PATTERNS AND DYNAMICS OF PELAGIC STREAM MICROBIAL COMMUNITY ASSEMBLY

by

NORMAN H. HASSELL

(Under the Direction of Elizabeth Ottesen)

ABSTRACT

The pelagic microbial communities of streams represent ideal populations to study community assembly along a natural continuum. We assessed multiple physiochemical and land-use parameters along the network, but the primary predictor of microbial community structure was the communities' position relative to the stream continuum. Successional patterns were identified where taxonomic richness and compositional diversity decreased while the proportion of known freshwater taxa increased with increasing stream size. However, the observed trend was not present in two observation sets. In these samplings, streams exhibited uniformly high microbial diversity across the watershed, and the fraction of freshwater taxa showed no correlation to network position. Our work has revealed that these communities in streams exhibit a natural selection gradient under normal conditions but that this process is highly dynamic and can be disrupted, potentially in response to major environmental variation.

INDEX WORDS: Microbial ecology; Communities; Streams; Microbial diversity; Community assembly; Bioinformatics; Environmental

PATTERNS AND DYNAMICS OF PELAGIC STREAM MICROBIAL COMMUNITITY ASSEMBLY

by

NORMAN H. HASSELL

B.S., University of New Hampshire, 2010

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2016

© 2016

Norman H. Hassell

All Rights Reserved

PATTERNS AND DYNAMICS OF PELAGIC STREAM MICROBIAL COMMUNITITY ASSEMBLY

by

NORMAN H. HASSELL

Major Professor: Elizabeth Ottesen
Committee: Mary Ann Moran
Liang Liu
Zaid Abdo

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
August 2016

DEDICATION

This thesis is dedicated to my wife and parents. I would not have made it through this process without the help, encouragement, support, and friendship of my other half. I also wouldn’t have made it this far without the sacrifices my parents have made in order to give me a decent shot in this world. You all do more than I could ever ask and I rarely ever have to. You are the people that motivate me every day.

ACKNOWLEDGEMENTS

I would like to acknowledge the boundless support and mentorship given by my thesis advisor Dr. Elizabeth Ottesen, who allowed me the opportunity to contribute at the beginning of a developing project. I am forever grateful for your enthusiasm and patience. I would also like to acknowledge my committee members Dr. Mary Ann Moran, Dr. Zaid Abdo, and Dr. Liang Liu for their critical help and guidance. Additionally, I would like to thank the Upper Oconee Watershed network and community volunteers for all of their assistance in the coordination of sampling efforts. Last but not least, I would like to thank all of the members of the Ottesen lab, who were critical not only in the lab-wide effort to make this project possible, but were also constant sources of encouragement. I can honestly say you are the best group of people I have ever had the pleasure to work with.

TABLE OF CONTENTS

Page

ACKNOWLEDGEMENTS v

LIST OF TABLES vii

LIST OF FIGURES viii

CHAPTER

1 LITERATURE REVIEW 1

Stream Biogeochemistry 3

Niche and Neutral processes in streams 4

Benthic and pelagic stream microbial communities 5

References 8

2 MICROBIAL COMMUNITY ASSEMBLY AND DYNAMICS OF A MIXED-USE TEMPERATE WATERSHED 16

Abstract 17

Introduction 18

Results 20

Discussion 24

Methods 27

References 34

3 CONCLUSIONS 57

LIST OF TABLES

Page

Table 2.1: Primers and barcodes used in the generation of Illumina sequencing libraries............56

LIST OF FIGURES

Page

Figure 2.1: Map of Study collection sites 40

Figure 2.2: Total dissolved nitrogen and total dissolved phosphorus comparisons 41

Figure 2.3: Proportional abundance of major bacterial taxa for Spring 2013 sample collection ..42

Figure 2.4: Proportional abundance of bacterial taxa across seasons. 43

Figure 2.5: NMDS analyses of microbial communities and correlated variables 44

Figure 2.6: Map of land-useage of the cumulative catchment area of the study 45

Figure 2.7: 16S rRNA gene data from all study samplings 46

Figure 2.8: β -Diversity Cross-stream comparison 47

Figure 2.9: Community compositional β-Diversity as a function of cumulative stream length....48

Figure 2.10: Major Phylum/Class abundances between seasons. 49

Figure 2.11: Proportional abundance of selected freshwater lineages across sampling surveys...50

Figure 2.12: Spearman correlation histograms for each seasonal watershed survey.....................51

Figure 2.13: Proportional abundance of OTUs positively correlated with increasing stream size in the Spring 2013 sampling 52

Figure 2.14: Proportional abundance of correlated OTUs in all samplings 53

Figure 2.15: Impact of site selection and sampling depth on observed trends in species richness for quarterly surveys 54

Figure 2.16: Precipitation, stream discharge, and ambient temperature during the study period..55

CHAPTER 1

LITERATURE REVIEW

Over the past century the world has experienced substantial changes in land use, increased industrialization, and destruction of natural habitats. When assessing global biogeochemical nutrient budgets in the face of these changes, focus is often put on understanding the exchanges within marine and terrestrial environments [1, 2]. However, inland waters play critical roles in global biogeochemistry and evade disproportionate amounts of nutrients compared to their overall area [3-5]. Our relationship with these inland waters and their hydrological cycles has become increasingly strained with rises in population density and increased environmental pollution [6, 7]. These factors often have unintended consequences that affect the biotic processes and ecosystem composition of organisms within freshwater environments [8, 9]. Microorganisms are some of the primary players in these processes, being involved at the base level of carbon, nitrogen, and phosphorus cycling [4, 5, 10]. It has been shown that the composition of microbial biomass within freshwater can have direct effects on total system metabolism [11-13]. Considering these factors, it is crucial to study the microbial ecology of freshwater environments to understand their functional capacity, dynamics, and factors affecting microbial community composition.

Of the freshwater environments, streams exhibit an ideal natural laboratory to study both community ecology and community assembly. One of the overall goals of community ecology is to address the relative weight of processes that influence the assembly of a given community. The Lagrangian perspective of stream networks facilitates longitudinal examination of microbial

communities with space-as-time resolution, allowing the study of assembly patterns relative to positions along the network. This physical property of streams, combined with the naturally high turnover rate of microorganisms, enables the assessment of multiple factors shown to impact community assembly. These factors include, but are not limited to: light, temperature, elevation, dispersal, and nutrient gradients in relation to the changes in community structure over time. In this sense, streams are active conduits of nutrients [4, 14] and organisms [15-17] that can be examined along a natural continuum. Streams also represent a transitional interface between terrestrial, freshwater, and marine ecosystems. This enables the study of communities' initial interactions with soils at stream sources down to stream outflow into ocean environments.

Previous studies of stream microbial communities provide evidence that downstream communities are shaped directly by the dispersal of terrestrial microorganisms from headwaters where soil-water admixture is highest [18-21]. After initial dispersal, communities establish rapidly within the water column [22] and stream benthos [23, 24]. As these streams progress to rivers, changes in community structure become more gradual and overall community diversity decreases. The transition of these communities over time involves constant interaction between terrestrial, free-living, particle associated, and benthic microorganism fractions [22]. Large overlaps between all of these community niches are present, yet all maintain distinct subsets of taxa that differentiate the fractions [21, 22, 25]. It has been shown that the community assembly process involves both mass-effects through dispersal and in stream environmental filtering ("species sorting") [16, 17, 25-27]. The balance between these affects shifts with the network continuum. Initially, it is thought that mass effects play the biggest role in the selection of organisms within headwaters, but as these streams progress to rivers and community residence time increases, in-stream species sorting plays a larger role while dispersal effects diminish.

Although derived community residence time (from cumulative catchment area or stream length) is often the most correlated variable to community assembly gradients in streams [16, 17, 24], other system parameters have also been shown to influence stream community composition. These include but are not limited to nutrient concentrations [28, 29], pH [30], hydrology [27, 31, 32], metal contamination [33], organic matter bioavailability [34], land-use [35], and temperature variation [36]. However, the noticeable effect of each of these influencing variables is almost always dependent on the specific network system being studied, with little universal agreement across observation sets. It is therefore important to evaluate these biological networks within a local context while trying to identify universal properties of stream ecosystems that influence community assembly.

Stream Biogeochemistry

Within a given stream network, the metabolism of the stream ecosystem changes along a continuum. Initially, headwater stream function is dependent upon riparian vegetation which contributes allochthonous material and also provides shading for the stream [8, 37]. These headwater streams primarily exhibit heterotrophic communities of diverse compositions. As the streams progress towards rivers these communities transition from heterotrophy to autotrophy with the decrease in stream shading and increased water volumes. This transition holds generally true in most stream systems but the gradient of change is often influenced by stream turbidity and dissolved organic matter, which can impede primary production through light mitigation [8]. Other factors such as relative concentrations of dissolved organic carbon (DOC), total dissolved nitrogen (TDN), and total dissolved phosphorus (TDP) can result in more dramatic effects on ecosystem function such as system eutrophication [4, 38, 39]. Nutrient oversaturation often leads

to cyanobacterial blooms which disrupt the stability of the system and may result in macroorganism die-offs [8, 38, 40].

Recent studies have shown the role of streams in the cycling of carbon (C), nitrogen (N), and phosphorus (P) to be of even greater importance than previously thought [3-5, 14, 38, 41]. It is estimated that $60 \cdot 10^9$ moles y^{-1} of N and $20 \cdot 10^9$ moles y^{-1} of P are buried within sediments of streams [4]. Given this information, streams represent a critical source for excess nutrient deposition and storage. Conversely, over 1.9 Pg of C y^{-1} is released by inland waters [10], with the majority of C processing occurring within streams and rivers [3]. The relative amounts of these nutrients being retained, transported, and processed within a given system depends on several factors such as temperature [38], land use [39, 42], and population density [8, 39]. However, levels of nutrient input in streams have been shown to directly influence the structure and function of microbial communities within them, albeit with varied overall weights [36, 43, 44]. Although nutrient availability can cause changes in bacterial diversity, the physical structure of a given stream network and its corresponding hydrology is the most universal factor shown to influence microbial composition in streams [16, 17, 22, 27, 36, 44].

Niche and Neutral processes in streams

The role of niche versus neutral processes in community assembly has often been debated within the field of community ecology. Recent efforts have been made to unite these concepts into a continuum [45-47], yet disagreements as to the importance of each process exits. In stream ecosystems, both of these processes have been observed and are thought to act in balance along the network gradient. Within baseline or average environmental conditions, mass-effects from the dispersal of microorganisms are thought to be most important in shaping headwater

communities [21, 44, 48], whereas species-sorting is suggested to play the primary role in the progression of communities in large rivers [16, 17]. These ecological selection gradients can vary depending on the partition of the community being observed. Pelagic communities in streams, which are more variable and susceptible to disruption, are thought to exhibit slower selection gradient transitions [16, 17, 48, 49]. Benthic communities' residence times have been shown to be much higher than those of the pelagic and particle-associated communities within the water column [50, 51], allowing interspecific competition to occur much further up the network. This is thought to cause species-sorting to be the dominant selection mechanism over the progression of the benthic community downstream. It could be argued, however, that mass-effects at headwaters are the most critical processes because they 'seed' the communities that undergo further selection downstream [44].

In streams, the balance between members of the community is thought to transition from r-strategist (opportunistic, rapid growing) to k-strategist organism regimes (specialist, slow growing) [9, 16] as overall nutrient concentrations decrease and water volume increases along the network. This is supported by a comparison of growth rates of cultivable organisms found in high abundance in stream headwaters (*Flavobacteriales* and *Bacteroidales*) versus higher order rivers (*Actinobacteria*) [16, 17, 52-54]. Although the neutral to niche selection gradient is thought to be present within normal conditions, events such as flooding [23, 31], drought [55], and temperature fluctuation [26, 56, 57] can alter the selective dynamics within streams.

Benthic and pelagic stream microbial communities

Currently, the majority of work on aquatic ecosystem microbial ecology has been focused on the study of pelagic microorganisms [9, 57-60]. In the field of stream microbial ecology,

however, more emphasis has been put on characterizing the composition and function of communities within the stream benthos [18, 24, 51, 54, 61-63]. Benthic stream microbial communities usually form as complex aggregate biofilms with high densities of diverse organism taxa [24], including high proportions of bacterial and eukaryotictaxa, with smaller representation of archaeal taxa [62]. The role of these communities in ecosystem function is greatest at headwaters, where the microbial biomass of stream biofilms outnumbers the population within the water column [24, 62]. The communities of the stream benthos exhibit high densities of enzymatic activity, dissolved organism matter decomposition, and chemical processing [24, 51, 64]. As water volume accumulates with junctions down the network, decreases in nutrient availability, light, and temperature diminish the role of benthic biofilm processes as pelagic communities ultimately dominate system metabolism [37].

Pelagic communities are thought to play less of a functional role in headwaters than benthic communities as a result of their relative abundances. However, current evidence suggests that the assembly of both these niche populations along the continuum is dependent upon initial processes happening at the origins of the network [65, 66]. Work studying both stream ecosystem fractions has revealed that much of the downstream community is shaped by organisms dispersed into the network at headwaters and the rapid community selection that occurs within these reaches [21, 24, 48, 67]. Of the relatively few studies that have compared the diversity of pelagic and benthic communities within the same stream network, it has been found that the overall diversity of organisms is much higher in the pelagic fraction [65, 66]. This is most likely due to the greater effects of stochastic dispersal processes increasing the likelihood of organism immigration into stream water. Moreover, higher temporal fluctuations have been observed in pelagic communities than in benthic communities, further implicating the

importance of dispersal effects. Although there has been much progress in understanding the taxonomic diversity and composition of microorganisms within stream environments, the exact mechanisms that assemble and maintain these communities remains to be explained.

By examining the pelagic microbial communities of streams in the Upper Oconee watershed, we have assessed factors that influence the assembly of these communities at points across a temporal continuum. This study has allowed us to sample a wide variety of streams, from first-order headwaters to fifth-order rivers, at high spatial resolutions. The depth and diversity of sampling has enabled us to not only examine patterns of diversity, but also the variance in these communities in relation to fluctuating environmental, physiochemical, and landscape-level parameters.

References:

1. Lal, R., *Soil Carbon Sequestration Impacts on Global Climate Change and Food Security.* Science, 2004. **304**(5677): p. 1623-1627.
2. Le Quéré, C., et al., *Global carbon budget 2014.* Earth System Science Data, 2015. **7**(1): p. 47-85.
3. Cole, J.J., et al., *Plumbing the Global Carbon Cycle: Integrating Inland Waters into the Terrestrial Carbon Budget.* Ecosystems, 2007. **10**(1): p. 172-185.
4. Nixon, S.W., et al., *The Fate of Nitrogen and Phosphorus at the Land-Sea Margin of the North Atlantic Ocean.* Biogeochemistry, 1996. **35**(1): p. 141-180.
5. Tranvik, L.J., *Lakes and reservoirs as regulators of carbon cycling and climate.* Limnology and oceanography, 2009. **54**(6 part 2): p. 2298-2314.
6. Schwarzenbach, R.P., et al., *The Challenge of Micropollutants in Aquatic Systems.* Science, 2006. **313**(5790): p. 1072-1077.
7. Oki, T. and S. Kanae, *Global Hydrological Cycles and World Water Resources.* Science, 2006. **313**(5790): p. 1068-1072.
8. Dodds, W.K., *Eutrophication and trophic state in rivers and streams.* Limnology and Oceanography, 2006. **51**(1_part_2): p. 671-680.
9. Weinbauer, M.G. and M.G. Hofle, *Distribution and life strategies of two bacterial populations in a eutrophic lake.* Applied and environmental microbiology, 1998. **64**(10): p. 3776-83.
10. Raymond, P.A., et al., *Global carbon dioxide emissions from inland waters.* Nature, 2013. **503**(7476): p. 355-9.

11. Christoffersen, K., et al., *Qualitative importance of the microbial loop and plankton community structure in a eutrophic lake during a bloom of cyanobacteria.* Microbial Ecology, 1990. **20**(1): p. 253-272.

12. Correll, D.L., *The Role of Phosphorus in the Eutrophication of Receiving Waters: A Review.* Journal of Environmental Quality, 1998. **27**(2): p. 261-266.

13. Zeglin, L.H., *Stream microbial diversity in response to environmental changes: review and synthesis of existing research.* Frontiers in microbiology, 2015. **6**(May): p. 454-454.

14. Meybeck, M., *Riverine transport of atmospheric carbon: Sources, global typology and budget.* Water, Air, and Soil Pollution, 1993. **70**(1-4): p. 443-463.

15. Fortunato, C.S., et al., *Spatial variability overwhelms seasonal patterns in bacterioplankton communities across a river to ocean gradient.* ISME J, 2012. **6**: p. 554-563.

16. Read, D.S., et al., *Catchment-scale biogeography of riverine bacterioplankton.* ISME J, 2015. **9**(2): p. 516-26.

17. Savio, D., et al., *Bacterial diversity along a 2600 km river continuum.* Environ Microbiol, 2015.

18. Beier, S., K.P. Witzel, and J. Marxsen, *Bacterial community composition in Central European running waters examined by temperature gradient gel electrophoresis and sequence analysis of 16S rRNA genes.* Appl Environ Microbiol, 2008. **74**(1): p. 188-99.

19. Sekiguchi, H., et al., *Succession of Bacterial Community Structure along the Changjiang River Determined by Denaturing Gradient Gel Electrophoresis and Clone Library Analysis.* Applied and Environmental Microbiology, 2002. **68**(10): p. 5142-5150.

20. Winter, C., et al., *Longitudinal changes in the bacterial community composition of the Danube River: a whole-river approach.* Appl Environ Microbiol, 2007. **73**(2): p. 421-31.

21. Crump, B.C., L.A. Amaral-Zettler, and G.W. Kling, *Microbial diversity in arctic freshwaters is structured by inoculation of microbes from soils.* ISME J, 2012. **6**(9): p. 1629-39.

22. Ruiz-Gonzalez, C., J.P. Nino-Garcia, and P.A. Del Giorgio, *Terrestrial origin of bacterial communities in complex boreal freshwater networks.* Ecol Lett, 2015.

23. Besemer, K., et al., *Complexity of bacterial communities in a river-floodplain system (Danube, Austria).* Appl Environ Microbiol, 2005. **71**(2): p. 609-20.

24. Besemer, K., et al., *Headwaters are critical reservoirs of microbial diversity for fluvial networks.* Proceedings. Biological sciences / The Royal Society, 2013. **280**: p. 20131760.

25. Lindstrom, E.S. and S. Langenheder, *Local and regional factors influencing bacterial community assembly.* Environ Microbiol Rep, 2012. **4**(1): p. 1-9.

26. Staley, C., et al., *Species sorting and seasonal dynamics primarily shape bacterial communities in the Upper Mississippi River.* Sci Total Environ, 2015. **505**: p. 435-45.

27. Freimann, R., et al., *Hydrologic linkages drive spatial structuring of bacterial assemblages and functioning in alpine floodplains.* Front Microbiol, 2015. **6**: p. 1221.

28. Olapade, O.A. and L.G. Leff, *Seasonal Response of Stream Biofilm Communities to Dissolved Organic Matter and Nutrient Enrichments.* Applied and Environmental Microbiology, 2005. **71**(5): p. 2278-2287.

29. Van Horn, D.J., et al., *Response of heterotrophic stream biofilm -communities to a gradient of resources.* Aquatic Microbial Ecology, 2011. **64**(2): p. 149-161.

30. Fierer, N., et al., *ENVIRONMENTAL CONTROLS ON THE LANDSCAPE-SCALE BIOGEOGRAPHY OF STREAM BACTERIAL COMMUNITIES.* Ecology, 2007. **88**(9): p. 2162-2173.

31. Luef, B., et al., *Impact of hydrology on free-living and particle-associated microorganisms in a river floodplain system (Danube, Austria).* Freshwater Biology, 2007. **52**(6): p. 1043-1057.

32. Niño-García, J.P., C. Ruiz-González, and P.A. del Giorgio, *Interactions between hydrology and water chemistry shape bacterioplankton biogeography across boreal freshwater networks.* The ISME Journal, 2016: p. 1-12.

33. Feris, K., et al., *Differences in Hyporheic-Zone Microbial Community Structure along a Heavy-Metal Contamination Gradient.* Applied and Environmental Microbiology, 2003. **69**(9): p. 5563-5573.

34. Findlay, S.E.G., et al., *Metabolic and structural response of hyporheic microbial communities to variations in supply of dissolved organic matter.* Limnol. Oceanogr, 2003. **48**(4): p. 1608-1617.

35. Van Rossum, T., et al., *Year-long metagenomic study of river microbiomes across land use and water quality.* Frontiers in Microbiology, 2015. **6**.

36. Zeglin, L., *Stream microbial diversity responds to environmental changes: Review and synthesis of existing research.* Frontiers in Microbiology, 2015. **6**.

37. Vannote, R.L., Minshall, G. W., Cummins, K. W., Sedell, J. R. & Cushing, C. E., *The River Coninuum Concept.* Canadian Journal of Fisheries and Aquatic Sciences, 1980(37): p. 130-137.

38. Schaefer, S.C. and M. Alber, *Temperature controls a latitudinal gradient in the proportion of watershed nitrogen exported to coastal ecosystems.* Biogeochemistry, 2007. **85**(3): p. 333-346.

39. Weston, N.B., J.T. Hollibaugh, and S.B. Joye, *Population growth away from the coastal zone: thirty years of land use change and nutrient export in the Altamaha River, GA.* Sci Total Environ, 2009. **407**(10): p. 3347-56.

40. Huang, J., et al., *Detecting the Dynamic Linkage between Landscape Characteristics and Water Quality in a Subtropical Coastal Watershed, Southeast China.* Environmental Management, 2013. **51**(1): p. 32-44.

41. Battin, T.J., et al., *Biophysical controls on organic carbon fluxes in fluvial networks.* Nature Geosci, 2008. **1**(2): p. 95-100.

42. Mulholland, P.J., et al., *Stream denitrification across biomes and its response to anthropogenic nitrate loading.* Nature, 2008. **452**(7184): p. 202-205.

43. Ruiz-González, C., et al., *Effects of large river dam regulation on bacterioplankton community structure.* FEMS microbiology ecology, 2013. **84**(2): p. 316-31.

44. Nino-Garcia, J.P., C. Ruiz-Gonzalez, and P.A. del Giorgio, *Interactions between hydrology and water chemistry shape bacterioplankton biogeography across boreal freshwater networks.* ISME J, 2016. **10**(7): p. 1755-1766.

45. Tilman, D., *Niche tradeoffs, neutrality, and community structure: A stochastic theory of resource competition, invasion, and community assembly.* 2004.

46. Gravel, D., et al., *Reconciling niche and neutrality: the continuum hypothesis.* Ecology Letters, 2006. **9**(4): p. 399-409.

47. Adler, P.B., J. HilleRisLambers, and J.M. Levine, *A niche for neutrality.* Ecology Letters, 2007. **10**(2): p. 95-104.

48. Crump, B.C., et al., *Biogeography of bacterioplankton in lakes and streams of an Arctic tundra catchment.* Ecology, 2007. **88**: p. 1365-78.

49. Kolmakova, O.V., et al., *Spatial biodiversity of bacteria along the largest Arctic river determined by next-generation sequencing.* FEMS microbiology ecology, 2014. **89**(2): p. 442-50.

50. Boano, F., et al., *Hyporheic flow and transport processes: Mechanisms, models, and bioghemical implications.* Reviews of Geophysics, 2014: p. 1-77.

51. Battin, T.J., et al., *Contributions of microbial biofilms to ecosystem processes in stream mesocosms.* Nature, 2003. **426**(6965): p. 439-442.

52. Šimek, K., et al., *Maximum growth rates and possible life strategies of different bacterioplankton groups in relation to phosphorus availability in a freshwater reservoir.* Environmental Microbiology, 2006. **8**(9): p. 1613-1624.

53. Ruiz-González, C., J.P. Niño-García, and P.A. Del Giorgio, *Terrestrial origin of bacterial communities in complex boreal freshwater networks.* Ecology letters, 2015.

54. Wilhelm, L., et al., *Altitudinal patterns of diversity and functional traits of metabolically active microorganisms in stream biofilms.* The ISME Journal, 2015. **9**(11): p. 2454-2464.

55. Rees, G.N., et al., *Variability in sediment microbial communities in a semipermanent stream: impact of drought.* Journal of the North American Benthological Society, 2006. **25**(2): p. 370-378.

56. Crump, B.C. and J.E. Hobbie, *Synchrony and seasonality in bacterioplankton communities of two temperate rivers.* Limnology and Oceanography, 2005. **50**(6): p. 1718-1729.

57. Lindström, E.S., M.P. Kamst-Van Agterveld, and G. Zwart, *Distribution of typical freshwater bacterial groups is associated with pH, temperature, and lake water retention time.* Appl Environ Microbiol, 2005. **71**(12): p. 8201-6.

58. Newton, R.J., et al., *A guide to the natural history of freshwater lake bacteria.* Microbiol Mol Biol Rev, 2011. **75**(1): p. 14-49.

59. Bendall, M.L., et al., *Genome-wide selective sweeps and gene-specific sweeps in natural bacterial populations.* ISME J, 2016.

60. Lindström, E.S., et al., *External control of bacterial community structure in lakes.* Limnology and Oceanography, 2006. **51**(1): p. 339-342.

61. Araya, R., et al., *Bacterial activity and community composition in stream water and biofilm from an urban river determined by fluorescent in situ hybridization and DGGE analysis.* FEMS Microbiology Ecology, 2003. **43**(1): p. 111-119.

62. Battin, T.J., et al., *The ecology and biogeochemistry of stream biofilms.* Nature Reviews Microbiology, 2016. **14**(4): p. 251-263.

63. Lear, G., et al., *The biogeography of stream bacteria.* Global Ecology and Biogeography, 2013. **22**(5): p. 544-554.

64. Araya, R., Tani, K., Takagi, T., Yamaguchi, N. & Nasu, M., *Bacterial activity and community composition in stream water and biofilm from an urban river determined by fluorescent in situ hybridization and DGGE analysis.* FEMS Microbiol Ecol, 2003(43): p. 111-119.

65. Besemer, K., et al., *Unraveling assembly of stream biofilm communities.* The ISME Journal, 2012. **6**(10): p. 1459-1468.

66. Wilhelm, L., et al., *Microbial biodiversity in glacier-fed streams.* The ISME Journal, 2013. **7**: p. 1651-1660.

67. Febria, C.M., et al., *Microbial responses to changes in flow status in temporary headwater streams: a cross-system comparison.* Frontiers in Microbiology, 2015. **6**.

CHAPTER 2

MICROBIAL COMMUNITY ASSEMBLY AND DYNAMICS OF A MIXED-USE TEMPERATE WATERSHED[1]

[1]Hassell, N.H, Tinker, K.A, Moore, T., and Ottesen, E.A. Submitted to *Environmental Microbiology* (currently in revision), 05/29/16.

Abstract:

Pelagic microbial communities of streams represent unique populations that are distinct from the microbial communities of the stream benthos and surrounding soil. In this study, we utilized high-throughput 16S rRNA gene sequencing to characterize microbial community composition within a temperate stream network. We found that the microbial communities of headwater streams are compositionally diverse, with low representation of known freshwater microbial taxa. In three out of five seasonal samplings, a successional pattern was identified in which phylotype richness and compositional heterogeneity (cross-stream and over time) decreased while the proportion of known freshwater taxa increased with increasing stream size. However, this trend was not observed in two seasons. Instead, streams exhibited uniformly high microbial diversity across the watershed, and the fraction of freshwater taxa showed no relationship with stream size or distance traveled. Our data suggests that longitudinal trends in microbial diversity in temperate watersheds are generated by microbial dispersal from surrounding environments into stream headwaters, followed by in-stream selection for specific freshwater-adapted microbes. However, this microbial succession appears to be a highly dynamic process that can be disrupted at landscape scales, potentially in response to variation in temperature and precipitation.

2.2 Introduction

Lotic systems represent a unique biome that is critically important to life on a global scale. Streams and rivers act as sources of drinking water, irrigation reservoirs for agriculture, commercial fishery habitats, transportation routes, and manufacturing process resources, among many other uses. These environments also represent the transitional link between terrestrial and marine environments, serving as transport pathways for nutrients [4], substrates [14], and biological organisms [37]. Recent studies have shown that inland waters are active participants in global carbon cycling, releasing 2.1 Pg C from CO_2 evasion per year, with 1.8 Pg C originating from stream and riverine systems [3, 10]. This represents approximately 4% of terrestrial primary production [10], which is especially remarkable considering that surface freshwater accounts for less than 0.01% of all water globally [68]. In freshwater systems, microorganisms serve as the processing and conversion base for the majority of nutrients [69, 70]. At headwaters, benthic communities are responsible for the majority of nutrient transformation, but as these small streams transition to rivers, the importance of pelagic community activity increases and organisms within the water column ultimately dominate ecosystem metabolism [37]. In this context, it is imperative to understand microbial community assembly within pelagic stream environments.

Recent studies have addressed bacterial community diversity within the water columns of large rivers [17, 26, 71], across the network of a river catchment [16], and among the benthic biofilms of streams [23, 24]. While some studies have found taxonomic shifts closely related to seasonality and physiochemical parameters [26, 72], many have found overall patterns of decreasing diversity along the flow path of the river [16, 17] or within a watershed (Crump et al 2007, Crump et al 2012). These studies have given rise to a model in which stream microbial

biodiversity is driven by both mass-effects and species sorting, with the later process playing an increasingly important role with downstream distance traveled and increased community residence time [16, 17, 21, 48].

Studies of streams and rivers have found numerically abundant microbial populations related to bacterial taxa typically found in lakes and other freshwater environments [57, 58, 73]. The freshwater *Actinobacteria* lineage acI, *Betaproteobacteria* lineage betII, and the *Alphaproteobacteria* lineage alfV appear to be consistently dominant groups of bacteria in lotic environments [16, 17, 74]. These bacteria have been identified as particularly abundant in downstream environments (Savio et al 2015), suggesting that freshwater-adapted microbes have a selective advantage during in-stream microbial community assembly.

Seasonal [26, 56, 75, 76], physiochemical [6, 72, 77], environmental [26, 78], and dispersal effects [21, 48] have been observed to shape pelagic communities in streams. However, published studies have generally been limited to observations at a single period of time or from relatively few sites along an individual river, typically under baseline flow conditions. In order to gain a better understanding of factors affecting community assembly and fluctuation at watershed scales, we conducted multiple high-resolution surveys of community structure across the Upper Oconee watershed in Northeast Georgia. In this study, we demonstrate network community structure across five seasonal samplings over the course of one year. This study spans a riverine network from headwaters to fifth-order rivers, providing valuable information relating to microbial community assembly within streams.

2.3 Results

Water was sampled from streams and rivers throughout the Upper Oconee Watershed near Athens, GA (Fig. 2.1). A total of 211 samples were collected across five seasonal samplings (74 samples in Spring 2013, 15 samples in Summer 2013, 28 samples in Fall 2013, 20 samples in Winter 2014, and 74 samples in Spring 2014, see supplemental attached table for details). Calculated cumulative dendritic distance upstream of sampling sites ranged from 0.09 - 3778.59 km for the complete dataset, with the maximum range for any individual season of 0.09 - 3778.59 km (Spring 2013) and a minimum range of 0.24 - 2119.36 km (Spring 2014). Nutrient content readings from sample site filtrates were analyzed to assess possible changes in total dissolved nitrogen (TDN) and phosphorus (TDP) concentrations within the watershed. Overall concentrations of nutrients from seasons did not significantly differ (Tukey HSD, p-vals ≥ 0.07) (Fig. 2.2A and C). However, there was sample-to-sample variance in TDN and TDP when select sites were compared across seasons (Fig. 2.2B and D). Smaller streams exhibited more variation in nutrient concentrations over time.

For community analysis, particles collected within the 0.22 µm-5µm size fraction from water samples were used in DNA extraction, 16S rRNA gene amplification, and sequencing. A total of 15 651 160 sequences passed quality checks and were binned into 350 333 OTUs at 97% identity. The major bacterial taxa present within these samples were Actinobacteria, Bacteriodetes, Proteobacteria (specifically class Betaproteobacteria), and Verrucomicrobia (Fig. 2.3, Fig. 2.4). All of these phyla have been documented as highly abundant within many freshwater environments [17, 21, 58, 79].

Relationships between environmental metadata, microbial diversity, and community composition across the watershed were assessed. The environmental parameters that showed the

strongest relationships with microbial community data were those associated with stream size and position within the watershed network, including stream order, catchment area, and cumulative dendritic stream length. As these parameters are correlated, we have selected a single metric, cumulative dendritic stream length, to present. Although some land-use estimates and physiochemical measurements did significantly correlate ($p < 0.05$) with community variation based on ordination analyses (Fig. 2.5), these parameters' axes of variation corresponded closely to sample stream size. The sampling of headwaters versus rivers within our study area was relatively biased, as almost all of the sampled headwater catchments were located in semi-developed to highly developed areas, whereas most of the larger streams originated from rural, relatively undeveloped regions (Fig. 2.6). Considering this information, it was inferred that these relationships were most likely due to autocorrelation and could not be distinguished as true affects.

In general, the OTU richness of streams decreased with increasing cumulative stream length (Fig. 2.7A). This was accompanied by a trend in which both cross-stream beta diversity (Fig. 2.7B, 2.8) and beta diversity over time (Fig. 2.7, 2.9) also decreased with increasing stream length. These observations are consistent with previously published results from individual streams [21, 48], and suggest a pattern in which stream headwaters (and possibly soil and/or groundwater reservoirs) are a key source of microbial diversity in streams, and that a strong environmental filter selects for a consistent community across streams and over time. Interestingly, this relationship was only apparent in three out of the five watershed surveys, with the remaining two surveys showing high alpha and beta diversity across the entire watershed.

These trends in microbial diversity were accompanied by a consistent trend in community composition (Fig. 2.7C,D). Decreases in alpha and beta diversity were accompanied by an

increase in the relative abundance of 216 "core community OTU's", defined as sequences that were present in $\geq$ 90% of all samples. Many of these sequences were found to belong to taxa previously identified in other surveys of freshwater environments. In order to examine this relationship in more depth, we created a database of "freshwater-associated taxa", sequences that were highly represented among published datasets of 16S rRNA gene sequences from freshwater environments. The relative abundance of these freshwater-associated OTUs increased in direct relationship to increasing cumulative stream length (Fig. 2.7D). These relationships were only apparent, however, in datasets that exhibited the alpha and beta diversity trends described above.

Phylum level abundance and diversity for streams across the catchment gradient was assessed to observe variance in taxonomic composition. At the phylum level, Actinobacteria and Bacteroidetes increased in abundance with increasing dendritic distance, while other major bacterial phyla did not increase (Fig. 2.10). When higher taxonomic resolutions were examined, Actinobacteria lineages (acI, acIII, and Luna1), Betaproteobacteria lineages (betI, betII, and betIV), and Bacteroidetes lineages (bacI, bacII, and bacIII) were found in high abundance across the study area (2.11), and were enriched with increasing stream size. These groups are typically abundant in both freshwater lakes [58] and riverine systems [16, 17, 74]. Relative proportions of abundant freshwater lineages within first order streams remained similar across all seasons (mean 30.3%). However, fourth order streams and above contained a significantly (Tukey HSD test, p-val $\leq$0.02) higher proportion (mean 74.5%) of abundant freshwater lineages in seasons with significant diversity trends than in seasons with no observed trends (mean 21.4%).

Relationships between individual OTU abundance and cumulative dendritic stream length were assessed to identify OTU's that were consistently enriched or depleted in downstream environments. Spearman's correlation coefficient was calculated by analyzing

proportional counts of each seasonal sample OTU against log-normalized cumulative stream length of each site (Fig. 2.12). For seasons with previously identified diversity trends, the 100 most abundant OTUs displayed bi-modal distributions in correlation coefficients, progressing to a left skewed distribution among less abundant OTUs. Sample OTU bins positively correlated with flow contained a majority of OTUs matching the core freshwater reference set within the 100 most abundant OTUs. In contrast, seasons with no identified trends in overall microbial diversity or community composition showed correlation coefficients (ρ) with skewed normal distributions throughout count levels.

When examining OTUs positively correlated with increasing cumulative stream length (Spearman $\rho \geq 0.4$, p-value < 0.05), samples from larger streams exhibited higher proportions of positively correlated OTUs. The most abundant representatives originated from freshwater lineages (Fig. 2.13). This pattern was only observed within seasons where correlation trends were significant for diversity and presence of core freshwater OTUs compared to increasing cumulative stream length (Fig. 2.14). The most abundant positively correlated OTUs within these seasons were consistently dominated by five freshwater lineages (acI, Luna1, bacI, bacIII, and betI). Neutrally distributed OTUs were more mixed in their composition (Betaproteobacteria being most dominant) and negatively correlated OTUs were comprised mainly of non-freshwater Betaproteobacteria and Gammaproteobacteria taxa.

As noted above, trends in species richness and community composition were absent in two out of the five seasonal samplings. Instead, phylotype richness and cross-stream beta diversity remained high across all stream sizes. These seasons also exhibited changes in microbial community composition. Freshwater lineages that were found positively correlated in the other samplings remained the most abundant taxa detected, but their abundance was lower

than observed in other seasons and no longer correlated with stream size (Fig. 2.14). The relative abundance of freshwater taxa was highly variable across streams, which may result from either variation in initial species pools for individual streams or from variation in the degree to which community assembly processes were disrupted at different locations.

Differences across seasons in microbial biodiversity and species distributions do not appear to be a function of the particular sites observed. Streams that did not exhibit size-dependent trends in biodiversity and microbial community composition on the summer and fall sample dates did exhibit stream size-dependent trends in biodiversity in other seasons (Fig. 2.15). This suggests that these relationships were controlled by external, landscape-scale variables, including possible seasonal variation and/or temporary deviations in temperature and/or precipitation. While causal relationships cannot be fully determined without additional data collection, the summer 2013 sampling date followed a period of high precipitation (Fig. 2.16), which may have increased transport of runoff-associated microbial taxa into downstream environments, while the winter 2014 sampling data correlated with low average temperatures, which may have caused increased microbial generation times relative to stream residence times.

2.4 Discussion

In this study, we examined microbial community composition across an entire watershed on five dates over the course of a year. The variety of analyzed sites and quarterly sampling approach has allowed us to gain new insights into diversity gradients within the watershed and how they change over time. In particular, we found strong successional patterns in microbial diversity and community composition within the watershed for three out of five quarterly surveys. However, these patterns were highly disrupted for the remaining two quarterly

watershed surveys. These results suggest that community assembly in streams is highly dynamic and subject to landscape-scale changes in the distribution of microbial taxa and biodiversity within the watershed.

For three out of five large-scale watershed surveys, we identified significant relationships between microbial diversity and distance traveled within the stream network. We found that both alpha and beta diversity decreased with decreasing stream size, with beta diversity decreasing both between sites at a single time point (Fig. 2.7A,B, 2.9) and over time for individual streams (Fig. 2.8). Previous studies have shown decreasing alpha and beta diversity among microbial communities along the length of river paths [21, 78], along the catchment area gradient of large watersheds [16, 17], and among pelagic and benthic microbial communities of individual streams [21, 24, 48]. These observations are consistent with a model in which stream headwaters host highly diverse microbial communities primarily shaped by dispersal from the soil/water interface while downstream environments are less diverse and more strongly impacted by within-stream species sorting [17, 21].

Dramatic shifts in community diversity also corresponded to sharp changes in overall community composition. Taxonomic community composition along streams corresponded closely to overall seasonal diversity trends. In surveys that exhibited strong diversity gradients, these diversity trends were accompanied by a consistent downstream enrichment of a small subset of bacterial OTU's from freshwater-associated lineages. This suggests that a highly specific pelagic bacterial population is selected for within streams from the Spring 2013, Fall 2013, and Spring 2014 samplings, with similar taxa prevailing among downstream environments over time. These taxa remained relatively abundant during samplings that did not exhibit

significant diversity trends, but their abundance was not significantly correlated with stream length at those time points.

The communities from our study exhibited downstream increases among both Actinobacteria and Bacteriodetes in seasons with significant diversity trends. Betaproteobacteria and Verrucomicrobia were relatively neutral in their distributions across streams at the phylum level, but exhibited downstream enrichment of specific freshwater-associated lineages. Gammaproteobacteria and Alphaproteobacteria decreased significantly in overall abundance with increasing stream size (2.10). In contrast, studies of larger watersheds and rivers [16, 17] recorded a transition of communities from upstream environments dominated by Bacteriodetes members, to Actinobacteria-dominated communities downstream. Additionally, although the most abundant positively correlated Alphaproteobacteria group corresponded to the alfV-LD12 clade, overall presence remained at relatively minimal levels compared to previous observations [16, 17, 74]. These differences in compositions may be due to locational and environmental specificity.

Longitudinal trends in microbial diversity and community composition were disrupted for two out of our five surveys, suggesting that other, landscape-scale processes may be impacting stream microbial community assembly. While we do not currently have sufficient data to infer specific causes for these disruptions, the winter sampling date was associated with unusually low temperatures, while the summer sampling date was associated with high precipitation and stream flow (Fig. 2.16). A previous time-course study of the Upper Mississippi River in Minnesota found that the large majority of bacterial community variance was associated with rainfall and temperature [26]. Although obvious diversity trends could not be observed within these seasons, specific freshwater lineages (acI, Luna1, bacI, betI, and betII) were still the most abundant taxa

of organisms present within these samplings. These observations align with previous hypotheses of a ‘core’ bacterial community for a specified riverine system with proportions that fluctuate according to system perturbations [78].

Our results suggest that microbial communities from stream headwaters are highly diverse and variable, and that downstream movement is associated with the enrichment of a distinct core community comprising specific freshwater-associated bacterial lineages. However, this successional pattern appears to be susceptible to environmental disruption, with both diversity and community composition trends absent for two out of five watershed-level surveys. Possible contributors to this disruption include increased precipitation and temperature minima. Through continued study it may be possible to further untangle the interface between environmental variability, community structure change, and cycling of nutrients by microorganisms within the water column.

2.5 Methods

Sample Collection.

Lab members and community volunteers performed sampling of stream sites in collaboration with the Upper Oconee Watershed Network (UOWN http://www.uown.org/). UOWN is a non-profit organization of community volunteers and volunteer professionals from the Athens, GA regional area that is dedicated to promoting citizen awareness of local watersheds through education and watershed monitoring. Prior to sampling, lab members gave volunteers instruction on proper sampling procedure. Topics highlighted included the collection of samples at mid-stream and mid-depth, avoiding collection of river bottom sediment, and rinsing containers three times with sample water prior to sample collection. All volunteers were given a labeled 4 L

nucleotide free acid-washed cubitainer, and a sample collection packet containing gloves, pencil, maps, and information sheets for each assigned sampling site. Volunteer groups were typically assigned 2-4 stream sites for collection. The information sheet contained sections to record the sample site name, sample collection time, volunteer names, volunteer e-mail addresses, whether or not any of the volunteers smoked, and a notes section for additional records. The information sheet also restated detailed instructions of proper sampling protocol. The collection packets contained printed driving directions to each of the sampling sites and GPS coordinates of sampling sites for use in mobile GPS navigation applications. Following collection, samples were returned to lab members for processing within a maximum of 5 hrs following collection.

In parallel, water samples were collected in Whirl-Pak bags for chemical and biological analysis by UOWN. Data collected is available from UOWN (http://uown.org/data.html) and included measurements of conductivity, turbidity, and pH from all sites and *E. coli* coliform counts and NO_3 measurements from selected sites.

To obtain the microbial cell fractions from the water column, samples were filtered using a Masterflex peristaltic pump (Cole-Parmer, Vernon Hills, IL, USA) with LS-15 platinum-cured silicone tubing (Cole-Parmer). Water was pumped through a 5.0 µm, 47 mm diameter Durapore pre-filter (Millipore, Billerica, MA, USA), and a 0.22 µm Sterivex filter (Millipore) for approximately 10 minutes at 80 RPM. Between samples, all tubing lines were washed with ~ 150 ml of 50% ethanol followed by ~ 500 ml of ddH_2O to ensure removal of cells, debris, and chemical compounds. Filtrate was collected from each sample after ~ 250 ml of flow-through in 60 ml acid-washed HDPE bottles for measurement of TDN and TDP. Filtrate volume from each sample was measured. Filters were stored dry at -80°C. Nutrient bottles were stored at -20°C until date of analysis.

DNA Extraction, Amplicon Library Preparation, Sequencing, and Nutrient Measurements. DNA was extracted from each Sterivex filter using a combination of custom protocol and DNA extraction kit. Filters were thawed at room temperature for approximately 20 min before processing. After thawing, 1.6 ml of filter sterile lysis buffer (40 mM EDTA, 50 mM Tris, 0.73 M sucrose in ddH$_2$O) plus 4 mg of lysozyme was added to each column and incubated at 37°C for 30 min while rotating. After incubation, 100 µl of lysis buffer + proteinase K (1.5 mg) and 200 µl of 10% SDS were added and the sample was incubated at 55°C for 2 hrs while rotating. Following the second incubation, the sample was vortexed vigorously at maximum speed for 30 sec to completely release DNA. Lysate from the sample was then transferred to a 5 ml Eppendorf tube using a sterile 3 ml syringe. An equal volume of phenol:chloroform:IAA (25:24:1; pH 8.0) was added to the lysate and mixed. The sample was centrifuged for 5 min at 3500xG to separate aqueous lysate and phenol phases. The upper aqueous phase was then transferred to a new 5 ml tube, and 1/25 sample volumes of 5M NaCl and 0.7 volumes of 99.9% isopropanol were added. This sample was mixed, incubated for 10 min at room temperature, and centrifuged for 15 min at 17,000xG. Supernatant was discarded and sample DNA pellet was resuspended in 400 µl of kit elution buffer (Omega BioTek, Norcross, GA, USA). Omega BioTek E.Z.N.A. Water DNA Kit (Omega Biotek, May 2013 version) was used to process the resuspended DNA extract starting from kit protocol step #13 onward. Samples were eluted in 50 µl of elution buffer. DNA concentrations and A260/280 from complete extractions was quantified using NanoDrop Lite spectrophotomer (Thermo Fisher, Waltham, MA, USA).

Sample extractions were aliquoted and normalized to 5 ng/µl for library preparation. Sample DNA was amplified in replicate using NEB Q5 Hot Start High-Fidelity DNA

Polymerase (New England Biolabs, Ipswich, MA, USA) and V4 variable region primers 515F (5'- GTGCCAGCMGCCGCGGTAA - 3') and 806R [80] (5'- GGACTACHVGGGTWTCTAAT - 3') in 10 µl PCR reactions (1X Q5 reaction buffer, 200 µM dNTPS, 0.5 µM 515-F, 0.5 µM 806-R, 10 ng DNA, 0.02 U/µl Q5 polymerase) under the following conditions: 98°C (30 sec), followed by 15 cycles of 98°C (10 sec), 52°C (10 sec), 72°C (10 sec), a final elongation step at 72°C (2 min), and a hold at 10°C. Initial amplification reactions were used in a secondary amplification/dual barcode annealing reactions. Forward and reverse dual hamming barcode primers (primers and barcodes with different reference indices) were designed based upon primers generated by Caporoso *et.al* [81, 82] (see Table 2.1). Secondary amplification reactions were prepared with 9 µl of primary amplification product in 30 µl reactions and run under the following conditions: 98°C (30 sec), followed by 4 cycles of 98°C (10 sec), 52°C (10 sec), 72°C (10 sec), followed by 6 cycles of 98°C (10 sec), 72°C (1 min), followed by a final extension of 72°C (2 min), and a hold at 10°C. Two PCR reactions with unique barcode sets were performed in parallel for each sample to generate two technical replicates of each sample. Replicate PCR reactions were pooled and purified using Omega BioTek E.Z.N.A. Cycle-Pure Kit and were eluted in 50 µl of elution buffer. Purified amplicon libraries were quantified as described above. Sample libraries were then normalized and pooled to a concentration of 10 nM based on a predicted total product size of ~ 400 bp. Final sample library pool concentration was assessed using Qubit HS dsDNA assay (Thermo Fisher) and final library size was quantified using Agilent Bioanalyzer 2100 DNA-HS assay (Agilent, Santa Clara, CA, USA) before submission to the University sequencing center (Georgia Genomics Facility). Sample libraries were run on an Illumina MiSeq PE-250 (v2 chemistry, 500 cycles). De-multiplexed data mapped to read indices was retrieved from Illumina BaseSpace for analysis.

Raw sequence data was submitted to the NCBI Sequence Read Archive under accession number SRP075852.

Sample nutrient analysis (TDN, TDP) was completed by the Joye lab at the University of Georgia. Total content was measured using a Shimadzu TOC-Vcph (Shimadzu, Kyoto, Japan) in combination with a Shimadzu TNM-1 total nitrogen unit. The TDP standard was potassium hydrogen phthalate and the TDN standard was glycine.

Sequence Processing and Analysis.

Sequences in fastq format were processed with mothur (v1.35.1) [83], following general guidelines recommended by program authors [83-85]. Paired-end sequences were assembled and trimmed to retain an average quality score of ≥ Q25 over a 50 bp window. Sequences with ambiguous bases, repeats of > 8 bp, or lengths not within the range of 200-275 bp were removed. Sequences were aligned to the rRNA GreenGenes reference alignment (May 1999 release) [86]. The sequence alignment was screened for chimeric sequences using uchime [87], flagged chimeras (abskew = 1.9, minh = 0.3, mindiv = 0.5, xn = 8.0, dn = 1.4, xa = 1, chunks = 4, minchunk =64, id smoothing window = 32, min smooth id = 0.95, maxp = 2 , de-replicate = T) were removed from the alignment. The final alignment was classified based on the Greengenes classification database (May 2013 release) appended with curated freshwater rRNA sequences (July 2012 release) [58, 88] using the Wang method [89]. Sequences that were classified as chloroplasts or mitochondria were removed. Remaining sequences were clustered into operational taxonomic units (OTUs) at a 97% similarity level using the average nearest neighbor algorithm. Sample technical replicate counts were combined within the final OTU table. Representative OTU sequences were obtained from the end processed fasta alignment using a

maximum abundance method. All data analysis, plot generation, and diversity measures were conducted using the R statistical software program [90] and the R package vegan [91].

Identification of Freshwater-associated OTUs.

In order to identify freshwater-associated taxa, a freshwater specific 16S rRNA gene database was created. The full SILVA non-repetitive 16S rRNA arb-compatible alignment database was downloaded (July 2014 release) [92, 93] and imported into ARB (v6.0.1) [94]. Aligned species sequences within the database were filtered by "isolation source" (see Document S1 for full list of queries) to contain only rRNA sequences from organisms isolated from freshwater environments. The database was manually curated post-filtering to ensure correct isolation sources. A total of 7838 database sequences with freshwater isolation sources were obtained in this way and were appended with curated freshwater rRNA sequences [58] to increase the reference sequence pool diversity (combined total of 8986 sequences). Sequences from this database were then processed in mothur (v1.35.1) [83] in order to generate a core set of freshwater OTUs. Duplicate sequences were removed, the dataset was aligned to GreenGenes, and sequences were trimmed to the V4 region as described above. Aligned sequences with ambiguous bases, repeats of > 8bp, or lengths < 200bp were removed. The reference dataset was clustered to generate OTUs at 95% I.D. (2430 OTUs). Representative OTU sequences were generated from the 95% I.D. reference dataset using the maximum abundance method. From these representatives, a core set of freshwater OTUs was generated by selecting OTUs with counts > 10 (103 in total). The proportion of sample OTUs matching the core freshwater OTU reference was calculated by aligning OTU representatives from the sample dataset against the

freshwater OTU representatives using USEARCH [95]. Sequences with ≥ 97% I.D. to the core freshwater OTUs were returned (34 829 sequence matches, ~ 10% of sample dataset OTUs).

Estimation of catchment area, cumulative stream length, and land use for sampling sites. Watersheds were delineated in ESRI's ArcMap™ 10.2 [96] geographic information systems (GIS) software using the tools available in the Hydrology toolbox, a one meter resolution digital elevation model (DEM), and sampling locations as latitude and longitude points. All subsequent raster layers created during this process were at a resolution of one meter. Sample coordinate points were snapped to the flow accumulation raster within two meters of the original points. The Watershed tool was used to delineate watersheds for each snapped sampling point. The delineated watersheds were manually checked to ensure proper location and, where needed, sampling points were moved and the Watershed tool was rerun to create accurate watersheds. Percent land cover was calculated for each watershed using the National Land Cover Database 2011 (NLCD 2011) [97]. The Tabulate Area tool was run on the NLCD 2011 layer using the watersheds as zones to calculate the area of each type of land cover contained within each watershed. Watershed area, stream length, and percent land cover were summed to obtain totals for watersheds at each sampling point catchment.

2.6 References

1. Nixon, S.W., et al., *The Fate of Nitrogen and Phosphorus at the Land-Sea Margin of the North Atlantic Ocean.* Biogeochemistry, 1996. **35**(1): p. 141-180.
2. Meybeck, M., *Riverine transport of atmospheric carbon: Sources, global typology and budget.* Water, Air, and Soil Pollution, 1993. **70**(1-4): p. 443-463.
3. Vannote, R.L., Minshall, G. W., Cummins, K. W., Sedell, J. R. & Cushing, C. E., *The River Coninuum Concept.* Canadian Journal of Fisheries and Aquatic Sciences, 1980(37): p. 130-137.
4. Raymond, P.A., et al., *Global carbon dioxide emissions from inland waters.* Nature, 2013. **503**(7476): p. 355-9.
5. Cole, J.J., et al., *Plumbing the Global Carbon Cycle: Integrating Inland Waters into the Terrestrial Carbon Budget.* Ecosystems, 2007. **10**(1): p. 172-185.
6. Shiklomanov, I., et al., *Water in crisis : a guide to the world's fresh water resources*. 1993, Oxford University Press: New York.
7. Madsen, E.L., *Microorganisms and their roles in fundamental biogeochemical cycles.* Curr Opin Biotechnol, 2011. **22**(3): p. 456-64.
8. Findlay, S., *Stream microbial ecology.* Journal of the North American Benthological Society, 2010. **29**(1): p. 170-181.
9. Jackson, C.R., et al., *Free-living and particle-associated bacterioplankton in large rivers of the Mississippi River Basin demonstrate biogeographic patterns.* Applied and environmental microbiology, 2014. **80**: p. 7186-7195.
10. Savio, D., et al., *Bacterial diversity along a 2600 km river continuum.* Environ Microbiol, 2015.

11. Staley, C., et al., *Species sorting and seasonal dynamics primarily shape bacterial communities in the Upper Mississippi River.* Sci Total Environ, 2015. **505**: p. 435-45.

12. Read, D.S., et al., *Catchment-scale biogeography of riverine bacterioplankton.* ISME J, 2015. **9**(2): p. 516-26.

13. Besemer, K., et al., *Headwaters are critical reservoirs of microbial diversity for fluvial networks.* Proceedings. Biological sciences / The Royal Society, 2013. **280**: p. 20131760.

14. Besemer, K., et al., *Complexity of bacterial communities in a river-floodplain system (Danube, Austria).* Appl Environ Microbiol, 2005. **71**(2): p. 609-20.

15. Hu, A., et al., *Response of bacterial communities to environmental changes in a mesoscale subtropical watershed, Southeast China.* Sci Total Environ, 2014. **472**: p. 746-56.

16. Crump, B.C., L.A. Amaral-Zettler, and G.W. Kling, *Microbial diversity in arctic freshwaters is structured by inoculation of microbes from soils.* ISME J, 2012. **6**(9): p. 1629-39.

17. Crump, B.C., et al., *Biogeography of bacterioplankton in lakes and streams of an Arctic tundra catchment.* Ecology, 2007. **88**: p. 1365-78.

18. Newton, R.J., et al., *A guide to the natural history of freshwater lake bacteria.* Microbiol Mol Biol Rev, 2011. **75**(1): p. 14-49.

19. Lindström, E.S., M.P. Kamst-Van Agterveld, and G. Zwart, *Distribution of typical freshwater bacterial groups is associated with pH, temperature, and lake water retention time.* Appl Environ Microbiol, 2005. **71**(12): p. 8201-6.

20. Zwart, G., et al., *Typical freshwater bacteria: an analysis of available 16S rRNA gene sequences from plankton of lakes and rivers.* Aquatic Microbial Ecology, 2002. **28**: p. 141-155.

21. Ghai, R., et al., *Metagenomics of the Water Column in the Pristine Upper Course of the Amazon River.* PLoS ONE, 2011. **6**(8): p. e23785.

22. Crump, B.C. and J.E. Hobbie, *Synchrony and seasonality in bacterioplankton communities of two temperate rivers.* Limnology and Oceanography, 2005. **50**(6): p. 1718-1729.

23. Zhang, M., et al., *Structure and seasonal dynamics of bacterial communities in three urban rivers in China.* Aquatic Sciences, 2011. **74**(1): p. 113-120.

24. Crump, B.C., et al., *Circumpolar synchrony in big river bacterioplankton.* Proc Natl Acad Sci U S A, 2009. **106**(50): p. 21208-12.

25. Schwarzenbach, R.P., et al., *The Challenge of Micropollutants in Aquatic Systems.* Science, 2006. **313**(5790): p. 1072-1077.

26. Lu, S., et al., *Sequencing Insights into Microbial Communities in the Water and Sediments of Fenghe River, China.* Arch Environ Contam Toxicol, 2016.

27. Staley, C., et al., *Application of Illumina next-generation sequencing to characterize the bacterial community of the Upper Mississippi River.* J Appl Microbiol, 2013. **115**(5): p. 1147-58.

28. Gilbert, J.A., et al., *The seasonal structure of microbial communities in the Western English Channel.* Environmental microbiology, 2009. **11**: p. 3132-9.

29. Caporaso JG, L.C., Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, and K.R. Fierer N, *Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample.* Proc. Natl. Acad. Sci., 2011. **108, supp 1**: p. 4516-4522.

30. Caporaso, J.G., et al., *Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms.* ISME J, 2012. **6**(8): p. 1621-4.

31. Hamady, M., et al., *Error-correcting barcoded primers allow hundreds of samples to be pyrosequenced in multiplex.* Nature methods, 2008. **5**(3): p. 235-237.

32. Schloss, P.D., et al., *Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities.* Appl Environ Microbiol, 2009. **75**(23): p. 7537-41.

33. Kozich JJ, W.S., Baxter NT, Highlander SK, Schloss PD, *Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform.* Applied and Environmental Microbiology, 2013. **79(17)**: p. 5112-5120.

34. Schloss, P., Westcott, SL. *MiSeq SOP*. 2014 December 19, 2014 [cited 2014 July 10, 2014]; Available from: http://www.mothur.org/wiki/MiSeq_SOP.

35. McDonald, D., et al., *An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea.* ISME J, 2012. **6**(3): p. 610-8.

36. Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C., Knight, R., *UCHIME improves sensitivity and speed of chimera detection.* Bioinformatics, 2011. **27**: p. 2194-2200.

37. McMahon, K.D., Newton, R.J. *FWMFG - Freshwater microbial field guide*. Sep 17, 2014 [cited 2014 October 9, 2014]; Available from: https://github.com/mcmahon-uw/FWMFG.

38. Wang, Q., et al., *Naïve Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy.* Applied and Environmental Microbiology, 2007. **73**(16): p. 5261-5267.

39. R Core Team, *R: A language and environment for statistical computing.* 2014, R Foundation for Statisical Computing: Vienna, Austria.

40. Oksanen, J., et al., *vegan: Community Ecology Package*. 2013.

41. Quast, C., et al., *The SILVA ribosomal RNA gene database project: improved data processing and web-based tools.* Nucleic Acids Research, 2013. **41**(D1): p. D590-D596.

42. Quast, C., et al. *Silva High Quality Ribosomal RNA Databases: File Repository*. [Online Database] 2015 July, 15, 2015 [cited 2015 February, 20]; File Repository]. Available from: http://www.arb-silva.de/no_cache/download/archive/release_119/ARB_files/.

43. Ludwig, W., et al., *ARB: a software environment for sequence data.* Nucleic Acids Research, 2004. **32**(4): p. 1363-1371.

44. Edgar, R.C., *Search and clustering orders of magnitude faster than BLAST.* Bioinformatics, 2010. **26**(19): p. 2460-2461.

45. ESRI, *ArcGIS Desktop: Release 10.2*. 2011, Environmental Systems Research: Redlands, CA.

46. Homer, C.G., Dewitz, J.A., Yang, L., Jin, S., Danielson, P., Xian, G., Coulston, J., Herold, N.D., Wickham, J.D., and Megown, K., *Completion of the 2011 National Land Cover Database for the conterminous United States-Representing a decade of land cover*

change information. Photogrammetric Engineering and Remote Sensing, 2015. **81**(5): p. 345-354.

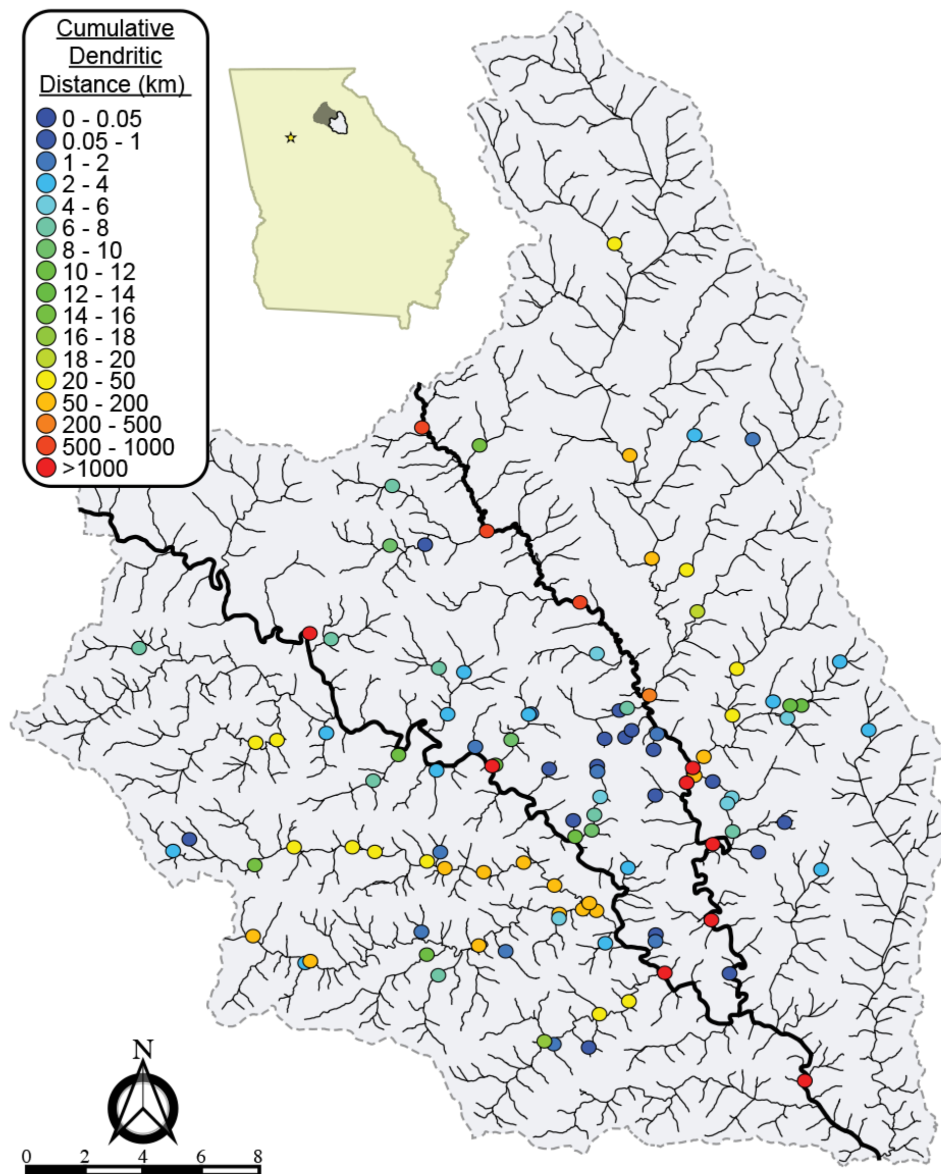

Figure 2.1. Map of Study collection sites. Points on the map represent locations of sampling sites and are colored according to cumulative dendritic distance upstream of each sampling site. River paths of the Middle Oconee River, North Oconee River, and Oconee River have been darkened to show the primary river network. Map inset displays the Georgia state border with the relative study area (light gray) and the cumulative catchment area of the study (shaded portion).

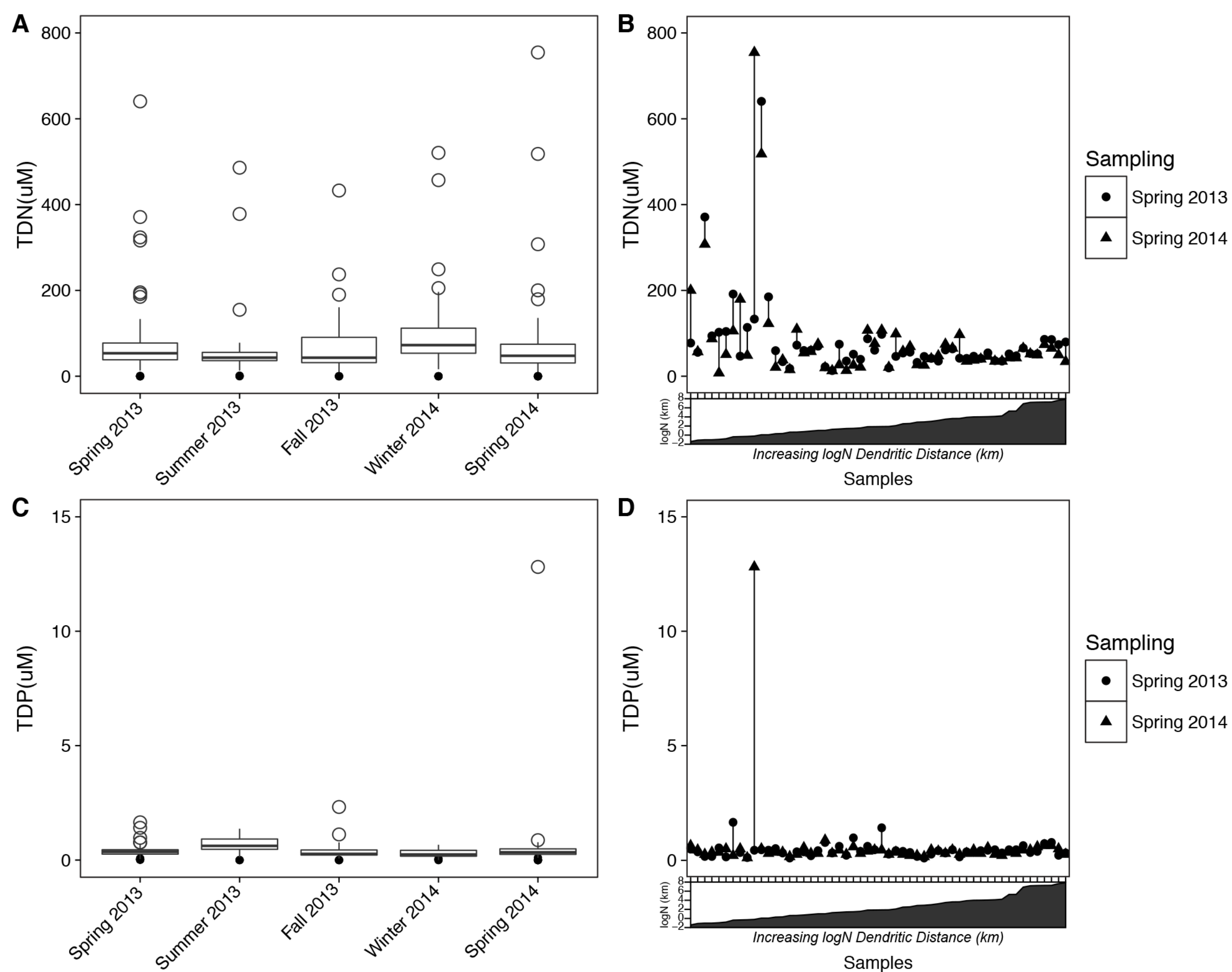

Figure 2.2. Total dissolved nitrogen (TDN) and total dissolved phosphorus (TDP) concentration comparisons between sites in all seasons (A,C) and sites in common between two Spring samplings (B,D). Outlier points are shown as open circles and experimental blanks (blanks of ddH2O filtered between stream samples) for each season are shown as solid points (A,C). Overall concentrations of nutrients did not significantly differ across seasons (Tukey HSD, p-vals ≥ 0.07) (A,C). For between-season comparisons of TDN and TDP (B,C) sample sites have been sorted according to cumulative stream length.

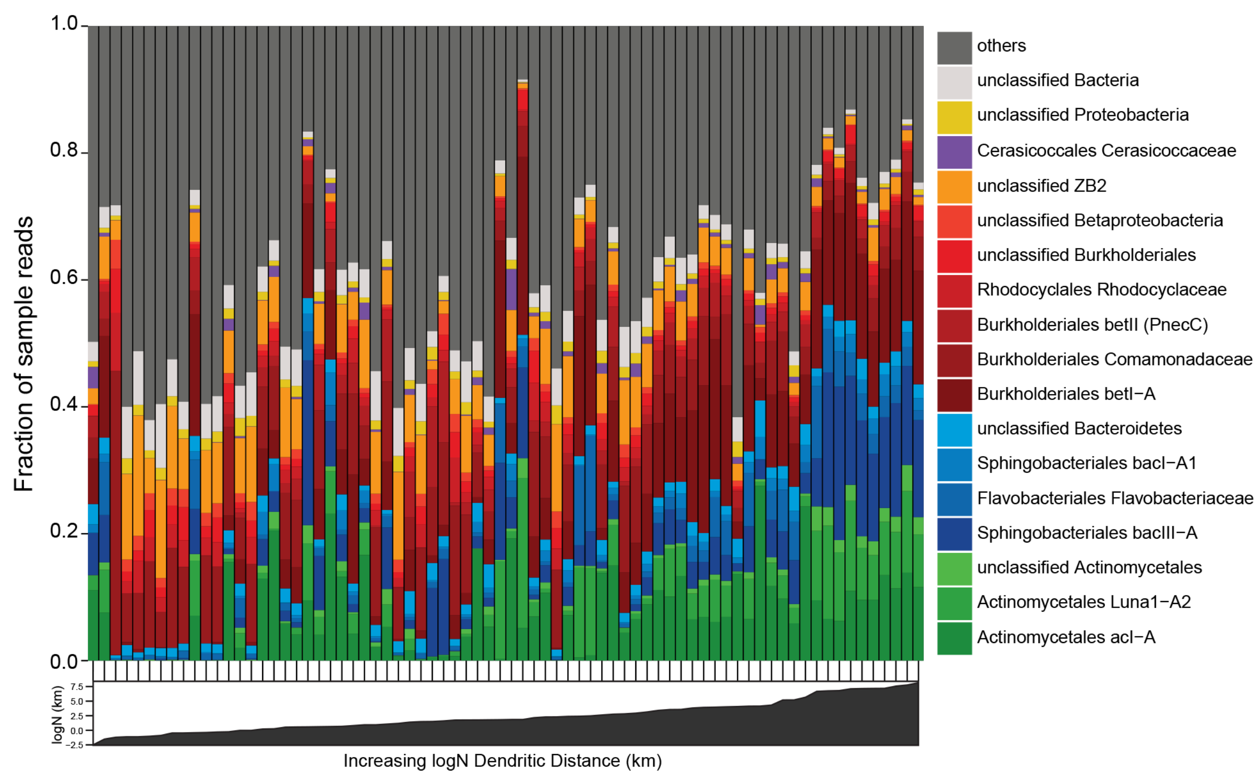

Figure 2.3. Proportional abundance of major bacterial taxa for Spring 2013 sample collection. Families at ≥1% abundance are shown, all other OTUs are grouped into the "others" category. Fractional abundance is calculated relative to all sequences passing quality filters. Samples have been sorted according to increasing cumulative stream length. All other study dates are shown in Fig. 2.4.

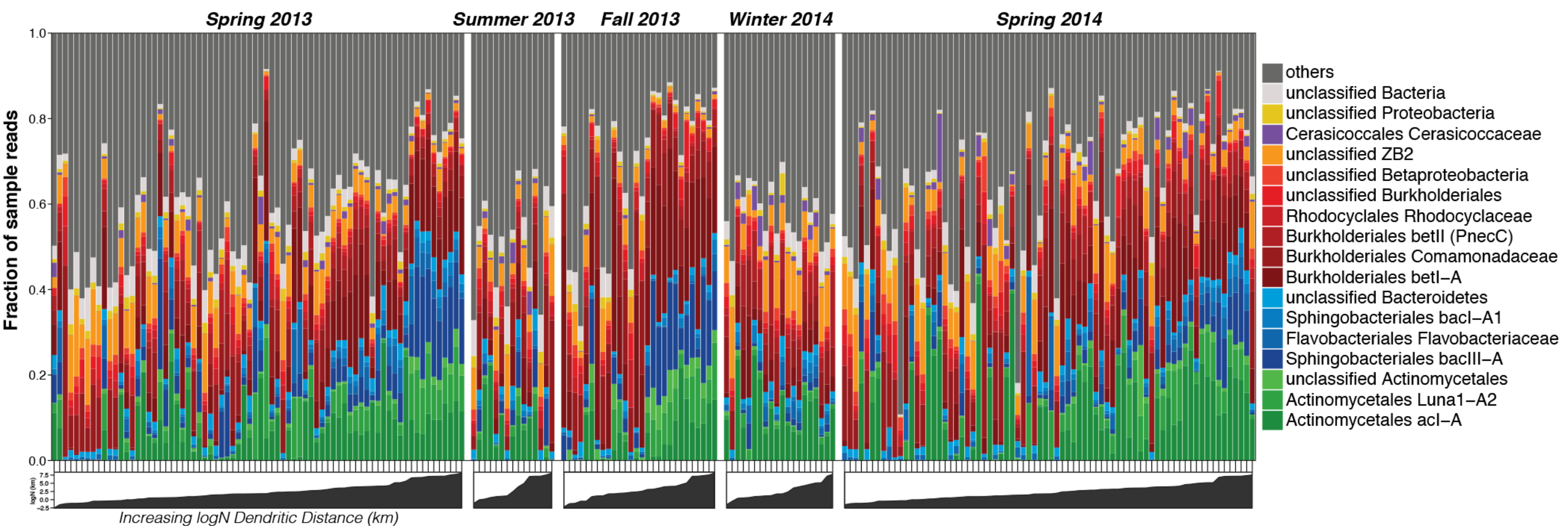

Figure 2.4. Proportional abundance of bacterial taxa across seasons. Families at ≥1% abundance are shown, all other OTUs are grouped into the "others" category. Samples in each season have been sorted according to increasing cumulative stream length.

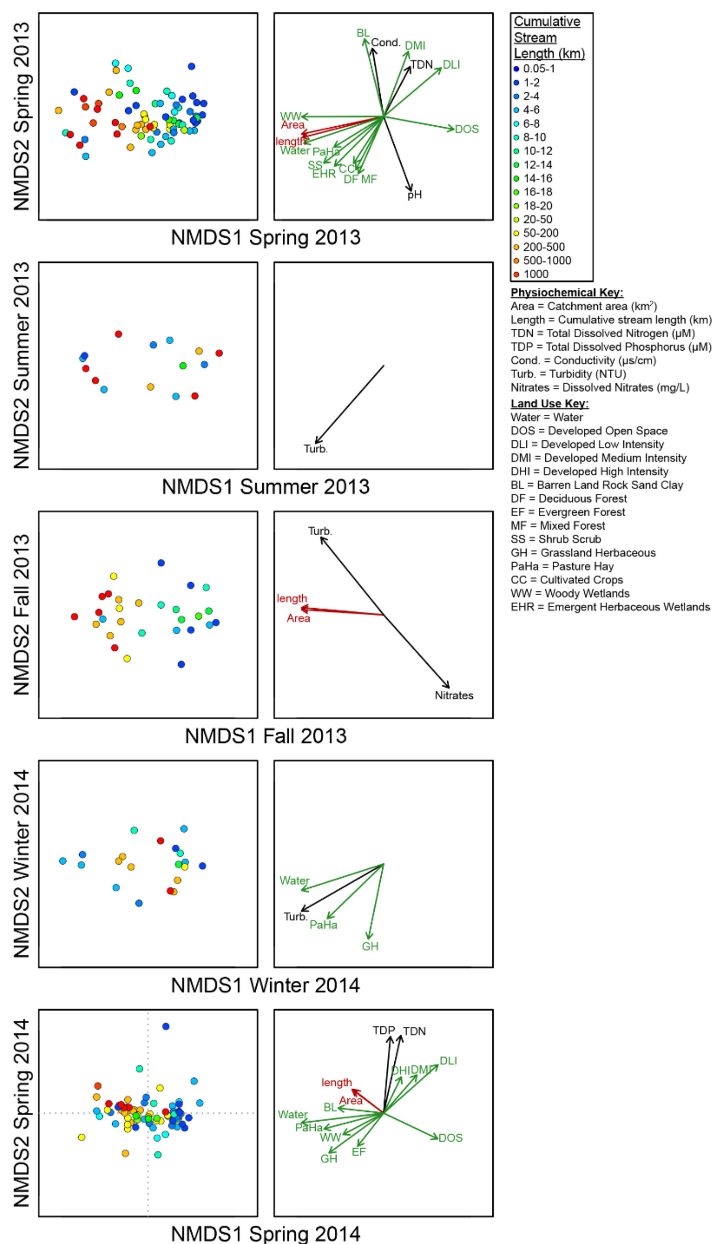

Figure 2.5. Non metric multidimensional scaling (NMDS) analyses of sampling microbial communities and correlated physiochemical and land-usage variables ($p < 0.05$). Points in NMDS are colored according to cumulative stream length. All correlated physiochemical parameters (except for sample catchment area and cumulative stream length, highlighted in red) are plotted in black, while correlated land-use variables are plotted in green. All samples have been subsampled to an even depth of 3,546 sequences.

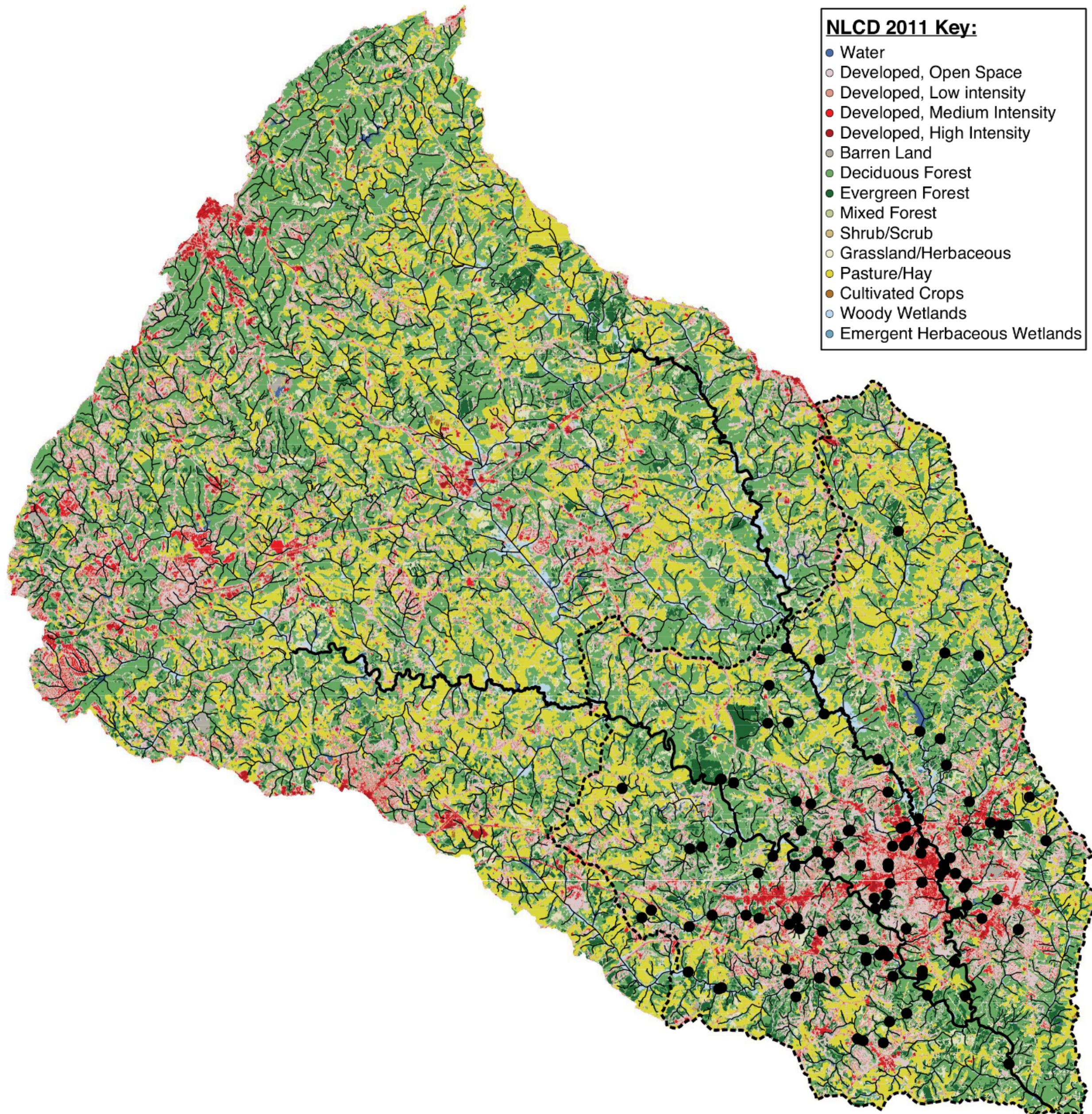

Figure 2.6. Map of land-usage of the cumulative catchment area of the study and covered sampling area (dashed section of map). Land usage is colored according to the USGS National Land Cover Dataset year 2011 (NLCD 2011). Sampling points within the study are plotted as points on the map. River paths of the Middle Oconee River, North Oconee River, and Oconee River have been darkened to show the primary river network.

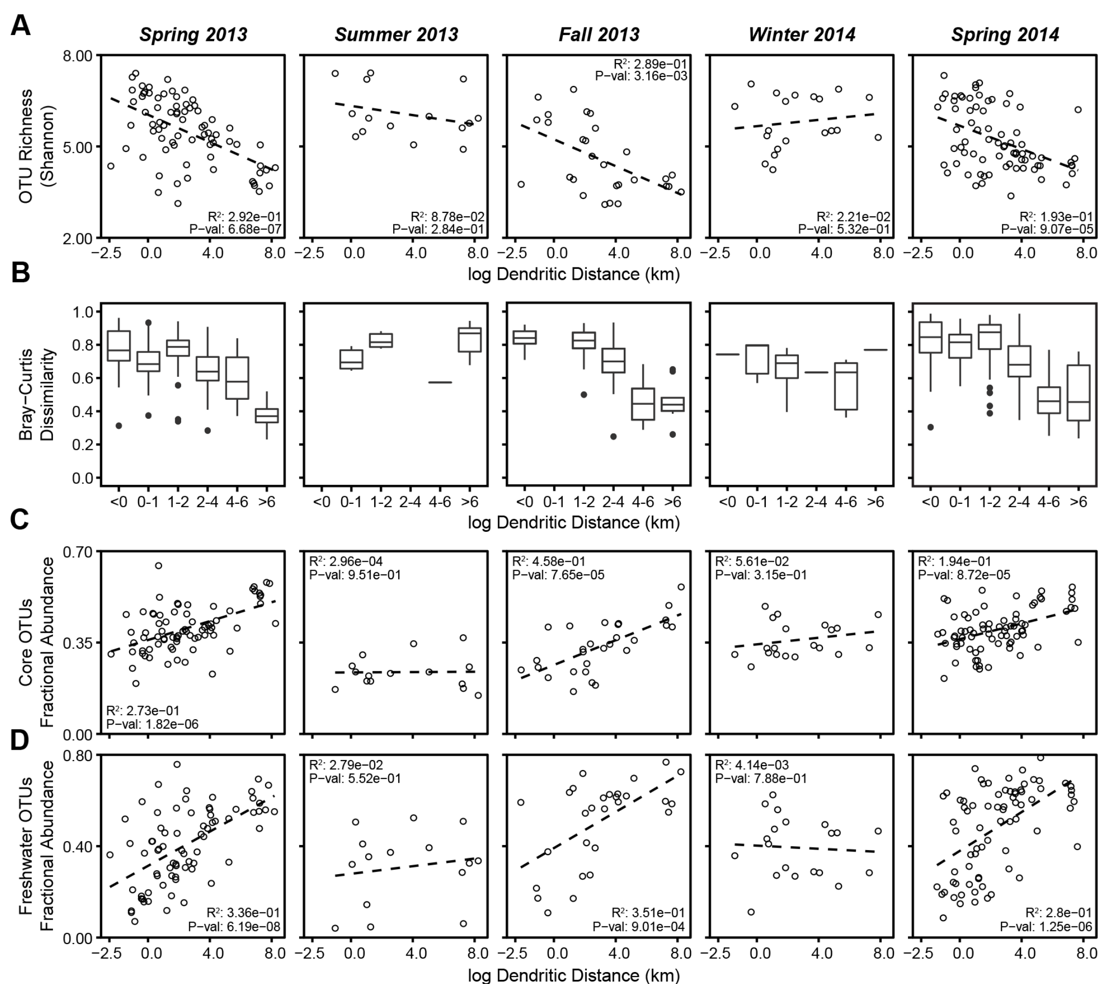

Figure 2.7. 16S rRNA gene data from all study samplings (97% I.D. OTUs, all samples resampled to 3546 sequences). Alpha diversity (Shannon index) (**A**), fraction of 'core' OTUs (**C**), and the fraction of taxa matching a core set of freshwater OTUs (**D**) are plotted against log dendritic distance as a proxy for sample site stream size. Boxplots showing beta-diversity (Bray-Curtis dissimilarity) are shown using size class groupings of stream sample comparisons (**B**). Surface plots showing beta diversity relationships with cumulative dendritic distance for all pairwise site combinations are shown in Fig. 2.9. 'Core' OTUs (**C**) were defined as OTUs present in ≥90% of all samples within the entire dataset. Bioinformatic identification of freshwater-associated OTUs (**D**) is detailed in the methods section. All trendlines were calculated using linear regression.

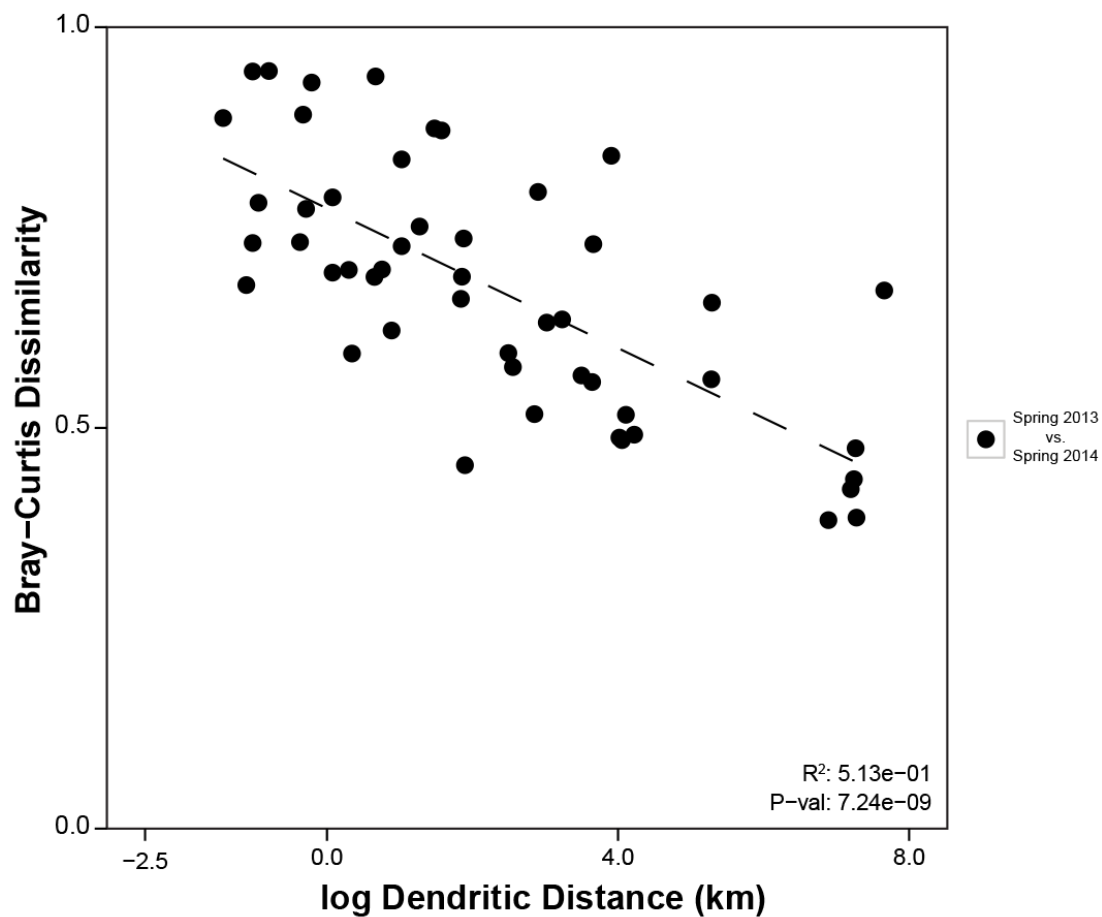

Figure 2.8. Beta-Diversity comparison (Bray-Curtis dissimilarity) of samples in common between the Spring 2013 and Spring 2014 samplings (all samples resampled to 3546 sequences) plotted against log normalized site dendritic distance. The dissimilarity measure has been used to assess samples collected from the same location within different sampling periods.

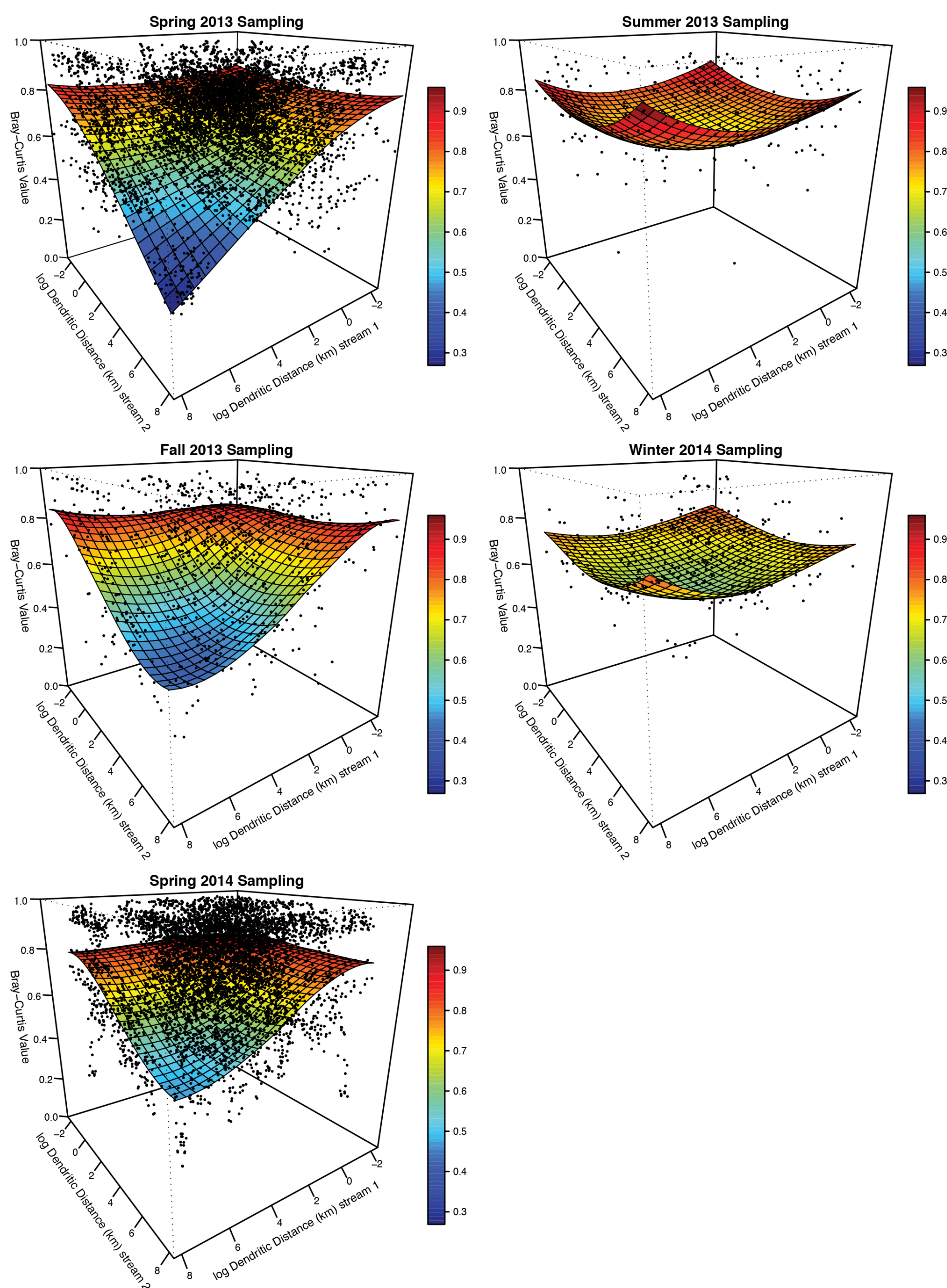

Figure 2.9. Community compositional dissimilarities (Bray-Curtis Dissimilarity) as a function of cumulative stream length (all samples resampled to 3546 sequences) from individual sampling periods. Each point represents a pairwise dissimilarity comparison (z axis) plotted against log-normalized cumulative dendritic distance of each stream (x and y axis). A locally weighted surface fit is shown and is colored according to dissimilarity values.

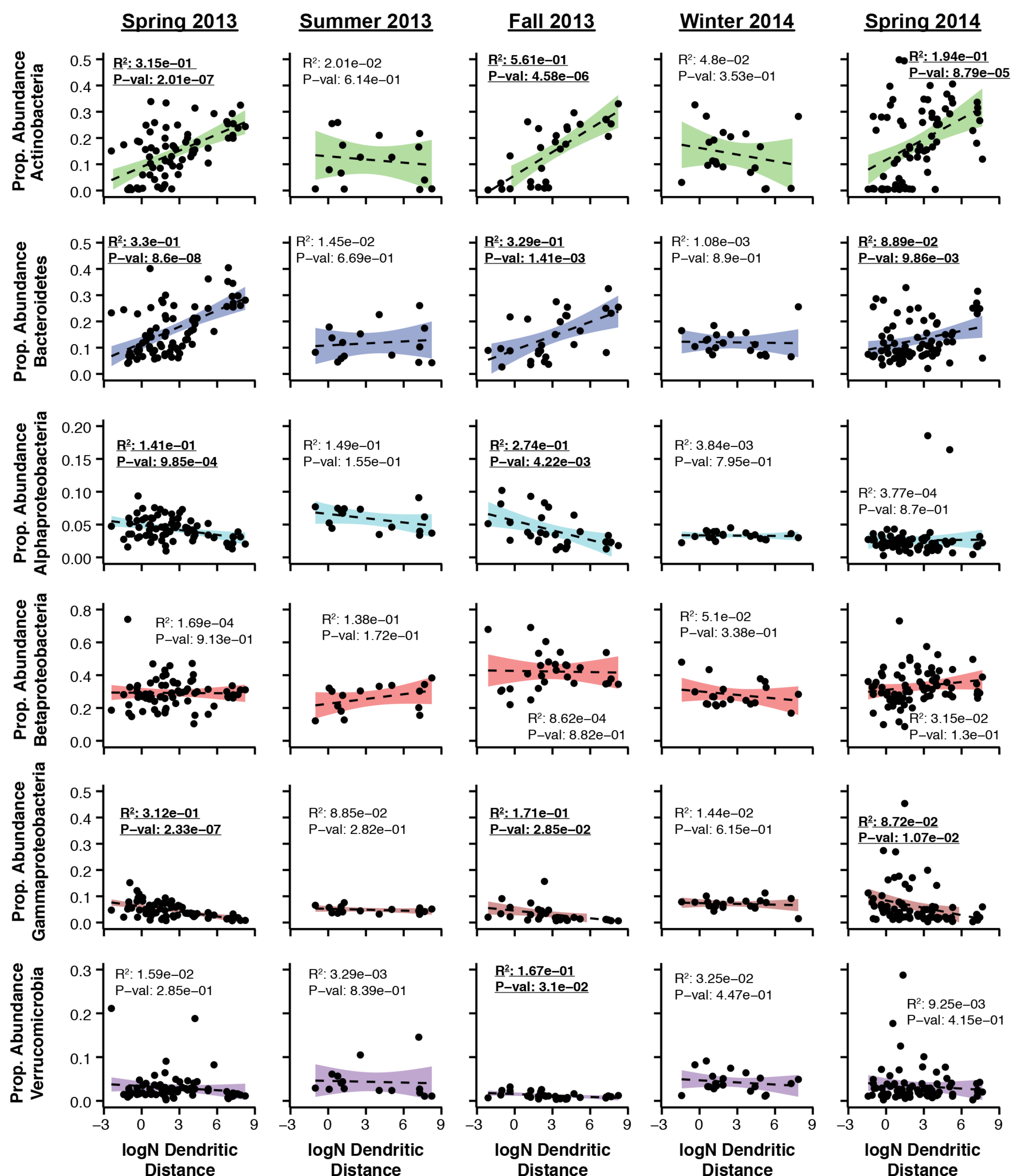

Figure 2.10. Major Phylum/Class abundances between seasons. The fraction of sample reads associated with each Phylum/Class vs. sample dendritic distance has been compared across seasons. Linear correlation fits have been calculated for each graph. Correlations significant at α=0.05 have been highlighted.

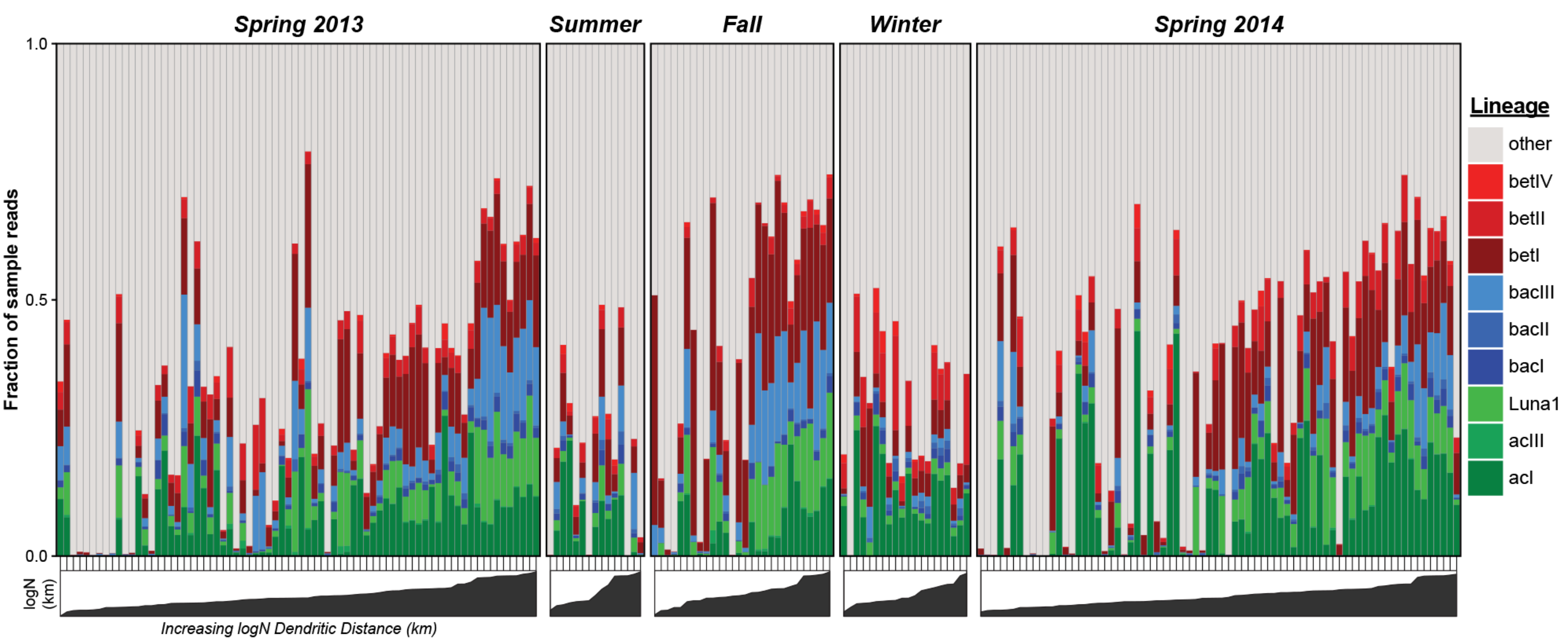

Figure 2.11. Proportional abundance of selected freshwater lineages across all quarterly watershed surveys. For each sampling date, sites have been sorted by increasing cumulative dendritic distance as a proxy for stream size. Fractional abundance is calculated relative to all sequences passing quality filters.

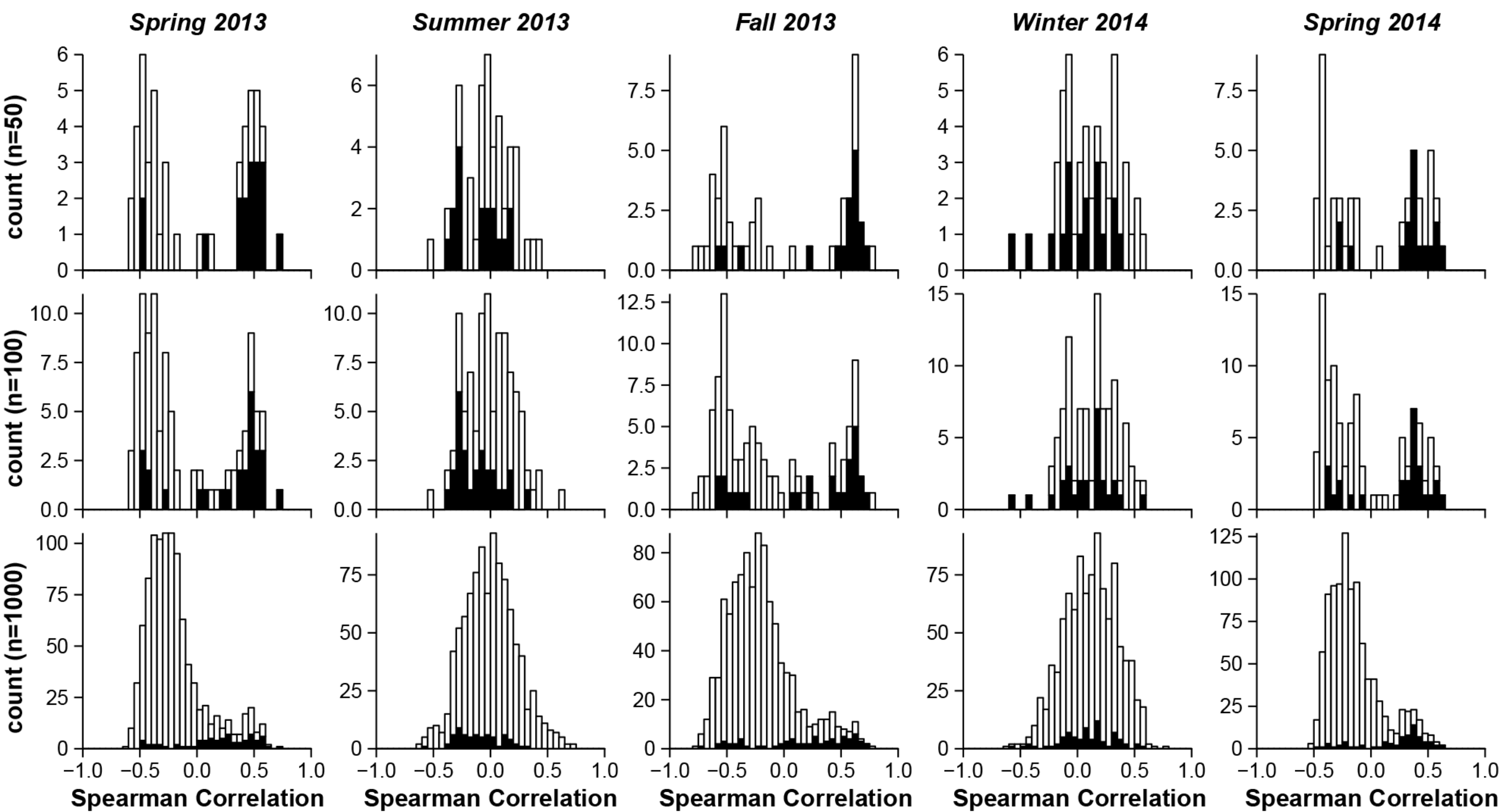

Figure 2.12. Spearman correlation histograms for each seasonal watershed survey. Spearman rank correlations were calculated by comparing proportional abundances of individual OTUs versus log normalized cumulative stream length in each sampling. Histograms show the distribution of correlation coefficients for 50, 100, and 1000 OTUs with the highest average abundance across the full dataset. OTUs identified as 'core' freshwater sequences have been have been colored in black.

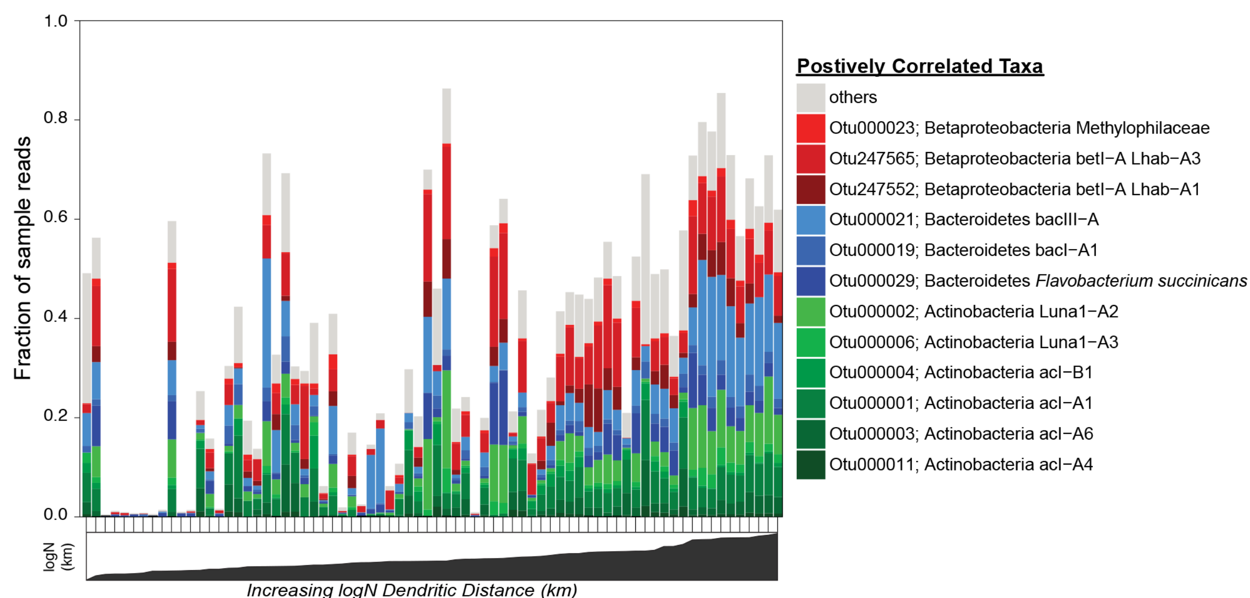

Figure 2.13. Proportional abundance of OTUs that are positively correlated with increasing stream size in the Spring 2013 sampling (Spearman $\rho \geq 0.4$, p-value < 0.05). OTUs of $\geq$ 1% of all positively correlated OTUs within the season are shown, all other positively correlated OTUs are grouped into the "others" category. Fractional abundance is calculated relative to all sequences passing quality filters. Samples have been sorted according to increasing cumulative stream length.

Figure 2.14. Proportional abundance of OTUs positively correlated (Spearman $\rho \geq 0.4$, p-value < 0.05) (A), neutrally associated (Spearman $\rho < 0.1$ and > -0.1, p-value > 0.05) (B), and negatively correlated (Spearman $\rho \leq -0.4$, p-value < 0.05) (C) with increasing stream size in all samplings. OTUs that account for ≥ 1% of all OTUs within each category (positively correlated, neutral, negatively correlated) in each sampling are shown with the exception of negatively correlated OTUs in the Spring of 2014 where OTUs at ≥ 2% of the category were shown (for visual reasons). All other OTUs in each graph are binned together in the “others” category. Samples have been sorted according to increasing cumulative stream length in each seasonal sampling.

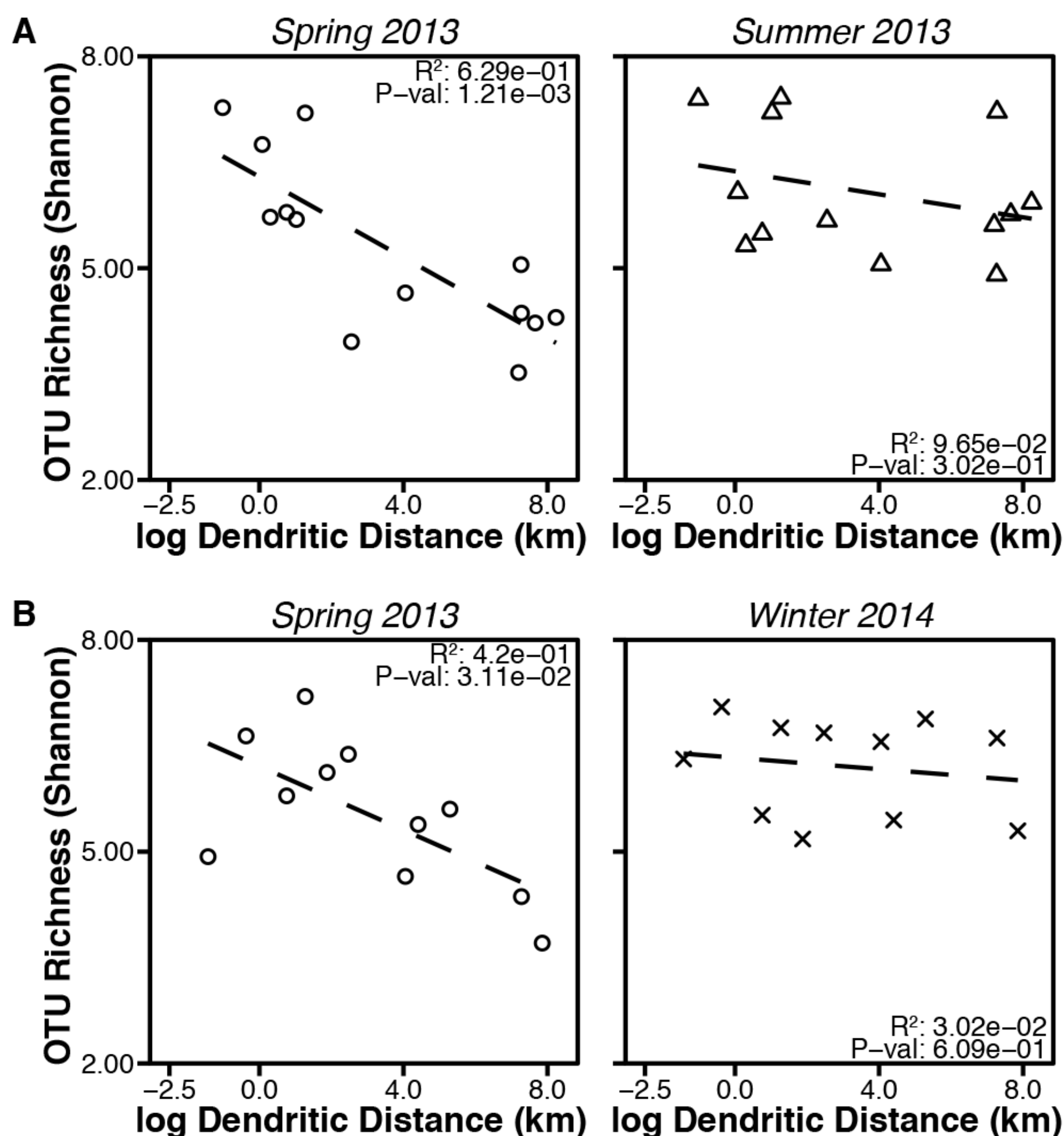

Figure 2.15. Impact of site selection and sampling depth on observed trends in species richness quarterly surveys. Relationships between cumulative dendritic distance and species richness (Shannon index, 97% I.D. OTUs, all samples subsampled to 3546 sequences) are shown using only sites sampled in both Spring 2013 and Summer 2013 (A) and sites in common between Spring 2013 and Winter 2014 (B). Spring 2013 sites show significant correlations for variable comparisons in both panels (even though sample size has been drastically limited) while Summer 2013 and Winter 2014 do not.

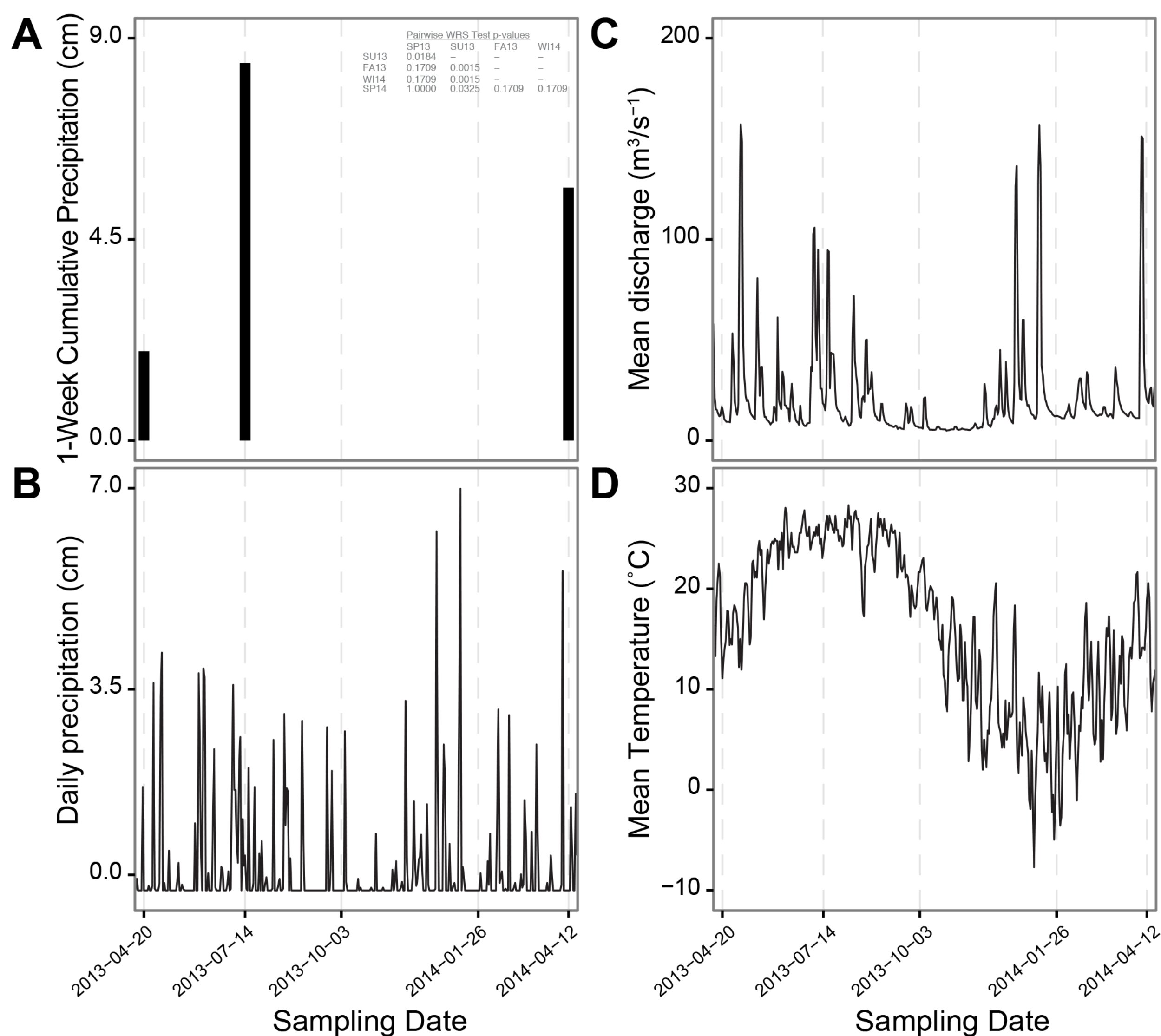

Figure 2.16. Precipitation, stream discharge, and ambient temperature during the study period. Precipitation a week prior to each sampling date was assessed from data queried from a NOAA long term site central to the sampling area (ID: GHCND:USW00013873, latitude: 33.948, longitude: -83.3275) (A), along with daily precipitation values (B), and daily ambient temperature averages (D). Stream discharge from a USGS stream monitoring location on the Middle Oconee River (USGS 02217500, latitude: 33.946641, longitude: -83.422887) is also shown (C) as a reference for water flow rates over time.

Table 2.1. Full list of primers and barcodes used in the generation of Illumina sequencing libraries.

Forward Barcode[1]	Reverse Barcode[2]
1. AACCAACC	1. GTGTGTGT
2. CCAACCAA	2. AACGAACG
3. GGTTGGTT	3. TGTCTCAC
4. TTGGTTGG	4. CCAACGTA
5. AGTCGACT	5. CGTAGCAT
6. CCATCCTA	6. TTCGTTCG
7. GTCAAGAG	7. ACACAGTC
8. TAGGTTGC	8. GAGTCAGA
9. AAGCAAGC	9. CGATGGTT
10. CGTTCGTT	10. ATCGTTGG
11. GCAAGCAA	11. TAGCAACC
12. TTCGTTCG	12. GCTACCAA
13. AGGTGAAC	13. CACTGAGT
14. CTACAGCA	14. AGTGTCTG
15. GACACTGT	15. TCACAGAC
16. TCTGTGTC	16. GTGACTCA

[1] 515F Barcode Primer: 5’ - AATGATACGGCGACCACCGAGATCTACAC XXXXXXXX TATGGTAATT CA GTGCCAGCMGCCGCGGTAA – 3’;

[2] 806R Barcode Primer: 5’ - CAAGCAGAAGACGGCATACGAGAT XXXXXXXX AGCAGCTCCAG AC GGACTACHVGGGTWTCTAAT – 3’

Initial amplification done with 515F and 806R primers:

515F V4 Primer: 5' - GTGCCAGCMGCCGCGGTAA - 3'

806R V4 Primter: 5' - GGACTACHVGGGTWTCTAAT - 3'

CHAPTER 3

CONCLUSIONS

To our knowledge, this is the first study to conduct successive high-resolution surveys of bacterioplankton communities within a temperate stream network. We have gained unique insights into longitudinal trends in microbial diversity and community composition and identified landscape-scale dynamics in stream microbial community assembly. Our evidence suggests that although headwaters remain taxonomically diverse, as these streams progress to rivers, a consistent subset of known freshwater taxa increase within the water column while overall diversity decreases. However, this selection is subject to disruption and was shown to be highly dynamic.

Based on the information gathered in our study of the pelagic bacterial communities in streams from the Upper Oconee watershed, we have observed that the most important factors influencing the assembly of these communities are the physical structure of the network and its associated hydrology. We assessed multiple physiochemical parameters and land-use estimates for every site within the sampling periods, but the most consistent variable associated with a given community was its position along the stream flow path. Although some land-use estimates and physiochemical measurements correlated with community variation, these parameters' axes of variation corresponded closely to sample stream size. These relationships were attributed to autocorrelation due to land use sampling bias and could not be distinguished as true affects.

The majority of samplings exhibited declining taxonomic diversity gradients and enrichment of a specific subset of known freshwater taxa with increasing cumulative stream

length. However, two sample sets deviated from this pattern. One of these samplings (Summer 2013) was associated with increased discharge rates from precipitation while the other was associated with lower than average temperature (Winter 2014). We have hypothesized that precipitation resulted in an increased rate of organism dispersal via increased runoff and discharge, which negated normal selection processes. In contrast, the sampling associated with lower temperatures (Winter 2014) may have been affected by slower turnover rates of organisms within the system, causing neutral processes to dominate and nullify selection. However, more sampling observations under similar conditions are needed to make any definitive conclusions on these hypotheses.

We have found that pelagic microbial communities are highly responsive to landscape scale variables and gained new insights into the assembly of these communities in streams. The community residence time, inferred from cumulative stream length, is the strongest observed predictor of microbial community assembly. However, it may be possible that other parameters play subtler roles in shaping community structure and are masked by the overwhelming effect of in-stream community residence time. It may also be possible that other parameters that were not measured within the study (light gradients, metal concentrations, groundwater inputs, C source types present, etc.) contributed significantly to community variance in the network. Further work on the system should include additional sampling time points, measurements of more extensive environmental metadata, estimations of pelagic community transcriptional activity along the network via transcriptomics, and microcosm experiments of stream water to test effects of isolated variables on community structure. This added data may provide more meaningful information into the intricacies of pelagic microbial community assembly in streams.