6.58	18	d__Bacteria, p__Verrucomicrobiota, c__Verrucomicrobiae, o__uncultured, f__uncultured, g__uncultured, s__uncultured_Opitutae
17	6.58	20	d__Bacteria, p__Bacteroidota, c__Bacteroidia, o__Flavobacteriales, f__Crocinitomicaceae, g__Fluviicola, s__unclassified_g__Fluviicola
18	6.52	23	d__Bacteria, p__Bacteroidota, c__Bacteroidia, o__Flavobacteriales, f__Flavobacteriaceae, g__Flavobacterium, s__unclassified_g__Flavobacterium

19	6.5	18	d__Bacteria, p__Proteobacteria, c__Gammaproteobacteria, o__Burkholderiales, f__Comamonadaceae, g__unclassified_f__Comamonadaceae, s__unclassified_f__Comamonadaceae
20	6.46	18	d__Bacteria, p__Bacteroidota, c__Bacteroidia, o__Flavobacteriales, f__Flavobacteriaceae, g__Flavobacterium, s__unclassified_g__Flavobacterium
21	6.45	19	d__Bacteria, p__Actinobacteriota, c__Actinobacteria, o__Frankiales, f__Sporichthyaceae, g__hgcI_clade, s__unclassified_g__hgcI_clade
22	6.43	23	d__Bacteria, p__Proteobacteria, c__Gammaproteobacteria, o__Burkholderiales, f__Methylophilaceae, g__Candidatus_Methylopumilus, s__unclassified_g__Candidatus_Methylopumilus
23	6.42	21	d__Bacteria, p__Actinobacteriota, c__Actinobacteria, o__Frankiales, f__Sporichthyaceae, g__hgcI_clade, s__unclassified_g__hgcI_clade
24	6.42	20	d__Bacteria, p__Bacteroidota, c__Bacteroidia, o__Chitinophagales, f__Chitinophagaceae, g__Sediminibacterium, s__unclassified_g__Sediminibacterium
25	6.39	31	d__Bacteria, p__Bacteroidota, c__Bacteroidia, o__Sphingobacteriales, f__NS11-12_marine_group, g__NS11-12_marine_group, s__uncultured_Flexibacteraceae
26	6.3	11	d__Bacteria, p__Proteobacteria, c__Gammaproteobacteria, o__Burkholderiales, f__Burkholderiaceae, g__Polynucleobacter, s__Polynucleobacter_cosmopolitanus
27	6.29	17	d__Bacteria, p__Proteobacteria, c__Gammaproteobacteria, o__Methylococcales, f__Methylomonadaceae, g__Methylobacter, s__unclassified_g__Methylobacter
28	6.26	27	d__Bacteria, p__Proteobacteria, c__Gammaproteobacteria, o__Cellvibrionales, f__Cellvibrionaceae, g__Cellvibrio, s__uncultured_Cellvibrio
29	6.25	20	d__Bacteria, p__Verrucomicrobiota, c__Verrucomicrobiae, o__Pedosphaerales, f__Pedosphaeraceae, g__SH3-11, s__uncultured_bacterium
30	6.2	29	d__Bacteria, p__Proteobacteria, c__Gammaproteobacteria, o__Aeromonadales, f__Aeromonadaceae, g__Aeromonas, s__unclassified_g__Aeromonas
31	6.19	19	d__Bacteria, p__Proteobacteria, c__Gammaproteobacteria, o__Burkholderiales, f__Methylophilaceae, g__Methylotenera, s__unclassified_g__Methylotenera
32	5.97	20	d__Bacteria, p__Bacteroidota, c__Bacteroidia, o__Chitinophagales, f__Chitinophagaceae, g__Sediminibacterium, s__unclassified_g__Sediminibacterium
33	5.94	27	d__Bacteria, p__Proteobacteria, c__Gammaproteobacteria, o__Burkholderiales, f__Chitinibacteraceae, g__Deefgea, s__Deefgea_sp.
34	5.9	20	d__Bacteria, p__Actinobacteriota, c__Actinobacteria, o__Frankiales, f__Sporichthyaceae, g__hgcI_clade, s__unclassified_g__hgcI_clade
35	5.9	21	d__Bacteria, p__Verrucomicrobiota, c__Verrucomicrobiae, o__Pedosphaerales, f__Pedosphaeraceae, g__uncultured, s__uncultured_bacterium
36	5.85	23	d__Bacteria, p__Proteobacteria, c__Gammaproteobacteria, o__Alteromonadales, f__Alteromonadaceae, g__Rheinheimera, s__unclassified_g__Rheinheimera
37	5.75	15	d__Bacteria, p__Bacteroidota, c__Bacteroidia, o__Cytophagales, f__Spirosomaceae, g__Arcicella, s__uncultured_bacterium
38	5.72	27	d__Bacteria, p__Proteobacteria, c__Alphaproteobacteria, o__SAR11_clade, f__Clade_III, g__Clade_III, s__Candidatus_Fonsibacter
39	5.71	16	d__Bacteria, p__Proteobacteria, c__Gammaproteobacteria, o__Burkholderiales, f__Comamonadaceae, g__Rhodoferax, s__unclassified_g__Rhodoferax
40	5.64	14	d__Bacteria, p__Bacteroidota, c__Bacteroidia, o__Chitinophagales, f__Chitinophagaceae, g__Edaphobaculum, s__uncultured_bacterium

41	5.62	16	d__Bacteria, p__Bacteroidota, c__Bacteroidia, o__Cytophagales, f__Spirosomaceae, g__Flectobacillus, s__Flectobacillus_fontis
42	5.58	23	d__Bacteria, p__Bacteroidota, c__Bacteroidia, o__Flavobacteriales, f__NS9_marine_group, g__NS9_marine_group, s__unclassified_g__NS9_marine_group
43	5.56	21	d__Bacteria, p__Bacteroidota, c__Bacteroidia, o__Chitinophagales, f__Saprospiraceae, g__uncultured, s__uncultured_bacterium
44	5.53	10	d__Bacteria, p__Bacteroidota, c__Bacteroidia, o__Sphingobacteriales, f__NS11-12_marine_group, g__NS11-12_marine_group, s__uncultured_Flexibacteraceae
45	5.5	28	d__Bacteria, p__Proteobacteria, c__Gammaproteobacteria, o__Xanthomonadales, f__Xanthomonadaceae, g__Arenimonas, s__unclassified_g__Arenimonas
46	5.46	21	d__Bacteria, p__Actinobacteriota, c__Actinobacteria, o__Frankiales, f__Sporichthyaceae, g__hgcI_clade, s__uncultured_actinobacterium
47	5.43	7	d__Bacteria, p__Proteobacteria, c__Gammaproteobacteria, o__Enterobacterales, f__Yersiniaceae, g__unclassified_f__Yersiniaceae, s__unclassified_f__Yersiniaceae
48	5.42	11	d__Bacteria, p__Proteobacteria, c__Gammaproteobacteria, o__Burkholderiales, f__Comamonadaceae, g__unclassified_f__Comamonadaceae, s__unclassified_f__Comamonadaceae
49	5.41	8	d__Bacteria, p__Cyanobacteria, c__Cyanobacteriia, o__unclassified_c__Cyanobacteriia, f__unclassified_c__Cyanobacteriia, g__unclassified_c__Cyanobacteriia, s__unclassified_c__Cyanobacteriia
50	5.38	12	d__Bacteria, p__Proteobacteria, c__Alphaproteobacteria, o__Rhodobacterales, f__Rhodobacteraceae, g__unclassified_f__Rhodobacteraceae, s__unclassified_f__Rhodobacteraceae
51	5.38	9	d__Bacteria, p__Bacteroidota, c__Bacteroidia, o__Cytophagales, f__Cyclobacteriaceae, g__Algoriphagus, s__unclassified_g__Algoriphagus
52	5.36	18	d__Bacteria, p__Proteobacteria, c__Gammaproteobacteria, o__Cellvibrionales, f__Halieaceae, g__OM60(NOR5)_clade, s__unclassified_g__OM60(NOR5)_clade
53	5.35	10	d__Bacteria, p__Proteobacteria, c__Gammaproteobacteria, o__Burkholderiales, f__Comamonadaceae, g__unclassified_f__Comamonadaceae, s__unclassified_f__Comamonadaceae
54	5.26	16	d__Bacteria, p__Verrucomicrobiota, c__Verrucomicrobiae, o__Verrucomicrobiales, f__Rubritaleaceae, g__Luteolibacter, s__unclassified_g__Luteolibacter
55	5.22	5	d__Bacteria, p__Proteobacteria, c__Gammaproteobacteria, o__Burkholderiales, f__Comamonadaceae, g__unclassified_f__Comamonadaceae, s__unclassified_f__Comamonadaceae
56	5.19	12	d__Bacteria, p__Cyanobacteria, c__Cyanobacteriia, o__Cyanobacteriales, f__Xenococcaceae, g__Pleurocapsa_PCC-7319, s__unclassified_g__Pleurocapsa_PCC-7319
57	5.18	8	d__Bacteria, p__Proteobacteria, c__Gammaproteobacteria, o__Burkholderiales, f__Comamonadaceae, g__Hydrogenophaga, s__unclassified_g__Hydrogenophaga
58	5.17	9	d__Bacteria, p__Proteobacteria, c__Alphaproteobacteria, o__Sphingomonadales, f__Sphingomonadaceae, g__Sphingorhabdus, s__unclassified_g__Sphingorhabdus
59	5.15	14	d__Bacteria, p__Verrucomicrobiota, c__Verrucomicrobiae, o__Pedosphaerales, f__Pedosphaeraceae, g__uncultured, s__metagenome
60	5.15	17	d__Bacteria, p__Bacteroidota, c__Bacteroidia, o__Cytophagales, f__Cyclobacteriaceae, g__Algoriphagus, s__uncultured_Hongiella
61	5.15	13	d__Bacteria, p__Bacteroidota, c__Bacteroidia, o__Cytophagales, f__Cytophagaceae, g__Cytophaga, s__unclassified_g__Cytophaga
62	5.12	10	d__Bacteria, p__Bacteroidota, c__Bacteroidia, o__Sphingobacteriales, f__env.OPS_17, g__env.OPS_17, s__uncultured_bacterium

63	5.09	8	d__Bacteria, p__Verrucomicrobiota, c__Verrucomicrobiae, o__uncultured, f__uncultured, g__uncultured, s__unclassified_g__uncultured
64	5.08	8	d__Bacteria, p__Proteobacteria, c__Gammaproteobacteria, o__Burkholderiales, f__Comamonadaceae, g__uncultured, s__unclassified_g__uncultured
65	5.06	10	d__Bacteria, p__Proteobacteria, c__Gammaproteobacteria, o__Burkholderiales, f__Burkholderiaceae, g__Polynucleobacter, s__uncultured_bacterium
66	5.06	14	d__Bacteria, p__Proteobacteria, c__Gammaproteobacteria, o__Cellvibrionales, f__Halieaceae, g__OM60(NOR5)_clade, s__unclassified_g__OM60(NOR5)_clade
67	5.03	8	d__Bacteria, p__Proteobacteria, c__Gammaproteobacteria, o__Burkholderiales, f__Comamonadaceae, g__Rhizobacter, s__unclassified_g__Rhizobacter
68	5.02	13	d__Bacteria, p__Proteobacteria, c__Gammaproteobacteria, o__Burkholderiales, f__Oxalobacteraceae, g__CM1G08, s__unclassified_g__CM1G08
69	4.96	20	d__Bacteria, p__Bacteroidota, c__Bacteroidia, o__Cytophagales, f__Spirosomaceae, g__Lacihabitans, s__uncultured_bacterium
70	4.94	18	d__Bacteria, p__unclassified_d__Bacteria, c__unclassified_d__Bacteria, o__unclassified_d__Bacteria, f__unclassified_d__Bacteria, g__unclassified_d__Bacteria, s__unclassified_d__Bacteria
71	4.89	13	d__Bacteria, p__Proteobacteria, c__Gammaproteobacteria, o__Burkholderiales, f__Comamonadaceae, g__Polaromonas, s__unclassified_g__Polaromonas
72	4.88	9	d__Bacteria, p__Proteobacteria, c__Gammaproteobacteria, o__Burkholderiales, f__Comamonadaceae, g__unclassified_f__Comamonadaceae, s__unclassified_f__Comamonadaceae
73	4.84	12	d__Bacteria, p__Bacteroidota, c__Bacteroidia, o__Flavobacteriales, f__Crocinitomicaceae, g__Fluviicola, s__unclassified_g__Fluviicola
74	4.8	8	d__Bacteria, p__Proteobacteria, c__Gammaproteobacteria, o__Burkholderiales, f__Methylophilaceae, g__unclassified_f__Methylophilaceae, s__unclassified_f__Methylophilaceae
75	4.76	7	d__Bacteria, p__Verrucomicrobiota, c__Verrucomicrobiae, o__Opitutales, f__Opitutaceae, g__Lacunisphaera, s__unclassified_g__Lacunisphaera
76	4.71	12	d__Bacteria, p__Bacteroidota, c__Bacteroidia, o__Flavobacteriales, f__Crocinitomicaceae, g__Fluviicola, s__unclassified_g__Fluviicola
77	4.64	9	d__Bacteria, p__Bacteroidota, c__Bacteroidia, o__Chitinophagales, f__Saprospiraceae, g__uncultured, s__uncultured_bacterium
78	4.61	15	d__Bacteria, p__Proteobacteria, c__Alphaproteobacteria, o__Rhizobiales, f__Rhizobiales_Incertae_Sedis, g__uncultured, s__metagenome
79	4.58	13	d__Bacteria, p__Bacteroidota, c__Bacteroidia, o__Bacteroidales, f__Prolixibacteraceae, g__uncultured, s__uncultured_prokaryote
80	4.57	18	d__Bacteria, p__Bacteroidota, c__Bacteroidia, o__Sphingobacteriales, f__env.OPS_17, g__env.OPS_17, s__unclassified_g__env.OPS_17
81	4.53	10	d__Bacteria, p__Campilobacterota, c__Campylobacteria, o__Campylobacterales, f__Arcobacteraceae, g__Pseudarcobacter, s__uncultured_bacterium
82	4.47	11	d__Bacteria, p__Bacteroidota, c__Bacteroidia, o__Sphingobacteriales, f__AKYH767, g__AKYH767, s__uncultured_Sphingobacteriales
83	4.45	17	d__Bacteria, p__Bacteroidota, c__Bacteroidia, o__Chitinophagales, f__Saprospiraceae, g__uncultured, s__uncultured_bacterium
84	4.41	13	d__Bacteria, p__Bacteroidota, c__Bacteroidia, o__Cytophagales, f__Spirosomaceae, g__Emticicia, s__unclassified_g__Emticicia

85	4.37	12	d__Bacteria, p__Bacteroidota, c__Bacteroidia, o__Chitinophagales, f__Saprospiraceae, g__uncultured, s__uncultured_bacterium
86	4.34	10	d__Bacteria, p__Verrucomicrobiota, c__Verrucomicrobiae, o__Verrucomicrobiales, f__Rubritaleaceae, g__Luteolibacter, s__unclassified_g__Luteolibacter
87	4.33	7	d__Bacteria, p__Verrucomicrobiota, c__Verrucomicrobiae, o__Verrucomicrobiales, f__Rubritaleaceae, g__Luteolibacter, s__unclassified_g__Luteolibacter
88	4.23	12	d__Bacteria, p__Actinobacteriota, c__Actinobacteria, o__Frankiales, f__Sporichthyaceae, g__Candidatus_Planktophila, s__unclassified_g__Candidatus_Planktophila
89	4.23	6	d__Bacteria, p__Verrucomicrobiota, c__Verrucomicrobiae, o__Pedosphaerales, f__Pedosphaeraceae, g__uncultured, s__uncultured_bacterium
90	4.19	11	d__Bacteria, p__Proteobacteria, c__Gammaproteobacteria, o__Cellvibrionales, f__Cellvibrionaceae, g__Cellvibrio, s__unclassified_g__Cellvibrio
91	4.13	10	d__Bacteria, p__Proteobacteria, c__Gammaproteobacteria, o__B2M28, f__B2M28, g__B2M28, s__unclassified_g__B2M28
92	4.11	10	d__Bacteria, p__Cyanobacteria, c__Cyanobacteriia, o__Leptolyngbyales, f__Leptolyngbyaceae, g__uncultured, s__uncultured_cyanobacterium
93	4.09	11	d__Bacteria, p__Proteobacteria, c__Alphaproteobacteria, o__Rhodobacterales, f__Rhodobacteraceae, g__Rhodobacter, s__uncultured_Alphaproteobacteria
94	4.06	11	d__Bacteria, p__Bacteroidota, c__Bacteroidia, o__Cytophagales, f__Microscillaceae, g__OLB12, s__uncultured_bacterium
95	3.99	10	d__Bacteria, p__Planctomycetota, c__Planctomycetes, o__Pirellulales, f__Pirellulaceae, g__Pirellula, s__unclassified_g__Pirellula
96	3.95	15	d__Bacteria, p__Bacteroidota, c__Bacteroidia, o__Cytophagales, f__Spirosomaceae, g__Emticicia, s__unclassified_g__Emticicia
97	3.93	12	d__Bacteria, p__Bacteroidota, c__Bacteroidia, o__Chitinophagales, f__Saprospiraceae, g__uncultured, s__uncultured_bacterium
98	3.91	9	d__Archaea, p__Crenarchaeota, c__Nitrososphaeria, o__Nitrosopumilales, f__Nitrosopumilaceae, g__Nitrosarchaeum, s__uncultured_archaeon
99	3.91	12	d__Bacteria, p__Fibrobacterota, c__Fibrobacteria, o__Fibrobacterales, f__unclassified_o__Fibrobacterales, g__unclassified_o__Fibrobacterales, s__unclassified_o__Fibrobacterales
100	3.76	11	d__Bacteria, p__Desulfobacterota, c__Desulfuromonadia, o__Bradymonadales, f__Bradymonadales, g__Bradymonadales, s__uncultured_bacterium
101	3.66	9	d__Bacteria, p__Proteobacteria, c__Alphaproteobacteria, o__Sphingomonadales, f__Sphingomonadaceae, g__Sphingomonas, s__unclassified_g__Sphingomonas
102	3.22	11	d__Bacteria, p__Proteobacteria, c__Gammaproteobacteria, o__Cellvibrionales, f__Spongiibacteraceae, g__BD1-7_clade, s__uncultured_bacterium

[a] d – Domain; p – Phylum; c – Class; o – Order; g – Genus; s - species

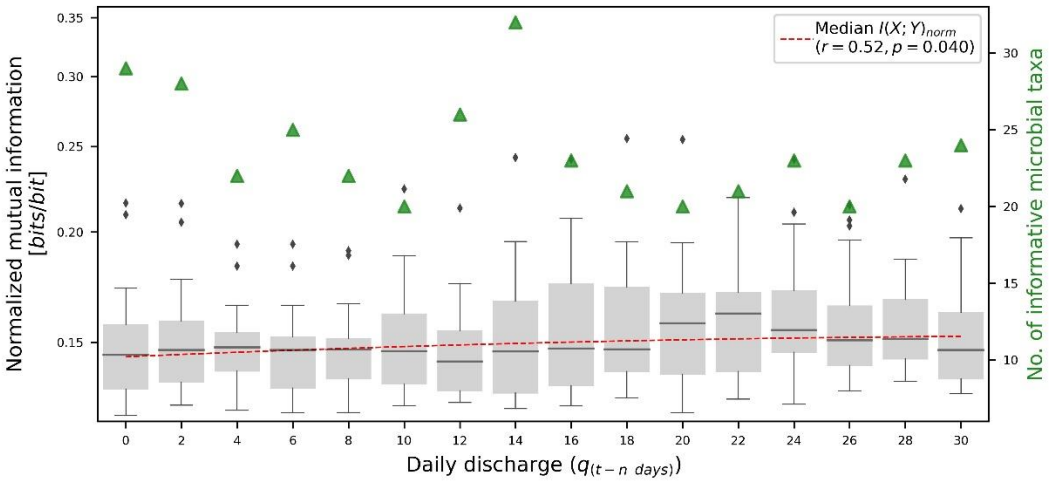

Figure 3.S1. Median value of normalized mutual information shared between informative streamwater microbial amplified sequence variants (ASVs) and daily mean discharge per unit area at different time lags ($q_{(t-n\ days)}$) for study streams in Oregon, USA. Boxes show medians and interquartile ranges; whiskers show values within 1.5 times the interquartile range. Green triangles indicate the number of ASVs with statistically significant mutual information (99% confidence). Dashed red line shows best fit curve between time lag and median value of normalized mutual information, with Pearson correlation (*r*) and *p*-value indicated in legend. We observed no significant relationship between time lag and number of informative ASVs. We collected microbial DNA samples in summer in the Willamette (2017), Deschutes (2017), and John Day (2018) basins.

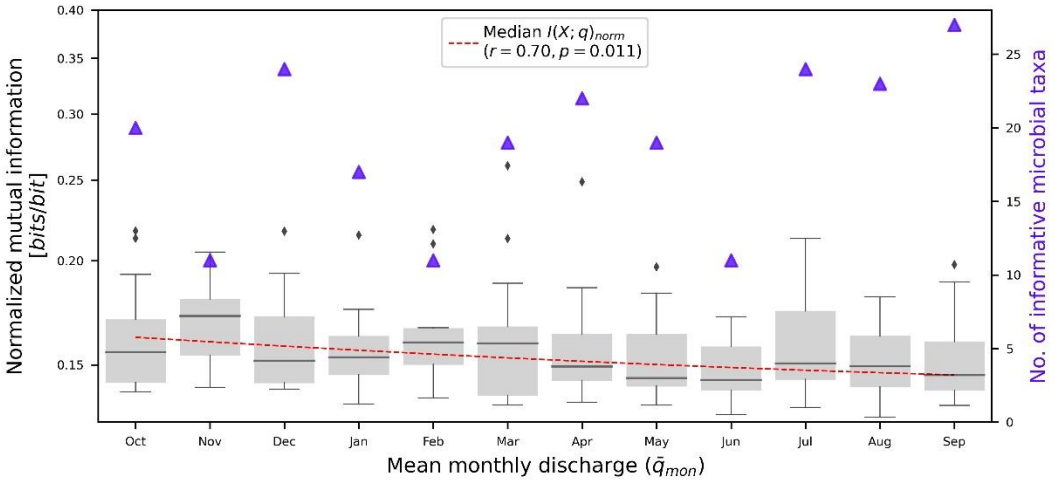

Figure 3.S2. Median value of normalized mutual information shared between informative unique streamwater microbial amplified sequence variants (ASVs) and mean monthly discharge per unit area of study streams in Oregon, USA. Boxes show medians and interquartile ranges; whiskers show values within 1.5 times the interquartile range. Purple triangles indicate the number of ASVs with statistically significant mutual information (99% confidence). Dashed red line shows best fit curve between month and median value of normalized mutual information, with Pearson correlation (*r*) and *p*-value indicated in legend. We observed no significant relationship between month and number of informative ASVs. We collected microbial DNA samples in July and August in the Willamette (2017), Deschutes (2017), and John Day (2018) basins.

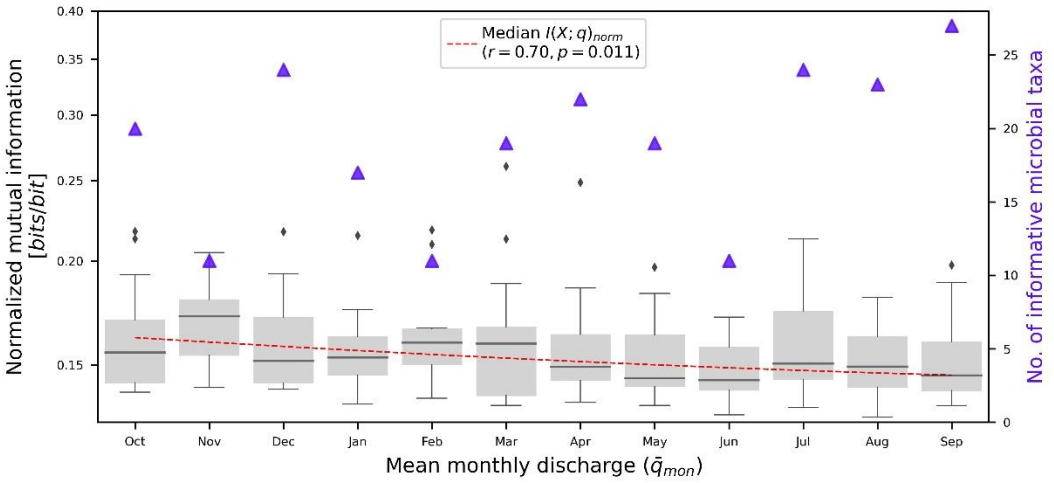

Figure 3.S3. Normalized mutual information shared between unique streamwater microbial amplified sequence variants (ASVs) and seasonal high and low flow durations per unit area ($q_{P,s}$ for P = 5- and 95-percent exceedance probability for seasons s = fall [*OND*], winter [*JFM*], spring [*AMJ*], summer [*JAS*], and annually [*Ann*]) at study streams in Oregon, USA. Boxes show medians and interquartile ranges; whiskers show values within 1.5 times the interquartile range. Blue triangles indicate the number of ASVs with statistically significant mutual information (99% confidence) for high flows (dark) and low flows (light). We collected microbial DNA samples in summer in the Willamette (2017), Deschutes (2017), and John Day (2018) basins.

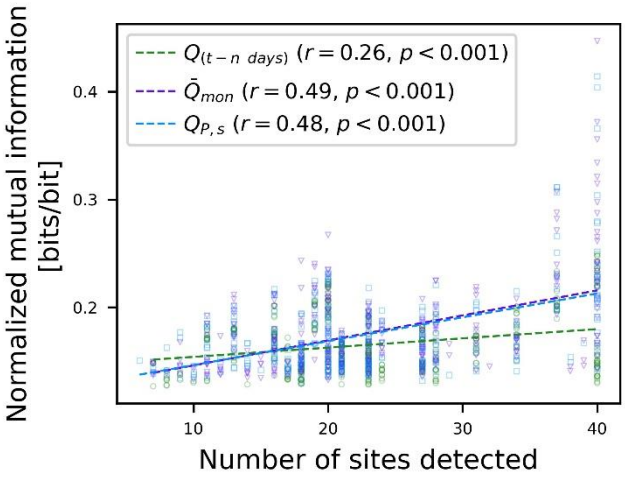

Figure 3.S4. Mutual information shared between hydrologic metrics and streamwater microbial taxa versus the number of sites at which taxa were detected in streams across Oregon, USA. We collected microbial DNA samples in summer in the Willamette (2017), Deschutes (2017), and John Day (2018) basins. Hydrologic metrics include daily discharge at time lags up to n = 30 days prior to DNA sample collection ($Q_{(t-n\ days)}$; green circles), mean monthly discharge ($\bar{Q}_{mon}$; purple triangles), and seasonal high and low flow durations ($Q_{P,s}$; 5- and 95-percent exceedance probability for all seasons and annually; blue squares). Legend shows Pearson's correlation (r) and p-value of the linear regression.

SHIFTS IN STREAMWATER MICROBIAL DIVERSITY TRACK STORM HYDROGRAPH DYNAMICS

Dawn R URycki, Stephen P Good, Byron C Crump, J Renée Brooks and Natalie Ceperley

4 Shifts in Streamwater Microbial Diversity Track Storm Hydrograph Dynamics

4.1 Abstract

A thorough understanding of watershed response to precipitation events is critical as our climate shifts to produce increasingly extreme precipitation and hydrologic events. Hydrogeochemical tools, such as stable isotope analysis, are a common approach for tracking precipitation and identifying the source of surface water in catchments, however in some cases the stable isotope signature may be difficult to interpret, as it integrates the multitude of complex processes involved in streamflow generation and water storage into a one- or two-dimensional datapoint. In contrast, aquatic microbial communities in streams are composed of thousands of taxa, originating from a variety of sources, including groundwater, sediment, stable upstream communities, and the upslope terrestrial environment. In this study, we explore the dynamics of the streamwater microbial community response to a precipitation event on the Marys River in Oregon, USA. We collected daily DNA samples from the Marys River before, during, and after a large, isolated precipitation event. Stable water isotope ratios ($\delta^{18}O$ and $\delta^{2}H$) were also analyzed. Though isotopes signatures exhibited little variation, prior work in the catchment suggests that distinct pre-event, early-event, and late-event water sources are visible. DNA samples were translated into the relative abundance of different distinct taxa (620 in total) using 16S amplicon sequencing. Cluster analysis of the microbial composition similarly reveals coherent pre-event, early-event, and late-event microbial communities. Shifts in microbial diversity reflect changes in discharge over the course of the storm, and abundance-discharge relationships (analogous to a concentration-discharge geochemical analysis) reveal that some taxa are mobilized and others diluted over the course of the event. This study provides an approach for integrating information from DNA suspended in the water column into an investigation of a hydrologic response that incorporates tools from both hydrology and microbiology and demonstrates that microbial DNA is useful not only as an indicator of biodiversity but also as an innovative hydrologic tracer.

4.2 Introduction

Managing increasing pressure on water resources and competing demands between ecosystems and societal needs requires detailed understanding of the response of water resources to perturbations and climate change effects. Whereas some watersheds are resilient to dramatic events, others are flashier and prone to more extreme flows. The way in which water resources respond to precipitation events depends on event characteristics, including precipitation volume and intensity, as well as the complex dynamics of water storage and release in the catchment basin.

Geochemical tracer analyses are commonly employed to analyze streamflow response to precipitation events. Most notable among the variety of geochemical tracers employed in hydrologic studies are stable water isotopes ^{2}H and ^{18}O. Stable isotopes are usually based on a mass balance approach (Pinder & Jones, 1969) and have led to transformative insights in catchment hydrology, including the observation that pre-event "old" water dominates streamflow during precipitation events in many systems (Sklash et al., 1986; Klaus and McDonnell, 2013), although the mechanisms driving a predominately old water stream response remain unknown. Furthermore, isotopic signatures in some systems may be difficult to discern. For example, in the Marys River Basin in Oregon, USA, Brooks et al., (2012) found that isotopic variation between tributary streams was explained primarily by differences in mean basin elevation among the tributaries. However, in a subsequent study in the same basin, Nickolas et al. (2017) observed isotopic ratios that were considerably more enriched than expected and did not demonstrate an elevation signal; differences in lithology have been proposed as an explanation for the unusual isotopic signature, but the question remains unresolved.

Water transport through catchments is governed by characteristics of the subsurface, which are not only physically and chemically complex, but also biologically complex. Whereas stable isotopes reflect some physical and chemical properties of the subsurface, biological interactions are not adequately captured by these geochemical tracers. New approaches are necessary to gain a more complete understanding of the dynamic and interrelated processes shaping hydrologic regimes.

Aquatic microbial communities in streams are composed of thousands of taxa from a variety of sources, including groundwater, sediment, stable upstream communities, and the upslope terrestrial environment. Streamwater microbial community composition is also strongly related to stream discharge (Crump and Hobbie, 2005; Doherty et al., 2017), as well as other hydrologic properties related to water residence time such as cumulative river distance upstream (Read et al., 2015), river kilometer, and catchment size (Savio et al., 2015), and responds dynamically to changes in environmental conditions (Thompson et al., 2017). Given the diversity of microbial community constituents and their sources, and the demonstrated relationships between microbial communities and watershed hydrology, we sought to investigate the precipitation event-scale dynamics between streamwater microbial communities and stream discharge.

We collected daily DNA samples over the course of an isolated precipitation event on the Marys River in Oregon, USA, starting prior to the onset of precipitation and continuing through the duration of the stream response (Fig. 4.1). We assess microbial community diversity in relation to the stream discharge event response. We compare the microbial community response to a common set of hydrologic tracers, stable isotope ratios of ^{2}H and ^{18}O. We furthermore characterize microbial taxonomic groups as mobilized or diluted based on whether abundance is correlated or anticorrelated, respectively, with stream discharge. While advances in technology have allowed for microbiome analyses in an expanding array of research areas including, more recently, hydrology, we are not aware of any microbial study with this temporal structure. Results of this analysis carry important implications for both hydrology and microbiology, as well as many fields at the intersection of the two.

4.3 Results

4.3.1 Microbial Community Diversity

Streamwater microbial diversity demonstrates a strong response to stream event flow (Fig 4.2a). Alpha diversity closely tracks stream discharge, in terms of both richness (the number of unique taxa) and Shannon index, which reflects richness as well as evenness. Richness rises steeply from a mean of 80.7 (SD = 10.8) taxa in the three days leading up to and including the first day of precipitation on 9 October, rising

steeply with discharge to an event maximum of 178 taxa on 14 October, and falling again to a post-event mean of 45.3 (SD = 2.5) taxa in the final three samples 21-25 October (Fig 4.2a). Shannon index follows a similar pattern, rising from a mean of 4.26 (SD = 0.12) 7-9 October, to a maximum of 5.26 on 14 October, and falling to a post-event mean of 3.71 (SD = 0.03) 21-25 October (Fig 4.2a). Both metrics of alpha diversity are correlated with one-day lagged discharge (richness: Pearson's $\rho = 0.45, p = 0.07$; Shannon index $\rho = 0.56, p = 0.02$). The relationship is even stronger when the singularly high value of alpha diversity in the first sample on 6 October is not considered (richness: $\rho = 0.73, p = 0.001$; Shannon index $\rho = 0.77, p < 0.001$). Alpha diversity is not correlated with precipitation or with one-day or two-day lagged precipitation.

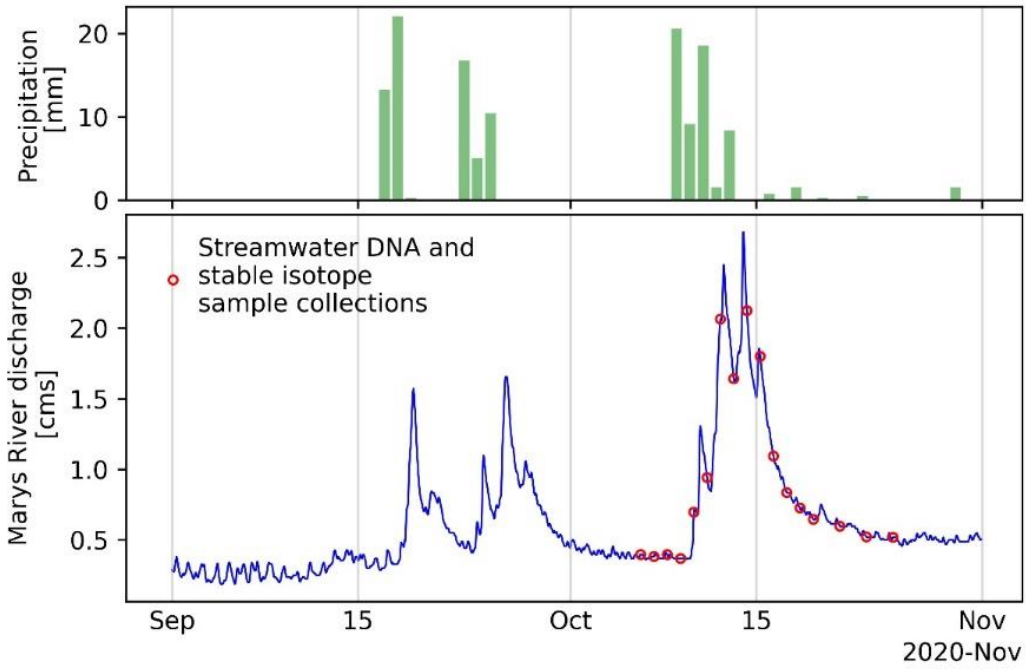

Figure 4.1. Daily precipitation, discharge, and streamwater DNA and stable water isotope (^{18}O, ^{2}H) sample collections on the Marys River before and after an isolated precipitation event in October 2020. Sampling began on 6 October; precipitation began on 9 October.

In addition to changes in microbial alpha diversity with storm discharge, the microbial communities prior to the storm distinctly differ from those after the storm. Beta diversity, as measured by Bray-Curtis (BC) dissimilarity, is higher among pairs

of pre- and post-event samples than among pairs of pre-event samples taken 6-9 October (Fig 4.2b). Between pairs of pre-and post-event samples, mean BC dissimilarity is 0.27 (SD = 0.04), whereas BC among pre-event samples is, on average, 0.13 (SD = 0.02). As with alpha diversity, dissimilarity from the pre-event community tracks the storm discharge hydrograph, with the most dissimilar community, relative to the pre-event communities, occurring with peak discharge on 13 October (BC dissimilarity = 0.39, SD = 0.01). The shift away from the pre-event microbial community can be observed in a principle coordinates analysis (Fig. 4.4), where a storm-discharge-associated community differs most drastically on 13 October (day of year [doy] 287) from several similar pre-event samples (doy 280-283); although the community is less dynamic after the event (doy >292), the community distinctly differs from the pre-event community.

The isotopic signature of the stream changes little over the course of the storm event and stream response, despite highly enriched precipitation inputs (Fig. 4.2c). Mean pre-event deuterium ratio (δ^2H) is -53.9 (SD = 0.2) permil, which increased to an event maximum of δ^2H = -51.5 permil with peak stream discharge on 13 October. Aggregated precipitation ratios are much higher; δ^2H = -38.7 permil for the two-week period ending on 13 October and δ^2H = -23.3 permil for the following two-week period (Fig. 4.2c). Similarly, mean pre-event ^{18}O ratio ($\delta^{18}O$) is -7.9 (SD = 0.0) permil and rises to an event maximum of $\delta^{18}O$ = -7.7 permil. Aggregated precipitation ratios are $\delta^{18}O$ = -6.3 permil and $\delta^{18}O$ = -5.0 permil for the periods ending on 13 October and beginning 14 October, respectively (Fig 2c).

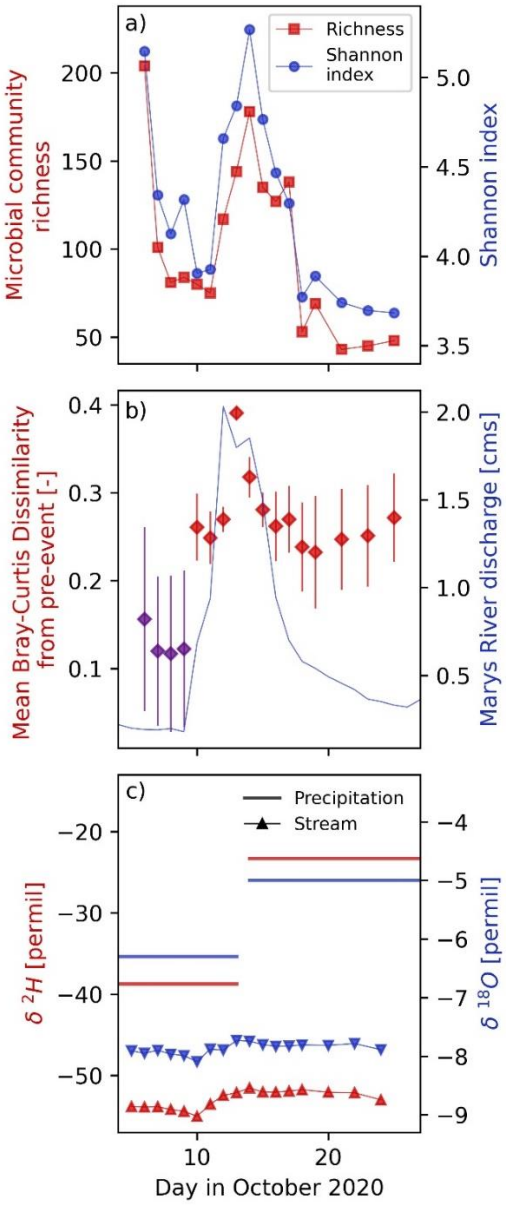

Figure 4.2. Responses of streamwater microbial community diversity and stable isotopes of water to an isolated precipitation event beginning 9 October 2020 on the Marys River, Oregon, USA. a) Microbial community alpha diversity, including taxonomic richness (number of unique amplified sequence variants; red squares) and Shannon index (blue circles). b) Marys River stream discharge [cms] and mean microbial community Bray-Curtis dissimilarity (diamonds) from pre-event samples (purple). Error bars indicate one standard deviation. c) Stable isotope ratios δ^2H (red) and $\delta^{18}O$ (blue) measured in the stream (triangles) and in approximately 2-week aggregated precipitation (lines)**.**

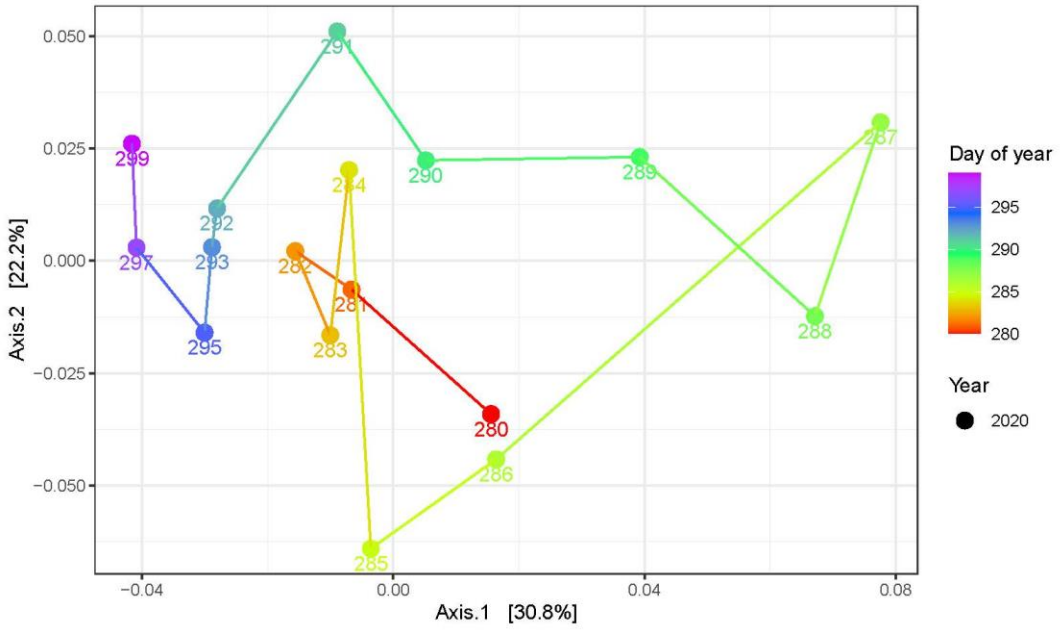

Figure 4.3 Principle coordinates analysis of streamwater microbial diversity over the course of a precipitation event and associated stream response 6-25 October 2020 on the Marys River, Oregon, USA. Day of year 280 corresponds to 6 October 2020.

4.3.2 Abundance – Discharge Relationships

We explored whether microbial taxonomic abundance was related to discharge dynamics by conducting an abundance-discharge analysis, analogous to a concentration-discharge (C-Q) analysis commonly employed in chemical tracer studies. Of the 68,000 sequences belonging to 620 unique taxa we analyzed, we were able to characterize 65,760 sequences from 111 taxa as mobilized (7,724 sequences across 17 unique taxa), diluted (41,936 sequences across 14 unique taxa), or static (16,550 sequences across 80 unique taxa) in response to changes in stream discharge (Fig. 4.4). Diluted microbes compose the greatest proportion of the community on all sample days, accounting for an average of 61% (SD = 11%) of microbial sequences detected on a given day. The fraction of the streamwater microbial community comprised of diluted taxa is inversely correlated with stream discharge (Pearson $\rho = -0.62, p = 0.01$), with a maximum composition of 76% of the community on 25 October, following the hydrograph recession, and a minimum of 42% on 13 October, the day of peak event discharge (Fig. 4.4). Diluted sequences are primarily classified as Gammaproteobacteria, accounting for an average of 34% of the overall stream

microbial community, Bacteroidota (21%), and Actinobacteriota (7%). Diluted Verrucomicrobiota and sequences classified as Other composed <1% of the microbial community.

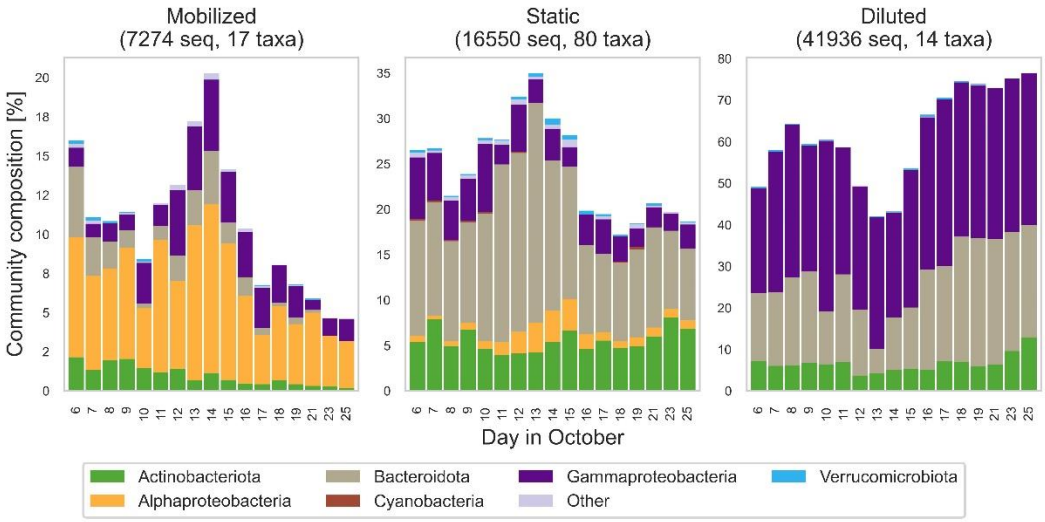

Figure 4.4. Fraction of streamwater microbial sequences that were mobilized, static, and diluted with stream discharge dynamics in response to a precipitation event beginning 6 October 2020 on the Marys River, Oregon, USA. Fraction is of the total abundance (number of sequences [n]) identified in at least three samples and that were positively correlated (mobilized), negatively correlated (diluted), or not correlated (static) with stream discharge ($p < 0.1$). Number of taxa represented

Mobilized microbes compose a much smaller fraction of the streamwater microbial community, accounting for an average of 11% (SD = 4.5%) of microbial sequences detected on a given day (Fig 4.4). The fraction of the microbial community composed of mobilized taxa is correlated with stream discharge ($\rho = 0.55, p = 0.02$), with a maximum composition of 20% of the community on 14 October and a minimum of 5% on 25 October. Mobilized sequences are primarily classified as Alphaproteobacteria, accounting for an average of 6% of the overall stream microbial community, Gammaproteobacteria (2%), Bacteroidota (1%); Actinobacteriota, Verrucomicrobiota, and Other each account for <1% of the community.

Static sequences are those that were not statistically correlated with discharge, and they account for a mean of 24% (SD = 5%) of the streamwater microbial community on a given day (Fig 4.4). The fraction of the overall microbial community composed of static sequences was the group most strongly correlated with discharge ($\rho = 0.73, p < 0.001$), with a peak composition of 35% on 13 October and the nadir

(17%) on 18 October. Static sequences are primarily classified as Bacteroidota, accounting for an average of 13% of the overall microbial community, Actinobacteriota (6%), Gammaproteobacteria (4%), and Alphaproteobacteria (1%); Other, Verrucomicrobiota, and Cyanobacteria each account for <1% of the community.

The fraction of sequences characterized as mobilized, diluted, or static varies within each of the phylogentic groups, but each group is generally dominated by either diluted or mobilized sequences (Fig. 4.5). For example, 84% of the 27,621 sequences classified as Gammaproteobacteria are characterized as diluted, whereas only 5% are mobilized (9% are static and 2% remain uncharacterized). Similar patterns are observed among Bacteroidota, in which 58% of 5,240 sequences are diluted and only 4% are mobilized (36% are static), and Actinobacteriota, in which 50% of 9,012 sequences are diluted and 7% are mobilized (42% are static). Alphaproteobacteria, on the other hand, is dominated by mobilized sequences: 78% of sequences are mobilized but none are diluted (19% are static and 3% are uncharacterized).

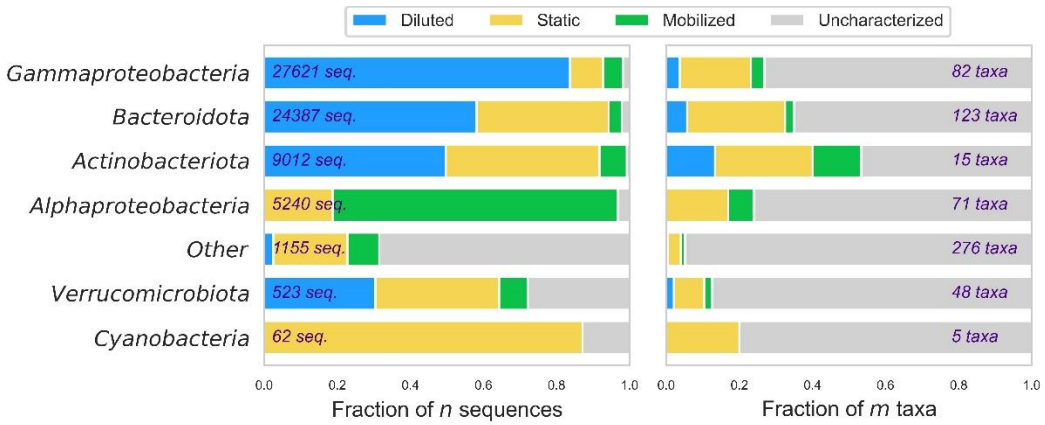

Figure 4.5. Microbial taxonomic groups identified as mobilized, static, and diluted, as a fraction of the total abundance of each amplified sequence variant over the sampling period (left) and as a fraction of the number of unique taxa (right), in response to a precipitation event beginning 6 October 2020 on the Marys River, Oregon, USA. Diluted, static, and mobilized ASVs are those that were identified in at least three samples and were negatively correlated, not correlated, or positively correlated with stream discharge ($p < 0.1$), respectively.

4.4 Discussion

The streamwater microbial community exhibited a rich and dynamic response to discharge dynamics following an isolated precipitation event on the Marys River, in

contrast to the dampened, bivariate response in the isotopic signature of the stream. Microbial community diversity was tightly coupled to the storm hydrograph, with number of unique microbial taxa increasing more than twofold as stream discharge increased, then falling to levels lower than those prior to the event (Fig. 4.2a). Streamwater istopic ratios ^{2}H and ^{18}O, on the other hand, registered only a modest response to this major precipitation event, especially considering the much higher ratios in the precipitation during the same period.

Even after the stream discharge and microbial diversity recede, the post-event community remained distinctly different than the pre-event community (Fig. 4.2b). A principle coordinates analysis illustrates how the community changes most drastically with the rising and falling limbs of the hydrograph while transitioning to a final community that is less dissimilar each day but still distinct from the pre-event community (Fig 4.3). This dynamic response is attributable to the many thousands of diverse constituents that comprise the microbial community and contrasts sharply with the two constituents that comprise the isotopic signature.

Beyond the scale and complexity of their response, the microbial community offers yet another feature for analysis: the unique characteristics of each of its diverse constituents. Analysis of the relationships of microbial abundance to discharge (analogous to concentration – discharge [C-Q] analyses in chemical tracer studies) reveal that some taxa increase in abundance (are mobilized), while others decrease (are diluted), in response to increasing discharge. These different responses to discharge likely reflect the varied sources of these taxonomic constituents, which are known to originate from groundwater as well as soil water, and may indicate shifting sources of water to the stream during an event response. For example, we identified most Alphaproteobacteria sequences as mobilized with increasing discharge. Alphaproteobacteria are characteristically oligotrophic, originating in low-nutrient environments, and a growing constituency of this group likely signals an increasing contribution of water from a low-nutrient water pool, such as groundwater. As such, the microbial community has the potential to offer key insights toward better understanding complex hydrologic processes, such as shifting sources of streamflow.

This investigation represents a transdisciplinary effort, employing analytical approaches and domain knowledge from both microbial ecology and hydrology and with important implications for both fields. First, in analyzing microbial community composition in the context of a precipitation event and associated stream response, we find that the microbial community is highly sensitive, at daily timescales, to discharge dynamics, in terms of both diversity and community composition. Thus, the hydrologic context may be an important consideration in studies involving analysis of surface water microbial communities. Additionally, our abundance – discharge analysis, an analytical tool drawn from chemical hydrology, identified patterns in the response of microbial phylogenetic groups to hydrologic regime, which may offer new insight as well as a new tool for microbial ecology studies. For hydrologic studies, we have demonstrated that the microbial community provides a new information-rich dataset with new tools for advancing process-based hydrologic understanding.

4.5 Materials and Methods

4.5.1 Datasets

We collected daily DNA samples at the Marys River stream gage beginning three days prior to a precipitation event, continuing through the streamflow response, and concluding after the storm hydrograph recession; sample dates were 6-19 October, and 21, 23, and 25 October (Fig. 4.1). We collected all samples at approximately 0800 hours (range = 0730-0848 hours). To collect DNA samples, we obtained a water sample from the approximate center of the stream using a plastic bucket lowered from a bridge; we filtered and extracted DNA from streamwater as described in Crump, Kling, Bahr, & Hobbie (2003). We isolated the DNA and prepared the library following common accepted protocols. A detailed description of equipment preparation and sampling methods is found in URycki et al. (2020).

We sequenced 16S rRNA with Illumina Miseq V.2 paired end 250bp sequencing. We classified sequence taxonomy with the SILVA 16S rRNA gene database v.132 (Quast et al., 2013) and removed ASVs classified as chloroplasts or mitochondria, or if they were not classified to the domains Bacteria and Archaea. The resultant dataset consisted of 248,286 sequences of 807 unique amplified sequence variants (ASVs) across the 17 stream samples. Each sample was composed of a median

of 11,460 (range = 4,333-37,740) sequences. To standardize sampling depth across samples, we rarefied the dataset to 4,000 sequences per sample, resulting in a dataset of 4,000 x 17 = 68,000 counts of 620 unique ASVs. We analyzed these data as a matrix of the count of each unique ASV detected on each sample day. We further grouped ASVs into the following eight categories based on taxonomy: phyla Actinobacteriota, Bacteroidota, Cyanobacteria, Planctomycetota, Verrucomicrobiota; orders Gammaproteobacteria and Alphaproteobacteria (both of Phylum Proteobacteria); and Other, for ASVs that did not classify into any of the other seven groups. In microbiome research, individual ASVs are usually considered distinct taxonomic groups, and, as such, we use the terms ASV and taxon interchangeably throughout this analysis.

Concurrent with DNA sample collections, we recorded stream temperature with a water quality sampling probe (YSI Incorporated, Yellow Springs, OH, USA) and collected streamwater samples for stable isotope analysis. We also obtained average precipitation water stable isotope concentrations for the study period. We obtained streamflow records for the Marys River stream gage managed by the United States Geological Survey (U.S. Geological Survey, 2016). We aggregated sub-hourly data to daily mean discharge (ft^3/s; converted to m^3/s) for the analysis period.

4.5.2 Microbial Community Diversity

Biodiversity is a characteristic measure of biological communities. We calculated alpha diversity, or the diversity within the community, for each sample day using two common metrics: taxonomic richness and Shannon's index (Fig. 4.3). Taxonomic richness is the total number of unique microbial taxa. Shannon's index (*H*) is the entropy of taxa within a community, calculated as

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

where p_i is the proportion of the total number of ASVs (*s*) represented by each unique ASV *i* (Shannon and Weaver, 1949). The base for the logarithm is arbitrary, however, when base 2 is used, as here, the resultant quantity is described in units of bits. Shannon's index thus accounts for richness as well as evenness, or the relative abundance of each taxa. Larger values of *H* indicate greater diversity, and *H* is maximized when all taxa have the same relative abundance.

We then analyzed the extent to which Marys River post-event microbial communities differed from the pre-event communities. To quantify the difference between communities, termed beta diversity, we employed the Bray-Curtis dissimilarity metric (Bray and Curtis, 1957), common in ecological studies. We calculated the Bray-Curtis dissimilarity, BC(*u*,*v*), between pairs of samples *u* and *v* as

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

where $n_{i,u}$ is the relative abundance of ASV i in sample u (Bray and Curtis, 1957). Bray-Curtis dissimilarity ranges [0, 1], with unity indicating identical communities. We calculated the Bray-Curtis dissimilarity between pairs of samples (u_l, v_m), for all samples l = 6, 7, 8, …, 25 October and pre-event samples m = 6, 7, 8, 9 October. We analyze the difference between each sample community and the pre-event community as the mean Bray-Curtis distance for each sample from the four pre-event samples as:

$$BC(u_l) = \frac{\sum_m BC(u_l, v_m)}{4}.$$

4.5.3 Abundance – Discharge Relationships

We performed a regression analysis to characterize how individual microbial taxa respond to changes in discharge. Analogous to a concentration-discharge (C-Q) analysis, we assessed whether each microbial taxa exhibited a mobilization, dilution, or stasis response. To characterize taxa into these three groups, we performed simple linear regression between taxonomic abundance (ASV counts) and stream discharge for all taxa identified in at least three samples (i.e., days) across the entire study period 6-25 October. We classified taxa as mobilized if the relationship exhibited a positive slope, diluted if the slope was negative, and static if the slope was not significant at a 90% confidence level. Uncharacterized taxa are those detected on too few (<3) days for linear regression.

4.6 Funding

This work was supported by the National Science Foundation grant EAR 1836768. The first author acknowledges STEM Scholarship support from NSF grant 1153490.

4.7 Author Contributions

Author contributions are as follows: DRU, conceptualization, data curation, analysis, writing; SPG, conceptualization, analysis, funding acquisition, resources; BCC, data curation, analysis, funding acquisition, resources; NC, analysis; RB, data curation, analysis.

4.8 Acknowledgments

DRU, SPG, and BCC acknowledge that Oregon State University in Corvallis, Oregon is located within the traditional homelands of the Ampinefu Band of Kalapuya. Following the Willamette Valley Treaty of 1855, Kalapuya people were forcibly removed to reservations in western Oregon. Today, living descendants of these people are a part of the Confederated Tribes of Grand Ronde Community of Oregon and the Confederated Tribes of the Siletz Indians.

4.9 References

Brooks, J. R., Wigington, P. J., Phillips, D. L., Comeleo, R., and Coulombe, R. (2012). Willamette River Basin surface water isoscape (δ 18 O and δ 2 H): temporal changes of source water within the river. *Ecosphere* 3, art39. doi:10.1890/ES11-00338.1.

Crump, B. C., and Hobbie, J. E. (2005). Synchrony and seasonality in bacterioplankton communities of two temperate rivers. *Limnol. Oceanogr.* 50, 1718–1729. doi:10.4319/lo.2005.50.6.1718.

Crump, B. C., Kling, G. W., Bahr, M., and Hobbie, J. E. (2003). Bacterioplankton community shifts in an Arctic lake correlate with seasonal changes in organic matter source. *Appl. Environ. Microbiol.* 69, 2253–2268. doi:10.1128/AEM.69.4.2253-2268.2003.

Doherty, M., Yager, P. L., Moran, M. A., Coles, V. J., Fortunato, C. S., Krusche, A. V, et al. (2017). Bacterial Biogeography across the Amazon River-Ocean Continuum. *Front. Microbiol.* 8, 882. doi:10.3389/fmicb.2017.00882.

Klaus, J., and McDonnell, J. J. (2013). Hydrograph separation using stable isotopes: Review and evaluation. *J. Hydrol.* 505, 47–64. doi:10.1016/j.jhydrol.2013.09.006.

Nickolas, L. B., Segura, C., and Brooks, J. R. (2017). The influence of lithology on surface water sources. *Hydrol. Process.* 31, 1913–1925. doi:10.1002/HYP.11156.

Pinder, G. F., and Jones, J. F. (1969). Determination of the ground-water component of peak discharge from the chemistry of total runoff. *Water Resour. Res.* 5, 438–

445. doi:10.1029/WR005i002p00438.

Read, D. S., Gweon, H. S., Bowes, M. J., Newbold, L. K., Field, D., Bailey, M. J., et al. (2015). Catchment-scale biogeography of riverine bacterioplankton. *ISME J.* 9, 516–26. doi:10.1038/ismej.2014.166.

Savio, D., Sinclair, L., Ijaz, U. Z., Parajka, J., Reischer, G. H., Stadler, P., et al. (2015). Bacterial diversity along a 2600?km river continuum. *Environ. Microbiol.* 17, 4994–5007. doi:10.1111/1462-2920.12886.

Sklash, M. G., Stewart, M. K., and Pearce, A. J. (1986). Storm Runoff Generation in Humid Headwater Catchments: 2. A Case Study of Hillslope and Low-Order Stream Response. *Water Resour. Res.* 22, 1273–1282. doi:10.1029/WR022i008p01273.

Thompson, L. R., Sanders, J. G., McDonald, D., Amir, A., Ladau, J., Locey, K. J., et al. (2017). A communal catalogue reveals Earth's multiscale microbial diversity. *Nature* 551, 457–463. doi:10.1038/nature24621.

5 Conclusion

The objective of this research was to explore streamwater microbial communities, as characterized by 16S rRNA gene sequencing of DNA samples, as an information-rich dataset to advance hydrologic knowledge, which may in some cases be effectively limited by the types of data and approaches currently available for hydrologic studies. The analyses described here are the result of a distinctly transdisciplinary effort, drawing on datasets, domain knowledge, literature, and analytical methods from across the spectrum of hydrologic and microbial research. The implications of this research likewise span disciplines.

Using common tools of ecological research, we analyzed beta diversity with Bray-Curtis dissimilarity and demonstrated that, in our study area, aquatic microbial communities are influenced by landscape-scale climatological and geomorphic characteristics of the drainage basin, characteristics that also shape hydrologic regime, lending support to the hypothesis these shared drivers might result in hydrologic signals being encoded within the complex microbial community structure. Relationships between microbial communities and catchment characteristics were stronger in headwater streams, suggesting that the drivers of microbial community composition shift in lower stream reaches.

We applied advanced statistical approaches derived from communication and information theory to quantify relationships between individual microbial taxa and specific hydrologic metrics at different time and flow scales. We found that in some cases, observing the abundance of a particular microbial taxa could reduce the uncertainty of a hydrologic metric by >40%, even in an opposite season. Additionally, we were able to identify taxa that demonstrated especially interesting relationships with hydrologic metrics, such as the Bacteroidota of genus *Lacihabitans*, which exhibited a strong relationships to many hydrologic metrics but especially to daily discharge.

Finally, with a uniquely high temporal resolution microbial community dataset, we employed both ecological metrics and analytical tools from chemical hydrology and geochemical tracer analyses to analyze microbial community dynamics in the context of a precipitation and stream event response. We found that the microbial community exhibits a response to stream hydrograph dynamics that is both richer and greater in

magnitude than the isotopic response, and with patterns in phylogeny that may reflect shifting sources of event flow. With abundance – discharge analyses, adapted from concentration – discharge analyses in geochemical hydrology, we demonstrate a novel system of characterizing microbial taxa that may be useful in microbial ecology studies.

In service of our overarching research objective, and driven by the zeal that comes with an exciting new idea, we collected and analyzed more than 300 DNA samples from large and small streams across the state of Oregon, spanning four watersheds, three years, and all 12 months. The results we present here, based on a fraction of the samples we collected, hardly scratch the surface of the potential of this dataset. This dataset, and the foundational knowledge we have established here, will support a variety of studies across microbial ecology and hydrology and promoting fundamental process-based understanding of the interactions between the biosphere and hydrosphere and the implications for both.

Future investigations might begin by expanding the inferences proposed here. For instance, future studies might address the questions like:

- *How are winter microbial communities related to macroscale catchment characteristics and hydrology? How does the hydrologic information shared with summer microbial communities compare to that shared with winter microbial communities?*
- *How does the relationship between the microbial community and streamflow change for storm events of different magnitudes, at different points during the rainy season, or during different seasons?*
- *How does the streamwater microbial community relate to catchments and hydrology in different regions? What factors appear to influence these relationships when compared between regions and ecoclimatic conditions?*

Future studies might also build upon this work by addressing questions such as:

- *What information about hydrologic function can be gained from metagenomics analysis, and how does it compare to that gained from microbiome analysis?*
- *Are there patterns in the taxonomy of hydrologically informative microbes across seasons, flow conditions, or regions?*

- *What can we learn about ecohydrologic function from analyzing the functional roles and community dynamics of microbial taxa detected in streams?*
- *What analytical tools and methods can be used to effectively employ naturally-occurring DNA to strengthen process-based understanding of watershed function?*

The research described here establishes a foundation upon which studies of microbial DNA, or other types of genetic analysis, might launch in any number of new directions. Given the expanding availability of these types of datasets, and the tremendous amount of data they contain, there is much to be gained by continuing to collect, analyze, and employ genetic material in studies across the geosciences.