# MICROBIAL COMMUNITY COMPOSITION OF A NATURAL SEDIMENT SALINITY GRADIENT: TAXONOMIC AND METABOLIC PATTERNS AND CONTROLLING FACTORS

Thomas C. Jeffries

B.Sc. Hons.

A THESIS SUBMITTED FOR THE DEGREE DOCTOR OF PHILOSOPHY

School of Biological Sciences, Flinders University, Adelaide, Australia

August 2011

**Contents**

## Summary

This thesis investigates the distribution of microbial taxonomy and metabolism along a continuous natural gradient of salinity and nutrient concentration, the Coorong lagoon, Australia. By applying Next-Generation DNA sequencing techniques, I use this system as a model to observe the relative influence of local habitat variability on sediment microbial community structure. I also use the Coorong as a reference point to determine global scale determinants of metagenomic patterns in microbial diversity. My data demonstrated strong shifts in the abundance of both bacterial and archaeal taxonomic groups along the gradient coupled to an overrepresentation of genes involved in halotolerance and photosynthesis in the most hypersaline samples relative to the marine salinity samples used as a baseline. Whilst these gradient driven shifts indicate the influence of salinity and nutrient content on microbial community structure, the overall genomic signature of the community remained conserved along the gradient. When this signature was compared to other metagenomes from a variety of habitats and salinities, Coorong samples were most similar to other sediment and soil habitats which formed a discrete 'sediment' cluster regardless of salinity variation. This indicates for the first time the fundamental role of substrate type in determining microbial community metabolism and highlights the hierarchical nature of variables acting on different scales of community organization.

## Declaration

I declare that this thesis does not contain any material previously submitted for any diploma or degree in any university without acknowledgement, and that to the best of my knowledge it does not contain any material previously published by any other person except where due reference is made.

Thomas Charles Jeffries

**Acknowledgements**

Thank you to my Primary Supervisor, Assoc. Prof. Jim Mitchell for his guidance, mentoring and support during the course of this project and for giving me every opportunity to pursue my work and career at the highest (and most expensive) possible level. I appreciate it. You have taught me much. I am also deeply grateful to Dr. Justin Seymour for his guidance and advice on all things thesis and career. I could not have done this without your help. Also the entire laboratory deserves my debt of gratitude for their support and humor. A particular mention needs to go to Ben and Trish Roudnew for all of the chats, Renee Smith for working out a whole ton of software with me, and most of all to Kelly Newton who has been here with me since the beginning. Again, I could not have done this without you. I am also grateful to all of my colleagues, collaborators and co-authors for their help and feedback.

Also thank you to my parents who have supported my scientific endeavors since buying me my first microscope and chemistry set, and who have supported me in every way throughout my life and career, and to my entire family for everything. I could not ask for better parents, brothers, sisters, nieces, nephews and dogs.

The biggest thank you of all however goes to Kimberley. Your love, patience, faith, support and comfort means the world to me. I can't wait to share every day of the future with you.

Cheers all

This thesis is dedicated to Kimberley

*Now this is not the end. It is not even the beginning of the end. But it is, perhaps, the end of the beginning.*

Sir Winston Churchill 1942

*I do not know what I may appear to the world, but to myself I seem to have been only like a boy, playing on the seashore…whilst the great ocean of truth lay all undiscovered before me.*

Sir Isaac Newton 1642-1727

# GENERAL INTRODUCTION

## Microbial biogeography in the age of ecogenomics

Microbial biogeography is commonly conceptualized using the Baas-Becking hypothesis (1, 2), that "everything is everywhere but the environment selects", which is interpreted as meaning there is a cosmopolitan distribution of prokaryotic species from which certain taxa may become abundant in response to localized physiochemical parameters. This proposed cosmopolitan distribution implies that the overall diversity of microbes present is the same in all habitats, but that sampling detects the most abundant and active members of a community, that which in turn reflects those best adapted to the current ecological state. Effectively, this says the majority of species are present at an abundance level below the detection limit of traditional technologies, and at the detection limit of current technologies.

An exception to the Baas-Becking hypothesis appears to be extreme habitats (19, 26), where it appears that some taxa are not cosmopolitan, however, generally speaking the Baas-Becking hypothesis is applicable given a sufficient depth of sampling. For ocean microbial biogeography in particular, where given enough time Atlantic water becomes Pacific water, the reality is unlikely to be the binary concept of presence or absence, nor, to take the Baas-Becking hypothesis to a heuristic extreme, that all microbial species are found in a milliliter of seawater. Instead, it seems more likely that each species or strain dies out in many places while thriving in many others, which can be interpreted as continuing shifts in the relative abundance of operational taxonomic units or microbial genes in response to ecological conditions on varying scales, rather than presence or absence of given taxa in a habitat. One way to begin to resolve actual microbial dynamics is not to look at them in a uniform environment, but instead to

examine the dynamics across gradients that approach the biogeographical scale as is done in this thesis.

Only recently has serious investigation of the Baas-Becking hypothesis been possible. The recent development of high-throughput DNA sequencing platforms has led to a revolution in the extent to which a microbial community can be described, and has led to fundamental new insights into the biogeography of microorganisms. Deep-sequencing of the 16S ribosomal DNA gene subunit has allowed the application of this taxonomic marker to be extended beyond the dozen or so clones traditionally sequenced in libraries to allow for thousands to hundreds of thousands of sequences to be analyzed (25). This captures a wider breadth of the diversity of the microbes in a habitat and identifies rare organisms in the latent 'rare-biosphere' (23). At the extreme of attempting to capture the breadth of metabolic function as well as taxonomy is metagenomics, the shotgun sequencing of genomic DNA fragments from the collective 'metagenome' of the microbial community. This has determined the taxonomic structure and metabolic potential of assemblages (10, 12, 28) and has ushered in the possibility of genetic analysis of microbes at the ecosystem scale

These collective tools, often referred to as 'ecogenomics', have been used to compare the microbial community structure of different habitats and elucidate new biogeographical patterns in community composition. When these patterns are correlated to environmental parameters measured at the time of sampling, an explanatory and mechanistic view of how the 'environment selects' for genes and species can be elucidated.

Comparisons of metagenomes from a variety of habitats have shown that the overall functional potential of microbial communities is broadly determined by the biome from which the sample is

derived, with samples clustering into specific habitat groups (4, 27). This indicates that the local physiochemical parameters of the habitat are fundamental determinants of genetic profiles. Within the ocean biome, the most well studied habitat to date, global-scale spatial patterns in gene abundance correlate to differences in temperature and sunlight, indicating the role of climate in determining functional potential (20). Genes specific to phosphate utilization have also been shown to vary along nutrient gradients on this scale (21) and within the Pacific Ocean (11), however the overall functional signature within the ocean shows little variability along gradients reflecting the core processes central to life in the surface ocean, such as photosynthesis, DNA replication, protein synthesis and carbohydrate metabolism. On local scales however, individual metagenomic profiles show strong vertical zonation of taxonomic groups and specific metabolic categories, concurrent with stratified physiochemical parameters such as light, oxygen and temperature (3).

A detailed understanding of taxonomic patterns, which encompasses the rare organisms present in the sample, has been provided by high throughput sequencing of the 16S rDNA gene (25). Salinity appears to be the primary determinant of patterns in 16S rDNA phylotype distribution globally (16, 24) with the substrate type, whether a sample comes from water or sediment, also being an important factor. The role of salinity is potentially due to the requirement of cells to evolve specialist cellular machinery to survive osmotic stress (18).

The current view of microbial biogeography emerging through use of next-generation sequencing techniques is a complex one. Extreme habitats appear to show some endemism of taxa and community structure (19, 26). Some taxonomic patterns also demonstrate distance effects that can be explained by the legacy of historical processes such as dispersal limitation (8,

17). For most metagenomic and high-throughput sequenced 16S rDNA datasets investigated to date however, biogeographic patterns seem to be determined by the influence of various local contemporary conditions on varying scales (4, 5, 7, 9, 11, 14, 20, 21). I hypothesize that these various determinants of community composition are not mutually exclusive, and that the overall profile of the community represents the simultaneous influence of many variables on the overall signature of the metagenomes and on individual taxa and metabolic processes within that signature. Put in the context of the Baas-Becking hypothesis, the metagenome as a whole is a discrete unit on which 'the environment selects' (6) and individual genes and taxa within the community are also selected for by local conditions. In reality individual genes are passed among microbes creating continually changing gene sets rather than fixed units.

**The Coorong: a model system for microbial biogeography**

Physicochemical gradients provide natural model systems for investigating the influence of environmental variables on microbial community structure. A unique natural continuous salinity gradient, ranging from brackish to hypersaline salinities occurs in the Coorong, a temperate coastal lagoon located at the mouth of the Murray River, Australia's longest river system. In recent decades drought and increased irrigation demands from the Murray river have reduced freshwater flows at the estuarine end of the gradient, resulting in markedly increased salinity levels and a strong continuous salinity and nutrient gradient along the 100 km long lagoon (13, 15, 22). The lagoon is defined by a unique combination of water inputs that result in a mixture of fresh river water, groundwater, terrestrial runoff, coastal seawater and hypersaline brine. Thus, microbes are dispersed into the system from a variety of sources where they are then exposed to

the contemporary gradient in salinity and nutrients along the lagoon, providing an ideal habitat to investigate the influence of habitat variability on microbial community structure.

**Overview of the thesis**

In this thesis I use the Coorong lagoon as a model to observe the relative influence of local habitat variability on taxonomic and metabolic structure, using next-generation sequencing tools to access microbial diversity. I also use the Coorong as a reference point to determine global scale determinants of microbial metagenomic distribution.

Specifically the aims are as follows:

1. To determine the extent to which sediment microbial community taxonomic composition changes with physiochemical parameters along gradients of salinity and nutrients, and to identify which taxonomic groups demonstrate the largest shifts.

2. To determine the extent to which community composition shifts that do occur along the gradient are functionally driven by underlying shifts in the abundance of metabolic gene categories.

3. To provide novel insight into localized microbial adaptation to habitat variability at the genetic level by determining which metabolic categories shift in response to continuous gradients of salinity and nutrients.

Our four sampling points are reference stations within an overall sampling scheme employed by our laboratory and other groups from various institutions investigating the Coorong (e.g. 15). Previous work (22) has shown these sites to be characteristic of different physiochemical regions

of the lagoon and to harbour distinct pelagic microbial communities. Thus, the increment in salinity between each site is not uniform. Each sample thus represents a discrete habitat within the overall continuum of the physiochemical gradients present with clear but varying differences in salinity and other variables such as nutrient content and microbial abundance (Table 1).

Each chapter of the thesis is formatted as a manuscript for journal submission, each addressing a specific question and aim. Thus there is some redundancy in the introduction and methods of each chapter, which was necessary to make each a complete manuscript. Chapters 1 and 2 employee tag encoded FLX amplicon pyrosequencing of the 16S rDNA gene (TEFAP). There is a separate dataset for each chapter: a bacterial dataset for chapter 1 and an archaeal dataset for chapter 2. Chapters 3 and 4 utilize the same dataset: metagenomes from the four sampling sites. In chapter 3 differences between these four metagenomes are examined. In chapter 4, these metagenomes are compared to a plethora of metagenomes from diverse habitats. Chapter 5 then takes this larger dataset, from chapter four, and further explores the parameters which drive the relationships between habitats observed in the previous chapter. The thesis as a whole is conceptually divided into two sections; one that deals with the influence of salinity and nutrients on Coorong communities (chapters 1,2 and 3) and one that uses the Coorong as a model to investigate substrate partitioning within globally distributed metagenomes (chapters 4 and 5). The thesis is structured this way to elucidate the hierarchical controlling factors of Coorong community composition on the local and global scale.

## References


1. **Baas Becking, L. G. M**. 1934. Geobiologie of inleiding tot de milieukunde.

2. **De Wit, R., and T. Bouvier.** 2006. 'Everything is everywhere, but, the environment selects'; what did Baas Becking and Beijerinck really say? Environmental Microbiology **8**:755-758.

3. **DeLong, E. F., C. M. Preston, T. Mincer, V. Rich, S. J. Hallam, N. U. Frigaard, A. Martinez, M. B. Sullivan, R. Edwards, B. R. Brito, S. W. Chisholm, and D. M. Karl**. 2006. Community genomics among stratified microbial assemblages in the ocean's interior. Science **311**:496-503.

4. **Dinsdale, E. A., R. A. Edwards, D. Hall, F. Angly, M. Breitbart, J. M. Brulc, M. Furlan, C. Desnues, M. Haynes, L. L. Li, L. McDaniel, M. A. Moran, K. E. Nelson, C. Nilsson, R. Olson, J. Paul, B. R. Brito, Y. J. Ruan, B. K. Swan, R. Stevens, D. L. Valentine, R. V. Thurber, L. Wegley, B. A. White, and F. Rohwer**. 2008. Functional metagenomic profiling of nine biomes. Nature **452**:629-U8.

5. **Dinsdale, E. A., O. Pantos, S. Smriga, R. A. Edwards, F. Angly, L. Wegley, M. Hatay, D. Hall, E. Brown, M. Haynes, L. Krause, E. Sala, S. A. Sandin, R. V. Thurber, B. L. Willis, F. Azam, N. Knowlton, and F. Rohwer**. **2008**. Microbial Ecology of Four Coral Atolls in the Northern Line Islands. PLoS One **3**: e1584.

6. **Doolittle, W. F., and O. Zhaxybayeva. 2010**. Metagenomics and the Units of Biological Organization. Bioscience **60**:102-112.

7. **Frias-Lopez, J., Y. Shi, G. W. Tyson, M. L. Coleman, S. C. Schuster, S. W. Chisholm, and E. F. DeLong**. 2008. Microbial community gene expression in ocean surface waters. Proceedings of the National Academy of Sciences of the United States of America **105**:3805-3810.

8. **Fuhrman, J. A.** 2009. Microbial community structure and its functional implications. Nature **459**:193-199.

9. **Gianoulis, T. A., J. Raes, P. V. Patel, R. Bjornson, J. O. Korbel, I. Letunic, T. Yamada, A. Paccanaro, L. J. Jensen, M. Snyder, P. Bork, and M. B. Gerstein.** 2009. Quantifying environmental adaptation of metabolic pathways in metagenomics. Proceedings of the National Academy of Sciences. **106**: 1374–1379.

10. **Gilbert, J. A., and C. L. Dupont. 2011**. Microbial Metagenomics: Beyond the Genome. Annual Review of Marine Science **3**:347-371.

11. **Hewson, I., R. W. Paerl, H. J. Tripp, J. P. Zehr, and D. M. Karl.** 2009. Metagenomic potential of microbial assemblages in the surface waters of the central Pacific Ocean tracks variability in oceanic habitat. Limnology and Oceanography **54**:1981-1994.

12. **Hugenholtz, P., and G. W. Tyson**. 2008. Microbiology: Metagenomics. Nature **455**:481-483.

13. **Kingsford, R. T., K. F. Walker, R. E. Lester, W. J. Young, P. G. Fairweather, J. Sammut, and M. C. Geddes**. 2011. A Ramsar wetland in crisis – the Coorong, Lower Lakes and Murray Mouth, Australia. Marine and Freshwater Research **62**:255-265.

14. **Kunin, V., J. Raes, J. K. Harris, J. R. Spear, J. J. Walker, N. Ivanova, C. von Mering, B. M. Bebout, N. R. Pace, P. Bork, and P. Hugenholtz.** 2008. Millimeter-scale genetic gradients and community-level molecular convergence in a hypersaline microbial mat. Mol Syst Biol **4**:198.

15. **Lester, R. E., and P. G. Fairweather**. 2009. Modelling future conditions in the degraded semi-arid estuary of Australia's largest river using ecosystem states. Estuarine Coastal and Shelf Science **85**:1-11.

16. **Lozupone, C. A., and R. Knight**. 2007. Global patterns in bacterial diversity. Proceedings of the National Academy of Sciences of the United States of America **104**:11436-11440.

17. **Martiny, J. B. H., B. J. M. Bohannan, J. H. Brown, R. K. Colwell, J. A. Fuhrman, J. L. Green, M. C. Horner-Devine, M. Kane, J. A. Krumins, C. R. Kuske, P. J. Morin, S. Naeem, L. Ovreas, A. L. Reysenbach, V. H. Smith, and J. T. Staley**. 2006. Microbial biogeography: putting microorganisms on the map. Nature Reviews Microbiology **4**:102-112.

18. **Oren, A**. 2008. Microbial life at high salt concentrations: phylogenetic and metabolic diversity. Saline Systems **4**:1-13.

19. **Papke, R. T., N. B. Ramsing, M. M. Bateson, and D. M. Ward.** 2003. Geographical isolation in hot spring cyanobacteria. Environmental Microbiology **5**:650-659.

20. **Raes, J., I. Letunic, T. Yamada, L. J. Jensen, and P. Bork.** 2011. Toward molecular trait-based ecology through integration of biogeochemical, geographical and metagenomic data. Molecular Systems Biology **7**. 473

21. **Rusch, D. B., A. L. Halpern, G. Sutton, K. B. Heidelberg, S. Williamson, S. Yooseph, D. Y. Wu, J. A. Eisen, J. M. Hoffman, K. Remington, K. Beeson, B. Tran, H. Smith, H. Baden-Tillson, C. Stewart, J. Thorpe, J. Freeman, C. Andrews-Pfannkoch, J. E. Venter, K. Li, S. Kravitz, J. F. Heidelberg, T. Utterback, Y. H. Rogers, L. I. Falcon, V. Souza, G. Bonilla-Rosso, L. E. Eguiarte, D. M. Karl, S. Sathyendranath, T. Platt, E. Bermingham, V. Gallardo, G. Tamayo-Castillo, M. R. Ferrari, R. L. Strausberg, K. Nealson, R. Friedman, M. Frazier, and J. C. Venter.** 2007. The Sorcerer II Global Ocean Sampling expedition: Northwest Atlantic through Eastern Tropical Pacific. PLoS Biology **5**:398-431.

22. **Schapira, M., M. J. Buscot, S. C. Leterme, T. Pollet, C. Chapperon, and L. Seuront.** 2009. Distribution of heterotrophic bacteria and virus-like particles along a salinity gradient in a hypersaline coastal lagoon. Aquatic Microbial Ecology **54**:171-183.

23. **Sogin, M. L., H. G. Morrison, J. A. Huber, D. Mark Welch, S. M. Huse, P. R. Neal, J. M. Arrieta, and G. J. Herndl.** 2006. Microbial diversity in the deep sea and the underexplored "rare biosphere". Proceedings of the National Academy of Sciences of the United States of America **103**:12115-12120.

24. **Tamames, J., J. Abellan, M. Pignatelli, A. Camacho, and A. Moya.** 2010. Environmental distribution of prokaryotic taxa. BMC Microbiology **10**:85.

25. **Tringe, S. G., and P. Hugenholtz.** 2008. A renaissance for the pioneering 16S rRNA gene. Current Opinion in Microbiology **11**:442-446.

26. **Whitaker, R. J., D. W. Grogan, and J. W. Taylor.** 2003. Geographic barriers isolate endemic populations of hyperthermophilic archaea. Science **301**:976-978.

27. **Willner, D., R. V. Thurber, and F. Rohwer.** 2009. Metagenomic signatures of 86 microbial and viral metagenomes. Environmental Microbiology **11**:1752-1766.

28. **Wooley, J. C., A. Godzik, and I. Friedberg.** 2010. A Primer on Metagenomics. PLoS Computational Biology **6**:e1000667.

**Table 1.** E**nvironmental data for Coorong sampling sites.**

| Sampling Site | 37 PSU | 109 PSU | 132 PSU | 136 PSU |
|---|---|---|---|---|
| Salinity (PSU) | 37 | 109 | 132 | 136 |
| pH | 8.25 | 7.85 | 7.79 | 8.05 |
| Temperature (°C) | 21 | 25 | 27 | 24 |
| Ammonia concentration (mgN/L) | 0.23 (±0.15) | 0.21 (±0.09) | 0.96 (±0.31) | 3.10 (±0.84) |
| Phosphate concentration (mgP/L) | 0.05 (±0.01) | 0.11 (±0.02) | 0.12 (±0.03) | 0.27 (±0.09) |
| Porewater bacteria concentration (per mL) | $4.8 \times 10^6$ ($\pm 6.3 \times 10^5$) | $7.4 \times 10^7$ ($\pm 8.4 \times 10^6$) | $7.2 \times 10^7$ ($\pm 4.2 \times 10^6$) | $1.5 \times 10^8$ ($\pm 1.4 \times 10^7$) |
| Porewater virus concentration (per mL) | $1.5 \times 10^7$ ($\pm 5.8 \times 10^6$) | $2.3 \times 10^8$ ($\pm 3.1 \times 10^7$) | $1.8 \times 10^8$ ($\pm 1.5 \times 10^7$) | $4.2 \times 10^8$ ($\pm 3.1 \times 10^7$) |
| Turbidity of water column (NTU) | 7 | 16 | 16 | 10 |
| Dissolved Oxygen in water column (%) | 93 | 140 | 134 | 89 |

All data was measured in sediment interstitial porewater with the exception of turbidity and dissolved oxygen which were measured in the overlying water column. ± indicates Standard error of the mean (n=3 for nutrient measures, n=5 for microbial abundances). N=nitrogen, P=phosphate, PSU=practical salinity units, NTU=Nephelometric Turbidity Units.