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)	4.8×10^6 ($\pm 6.3 \times 10^5$)	7.4×10^7 ($\pm 8.4 \times 10^6$)	7.2×10^7 ($\pm 4.2 \times 10^6$)	1.5×10^8 ($\pm 1.4 \times 10^7$)
Porewater virus concentration (per mL)	1.5×10^7 ($\pm 5.8 \times 10^6$)	2.3×10^8 ($\pm 3.1 \times 10^7$)	1.8×10^8 ($\pm 1.5 \times 10^7$)	4.2×10^8 ($\pm 3.1 \times 10^7$)
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.

CHAPTER I

Profiling of bacterial phylogenetic transitions along a saline sediment gradient using Tag-Encoded FLX Amplicon Pyrosequencing (TEFAP)

Formatted for *PLoS One*

Abstract

Salinity is a primary determinant of global microbial community composition, however, little information exists regarding the relative abundance shifts of individual taxa along continuous salinity gradients, particularly in regard to rare taxa, which are not accessed by traditional clone libraries. We have employed Tag-Encoded FLX Amplicon Pyrosequencing (TEFAP) of the bacterial 16S rDNA gene to determine shifts in sediment microbial community structure along steep salinity and nutrient gradients in the Coorong lagoon, Australia. The overall community structure showed variation along the gradient using measures of ecological and evolutionary similarity, which was concordant with changes in salinity and nutrient concentration. The *Proteobacteria* genera *Roseobacter* and *Roseovarius*, in particular, peaked in abundance at intermediate salinities and the cyanobacterial genus *Euhalothece* dominated the community at the most saline and nutrient rich site. Overall, the Coorong was ecologically and evolutionarily distinct when compared to other habitats, indicating that the overall nature of the habitat had a role in determining community structure beyond the salinity and nutrient gradients. This was the first application of high throughput pyrotag sequencing to understanding the dynamics of microbial communities along a continuous sediment salinity gradient from marine to hypersaline salinities.

Introduction

The ribosomal RNA gene (16S rRNA/DNA) of prokaryotes provides microbial ecologists with a phylogenetic marker that is applicable to determining the structure of microbial communities without the biases introduced by culture [1]. Since the first applications of the polymerase chain reaction to describe the diversity of this gene in natural habitats [2,3] 16S rDNA profiles have been generated in almost all known microbial habitats, revealing extensive microbial diversity.

Recent meta-analyses of the global distribution of 16S rDNA phylotypes have revealed partitioning of community composition by biome and have identified salinity as a major determinant of phylotype distribution [4-6], with saline sediments being among the most diverse habitats sequenced. These studies, however, have applied broad comparisons of salinity, for example saline and non saline, losing the high resolution provided by salinity gradients in nature, such as estuaries and salterns. Microbial community composition, as defined by the 16S rDNA gene, shifts with salinity [7,8] with halotolerant and halophilic taxa becoming dominant in more extreme salinities. Shifts in microbial community structure have also been observed along estuaries [9-12] and in saline sediments [13, 14], suggesting the important selective role of salinity on local scales. However, the majority of these studies have investigated either estuarine habitats or extreme hypersaline environments (e.g. solar salterns) without addressing brackish to hypersaline gradients. In this context, the Coorong lagoon, in South Australia, provides a unique model system of a continuous, natural salinity gradient from estuarine to hypersaline salinities [15,16] in which to investigate shifts in bacterial phylotype abundance.

Recent advances in next generation sequencing have resulted in a much greater depth of sequencing of the 16S rDNA gene and have demonstrated the existence of the 'rare biosphere';

diverse phylotypes that occur at a low abundance [17], in addition to providing detailed resolution of the diversity of more abundant community members as well as rare phylotypes [e.g. 17]. Tag-Encoded FLX Amplicon Pyrosequencing (TEFAP) [19-21] represents one technique of using the 454 FLX sequencing platform to sequence muliplexed samples, identified by unique sample-specific tag sequences, in a highly parallel fashion, and has been used to examine microbial community shifts in clinical samples [21], the animal rumen [20] and sediment [14].

Here we employ TEFAP sequencing to test the hypothesis that local habitat variability will result in shifts in the abundance of certain phylotypes along an example of a continuous sediment salinity gradient, and compare the community composition of Coorong sediment with other habitats.

Methods

Sample Collection

Sampling was conducted at four reference stations along the Coorong lagoon, South Australia, in January 2008, during the Austral summer. Salinity notably varied by 99 practical salinity units (PSU) across stations. Names and GPS coordinates for the stations were as follows: 37PSU (-35.5, 138.8), 109PSU (-35.7, 139.3), 132 PSU (-35.9, 139.4) & 136 PSU (-36.1, 139.6). Each site was defined by different nutrient concentrations and microbial abundances [22].

At each site, 10g of submerged sediment in approximately 2m deep water was sampled using a sterile corer. This equated to a core containing the upper 10cm of sediment. These cores consisted of dark brown and black mud with an approximately 2cm layer of pale sand at the

surface, and were similar in sediment characteristics at all 4 sites. Samples were stored on ice prior to DNA extraction which was performed within 8 hours of collection.

DNA was extracted from 10g of homogenized sediment using a bead beating and chemical lysis procedure (Powersoil, MoBio) and dispatched to the Research and Testing Laboratory (RTL, Lubbock, Texas USA) for library construction and sequencing.

Tag-Encoded FLX Amplicon Pyrosequencing

Amplification, pyrosequencing and annotation was performed using previously described tag encoded FLX pyrosequencing method [19,20,21]. Briefly, samples were amplified with the primer 27F and amplicons were sequenced using Roche 454 Titanium chemistry, generating reads in the forward direction from the primer. Reads were sorted using DNA sequence barcode tags and filtered based on sequence quality by the Research and Testing Laboratory. Sequencing yielded 6359, 17339, 20121 and 20053 bacterial sequences at sites 37 PSU, 109 PSU, 132 PSU and 136 PSU, respectively.

Sequence annotation and community visualization

Amplicon sequences were annotated using the Ribosomal Database Project (RDP) release 10 pyrosequencing pipeline [23]. FASTA and quality files supplied by the RTL were de-multiplexed using the sequence tags, trimmed to remove the sequence tags and primer sequences. Further quality control was conducted using the RDP pre-processing pipeline. Sequences of low quality, defined using the quality score provided by RTL, a read length of <150bp and the presence of ambiguous bases (N), were removed. . Outputted FASTA files were then classified using the RDP classifier [24] which uses a naïve Bayesian rRNA classifier to

assign sequences to the higher-order prokaryotic taxonomy. We used a confidence threshold of 80% [23]. Output files from this program were visualized using the MEtaGenomics ANalyzer (MEGAN) software package [25] to display and compare phylotypes as mapped against the NCBI taxonomic hierarchy. Additionally, taxonomic assignments resultant from a megaBLAST [26] search against NCBI 16S rDNA sequences ($E<10^{-5}$) supplied by RTL supplemented the RDP analysis.

Statistical comparison of Coorong communities with other biomes

Publicly available FASTA files of 16S rDNA sequence libraries generated using 454 tag encoded pyrosequencing were exported from two databases: the Visualization and Analysis of Microbial Population Structures (VAMPS) project (http://vamps.mbl.edu/overview.php) and the MetaGenomics using Rapid Annotation of Subsystems Technology (MG-RAST) pipeline (http://metagenomics.anl.gov/) [27]. These samples were chosen to represent several biomes, substrate types (water or sediment) and salinities and consisted of between 1555 and 24675 sequences (Table 1). Sequences were then annotated to the RDP database as above and imported into MEGAN [25] to perform statistical analyses of RDP annotations mapped to the NCBI taxonomy tree. Two statistical indices were employed to determine the similarity between samples: Goodall's index [26] and UniFrac [29]. Goodall's similarity index is an non-parametric measure which by definition is particularity well suited to pyrosequenced phylotype data as it is weighted to consider rare taxa, a characteristic of these datasets [30]. The UniFrac metric is based on the distance between communities as a fraction of branch length in a phylogenetic tree that leads to descendants of members of each community but not both [29]. This measure of environment specific evolution was specifically designed for 16S rDNA datasets and has been

widely applied to next-generation sequencing datasets [31, 4]. To visualize relationships between samples we used the neighbour-net algorithm [32] and UPGMA clustering [33] within MEGAN.

Results

Overall community structure

At phyla level, all communities were dominated by the *Proteobacteria* with the *Bacteroidetes* also predominating in all libraries (Fig. 1). The phylum *Firmicutes* was a major contributor to the 37 PSU library but was less abundant in the hypersaline environments. *Cyanobacteria* was the second most abundant phylum in the 136 PSU library (Fig. 1) but was less represented in all other libraries. Several classes showed abundance shifts along the salinity gradient (Fig. 2). Within the *Firmicutes*, Anaerobic sulfite-reducing *Clostrida* and the class *Bacilli* were predominant in the 37 PSU library but were rare in hypersaline samples. Representation of *β-Proteobacteria* also consisted largely of matches from the 37 PSU sample. The *α-Proteobacteria, Actinobacteria* and *Sphingobacteria* were prominent in the intermediate 109 PSU and 132 PSU libraries but were rare at 37 PSU.

Taxonomic transitions along the salinity gradient

The relative abundance of phylotypes, resultant from a megaBLAST [25] comparison of sequences against a database of NCBI prokaryotic 16S rDNA genes, showed marked shifts in species abundance along the salinity gradient (Fig. 3). Specifically, *Pseudomonas* sp. and *Cytophaga* sp., which belong to the *γ-Proteobacteria* and *Bacteroidetes* respectively, showed steep reductions in abundance between the 37 PSU sample and hypersaline samples (Fig. 3A). Phylotypes belonging to the *α-Proteobacteria* genera *Roseobacter* and *Rosiovarius* peaked in

abundance at 109 PSU whilst the *δ-Proteobacteria Desulfosarcina* sp. peaked at the 132 PSU site (Fig. 3B). The 136 PSU library was dominated by the cyanobacterial *Euhalothece* sp. which increased by 30% from the 109 PSU sample and was not present in the 37 PSU library (Fig. 3C). Whilst the taxa present in Fig. 3, which showed the greatest shifts in abundance from a pool of 1174 phylotypes, represented different abundance categories in the overall community, these groups tended to be the dominant taxon in the particular library in which they peaked, for example *Pseudomonas* sp., *Roseobacter* sp., *Desulfosarcina* sp. and *Euhalothece* sp. were the most abundant taxa at 37 PSU, 109 PSU, 132 PSU and 136 PSU respectively.

Overall ecological and evolutionary similarity between communities

To determine the overall similarity of communities along the Coorong salinity gradient at several phylogentic levels, we compared the TEFAP 16S rDNA profiles using two indices: an ecological metric, Goodall's index [28], and an evolutionary metric, UniFrac [29]. Using Goodall's index, at class level, hypersaline samples clustered together relative to the 37 PSU sample. In particular the 109 PSU and 132 PSU libraries were indiscriminant from each other (Fig. 4A). This similarity between 109 PSU and 132 PSU was maintained at the species rank of the NCBI Taxonomy (Fig. 4B), however this intermediate group was more closely related to the 37 PSU sample than to the most hypersaline 136 PSU sample. The 37 PSU sample was evolutionarily distinct from the hypersaline samples, which clustered more closely together in terms of the amount of environment specific evolution within the overall phylogenetic tree (Fig. 5). The 109 PSU and 132 PSU samples were evolutionarily more similar to each other than to the 37 PSU or 136 PSU libraries.

Ecological and evolutionary similarity between Coorong communities and other habitats

To determine the overall similarity of Coorong communities to samples from other habitats, we used network analysis to compare the 16S rDNA from our libraries to other publically available tag pyrosequenced libraries using two indices: an ecological metric, Goodall's index, and an evolutionary metric, UniFrac. Ecologically, Coorong samples were more similar to each other and to a marine sand sample and were least similar to marine water samples (Fig. 6A). In terms of UniFrac distance (Fig. 6B) the Coorong samples were distinct from other habitats, showing a high degree of shared evolutionary history. The most dissimilar habitats were marine water samples and the Dead Sea. The 37 PSU sediment library shared an edge with marine sand and the 136 PSU sample shared an edge with tropical soil and river plume libraries.

Discussion

Taxonomic transitions along a physiochemical gradient

Our libraries showed clear clustering patterns based on both ecological and evolutionary similarity. At class level, the higher degree of Goodall's index similarity displayed by hypersaline libraries to each other relative to the 37 PSU sample indicated that salinity is a structuring variable of the overall community composition. This is in contrast to the similarity at species level, where the moderately hypersaline 109 and 132 PSU clusters were more similar to the marine 37 PSU library than to the other hypersaline 136 PSU sample. The dissimilarity between the 132 PSU sample and the 137 PSU sample is unlikely due to the salinity difference of 4 PSU but could be a result of the elevated nutrient concentrations and microbial abundance at the 137 PSU sampling site. Evolutionary similarity, determined using the UniFrac metric,

indicated that there was more shared environment specific evolution within the hypersaline libraries, which clustered together, than between the marine 37 PSU sample and the hypersaline samples. This is consistent with previous studies which highlight salinity as the main determinant of UniFrac distance between phylotype clusters [4]. Overall the shifts in community composition along the salinity gradient are consistent with previous literature [7,8], however, we are aware of the potential influence of other parameters such as nutrient concentration, which co-vary with salinity along the gradient.

Within the overall community structures, the abundance of specific taxa showed sharp peaks at different locations along the gradient indicating that each site is characterized by different abundant taxa reflecting ecological transitions along the lagoon. The 109 PSU salinity library was defined by a dominance of taxa belonging to the *Roseobacter* clade, specifically the genera *Roseobacter* and *Roseovarius*, which increased by in abundance from the 37 PSU sample to peak in abundance at 109 PSU. The *Roseobacter* clade is a phylogenetically coherent but metabolically diverse group which can represent up to 25% of marine communities [34,35] and are of ecological significance due to their role in the DMSP cycle and alternative aerobic anoxygenic photosynthesis [34,35]. Within the clade several lineages are adapted to hypersaline conditions, and a requirement for sodium ions is a general trait of the clade [34,36]. Whilst reported in hypersaline lakes, soil and microbial mats [37-39], few studies have actually elucidated the spatial dynamics of this clade in response to physiochemical heterogeneity. Our results indicate that certain lineages of the *Roseobacter* and *Roseovarius* genera, including strains of *R. mucosus* and *R. pacificus*, respond strongly to salinity and nutrient gradients and proliferate at moderately hypersaline sites, but are less abundant members of marine and strongly hypersaline communities. Interestingly other *Roseobacter* taxa did not replace these groups at

other salinities indicating their ecological niche was transient along the gradient or was filled by other taxa. Whilst found in diverse habitats and displaying diverse phenotypes [35], the described members of the genus have the ability to perform anoxygenic photosynthesis [35,36], have been found in symbiotic associations with algae [35] and have been found in hypersaline mats to be capable of growth on the exudates of primary producers [39]. *Roseobacter* appear to play a similar role in the marine environment where they degrade the algal osmolyte dimethyl sulfoniopropionate (DMSP) [34], thus their niche in the Coorong could be related to an association with primary producers.

The 136 PSU site was characterized by a dominance of the *Cyanobacteria* genus *Euhalothece*. This halotolerant genus can grow over a wide salinity range but has been shown to increase with salinity along a benthic salinity gradient being restricted to the highest salinity sites [40]. This trend in our data is potentially the result of a reduction in interactions with competitors or grazers with salinity, but is more likely a result of elevated nutrient levels at the 136 PSU site as the order *Chroococcales*, to which the *Euhalothece* belong, has been shown to increase in abundance with nitrogen and phosphorus levels [41].

Ecological and evolutionary similarity between Coorong communities and other habitats

The observation that ecologically Coorong samples were more similar to each other and to a marine sand sample, despite salinity variability, and were least similar to marine water samples indicates that the overall habitat type, e.g. sediment or sand, could be an important determinant of community composition in addition to salinity. In terms of UniFrac distance the Coorong samples were distinct from other habitats indicating a high degree of shared evolutionary history between these libraries despite their variability in physiochemical characteristics. Evolutionary

similarity could be a result of geographic proximity and shared environmental history [42] and similar broad habitat characteristics, such as the sedimentary nature of the samples. Indeed the distinction between sediment and water was found to be the second most important determinant of 16S rRNA phylotype distribution following salinity [4], suggesting that several factors are acting as controls of community composition simultaneously. We note that the samples chosen for comparison are not an exhaustive selection of different habitats and are representative of habitats that could be expected to show similarity to the Coorong based on salinity. They were also chosen as they were sequenced using amplicon pyrosequencing and had a similar number of reads. Further analysis incorporating more habitat variability would provide more context in which to evaluate community similarity.

Our dataset represents the first TEFAP profile of taxonomic composition along a continuous sediment salinity gradient from marine to hypersaline conditions. Ecological and evolutionary dissimilarity along the gradient was reflected in the transition from marine salinity to an intermediate hypersaline community in which members of the *Roseobacter* clade peaked in abundance and were the most abundant community members. The most hypersaline site, which also had the highest nutrient concentration and microbial abundance, was dominated by the *Cyanobacteria* genus *Euhalothece*. Despite this variability however, the Coorong was unique from other habitats sharing an overall ecological and evolutionary signature. The Coorong thus provides a unique model system into which to investigate how environmental variability along a salinity gradient influences microbial genotypes that will provide an excellent resource in which to apply other next-generation sequencing applications, such as metagenomics, to determine functional as well as taxonomic responses to environmental heterogeneity.

References

1. Tringe SG, Hugenholtz P (2008) A renaissance for the pioneering 16S rRNA gene. Current Opinion in Microbiology 11: 442-446.

2. Woese CR (1987) Bacterial evolution. Microbiol Rev 51: 221-271.

3. Pace NR (1997) A Molecular View of Microbial Diversity and the Biosphere. Science 276: 734-740.

4. Lozupone CA, Knight R (2007) Global patterns in bacterial diversity. Proceedings of the National Academy of Sciences of the United States of America 104: 11436-11440.

5. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, et al. (2010) Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proceedings of the National Academy of Sciences. doi:10.1073/pnas.1000080107

6. Tamames J, Abellan JJ, Pignatelli M, Camacho A, Moya A (2010) Environmental distribution of prokaryotic taxa. BMC Microbiology 10:

7. Estrada M, Henriksen P, Gasol JM, Casamayor EO, Pedros-Alio C (2004) Diversity of planktonic photo auto trophic microorganisms along a salinity gradient as depicted by microscopy, flow cytometry, pigment analysis and DNA-based methods. FEMS Microbiology Ecology 49: 281-293.

8. Benlloch S, Lopez-Lopez A, Casamayor EO, Ovreas L, Goddard V, et al. (2002) Prokaryotic genetic diversity throughout the salinity gradient of a coastal solar saltern. Environmental Microbiology 4: 349-360.

9. Bouvier TC, Giorgio PA del (2002) Compositional changes in free-living bacterial communities along a salinity gradient in two temperate estuaries. Limnology and Oceanography 47: 453-470.

10. Oakley BB, Carbonero F, Gast CJ van der, Hawkins RJ, Purdy KJ (2010) Evolutionary divergence and biogeography of sympatric niche-differentiated bacterial populations. ISME J 4: 488-497.

11. Bernhard AE, Donn T, Giblin AE, Stahl DA (2005) Loss of diversity of ammonia-oxidizing bacteria correlates with increasing salinity in an estuary system. Environmental Microbiology 7: 1289-1297.

12. Herlemann DPR, Labrenz M, Jurgens K, Bertilsson S, Waniek JJ, et al. (2011) Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. ISME J 5: 1571-1579.

13. Swan BK, Ehrhardt CJ, Reifel KM, Moreno LI, Valentine DL (2010) Archaeal and Bacterial Communities Respond Differently to Environmental Gradients in Anoxic Sediments of a California Hypersaline Lake, the Salton Sea. Applied and Environmental Microbiology 76: 757-76

14. Hollister EB, Engledow AS, Hammett AJM, Provin TL, Wilkinson HH, et al. (2010) Shifts in microbial community structure along an ecological gradient of hypersaline soils and sediments. ISME J 4: 829-838.

15. Lester RE, Fairweather PG (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. Schapira M, Buscot MJ, Leterme SC, Pollet T, Chapperon C, et al. (2009) Distribution of heterotrophic bacteria and virus-like particles along a salinity gradient in a hypersaline coastal lagoon. Aquatic Microbial Ecology 54: 171-183.

17. Sogin ML, Morrison HG, Huber JA, Mark Welch D, Huse SM, et al. (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.

18. Kirchman DL, Cottrell MT, Lovejoy C (2010) The structure of bacterial communities in the western Arctic Ocean as revealed by pyrosequencing of 16S rRNA genes. Environmental Microbiology 12: 1132-1143.

19. Sun Y, Wolcott RD, Dowd SE (2011) Tag-Encoded FLX Amplicon Pyrosequencing for the Elucidation of Microbial and Functional Gene Diversity in Any Environment. Methods In Molecular Biology Clifton NJ 733: 129-141.

20. Dowd SE, Callaway TR, Wolcott RD, Sun Y, McKeehan T, et al. (2008) Evaluation of the bacterial diversity in the feces of cattle using 16S rDNA bacterial tag-encoded FLX amplicon pyrosequencing (bTEFAP). BMC Microbiology 8. 125.

21. Bailey MT, Dowd SE, Parry NMA, Galley JD, Schauer DB, et al. (2010) Stressor Exposure Disrupts Commensal Microbial Populations in the Intestines and Leads to Increased Colonization by Citrobacter rodentium. Infection and Immunity 78: 1509-1519.

22. Jeffries TC, Seymour JR, Gilbert JA, Dinsdale EA, Newton K, et al. (2011) Substrate Type Determines Metagenomic Profiles from Diverse Chemical Habitats. PLoS ONE 6: e25173

23. Cole JR, Wang Q, Cardenas E, Fish J, Chai B, et al. (2009) The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. Nucleic Acids Research 37: D141-D145.

24. Wang Q, Garrity GM, Tiedje JM, Cole JR (2007) Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Applied and Environmental Microbiology 73: 5261-5267.

25. Huson DH, Mitra S, Ruscheweyh H-J, Weber N, Schuster SC (2011) Integrative analysis of environmental sequences using MEGAN4. Genome Research. Available: http://genome.cshlp.org/content/early/2011/06/17/gr.120618.111.abstract.

26. Zhang Z, Schwartz S, Wagner L, Miller W (2000) A greedy algorithm for aligning DNA sequences. Journal of Computational Biology 7: 203-214.

27. Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, et al. (2008) The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC Bioinformatics 9: 386.

28. Goodall DW (1966) A new similarity index based on probability. Biometrics 22: 882

29. Lozupone C, Knight R (2005) UniFrac: a New Phylogenetic Method for Comparing Microbial Communities. Applied and Environmental Microbiology 71: 8228-8235.

30. Mitra S, Gilbert JA, Field D, Huson DH (2010) Comparison of multiple metagenomes using phylogenetic networks based on ecological indices. 4: 1236-1242.

31. Lozupone C, Hamady M, Knight R (2006) UniFrac - An online tool for comparing microbial community diversity in a phylogenetic context. BMC Bioinformatics 7: 371.

32. Bryant D, Moulton V (2004) Neighbor-Net: An Agglomerative Method for the Construction of Phylogenetic Networks. Molecular Biology and Evolution 21: 255-265.

33. Sokal RR, Michener CD (1958) A statistical method for evaluating systematic relationships. University of Kansas Scientific Bulletin 28: 1409-1438.

34. Brinkhoff T, Giebel H-A, Simon M (2008) Diversity, ecology, and genomics of the Roseobacter; clade: a short overview. Archives of Microbiology 189: 531-539.

35. Wagner-Döbler I, Biebl H (2006) Environmental biology of the marine Roseobacter lineage. Annual Review of Microbiology 2006: 255-280.

36. Buchan A, Gonzalez JM, Moran MA (2005) Overview of the Marine Roseobacter Lineage. Applied and Environmental Microbiology 71: 5665-5677.

37. Labrenz M, Lawson PA, Tindall BJ, Collins MD, Hirsch P (2005) Roseisalinus antarcticus gen. nov., sp. nov., a novel aerobic bacteriochlorophyll a-producing α-proteobacterium isolated from hypersaline Ekho Lake, Antarctica. International Journal of Systematic and Evolutionary Microbiology 55: 41-47.

38. Martínez-Checa F, Quesada E, Martínez-Cánovas MJ, Llamas I, Béjar V (2005) Palleronia marisminoris gen. nov., sp. nov., a moderately halophilic, exopolysaccharide-producing bacterium belonging to the “Alphaproteobacteria”, isolated from a saline soil. International Journal of Systematic and Evolutionary Microbiology 55: 2525-2530.

39. Jonkers HM, Abed RMM (2003) Identification of aerobic heterotrophic bacteria from the photic zone of a hypersaline microbial mat. Aquatic Microbial Ecology 30: 127-133.

40. Nübel U, Garcia-Pichel F, Clavero E, Muyzer G (2000) Matching molecular diversity and ecophysiology of benthic cyanobacteria and diatoms in communities along a salinity gradient. Environmental Microbiology 2: 217-226.

41. Rejmánková E, Komárková J (2005) Response of cyanobacterial mats to nutrient and salinity changes. Aquatic Botany 83: 87-107.

42. Martiny JBH, Bohannan BJM, Brown JH, Colwell RK, Fuhrman JA, et al. (2006) Microbial biogeography: putting microorganisms on the map. Nature Reviews Microbiology 4: 102-112.

Figure Legends

Figure 1 Phyla composition of saline sediment TEFAP libraries.

Figure 2 Community profile of TEFAP ribotypes matching the NCBI Taxonomic tree. Phyla are expanded to class level where available.

Figure 3 Relative abundance of taxa along a salinity gradient which showed the greatest variation in abundance. Taxa A) decrease with salinity B) peak at intermediate salinities and C) peak at the most hypersaline site.

Figure 4 UMPGA clustering of Coorong communities based on Goodall's similarity index at A) class level and B) species level (NCBI Taxonomy).

Figure 5 UMPGA clustering of Coorong communities based on UniFrac distance at species level (NCBI Taxonomy).

Figure 6 Networks obtained using A) Goodall's index and B) UniFrac distance showing comparison between Coorong sediment metagenomes and other habitats. UniFrac was calculated at the species rank of **NC**BI taxonomy and Goodall's index at the class rank.

Table 1 16S rDNA libraries used to compare Coorong profiles to other habitats. FASTA files were downloaded from the VAMPS and MG-RAST pipelines as described in the materials and methods.

Sample ID	Habitat	Number of reads	Label
Samples from VAMPS			
CNE_0003_2003_10_11	Coastal seawater (New England)	18536	coastaL_ NE
ICM_FIS_Bv6__FIS_0001_2008_04_10	Coastal sand	21325	sand
ENV1	Salt marsh sediment	24675	salt_marsh
KCK_RIP_Bv6__RIP_0005_1992_03	Dead sea	22729	dead_sea
Samples from MG-RAST			
4441488.3	English channel	16923	english_channel
4455740.3	Tropical forest soil	1774	tropical soil
4455879.3	River delta plume (arsenic)	1555	river_plume


Synergistetes (0)
Not assigned (0)
No hits (0)
Spirochaetes (26)
Proteobacteria (43,573)
Actinobacteria (520)
Bacteroidetes (16,831)
Lentisphaerae (76)
Verrucomicrobia (647)
Chloroflexi (164)
Cyanobacteria (3,219)
Deferribacteres (0)
Deinococcus-Thermus (0)
Acidobacteria (760)
Firmicutes (7,172)
Fusobacteria (89)
Planctomycetes (723)
[1]37 PSU
Synergistetes (0)
Not assigned (0)
No hits (0)
Spirochaetes (905)
Proteobacteria (31,583)
Actinobacteria (888)
Bacteroidetes (13,530)
Lentisphaerae (46)
Verrucomicrobia (447)
Chloroflexi (121)
Cyanobacteria (488)
Deferribacteres (0)
Deinococcus-Thermus (88)
Acidobacteria (46)
Firmicutes (182)
Fusobacteria (0)
Planctomycetes (353)
[2]109 PSU
Synergistetes (0)
Not assigned (0)
No hits (0)
Spirochaetes (777)
Proteobacteria (26,228)
Actinobacteria (544)
Bacteroidetes (10,534)
Lentisphaerae (21)
Verrucomicrobia (284)
Chloroflexi (201)
Cyanobacteria (321)
Deferribacteres (5)
Deinococcus-Thermus (43)
Acidobacteria (100)
Firmicutes (177)
Fusobacteria (2)
Planctomycetes (325)
[3]132 PSU
Not assigned (0)
No hits (0)
Synergistetes (4)
Spirochaetes (865)
Proteobacteria (30,211)
Planctomycetes (713)
Fusobacteria (4)
Firmicutes (208)
Acidobacteria (82)
Deinococcus-Thermus (69)
Deferribacteres (61)
Actinobacteria (294)
Bacteroidetes (5,249)
Lentisphaerae (16)
Verrucomicrobia (105)
Chloroflexi (1,019)
Cyanobacteria (27,737)
[4]136 PSU
Actinobacteria
Bacteroidetes
Lentisphaerae
Verrucomicrobia
Chloroflexi
Cyanobacteria
Deferribacteres
Deinococcus-Thermus
Acidobacteria
Firmicutes
Fusobacteria
Planctomycetes
Proteobacteria
Spirochaetes
Synergistetes
Not assigned
No hits

Figure 1

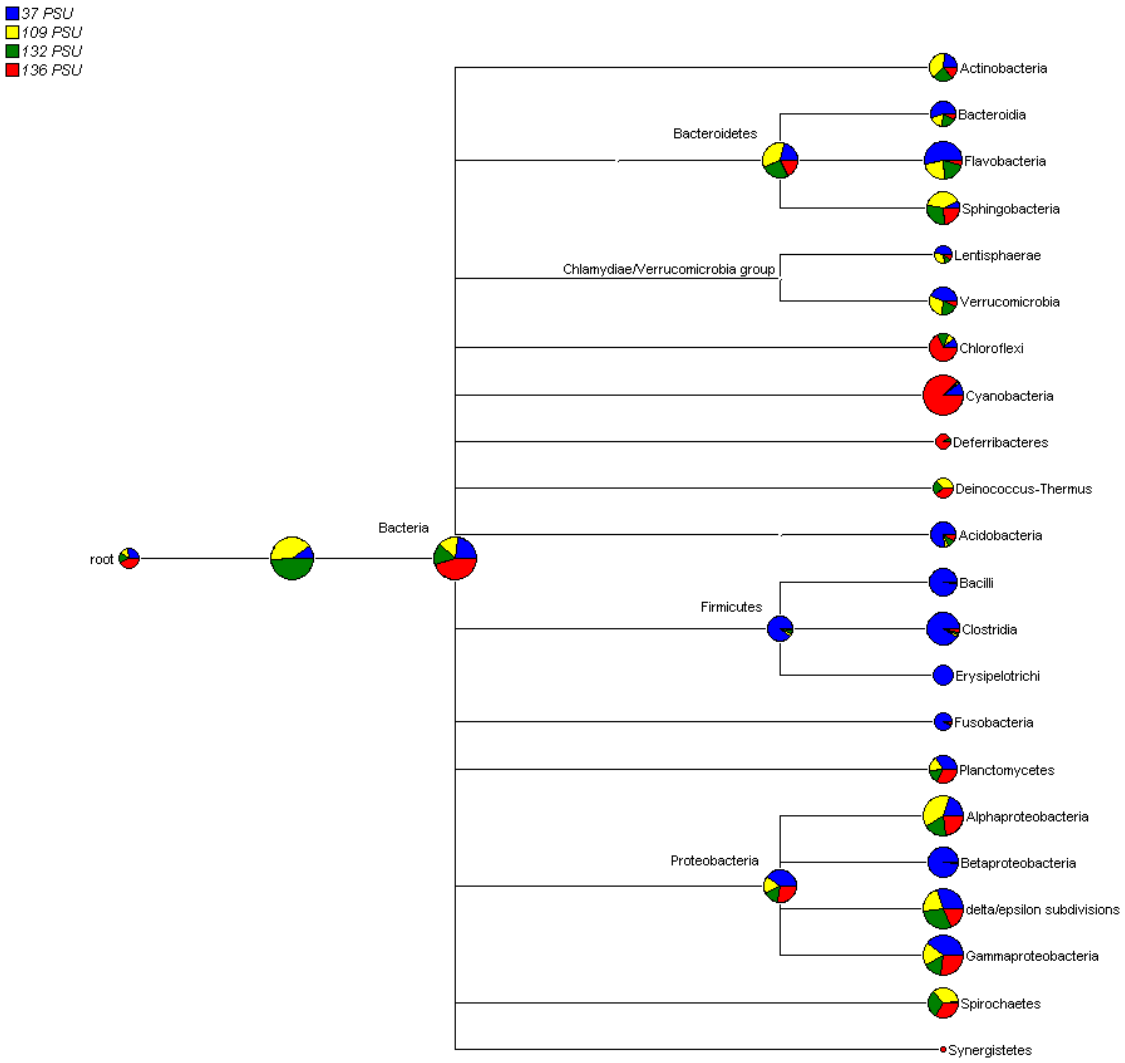
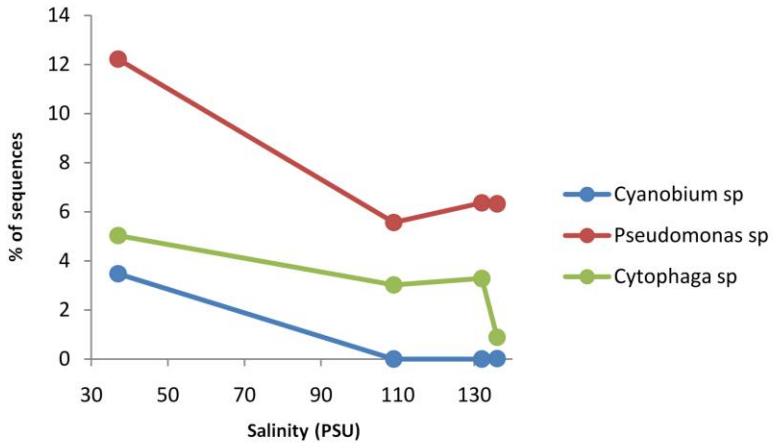
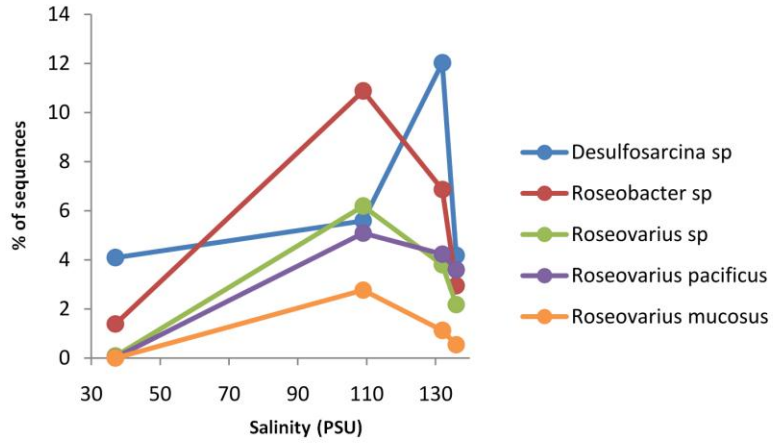

37 PSU
109 PSU
132 PSU
136 PSU
root
Bacteria
Actinobacteria
Bacteroidetes
Bacteroidia
Flavobacteria
Sphingobacteria
Chlamydiae/Verrucomicrobia group
Lentisphaerae
Verrucomicrobia
Chloroflexi
Cyanobacteria
Deferribacteres
Deinococcus-Thermus
Acidobacteria
Firmicutes
Bacilli
Clostridia
Erysipelotrichi
Fusobacteria
Planctomycetes
Proteobacteria
Alphaproteobacteria
Betaproteobacteria
delta/epsilon subdivisions
Gammaproteobacteria
Spirochaetes
Synergistetes

Figure 2

A)


14
12
10
8
6
4
2
0
% of sequences
30
50
70
90
110
130
Salinity (PSU)
Cyanobium sp
Pseudomonas sp
Cytophaga sp

B)


14
12
10
8
6
4
2
0
% of sequences
30
50
70
90
110
130
Salinity (PSU)
Desulfosarcina sp
Roseobacter sp
Roseovarius sp
Roseovarius pacificus
Roseovarius mucosus

C)

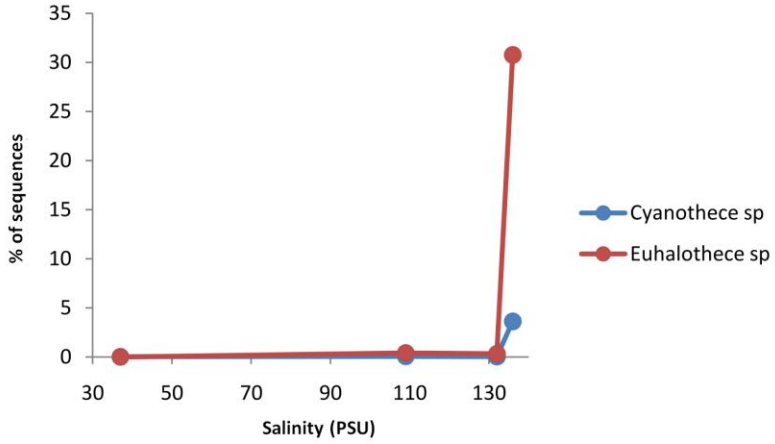

35
30
25
20
15
10
5
0
% of sequences
30
50
70
90
110
130
Salinity (PSU)
Cyanothece sp
Euhalothece sp

Figure 3

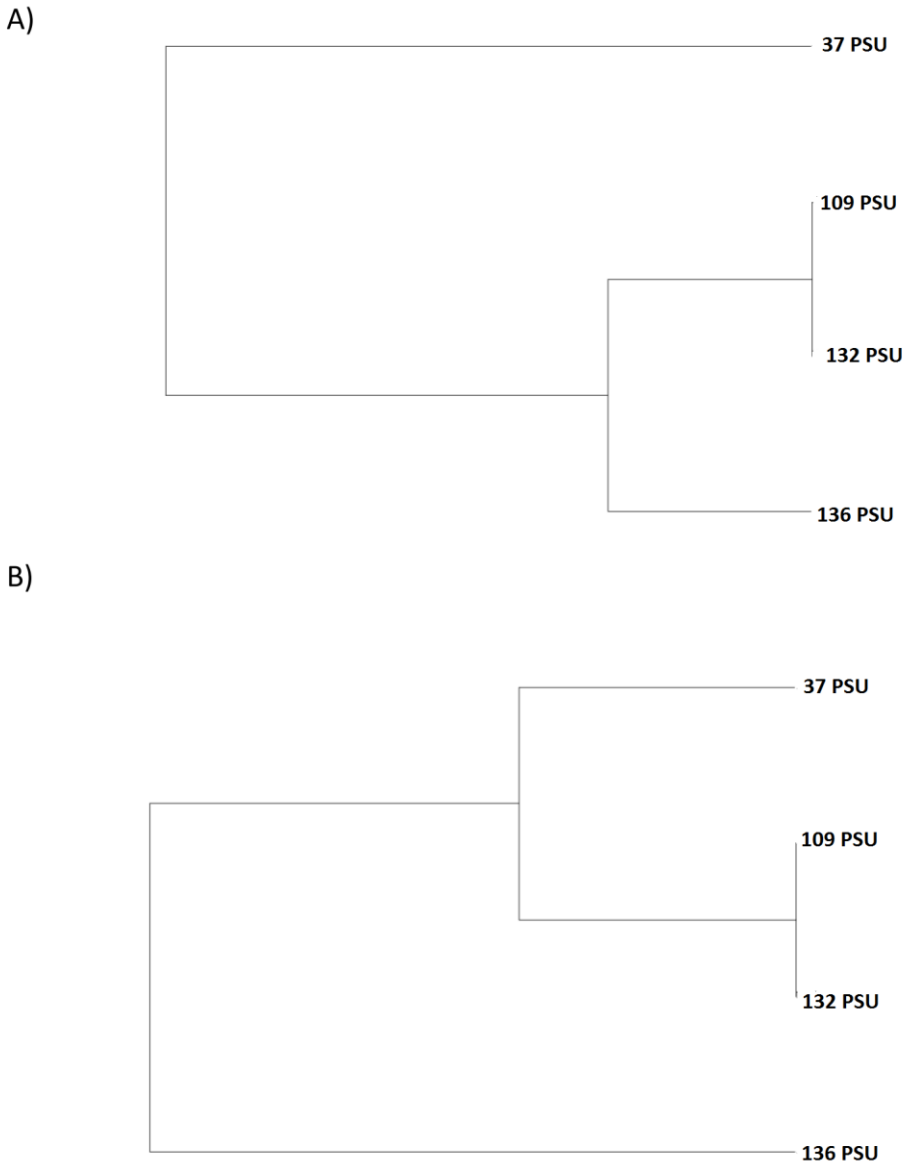


Figure 4

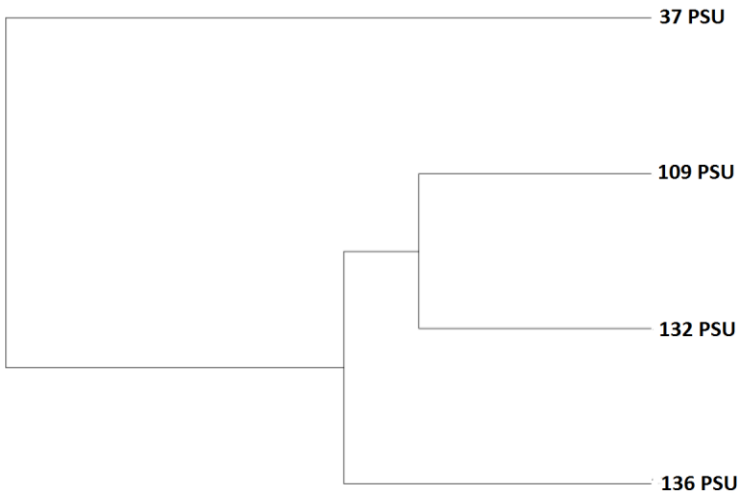
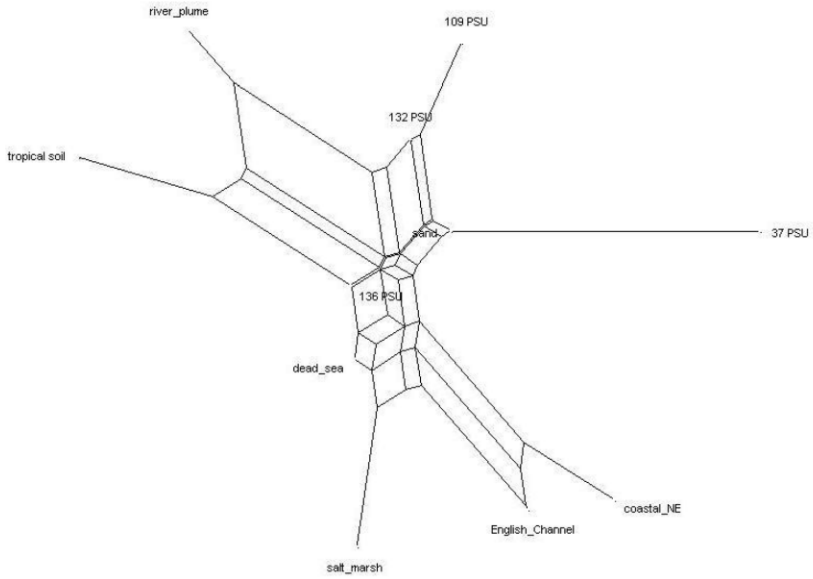


Figure 5

A)


river_plume
109 PSU
132 PSU
tropical soil
sand
37 PSU
136 PSU
dead_sea
coastal_NE
English_Channel
salt_marsh

B)

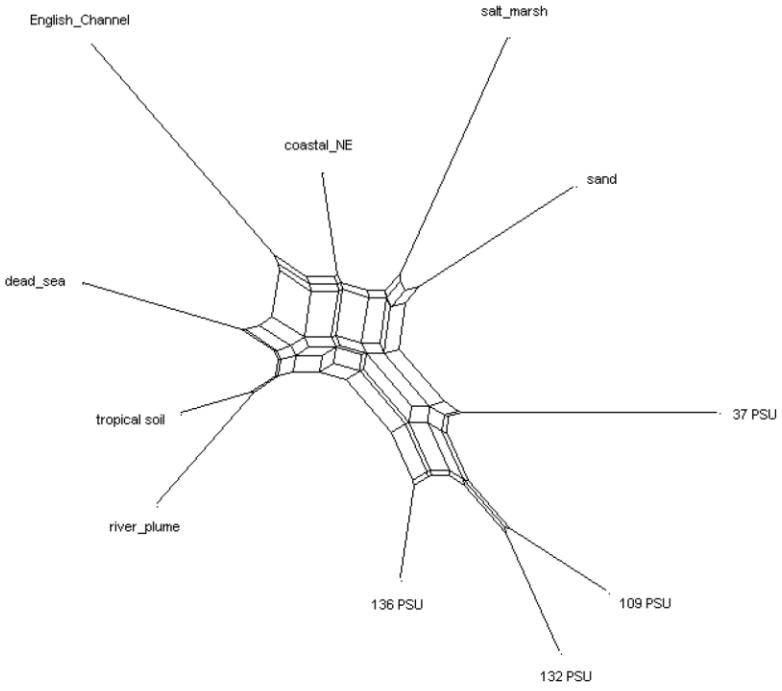

English_Channel
salt_marsh
coastal_NE
sand
dead_sea
37 PSU
tropical soil
river_plume
136 PSU
109 PSU
132 PSU

Figure 6

CHAPTER II

Strong shifts in sediment archaeal communities along an intense salinity gradient

Formatted as a short form paper for *Applied and Environmental Microbiology*

Abstract

Archaeal communities shifted from 70% *Crenarchaeota* at 37 PSU to 70% *Euryarchaeota* of the class *Thermoplasmata* at 136 PSU. At intermediate salinities rare archaeal groups rose in abundance 18 to 25 times. Our results show that archaeal communities are strongly structured by salinity.

Since *Archaea* were found in significant numbers in seawater (9, 13), their distributional dynamics and biogeochemical roles in aquatic systems have been widely examined (5-7, 10, 12, 16, 17, 21). Although *Archaea* define aquatic microbial communities at extreme hypersalinity (18), clear and consistent responses of individual lineages to salinity gradients remain elusive despite studies demonstrating changes in archaeal community structure with increasing salinity (3, 5, 6, 16).

Salinity gradient studies to date have analysed archaeal phylogenetic diversity by targeting universal regions of the 16S rDNA gene which are not specific to *Archaea* or have sequenced clone libraries of archaeal specific 16S rDNA sequences. The latter only captures the most abundant community members. We overcame these limitations by employing high-throughput, tag-encoded pyrosequencing of *Archaea*-specific 16S rDNA amplicons generating over 20,000 archaeal sequences from four samples along a continuous sediment salinity gradient. Our goal was to investigate how archaeal lineages in sediment respond to a salinity gradient.

Sediment cores of 10cm depth were taken at four sites along the Coorong, South Australia (35°47’ S, 139°19’E), a 100km long, shallow, temperate lagoon. At the time of sampling, sediment was covered by approximately 30cm of overlying water. The four sites had salinities of 37, 109, 132 and 136 Practical Salinity Units (PSU). Sediment cores were stored on ice, and DNA was extracted within 12 hours of sampling using bead beating and chemical lysis (Powermax soil, MoBio). Amplification, pyrosequencing and annotation were performed at the Research and Testing Laboratory (Lubbock, Texas USA) using previously described tag encoded FLX pyrosequencing methods (2, 11). Briefly, samples were amplified with the primer pair Arch 349F and 806R (24) and amplicons were sequenced using Roche 454 Titanium chemistry,

generating reads in the forward direction from Arch 359F. Reads were sorted using DNA sequence barcode tags and filtered based on sequence quality. Sequences were annotated using megaBLAST (26) against a database of NCBI prokaryotic 16S sequences with an e-value cut-off of $E<10^{-5}$. Sequencing yielded 4097, 4818, 6639 and 8569 archaeal sequences at sites 37 PSU, 105 PSU, 132 PSU and 136 PSU, respectively.

At the phylum level, all four libraries contained sequences matching the *Euryarchaeota* and *Crenarchaeota*, with the *Thaumarchaeota* only present at 37 PSU (Fig. 1A). The *Euryarchaeota* and *Crenarchaeota* typically dominated the archaeal community across a diverse range of terrestrial and aquatic habitats (21). The importance of both groups in hypersaline sediments has also been demonstrated (23). The *Thaumarchaeota* are a recently proposed deep branching phylum (4) potentially involved in marine ammonia oxidization (22). The absence of this phylum in the extreme salinity samples suggests that this biogeochemically significant taxon is intolerant of high salt concentrations.

Clear shifts in phyla level community structure were observed along the salinity gradient (Fig. 1A). Sequences matching the *Crenarchaeota* dominated at 37 PSU, representing 70% of archaeal phylotypes. The relative proportion of these decreased by ten-fold along the salinity gradient (Fig. 1A), indicating potential intolerance to high salt concentrations. This is consistent with results from clone libraries derived from a sediment core salinity gradient (19) and with the absence of *Crenarchaeota* in other hypersaline examples (3, 8), but contradicts the increase in *Crenarchaeota* sequence abundance with salinity observed in other hypersaline sediment (23) and soil (25) samples. Combined, these results indicate that the relationship between

Crenarchaeota abundance and salinity can differ in different habitats, potentially as a result of different co-variables and competition.

The decrease in *Crenarchaeota* sequence abundance along the salinity gradient was paralleled by a 73% increase in the proportion of sequences matching the *Euryarchaeota* (Fig. 1A). The phylum *Euryarchaeota* contains several lineages of halotolerant and halophilic taxa, which have been identified in hypersaline sediments from a variety of locations including saline lakes, salterns and saltmarsh (1, 16, 19, 23). *Euryarchaeota* representatives have also been found to predominate in deep Arctic waters (14) and in coastal waters influenced by high sediment loads (15), indicating that this group can demonstrate biogeographic patterns**.** In our data, the *Euryarchaeota* are largely composed of the class *Thermoplasmata* which increased to 71% of the overall community along the gradient (Fig. 1B). This indicates that members of the *Thermoplasmata* thrive in increasingly saline environments or fill the niche vacated by the *Crenarchaeota*. Whilst members of the *Thermoplasmata* have been observed in other environmental samples (3, 6, 8), such a clear and dramatic relationship between *Thermoplasmata* sequence abundance and salinity has not been observed previously. Sequences matching unclassified classes within the *Euryarchaeota* also increased by 23% at 105 PSU, before decreasing in the most hypersaline samples (Fig. 1B). This peak of uncharacterized *Euryarchaeota* sequence abundance at 105 PSU corresponds to a peak in the abundance of sequences matching the less abundant euryarchaeal classes *Halobacteria* and *Methanomicrobia* (Fig. 1C) which increased by18 to 25 times, indicating that salinity is also influencing the dynamics of less abundant taxa. Thus, at 105 PSU a more even community occurs, with peaks in the contribution of less abundant classes without the dominance of *Thermoplasmata* observed in the most extreme samples.

These clear shifts in the community structure occur across the transition zone from the extremes of an estuarine lagoon to a hypersaline lagoon (20). Between the two lagoons is a zone of fluctuating salinity. Our data (Fig. 1) show that three distinct archaeal communities exist: a *Crenarchaeota*-dominated community at 37 PSU, a *Euryarchaeota* (primarily *Thermoplasmata*)-dominated community at 136 PSU and a more even community in the transitional zone of salinity at 105 PSU and 132 PSU in which there is less dominance of either group, and in which rare taxa peak in abundance. This intermediate zone is potentially the product of transitional salinities in which an ephemeral community structure exists, resulting in higher diversity.

Our data has used tag encoded amplicon pyrosequencing to demonstrate a strong succession in the archaeal community composition along a sediment salinity gradient. This data provides a unique example of shifts in dominant taxa at salinity gradient endpoints. For archaeal classes that range in abundance from 0.1 to 2% of the community we identify strong abundance shifts. We designate this group the 'transitional biosphere'. They show large increases relative to their initial abundance, indicating the transitional biosphere is a highly dynamic part of the community. The observed sharp changes in community structure and dominance of key groups demonstrate the importance of salinity in structuring archaeal communities.

This work was supported by the Australian Research Council.

References

1. **Baati, H., S. Guermazi, N. Gharsallah, A. Sghir, and E. Ammar.** 2010. Novel prokaryotic diversity in sediments of Tunisian multipond solar saltern. Res. Microbiol. **161:**573-582.
2. **Bailey, M. T., S. E. Dowd, N. M. A. Parry, J. D. Galley, D. B. Schauer, and M. Lyte.** 2010. Stressor Exposure Disrupts Commensal Microbial Populations in the Intestines and Leads to Increased Colonization by Citrobacter rodentium. Infect. Immun. **78:**1509-1519.
3. **Benlloch, S., A. Lopez-Lopez, E. O. Casamayor, L. Ovreas, V. Goddard, F. L. Daae, G. Smerdon, R. Massana, I. Joint, F. Thingstad, C. Pedros-Alio, and F. Rodriguez-Valera.** 2002. Prokaryotic genetic diversity throughout the salinity gradient of a coastal solar saltern. Environ. Microbiol. **4:**349-360.
4. **Brochier-Armanet, C., B. Boussau, S. Gribaldo, and P. Forterre.** 2008. Mesophilic crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota. Nat. Rev. Microbiol. **6:**245-252.
5. **Casamayor, E. O., J. I. Calderon-Paz, and C. Pedros-Alio.** 2000. 5S rRNA fingerprints of marine bacteria, halophilic archaea and natural prokaryotic assemblages along a salinity gradient. FEMS. Microbiol. Ecol. **34:**113-119.
6. **Casamayor, E. O., R. Massana, S. Benlloch, L. Ovreas, B. Diez, V. J. Goddard, J. M. Gasol, I. Joint, F. Rodriguez-Valera, and C. Pedros-Alio.** 2002. Changes in archaeal, bacterial and eukaryal assemblages along a salinity gradient by comparison of genetic fingerprinting methods in a multipond solar saltern. Environ. Microbiol. **4:**338-348.
7. **Chaban, B., S. Y. M. Ng, and K. F. Jarrell.** 2006. Archaeal habitats - from the extreme to the ordinary. Can. J. Microbiol. **52:**73-116.

8. **Cytryn, E., D. Minz, R. S. Oremland, and Y. Cohen.** 2000. Distribution and diversity of archaea corresponding to the limnological cycle of a hypersaline stratified lake (Solar Lake, Sinai, Egypt). Appl. Environ. Microbiol. **66:**3269-3276.

9. **Delong, E. F.** 1992. Archaea in Coastal Marine Environments. Proc. Natl. Acad. Sci. USA **89:**5685-5689.

10. **DeLong, E. F., and N. R. Pace.** 2001. Environmental Diversity of Bacteria and Archaea. Systematic Biology **50:**470-478.

11. **Dowd, S. E., T. R. Callaway, R. D. Wolcott, Y. Sun, T. McKeehan, R. G. Hagevoort, and T. S. Edrington.** 2008. Evaluation of the bacterial diversity in the feces of cattle using 16S rDNA bacterial tag-encoded FLX amplicon pyrosequencing (bTEFAP). BMC Microbiol. **8:**125.

12. **Francis, C. A., J. M. Beman, and M. M. M. Kuypers.** 2007. New processes and players in the nitrogen cycle: the microbial ecology of anaerobic and archaeal ammonia oxidation. ISME J. **1:**19-27.

13. **Fuhrman, J. A., K. Mccallum, and A. A. Davis.** 1992. Novel Major Archaebacterial Group from Marine Plankton. Nature **356:**148-149.

14. **Galand, P. E., Casamayor, E. O., Kirchman, D. L., Potvin, M., and Lovejoy, C**. 2009. Unique archaeal assemblages in the Arctic Ocean unveiled by massively parallel tag sequencing, ISME J . **3**: 860-869,

15. **Galand, P. E., Lovejoy, C., Pouliot, J., and Vincent, W. F.** 2008**.** Heterogeneous archaeal communities in the particle-rich environment of an arctic shelf ecosystem, Journal of Marine Systems, **74**, 774-782.

16. **Hollister, E. B., A. S. Engledow, A. J. M. Hammett, T. L. Provin, H. H. Wilkinson, and T. J. Gentry.** 2010. Shifts in microbial community structure along an ecological gradient of hypersaline soils and sediments. ISME J. **4:**829-838.

17. **Jarrell, K. F., A. D. Walters, C. Bochiwal, J. M. Borgia, T. Dickinson, and J. P. J. Chong.** 2011. Major players on the microbial stage: why archaea are important. Microbiology **157:**919-936.

18. **Javor, B.** 1989. Hypersaline environments : microbiology and biogeochemistry. Springer-Verlag, Berlin ; New York.

19. **Jiang, H. C., H. L. Dong, B. S. Yu, X. Q. Liu, Y. L. Li, S. S. Ji, and C. L. L. Zhang.** 2007. Microbial response to salinity change in Lake Chaka, a hypersaline lake on Tibetan plateau. Environ. Microbiol. **9:**2603-2621.

20. **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. Estuar. Coast. Shelf. S. **85:**1-11.

21. **Schleper, C., G. Jurgens, and M. Jonuscheit.** 2005. Genomic studies of uncultivated archaea. Nat. Rev. Microbiol. **3:**479-488.

22. **Spang, A., R. Hatzenpichler, C. Brochier-Armanet, T. Rattei, P. Tischler, E. Spieck, W. Streit, D. A. Stahl, M. Wagner, and C. Schleper.** 2010. Distinct gene set in two different lineages of ammonia-oxidizing archaea supports the phylum Thaumarchaeota. Trends Microbiol. **18:**331-340.

23. **Swan, B. K., C. J. Ehrhardt, K. M. Reifel, L. I. Moreno, and D. L. Valentine.** 2010. Archaeal and Bacterial Communities Respond Differently to Environmental Gradients in

Anoxic Sediments of a California Hypersaline Lake, the Salton Sea. Appl. Environ. Microbiol. **76:**757-768.

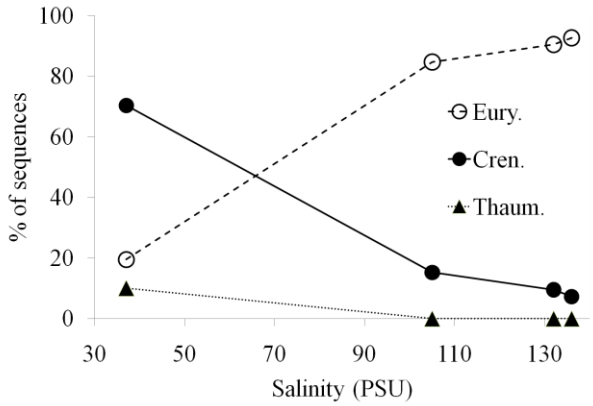
24. **Takai, K., and K. Horikoshi.** 2000. Rapid detection and quantification of members of the archaeal community by quantitative PCR using fluorogenic probes. Appl. Environ. Microbiol. **66:**5066-5072.

25. **Walsh, D. A., R. T. Papke, and W. F. Doolittle.** 2005. Archaeal diversity along a soil salinity gradient prone to disturbance. Environ. Microbiol. **7:**1655-1666.

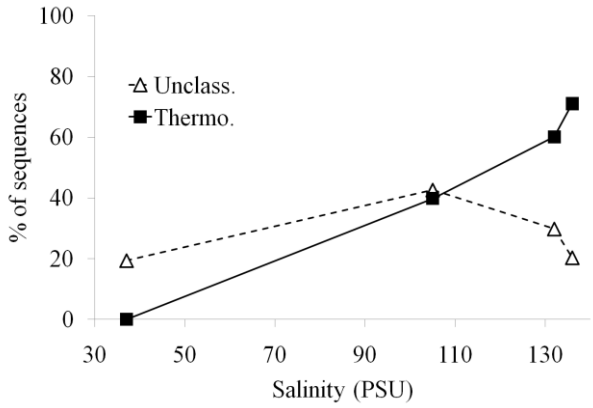
26. **Zhang, Z., S. Schwartz, L. Wagner, and W. Miller.** 2000. A greedy algorithm for aligning DNA sequences. J. Comput. Biol. **7:**203-214.

Figure 1. Shifts in archaeal community structure along a salinity gradient. A) phyla level B) abundant classes of *Euryarchaeota* C) low abundance classes of *Euryarchaeota*. Eury. = *Euryarchaeota*, Cren. = *Crenarchaeota*, Thaum. = *Thaumarchaeota*, Unclass. = unclassified *Euryarchaeota*, Thermo. = *Thermoplasmata*, Halo. = *Halobacteria*, Methanomic. = *Methanomicrobia*, Methano. = *Methanobacteria*.

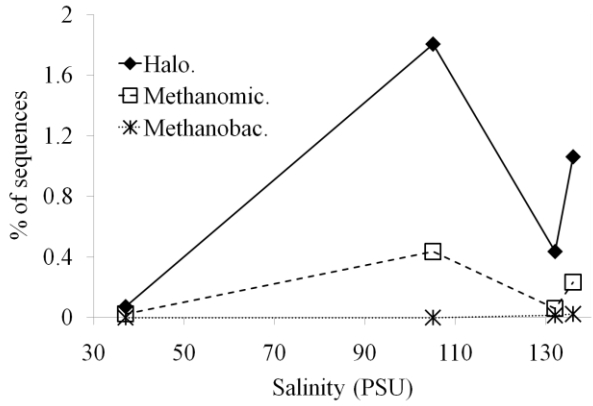
A)


100
80
60
40
20
0
% of sequences
Eury.
Cren.
Thaum.
30
50
70
90
110
130
Salinity (PSU)

B)


100
80
60
40
20
0
% of sequences
Unclass.
Thermo.
30
50
70
90
110
130
Salinity (PSU)

C)


2
1.6
1.2
0.8
0.4
0
% of sequences
Halo.
Methanomic.
Methanobac.
30
50
70
90
110
130
Salinity (PSU)

CHAPTER III

Increases in the abundance of microbial genes encoding halotolerance and photosynthesis along a sediment salinity gradient

A version of this manuscript is published as: Jeffries, T. C., Seymour, J. R., Newton, K., Smith, R. J., Seuront, L., and Mitchell, J. G. 2012.Increases in the abundance of microbial genes encoding halotolerance and photosynthesis along a sediment salinity gradient, Biogeosciences, 9, 815-825.

Abstract

Biogeochemical cycles are driven by the metabolic activity of microbial communities, yet the environmental parameters that underpin shifts in the functional potential coded within microbial community genomes are still poorly understood. Salinity is one of the primary determinants of microbial community structure and can vary strongly along gradients within a variety of habitats. To test the hypothesis that shifts in salinity will also alter the bulk biogeochemical potential of aquatic microbial assemblages, we generated four metagenomic DNA sequence libraries from sediment samples taken along a continuous, natural salinity gradient in the Coorong lagoon, Australia, and compared them to physical and chemical parameters. A total of 392483 DNA sequences obtained from four sediment samples were generated and used to compare genomic characteristics along the gradient. The most significant shifts along the salinity gradient were in the genetic potential for halotolerance and photosynthesis, which were more highly represented in hypersaline samples. At these sites, halotolerance was achieved by an increase in genes responsible for the acquisition of compatible solutes - organic chemicals which influence the carbon, nitrogen and methane cycles of sediment. Photosynthesis gene increases were coupled to an increase in genes matching Cyanobacteria, which are responsible for mediating CO_2 and nitrogen cycles. These salinity driven shifts in gene abundance will influence nutrient cycles along the gradient, controlling the ecology and biogeochemistry of the entire ecosystem.

1 Introduction

Biogeochemical cycles, over geological time, have fundamentally determined the chemical nature of the Earth's surface and atmosphere. Due to their high abundance and metabolic activities, microorganisms drive many global biogeochemical processes including the carbon, oxygen, nitrogen, hydrogen, sulfur and iron cycles (Falkowski et al., 2008;Fuhrman, 2009). The biochemical potential of the microbial inhabitants of an environment is determined by the community structure - the types of organisms present and their relative abundance, which is in turn largely determined by the physico-chemical conditions of the habitat, such as the need for cells to survive in highly saline environments by adjusting their internal salt concentrations (Oren, 2009). How microbial communities respond to and contribute to chemical gradients is a central question of microbial ecology and is essential to our understanding of biogeochemical cycling and biological adaptation to global change.

Salinity has an important influence on the global distribution of bacterial diversity (Lozupone and Knight, 2007). Salinity gradients occur in a wide variety of ecologically important habitats such as estuaries, wetlands, salt marshes and coastal lagoons. Many of these habitats are under increasing pressure from climate change, due to increased evaporation, reduced freshwater flows, and rising sea levels (Scavia et al., 2002; Schallenburg et al., 2003).

In high salinity environments, microbes must maintain their cellular osmotic balance via the acquisition of charged solutes (Roberts, 2005; Oren, 2009). This fundamental physiological requirement has led to the evolution of halotolerant specialists, with several studies in hypersaline habitats demonstrating that microbial diversity decreases with salinity (Estrada et al., 2004; Schapira et al., 2010; Pedrós-Alió et al., 2000; Benlloch et al., 2002) with halotolerant and

halophilic taxa becoming dominant in more extreme salinities. Shifts in microbial community structure have also been observed along estuaries (Bouvier and del Giorgio, 2002; Oakley et al., 2010; Bernhard et al., 2005) and in saline sediments (Swan et al., 2010; Hollister et al., 2010), with changes in the abundance of specific functional groups, such as ammonia-oxidizing (Bernhard et al., 2005) and sulfate-reducing bacteria (Oakley et al., 2010), and overall composition (Hollister et al., 2010; Swan et al., 2010; Bouvier and del Giorgio, 2002), suggesting the important selective role of salinity. However, it is not known how these taxonomic shifts will change the functional gene content involved in biogeochemical processes, with the majority of studies focusing on taxonomic marker genes or specific functional groups.

Metagenomics allows for the elucidation of the biochemical potential of microbial genomes present in a given environmental sample via direct sequencing of community DNA (Tyson et al., 2004;Wooley et al., 2010). Several metagenomic studies (Kunin et al., 2008; Rodriguez-Brito et al., 2010) have focused on specific hypersaline environments, but there has been no assessment of metabolic shifts along salinity gradients. Additionally, the majority of non-metagenomic studies have investigated either estuarine habitats that do not exceed 50 PSU salinity or extreme hypersaline environments, e.g. solar salterns.

In this context, the Coorong lagoon, in South Australia provides a unique model system of a continuous, natural salinity gradient from estuarine to hypersaline salinities (Lester and Fairweather, 2009; Schapira et al., 2009), which provides an opportunity to investigate shifts in the biogeochemical potential and function of microbial communities.

The Coorong lagoon is one of Australia's most significant wetlands and is listed under the Ramsar convention as a wetland of international significance (Kingsford et al., 2011). The

150km long, 2km wide system is contained between the last interglacial dune before the ocean and a modern peninsula that has been established from the mid-holocene. The system receives water inputs at one end from the Southern Ocean and the Murray River, Australia's largest freshwater system. These combined inputs result in an estuarine system at the mouth of the lagoon that becomes hypersaline along the gradient due to evaporation. In recent decades, reduced freshwater inputs due to agricultural practices and anthropogenic barriers, coupled with climate driven increases in evaporation and decreases in rainfall, have resulted in increasingly hypersaline conditions within the lagoon (Lester and Fairweather, 2009). This has led to a shift in the biogeochemical status of the system with increased nutrient levels, acidification, and degradation of the overall ecological condition of the wetland (Lester and Fairweather, 2009; Kingsford et al., 2011). A better knowledge of the response of microbial communities to these conditions is essential from the perspective of both (i) ecosystem management and (ii) as a model to understand the effect of increased salinity levels on microbially mediated biogeochemical cycles. While microbial and viral abundance and activity has been shown to increase along this salinity gradient (Schapira et al., 2009, Pollet et al., 2010; Schapira et al., Schapira et al., 2010), the identity and metabolic potential of the bacteria that drive particular steps in a biogeochemical cycle have not been characterized in this system.

We conducted a metagenomic survey of the Coorong lagoon as a model for continuous natural salinity and nutrient gradients, and describe the shifts in gene content of sediment microbial metagenomes along the salinity gradient from marine to hypersaline conditions. This provides a model for how environmental gradients can drive shifts in the biogeochemically important metabolic processes involved in salinity tolerance and in taxonomic groups involved in photosynthesis and nitrogen cycling.

2 Materials and Methods

2.1 Study sites and sample collection

Sampling was conducted at four reference stations along the Coorong lagoon, South Australia, in January 2008, during the Austral summer. Salinity varied by 99 practical salinity units (PSU) across stations. The sites were named by their salinity and defined by their GPS coordinates, which were as follows: 37 PSU (-35.551S, 138.883E), 109 PSU (-35.797S, 139.317E), 132 PSU (-35.938S, 139.488E) & 136 PSU (-36.166S, 139.651E). Ammonia concentrations at these sites ranged between 0.21 (±0.09) and 3.10 (±0.84) mgN/L, phosphate concentrations ranged between 0.05 (±0.01) and 0.27 (±0.09) mgP/L (Supplementary Fig. S1). Heterotrophic bacteria and virus like particles in porewater, as determined by flow cytometry (Marie et al., 1995; Seymour et al., 2005), increased from 4.8×10^6 ($\pm 6.3 \times 10^5$) to 1.5×10^8 ($\pm 1.4 \times 10^7$) bacteria per mL and 1.5×10^7 ($\pm 5.8 \times 10^6$) to 4.2×10^8 ($\pm 3.1 \times 10^7$) viruses per mL along the salinity gradient (Supplementary Fig. S1).

At each site, 10g of sediment, submerged in approximately 2m deep water was sampled using a sterile corer. This equated to a core containing the upper 10cm of sediment. This sampling approach averages out the vertical heterogeneity present in the sample, combining chemical gradients and pooling both oxic sand and black anaerobic mud. In each sample approximately 7cm of the core was dark grey and black mud overlaid by approximately 3cm of pale sand. Sediment cores on this scale demonstrate strong vertical gradients in Oxygen, Nitrogen, Carbon, and Sulfur (Paerl and Pickney, 1996). As our focus was on regional-scale rather than micro-scale shifts it was necessary to incorporate all of this heterogeneity in our sample to characterize the bulk metagenomic potential of the upper surface sediment, in a similar fashion to which

water metagenomic studies (e.g. Dinsdale et. al., 2008; Rusch et. al., 2007) and sediment 16S rDNA studies (e.g. Hollister et al, 2010) combine spatially heterogeneous samples to investigate regional scale shifts. Samples were stored on ice prior to DNA extraction which was performed within 8 hours of collection.

2.2 DNA extraction and sequencing

DNA was extracted from 10g of homogenized sediment using a bead beating and chemical lysis procedure (Powersoil, MoBio). Four shotgun metagenomic libraries were generated and sequenced using 454 GS-FLX pyrosequencing technology (Roche) at the Australian Genome Research facility. This sequencing yielded 68888 DNA sequences in the 37 PSU metagenome, 101003 sequences in the 109 PSU metagenome, 114335 sequences in the 132 PSU metagenome and 108257 sequences in the 136 PSU metagenome, with an average read length of 232bp. Sequence tags were trimmed by the Australian Genome Research Facility (AGRF) and subsequent sequence quality control was conducted as part of the MG-RAST submission pipeline (Meyer et al., 2008).

2.3 Bioinformatics and statistical analysis

Unassembled DNA sequences (environmental sequence tags) from each site were annotated using the MG-RAST pipeline (Meyer et al., 2008). MG-RAST implements the automated BLASTX annotation of DNA sequencing reads to the SEED non redundant database which is a database of genome sequences organized into cellular functions termed subsystems (Overbeek et al., 2005). Within MG-RAST, metabolic assignments were annotated to the SEED subsystems database (Overbeek et al., 2005) and taxonomic identification was determined based on the top

BLAST hit to the SEED taxonomy. The SEED is organized in three hierarchical levels for metabolism and six for taxonomy and allows for data to be exported at each level. The heat map function of MG-RAST version 3.0 was used to display the normalized abundance of sequences matching different categories with the Euclidian distance between profiles being displayed as a ward-based clustering dendogram. Taxonomic and metabolic reconstructions generated using MG-RAST version 2.0 with an E-value cutoff of 1×10^{-5} and a 50bp minimum alignment length were imported into the STatistical Analysis of Metagenomic Profiles (STAMP) package to test for statistically significant abundance differences in taxonomic and metabolic groupings (Parks and Beiko, 2010). These were investigated at the second and third level of the MG-RAST metabolic hierarchy and the third level of the MG-RAST taxonomic hierarchy. Fisher's exact test was used to determine the most significantly different categories, with a Storey's FDR multiple test correction applied (Agresti, 1990; Storey and Tibshirani, 2003). Confidence intervals were determined using a Newcombe-Wilson method (Newcombe, 1998). Results were filtered to display only categories with a q-value of <0.05. Given the uneven number of sequences between samples, data was normalized so that all statistical tests were conducted on proportions as is standard in both the MG-RAST (Meyer et al, 2008) and STAMP (Parks and Beiko, 2010) statistical packages, and we note that Fisher's exact test considers the difference in sample size in it's significance measure**.**

The Salinity tolerance of identified taxa were determined within the MEGAN software package (Huson et al., 2009) using the NCBI prokaryotic attributes table to display the results of a BLASTX search of our datasets against the NCBI non redundant database using CAMERA (Sun et al., 2011).

3 Results

3.1 Overall shifts in metagenomic profiles

To investigate the influence of salinity on the composition of the Coorong sediment metagenomes, we compared the abundance profiles of the metabolic potential (Fig. 1A) and the taxonomic identity of genes (Fig. 1B) sampled along the gradient. In both cases the metagenomic profiles demonstrated shifts in structure along the gradient. Metagenomes derived from hypersaline sites showed a higher degree of similarity to each other than to the 37 PSU metagenome for both function and taxonomic identity. The signature for metabolic potential was more conserved between samples than that for the phylogenetic identity of genes.

3.2 Shifts in functional potential along the salinity gradient

We further investigated shifts in the functional gene content of microbial communities along the salinity gradient using STAMP (Parks and Beiko, 2010) to determine which finer level metabolic processes were statistically over-represented in the hypersaline metagenomes relative to the 37 PSU metagenome (Fig. 2). This was investigated at the second level of the MG-RAST metabolic hierarchy.

Genes responsible for the synthesis of cell membrane bound ABC transporter proteins, predominantly composed of branched chain amino acid and oligopeptide transporters (Fig. 3A), were over-represented in the hypersaline metagenomes (Fig. 2), as were ATP synthase enzymes (Fig. 2A & 2C) and pathways responsible for the cellular response to osmotic stress. Osmotic stress genes were primarily involved in the synthesis and transport of the osmoprotectants choline, betaine, ectoine and periplasmic glucans (Fig. 3B). DNA metabolism genes and the

genes responsible for the metabolism of di- and oligosaccharide sugars were also significantly more abundant in the hypersaline metagenomes than in the 37 PSU metagenome.

Sequences related to photosynthesis and pigment synthesis were over-represented in all hypersaline metagenomes relative to the 37 PSU metagenome (Fig. 2). Specifically, the abundance of sequences matching tetrapyrrole synthesis (chlorophyll) and photosynthetic electron transport and photophosphorylation pathways were significantly higher in the hypersaline metagenomes than in the 37 PSU metagenome.

3.3 Shifts in taxonomic identity of genes along the salinity gradient

We further investigated the taxonomic identity of genes along the salinity gradient using STAMP (Parks and Beiko, 2010) to determine which finer level taxonomic groups were statistically different in abundance between the 37 PSU metagenome and the hypersaline metagenomes (Fig. 4, Supplementary Fig. S2). The cyanobacterial classes *Nostocales*, *Oscillatoriales* and *Chroococcales* were found to be over-represented in the most hypersaline metagenome (136 PSU) relative to the 37 PSU sample (Fig. 4), as was the photoheterotrophic bacterial class *Chloroflexi*, which contains the green non-sulfur bacteria .

Several archaeal taxa were over-represented in the 109 PSU, 132 PSU and 136 PSU metagenomes relative to the 37 PSU sample. Of these, the class *Methanomicrobia* was the most over-represented in all cases. The halophilic class *Halobacteria* were over-represented in the 136 PSU and 109 PSU metagenomes showing the highest increase in proportion in the most hypersaline metagenome (136 PSU) (Fig. 4; Supplementary Fig. S2).

We also observed shifts in the structure of the *Proteobacteria*. The class *δ/ε-Proteobacteria* were over-represented in hypersaline metagenomes, while the relative abundance of *γ-Proteobacteria, β-Proteobactera* and *α-Proteobacteria* were significantly higher in the 37 PSU metagenome. The classes *Bacteroidetes* and *Plactomycetacia* were also strongly over-represented in the 37 PSU metagenome.

To investigate how these shifts in taxon abundance were reflected in the salinity tolerance of members of the microbial community, we used MEGAN (Huson et al., 2009) to summarize taxonomic assignments of sequencing reads in NCBI's microbial attributes table. We found that the proportion of reads matching moderate halophiles and extreme halophiles increased by 5% and 6% respectively, and that the total number of moderate and extreme halophilic taxa increased from 15 to 32 in the 136 PSU metagenome relative to 37 PSU metagenome (Fig. 5). Overall, the majority of identifiable taxa in both of these communities were mesophilic and moderately halophilic.

4 Discussion

Our results comprise the first metagenomic survey of a model continuous natural salinity gradient and describe the shifts in gene content of sediment microbial metagenomes along the gradient from marine to hypersaline salinities. Overall shifts in the genetic composition of the metagenomes highlighted the substantial influence of salinity on the metabolic potential of microbial communities, which in turn has biogeochemical consequences. Taxonomic shifts may also reflect variation in other variables such as nutrient concentration and the relative amount of oxic and anoxic sediment present in each core, however the nature of metabolic shifts along the gradient indicate that salinity is a dominant factor, as does the increased representation of

halophiles along the gradient. The most significant differences along the gradient can be categorized into two biogeochemically important categories: osmotic stress tolerance, via acquisition of compatible solutes, and photosynthesis. Our data allows us to form several new hypotheses relating to how microbial communities may respond to increasing salinity levels in the environment, and influence the biogeochemistry of salinity gradient habitats.

4.1 Salinity tolerance via compatible solute acquisition and its influence on carbon and nutrient cycling

Many of the metabolic pathways over-represented in the hypersaline metagenomes (Fig. 2) are potentially involved in cellular halotolerance. Microorganisms can overcome the osmotic stress caused by increased salt concentration by two mechanisms: the accumulation of KCl, which requires heavy modification of the enzyme content of the cell, or by accumulating organic compatible solutes which requires less proteomic modification and allows adaptation to a broad salinity range (Oren, 2008). It is this 'organic solutes in' strategy that seems most prevalent in our data. Osmotic stress functional categories were over-represented in hypersaline metagenomes and these were largely composed of pathways responsible for choline, betaine and ectoine transport and synthesis, and the acquisition of periplasmic glucans. These solutes are common osmoprotectants in halotolerant and halophilic microorganisms. In particular, ectoine and betaine are important osmolytes in a wide range of taxonomic groups (Oren, 2008; Roberts, 2005) and betaine is an important characteristic of halotolerant *Cyanobacteria* and other phototrophic bacteria (Welsh, 2000). Choline is a precursor for betaine synthesis and its concentration has been shown to be salt dependant in halophilic bacteria (Roberts, 2005;Canovas

et al., 1998). Periplasmic glucans have been proposed to play a role as osmoprotectants in the cellular intermembrane space (Bohin, 2000).

Consistent with the osmoregulated accumulation of solutes, di- and oligosaccharide functional categories were over-represented in both hypersaline metagenomes (Fig. 2) and the biosynthesis of other sugars (galactoglycans/lipopolysaccharide) was also enriched in the most hypersaline metagenome. Many sugars act as osmoprotectants (Oren, 2008; Roberts, 2005) for example trehalose is a common compatible solute in a variety of halotolerant and halophilic microorganisms, and sucrose in halotolerant *Cyanobacteria* and *Proteobacteria* (Roberts, 2005). The presence of elevated sugar biosynthesis has the biogeochemical implications that microbially mediated cycling can occur at higher salinities and that there will be more energy available in the form of sugars to stimulate the metabolism of biogeochemically active heterotrophic bacteria.

Genes responsible for the synthesis of cell membrane bound ATP binding cassette (ABC) transporter proteins were over-represented in both hypersaline metagenomes and also potentially play a role in salinity tolerance. In our data, these enzymes were largely dominated by those involved in the transport of branched chain amino acids. Amino acids are common compatible solutes (Oren, 2008) and a branched chain amino acid ABC-transporter has been transcriptionally up-regulated during salt adaptation in the sediment bacteria *Desulfovibrio vulgaris* along with other ABC transporters responsible for betaine transport (He et al., 2010). The over-representation of sequences for ATP synthase enzymes is also potentially explained by halotolerance as these membrane bound pumps are up regulated in salt stressed yeast (Yale and

Bohnert, 2001) and a novel form of this enzyme plays a role in salinity tolerance in halotolerant *Cyanobacteria* (Soontharapirakkul et al., 2011).

In addition to providing survivability to the increasing biomass present in the hypersaline samples, which is reflected in the increase in halotolerant and halophilic taxa along the gradient, the increased synthesis and uptake of compatible solutes also has direct consequences for the nutrient cycling and greenhouse gas emissions of the sediment. The extent to which compatible solute metabolism influences primary production and provides key substrates for heterotrophic nutrition is still to be determined (Oren, 2009), but the release of osmoprotectants via diffusion, lysis and grazing provides a significant source of carbon, nitrogen and sulfur to heterotrophic microorganisms (Welsh, 2000; Howard et al. 2006). This process appears to be particularly important in hypersaline sediments and mats where the utilization of high concentration glycine-betaine, trehalose and sucrose represent a significant carbon source for microorganisms and where glycine betaine can represent up to 20% of the total nitrogen of the surface layers (Welsh, 2000; King, 1988). The potentially increased catabolism of betaine is particularly significant in hypersaline sediment where anaerobic degradation of this compound may result in methane as an end product (Welsh, 2000). Additionally, the climate regulating gas dimethylsulfide (DMS) precursor dimethylsulfoniopropionate (DMSP) is a structural analogue to betaine and shares a cellular transport system (Welsh, 2000), thus the increased abundance of betaine transport potential with salinity could also result in an increase in the accumulation of this solute which is central to global scale climate and sulfur cycles. Thus, the observation that metabolisms related to compatible solute metabolism are over-represented in hypersaline metagenomes directly links the halotolerant metabolic potential of the community to global scale nutrient cycles and climate

processes, and suggest that with increasing salinity, this influence will become further exaggerated.

4.2 Photosynthesis

The over-representation of sequences matching tetrapyrrole synthesis (chlorophyll) and photosynthetic electron transport and photophosphorylation pathways in the hypersaline metagenomes is consistent with the overrepresentation of *Cyanobacteria* in the 136 PSU metagenome. *Cyanobacteria* are abundant in hypersaline systems (Javor, 1989; Oren, 2002) particularly in the form of benthic microbial mats which drive primary productivity in hypersaline environments between 100 and 200 gl^{-1} salinity (Oren, 2009). The *Cyanobacteria* over-represented in our most hypersaline metagenome represent filamentous *Cyanobacteria*. Many taxa comprising mats are filamentous (Oren, 2002; Oren, 2009), however the sediment we sampled in this study did not show the laminated structure characteristic of cyanobacterial mats, but was sandy sediment dominated by non photosynthetic taxa. Our data indicate that increasing salinity could potentially increase the presence of filamentous *Cyanobacteria* without precipitating the transformation of porous sediment into laminated mats. Mats are associated with photosynthesis and nitrogen cycling but our results indicate that these processes occur significantly in sediments without the visual presence of stratified mat communities.

Salinity often co-varies with other parameters such as nutrient concentration and microbial/viral abundance (Schapira et al., 2009) thus other gradients in the system can be expected to influence the abundance of *Cyanobacteria* and determine their morphology, such as the increase in ammonia and phosphate concentration observed in our data (Supplementary Fig. S1). Larger cells with small surface to volume ratios, such as colonial and filamentous *Cyanobacteria*,

preferentially grow at higher nutrient concentrations and find a niche when protozoan grazing is high (Cotner and Biddanda, 2002; Pernthaler et al., 2004). Reduced grazing due to high salinity also facilitates the development of stratified mats (Oren, 2009) however grazing is still prevalent in the hypersaline Coorong (Newton, et al, submitted) potentially limiting the formation of these structures, but favoring filamentous morphologies.

The increase of photosynthetic metabolisms and taxa in the most hypersaline metagenome (136 PSU) has implications for the exchange of nutrients and CO_2 between the benthic and pelagic systems within the lagoon. Photosynthetic microbial mats and similar environments release dissolved organic carbon and oxygen to the environment and act as a sink for CO_2 (Ford, 2007). Photosynthetic benthic surfaces also provide energy for nitrogen fixation in underlying sediments as well as capturing phosphorous and sulfur from the overlying water (Ford, 2007). Whilst the extent of these environments in the Coorong remain unknown and the overall influence of salinity on production rates and nutrient flux remains undetermined (Ford, 2007), our data indicate that this habitat could become more common with further increases in salinity, which have been predicted to occur in environments such as the Coorong due to climate change (Hughes, 2003), altering the primary productivity and nutrient levels of lagoons and potentially altering mineral precipitation via changes in DOC concentration (Javor, 1989).

4.3 Concluding remarks

Our study comprises the first metagenomic characterization of a model hypersaline, continuous and natural salinity gradient and describes the shifts in gene content of sediment microbial metagenomes in the system. Shifts in the biochemical potential and identity of the microorganisms controlling the potential can be summarized as an increase in halotolerant and

benthic photosynthetic forms with salinity. This data provides the first direct observation of an increase in genes responsible for the acquisition of compatible solutes in a natural hypersaline environment as opposed to in culture. The biogeochemical implications of an increase in compatible solute acquisition and increased benthic photosynthesis potentially represent important drivers of the ecosystem biogeochemistry. Given the ecological and biogeochemical importance of salinity gradients and increased pressure on these systems from climate change and its associated effects, understanding microbial adaptation to increasing salinity at the community level is crucial to predicting how the biogeochemistry of aquatic habitats will change over space and time.

Acknowledgements

This work was supported by the Australian Research Council

References

Agresti, A.: Categorical data analysis, Wiley series in probability and mathematical statistics Applied probability and statistics, Wiley, New York, xv, 558 p. pp., 1990.

Benlloch, S., Lopez-Lopez, A., Casamayor, E. O., Ovreas, L., Goddard, V., Daae, F. L., Smerdon, G., Massana, R., Joint, I., Thingstad, F., Pedros-Alio, C., and Rodriguez-Valera, F.: Prokaryotic genetic diversity throughout the salinity gradient of a coastal solar saltern, Environ. Microbiol., 4, 349-360, 2002.

Bernhard, A. E., Donn, T., Giblin, A. E., and Stahl, D. A.: Loss of diversity of ammonia-oxidizing bacteria correlates with increasing salinity in an estuary system, Environ. Microbiol., 7, 1289-1297, 10.1111/j.1462-2920.2005.00808.x, 2005.

Bohin, J.-P.: Osmoregulated periplasmic glucans in Proteobacteria1, FEMS Microbiol. Lett., 186, 11-19, 10.1111/j.1574-6968.2000.tb09075.x, 2000.

Bouvier, T. C., and del Giorgio, P. A.: Compositional changes in free-living bacterial communities along a salinity gradient in two temperate estuaries, Limnol. Oceanogr., 47, 453-470, 2002.

Canovas, D., Vargas, C., Csonka, L. N., Ventosa, A., and Nieto, J. J.: Synthesis of Glycine Betaine from Exogenous Choline in the Moderately Halophilic Bacterium Halomonas elongata, Appl. Environ. Microbiol., 64, 4095-4097, 1998.

Cotner, J. B., and Biddanda, B. A.: Small Players, Large Role: Microbial Influence on Biogeochemical Processes in Pelagic Aquatic Ecosystems, Ecosystems, 5, 105-121, 10.1007/s10021-001-0059-3, 2002.

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

Estrada, M., Henriksen, P., Gasol, J. M., Casamayor, E. O., and Pedros-Alio, C.: Diversity of planktonic photo auto trophic microorganisms along a salinity gradient as depicted by microscopy, flow cytometry, pigment analysis and DNA-based methods, FEMS Microbiol. Ecol., 49, 281-293, 10.1016/j.femsec.2004.04.002, 2004.

Falkowski, P. G., Fenchel, T., and Delong, E. F.: The microbial engines that drive Earth's biogeochemical cycles, Science, 320, 1034-1039, DOI 10.1126/science.1153213, 2008.

Ford, P. W.: Biogeochemistry of the Coorong. Review and identification of future research requirements., Water for a Healthy Country Flagship, CSIRO, 33 pp., 2007.

Fuhrman, J. A.: Microbial community structure and its functional implications, Nature, 459, 193-199, Doi 10.1038/Nature08058, 2009.

He, Z., Zhou, A., Baidoo, E., He, Q., Joachimiak, M. P., Benke, P., Phan, R., Mukhopadhyay, A., Hemme, C. L., Huang, K., Alm, E. J., Fields, M. W., Wall, J., Stahl, D., Hazen, T. C., Keasling, J. D., Arkin, A. P., and Zhou, J.: Global Transcriptional, Physiological, and Metabolite Analyses of the Responses of Desulfovibrio vulgaris Hildenborough to Salt Adaptation, Appl. Environ. Microbiol., 76, 1574-1586, 10.1128/aem.02141-09, 2010.

Hollister, E. B., Engledow, A. S., Hammett, A. J. M., Provin, T. L., Wilkinson, H. H., and Gentry, T. J.: Shifts in microbial community structure along an ecological gradient of hypersaline soils and sediments, ISME J., 4, 829-838, 10.1038/ismej.2010.3, 2010.

Howard, E. C., Henriksen, J. R., Buchan, A., Reisch, C. R., Bürgmann, H., Welsh, R., Ye, W., González, J. M., Mace, K., Joye, S. B., Kiene, R. P., Whitman, W. B., and Moran, M. A.: Bacterial taxa that limit sulfur flux from the ocean, Science, 314, 649-652, 10.1126/science.1130657, 2006.

Hughes, L.: Climate change and australia: Trends, projections and impacts, Austral Ecology, 28, 423-443, 2003.

Huson, D. H., Richter, D. C., Mitra, S., Auch, A. F., and Schuster, S. C.: Methods for comparative metagenomics, BMC Bioinformatics, 10 Suppl 1, S12, 1471-2105-10-S1-S12 [pii] 10.1186/1471-2105-10-S1-S12, 2009.

Javor, B.: Hypersaline environments : microbiology and biogeochemistry, Brock/Springer series in contemporary bioscience, Springer-Verlag, Berlin ; New York, viii, 328 p. pp., 1989.

King, G. M.: Methanogenesis from Methylated Amines in a Hypersaline Algal Mat, Appl. Environ. Microbiol., 54, 130-136, 1988.

Kingsford, R. T., Walker, K. F., Lester, R. E., Young, W. J., Fairweather, P. G., Sammut, J., and Geddes, M. C.: A Ramsar wetland in crisis - the Coorong, Lower Lakes and Murray Mouth, Australia, Mar. Freshwater Res., 62, 255-265, 10.1071/mf09315, 2011.

Kunin, V., Raes, J., Harris, J. K., Spear, J. R., Walker, J. J., Ivanova, N., von Mering, C., Bebout, B. M., Pace, N. R., Bork, P., and Hugenholtz, P.: Millimeter-scale genetic gradients and community-level molecular convergence in a hypersaline microbial mat, Mol. Syst. Biol., 4, 198, msb200835 [pii10.1038/msb.2008.35, 2008.

Lester, R. E., and Fairweather, P. G.: Modelling future conditions in the degraded semi-arid estuary of Australia's largest river using ecosystem states, Estuar. Coast Shelf S., 85, 1-11, 10.1016/j.ecss.2009.04.018, 2009.

Lozupone, C. A., and Knight, R.: Global patterns in bacterial diversity, Proc. Natl. Acad. Sci. USA, 104, 11436-11440, DOI 10.1073/pnas.0611525104, 2007.

Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E. M., Kubal, M., Paczian, T., Rodriguez, A., Stevens, R., Wilke, A., Wilkening, J., and Edwards, R. A.: The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes, BMC Bioinformatics, 9, Artn 386 Doi 10.1186/1471-2105-9-386, 2008.

Marie, D., Brussaard, C. P. D., Thyrhaug, R., Bratbak, G., and Vaulot, D.: Enumeration of marine viruses in culture and natural samples by flow cytometry, Appl Environ Microb, 65, 45-52, 1999.

Newcombe, R. G.: Interval estimation for the difference between independent proportions: Comparison of eleven methods, Statistics in Medicine, 17, 873-890, 1998.

Oakley, B. B., Carbonero, F., van der Gast, C. J., Hawkins, R. J., and Purdy, K. J.: Evolutionary divergence and biogeography of sympatric niche-differentiated bacterial populations, ISME J., 4, 488-497, http://www.nature.com/ismej/journal/v4/n4/suppinfo/ismej2009146s1.html, 2010.

Oren, A.: Salts and Brines, in: The Ecology of Cyanobacteria, edited by: Whitton, B., and Potts, M., Springer Netherlands, 281-306, 2002.

Oren, A.: Microbial life at high salt concentrations: phylogenetic and metabolic diversity, Saline Systems, 4, 2, 1746-1448-4-2 [pii] 10.1186/1746-1448-4-2, 2008.

Oren, A.: Saltern evaporation ponds as model systems for the study of primary production processes under hypersaline conditions, Aquat. Microb. Ecol., 56, 193-204, 10.3354/ame01297, 2009.

Overbeek, R., Begley, T., Butler, R. M., Choudhuri, J. V., Chuang, H. Y., Cohoon, M., de Crecy-Lagard, V., Diaz, N., Disz, T., Edwards, R., Fonstein, M., Frank, E. D., Gerdes, S., Glass, E. M., Goesmann, A., Hanson, A., Iwata-Reuyl, D., Jensen, R., Jamshidi, N., Krause, L., Kubal, M., Larsen, N., Linke, B., McHardy, A. C., Meyer, F., Neuweger, H., Olsen, G., Olson, R., Osterman, A., Portnoy, V., Pusch, G. D., Rodionov, D. A., Ruckert, C., Steiner, J., Stevens, R., Thiele, I., Vassieva, O., Ye, Y., Zagnitko, O., and Vonstein, V.: The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes, Nucleic Acids Res., 33, 5691-5702, Doi 10.1093/Nar/Gki866, 2005.

Parks, D. H., and Beiko, R. G.: Identifying biologically relevant differences between metagenomic communities, Bioinformatics, 26, 715-721, DOI 10.1093/bioinformatics/btq041, 2010.

Paerl, H. W., and Pinckney, J. L.: A mini-review of microbial consortia: Their roles in aquatic production and biogeochemical cycling, Microbial Ecology, 31, 225-247, 10.1007/bf00171569, 1996.

Pedrós-Alió, C., Calderón-Paz, J. I., MacLean, M. H., Medina, G., Marrasé, C., Gasol, J. M., and Guixa-Boixereu, N.: The microbial food web along salinity gradients, FEMS Microbiol. Ecol., 32, 143-155, 10.1111/j.1574-6941.2000.tb00708.x, 2000.

Pernthaler, J., Zollner, E., Warnecke, F., and Jurgens, K.: Bloom of Filamentous Bacteria in a Mesotrophic Lake: Identity and Potential Controlling Mechanism, Appl. Environ. Microbiol., 70, 6272-6281, 10.1128/aem.70.10.6272-6281.2004, 2004.

Pollet, T., Schapira, M., Buscot, M. J., Leterme, S. C., Mitchell, J. G., and Seuront, L.: Prokaryotic aminopeptidase activity along a continuous salinity gradient in a hypersaline coastal lagoon (the Coorong, South Australia), Saline Systems, 6, 5, 1746-1448-6-5 [pii]10.1186/1746-1448-6-5, 2010.

Roberts, M. F.: Organic compatible solutes of halotolerant and halophilic microorganisms, Saline Systems, 1, 5, 1746-1448-1-5 [pii] 10.1186/1746-1448-1-5, 2005.

Rodriguez-Brito, B., Li, L., Wegley, L., Furlan, M., Angly, F., Breitbart, M., Buchanan, J., Desnues, C., Dinsdale, E., Edwards, R., Felts, B., Haynes, M., Liu, H., Lipson, D., Mahaffy, J., Martin-Cuadrado, A. B., Mira, A., Nulton, J., Pasic, L., Rayhawk, S., Rodriguez-Mueller, J., Rodriguez-Valera, F., Salamon, P., Srinagesh, S., Thingstad, T. F., Tran, T., Thurber, R. V., Willner, D., Youle, M., and Rohwer, F.: Viral and microbial community dynamics in four

aquatic environments, ISME .J, 4, 739-751, http://www.nature.com/ismej/journal/v4/n6/suppinfo/ismej20101s1.html, 2010.

Rusch, D. B., Halpern, A. L., Sutton, G., Heidelberg, K. B., Williamson, S., Yooseph, S., Wu, D., Eisen, J. A., Hoffman, J. M., Remington, K., Beeson, K., Tran, B., Smith, H., Baden-Tillson, H., Stewart, C., Thorpe, J., Freeman, J., Andrews-Pfannkoch, C., Venter, J. E., Li, K., Kravitz, S., Heidelberg, J. F., Utterback, T., Rogers, Y.-H., Falcón, L. I., Souza, V., Bonilla-Rosso, G., Eguiarte, L. E., Karl, D. M., Sathyendranath, S., Platt, T., Bermingham, E., Gallardo, V., Tamayo-Castillo, G., Ferrari, M. R., Strausberg, R. L., Nealson, K., Friedman, R., Frazier, M., and Venter, J. C.: The Sorcerer II Global Ocean Sampling Expedition: Northwest Atlantic through Eastern Tropical Pacific, PLoS Biol, 5, e77, 10.1371/journal.pbio.0050077, 2007.

Scavia, D., Field, J., Boesch, D., Buddemeier, R., Burkett, V., Cayan, D., Fogarty, M., Harwell, M., Howarth, R., Mason, C., Reed, D., Royer, T., Sallenger, A., and Titus, J.: Climate change impacts on U.S. Coastal and Marine Ecosystems, Estuaries and Coasts, 25, 149-164, 10.1007/bf02691304, 2002.

Schallenberg, M., Hall, C. J., and Burns, C. W.: Consequences of climate-induced salinity increases on zooplankton abundance and diversity in coastal lakes, Marine Ecology-Progress Series, 251, 181-189, 10.3354/meps251181, 2003.

Schapira, M., Buscot, M. J., Leterme, S. C., Pollet, T., Chapperon, C., and Seuront, L.: Distribution of heterotrophic bacteria and virus-like particles along a salinity gradient in a hypersaline coastal lagoon, Aquat. Microb. Ecol., 54, 171-183, Doi 10.3354/Ame01262, 2009.

Schapira, M., Buscot, M. J., Pollet, T., Leterme, S. C., and Seuront, L.: Distribution of picophytoplankton communities from brackish to hypersaline waters in a South Australian coastal lagoon, Saline Systems, 6, 2, 1746-1448-6-2 [pii] 10.1186/1746-1448-6-2, 2010.

Seymour, J. R., Patten, N., Bourne, D. G., and Mitchell, J. G.: Spatial dynamics of virus-like particles and heterotrophic bacteria within a shallow coral reef system, Mar Ecol-Prog Ser, 288, 1-8, 2005.

Soontharapirakkul, K., Promden, W., Yamada, N., Kageyama, H., Incharoensakdi, A., Iwamoto-Kihara, A., and Takabe, T.: Halotolerant Cyanobacterium Aphanothece halophytica Contains an Na+-dependent F1F0-ATP Synthase with a Potential Role in Salt-stress Tolerance, Journal of Biological Chemistry, 286, 10169-10176, 10.1074/jbc.M110.208892, 2011.

Storey, J. D., and Tibshirani, R.: Statistical significance for genomewide studies, Proc. Natl. Acad. Sci. USA, 100, 9440-9445, 10.1073/pnas.1530509100, 2003.

Sun, S. L., Chen, J., Li, W. Z., Altintas, I., Lin, A., Peltier, S., Stocks, K., Allen, E. E., Ellisman, M., Grethe, J., and Wooley, J.: Community cyberinfrastructure for Advanced Microbial Ecology

Research and Analysis: the CAMERA resource, Nucleic Acids Res., 39, D546-D551, 10.1093/nar/gkq1102, 2011.

Swan, B. K., Ehrhardt, C. J., Reifel, K. M., Moreno, L. I., and Valentine, D. L.: Archaeal and Bacterial Communities Respond Differently to Environmental Gradients in Anoxic Sediments of a California Hypersaline Lake, the Salton Sea, Appl. Environ. Microb., 76, 757-768, 10.1128/aem.02409-09, 2010.

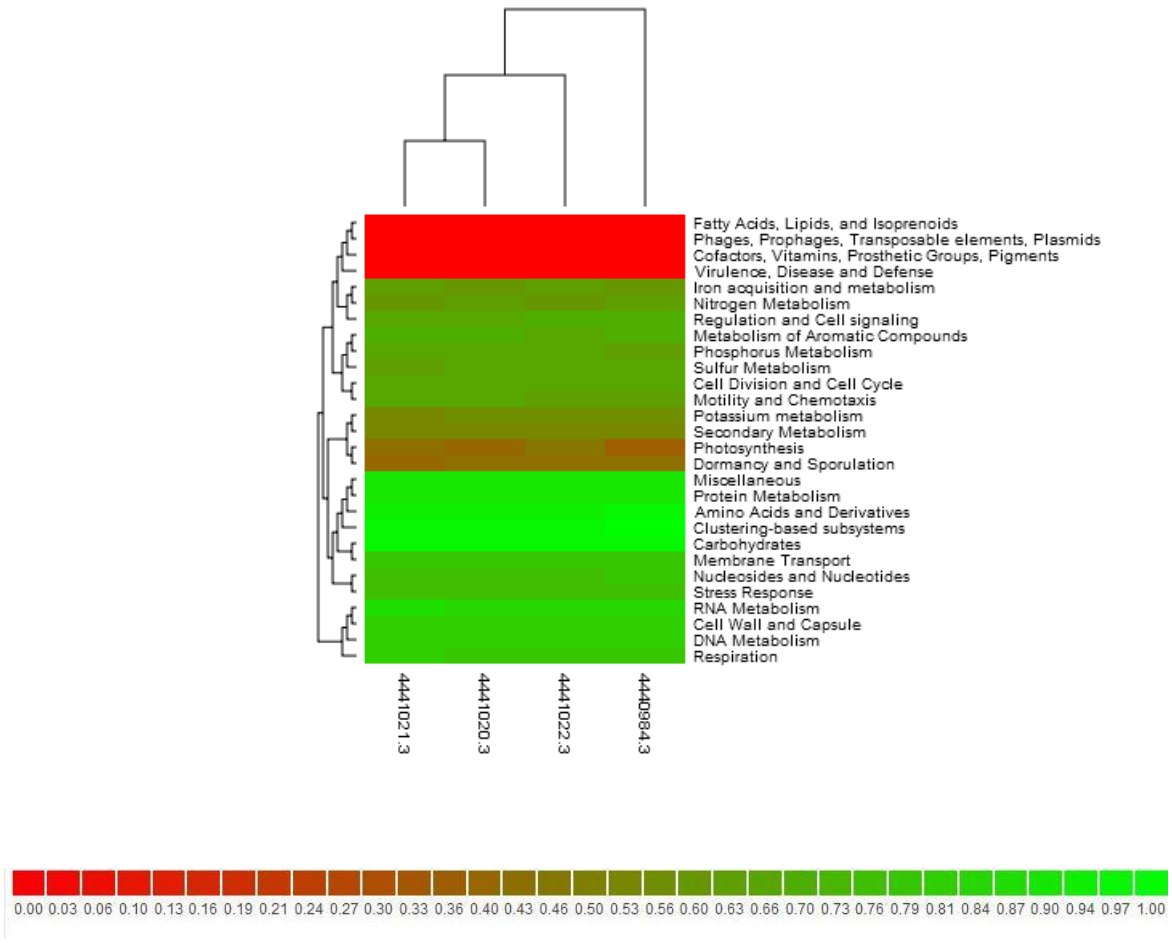
Tyson, G. W., Chapman, J., Hugenholtz, P., Allen, E. E., Ram, R. J., Richardson, P. M., Solovyev, V. V., Rubin, E. M., Rokhsar, D. S., and Banfield, J. F.: Community structure and metabolism through reconstruction of microbial genomes from the environment, Nature, 428, 37-43, Doi 10.1038/Nature02340, 2004.

Welsh, D. T.: Ecological significance of compatible solute accumulation by micro-organisms: from single cells to global climate, FEMS Microbiol. Rev., 24, 263-290, S0168-6445(99)00038-8 [pii], 2000.

Wooley, J. C., Godzik, A., and Friedberg, I.: A primer on metagenomics, PLoS Comput. Biol., 6, e1000667, 10.1371/journal.pcbi.1000667, 2010.

Yale, J., and Bohnert, H. J.: Transcript Expression in Saccharomyces cerevisiae at High Salinity, Journal of Biological Chemistry, 276, 15996-16007, 10.1074/jbc.M008209200, 2001.

A)


Fatty Acids, Lipids, and Isoprenoids
Phages, Prophages, Transposable elements, Plasmids
Cofactors, Vitamins, Prosthetic Groups, Pigments
Virulence, Disease and Defense
Iron acquisition and metabolism
Nitrogen Metabolism
Regulation and Cell signaling
Metabolism of Aromatic Compounds
Phosphorus Metabolism
Sulfur Metabolism
Cell Division and Cell Cycle
Motility and Chemotaxis
Potassium metabolism
Secondary Metabolism
Photosynthesis
Dormancy and Sporulation
Miscellaneous
Protein Metabolism
Amino Acids and Derivatives
Clustering-based subsystems
Carbohydrates
Membrane Transport
Nucleosides and Nucleotides
Stress Response
RNA Metabolism
Cell Wall and Capsule
DNA Metabolism
Respiration
4441021.3
4441020.3
4441022.3
4440984.3
0.00 0.03 0.06 0.10 0.13 0.16 0.19 0.21 0.24 0.27 0.30 0.33 0.36 0.40 0.43 0.46 0.50 0.53 0.56 0.60 0.63 0.66 0.70 0.73 0.76 0.79 0.81 0.84 0.87 0.90 0.94 0.97 1.00

B)

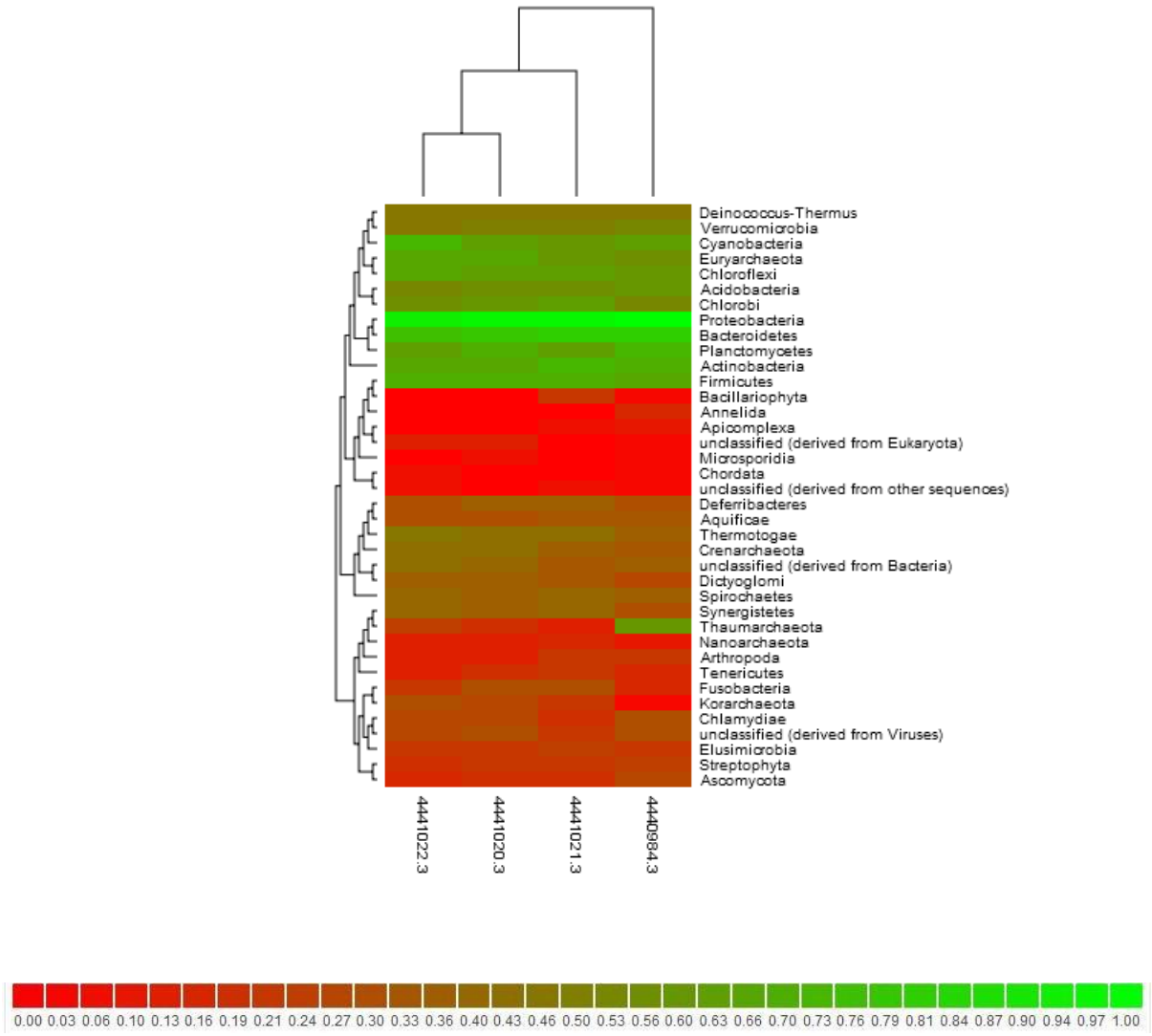
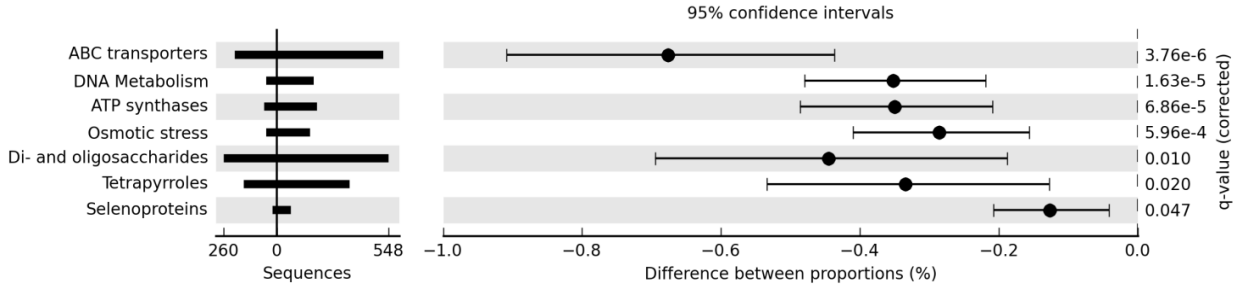
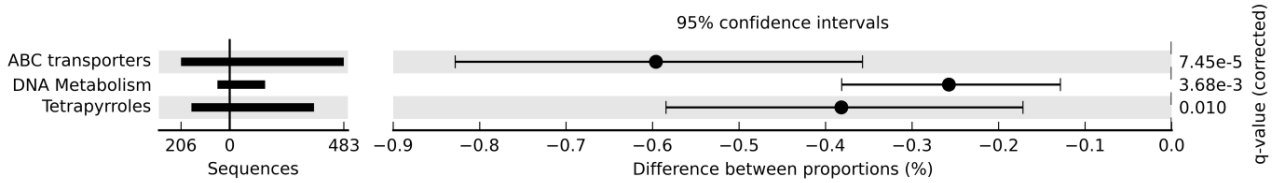


Figure 1. Disimilarity between metagenomic profiles. A) Functional potential B) Taxonomic composition. 4440984.3 = 37 PSU, 4441020.3 = 109 PSU, 4441021.3= 132 PSU, 4441022.3 = 136 PSU. Colour gradient represents proportion of sequences.

A)



B)



C)

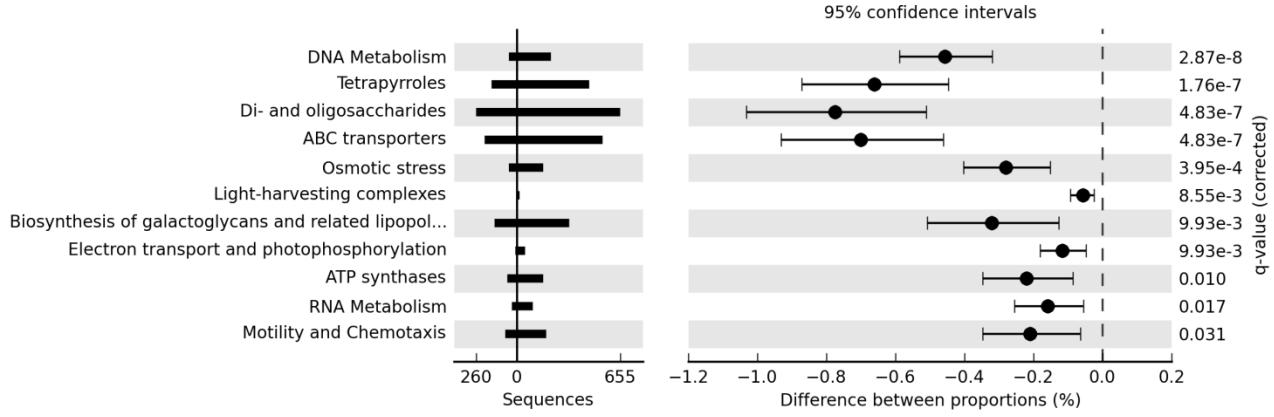
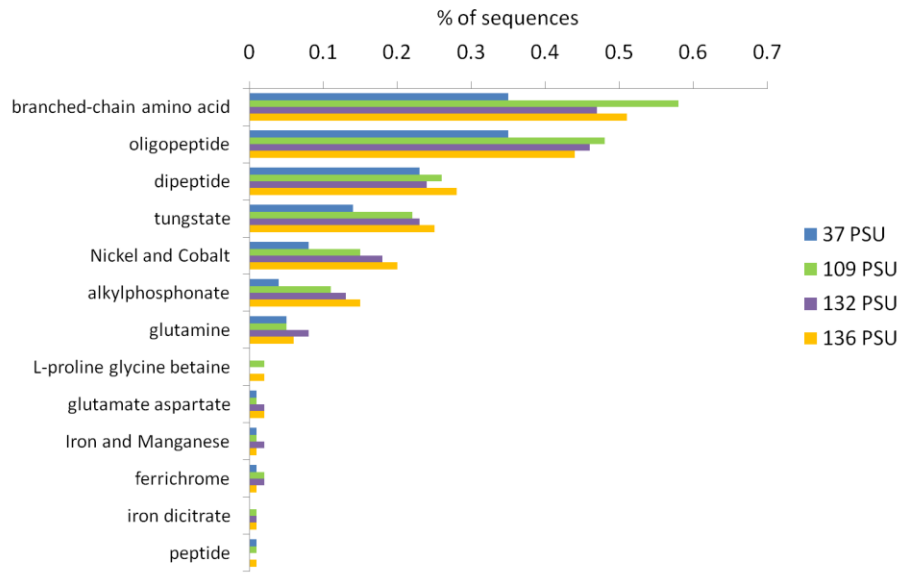


Figure 2 Metabolic processes over-represented in hypersaline metagenomes relative to the 37 PSU metagenome**.** A) 109 PSU B) 132 PSU C) 136 PSU. Corrected P-values were calculated using Storey's FDR approach.

A)



B)

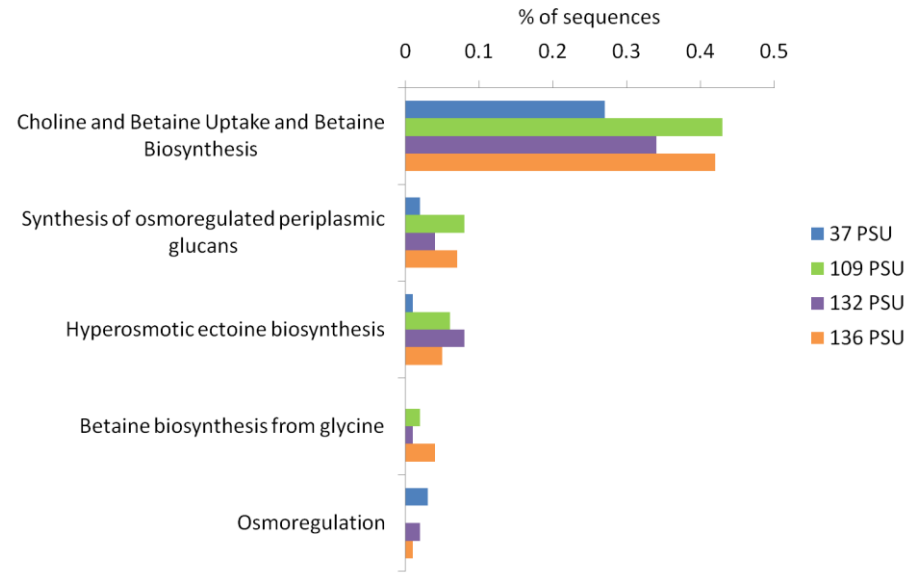


Figure 3 Breakdown of subsystem contribution to A) ABC transporter and B) osmotic stress categories. Subsystems are the third level of organization within the MG-RAST hierarchy.

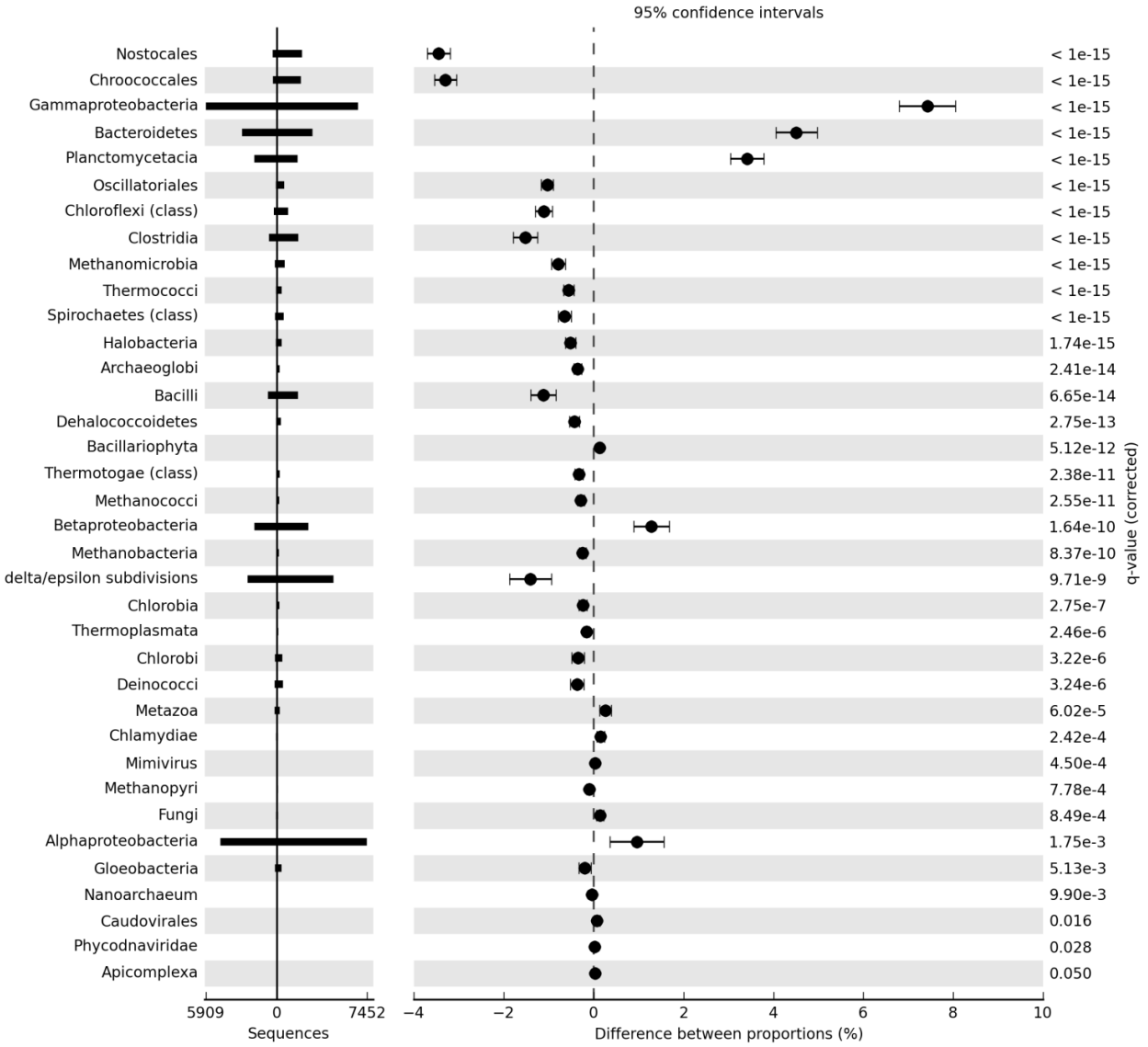
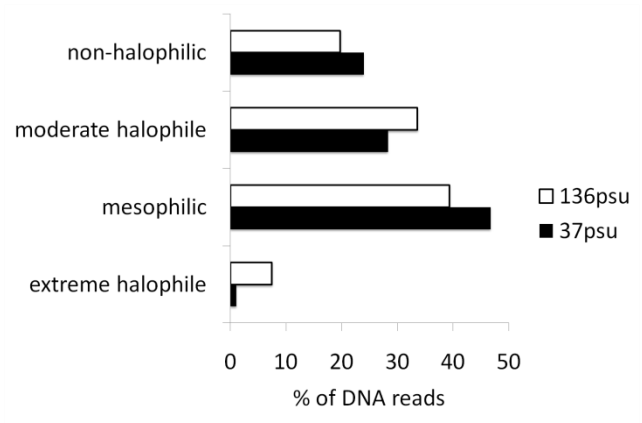


Figure 4 Taxa enriched or depleted in the 37 PSU and 136 PSU metagenomes. Corrected P-values were calculated using Storey's FDR approach. Taxa enriched in the marine (37 PSU) metagenomes have positive differences between proportions.

A)



B)

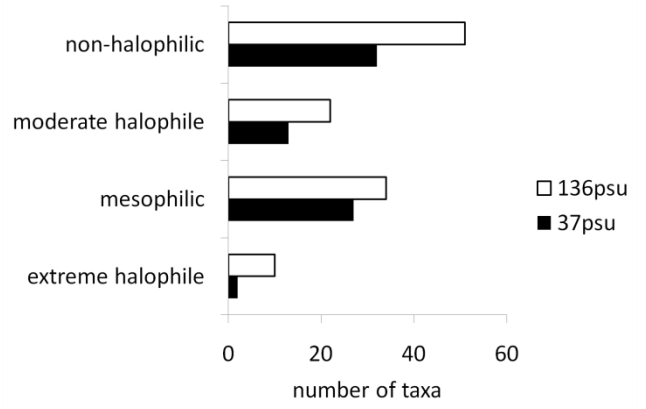
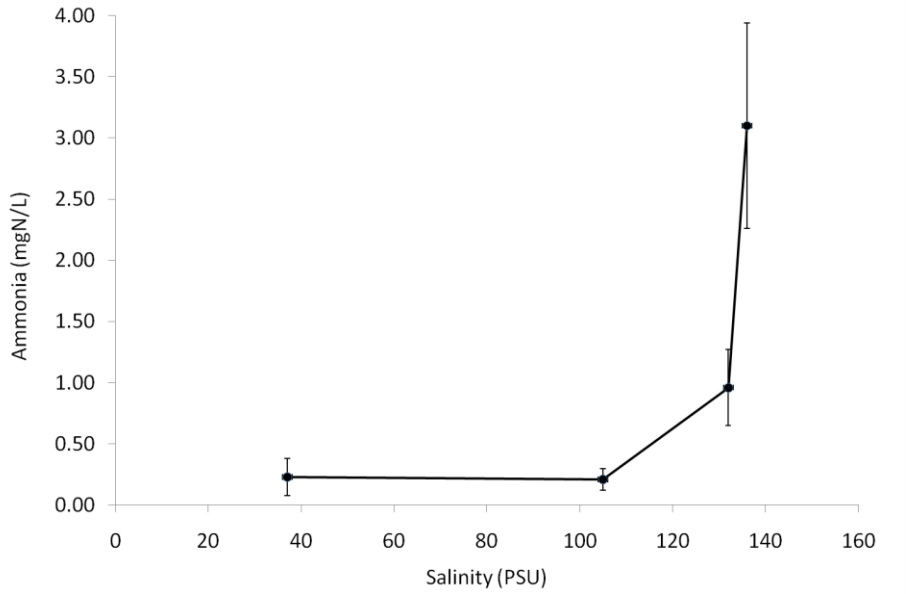


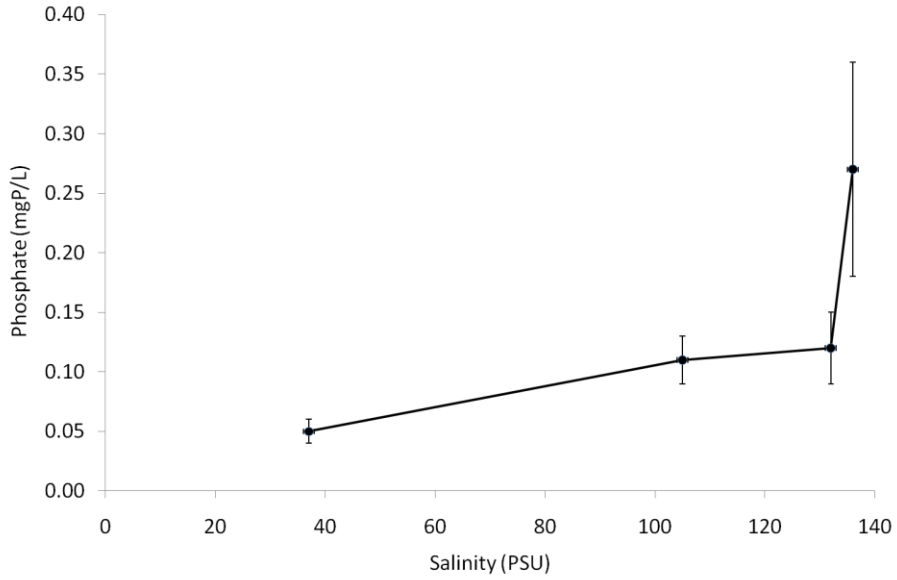
Figure 5 Representation of halophilic taxa in the 37 PSU and 136 PSU metagenome. A) % DNA reads matching taxa with a defined salinity tolerance B) number of taxa with a defined salinity tolerance

Supplementary information

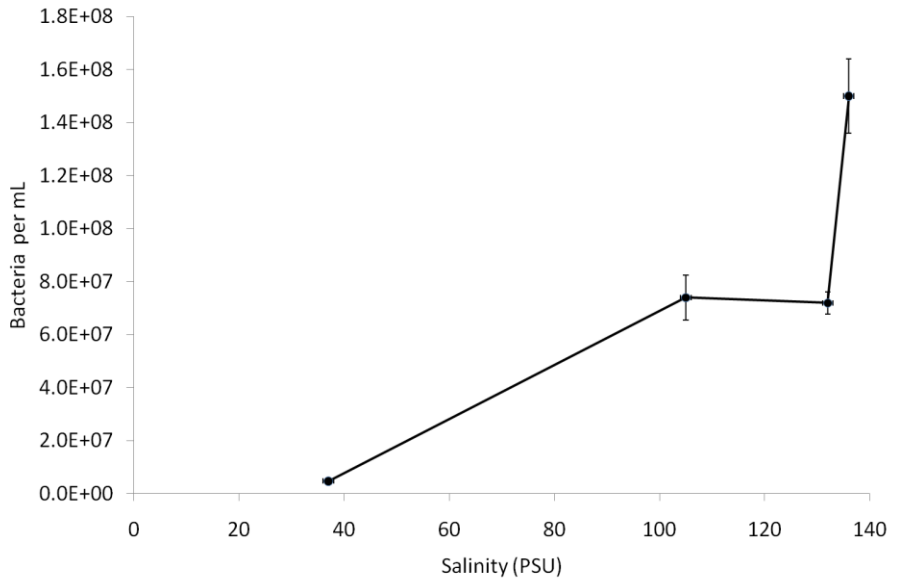
A)



B)



C)



D)

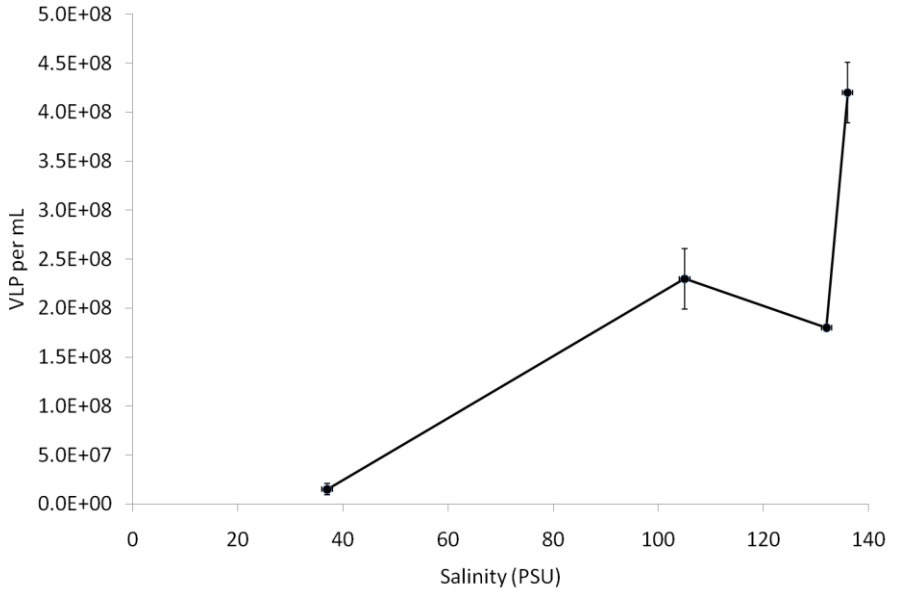
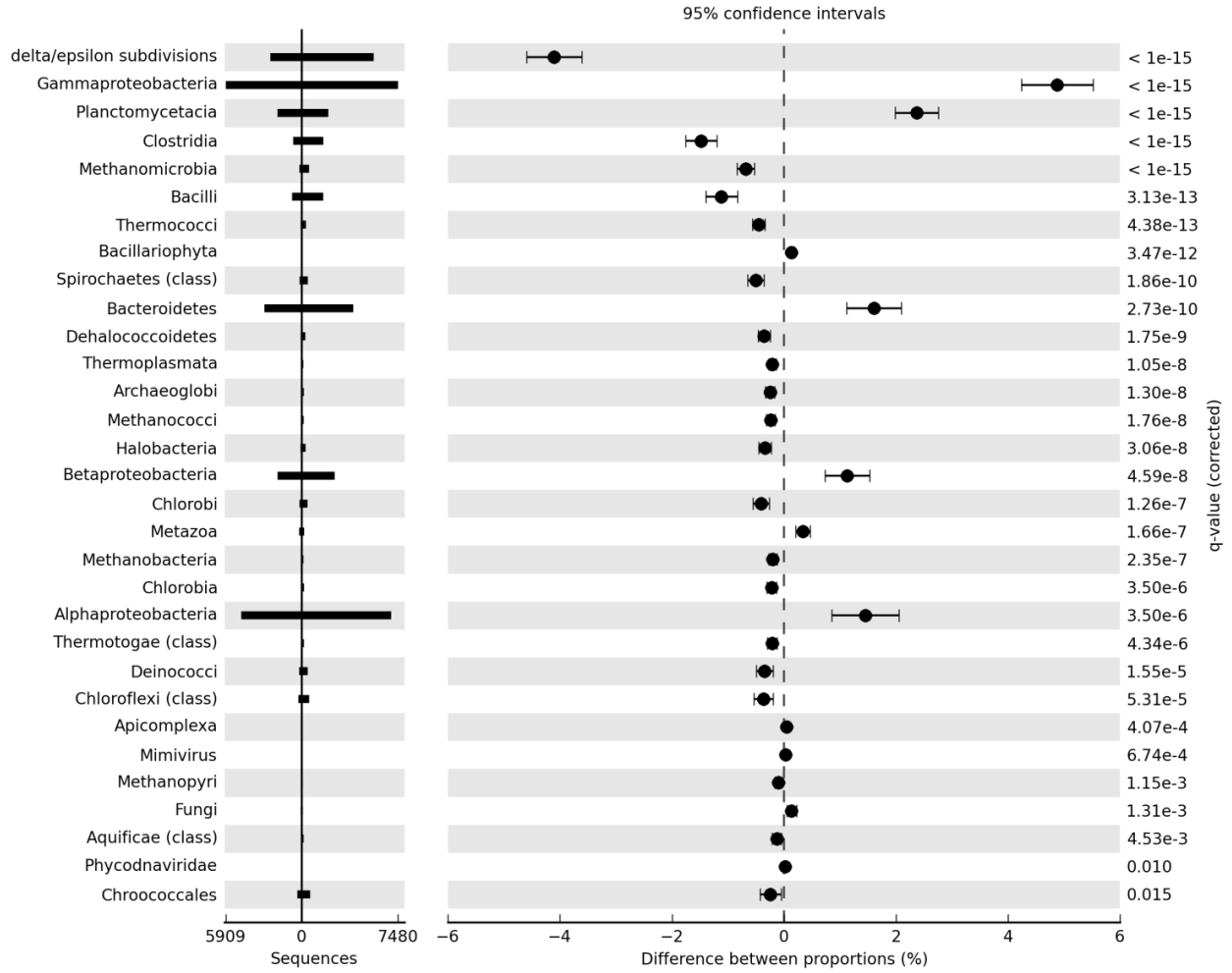


Figure S1 Environmental parameters of sampling sites A) ammonia concentration B) phosphate concentration C) heterotrophic bacteria abundance D) virus like particle (VLP) abundance. Error

bars represent standard error of the mean. Nutrient levels in porewater and were determined using a Lachat QuikChem 8500 nutrient analyzer. Microbial counts were conducted using flow cytometry as referenced in the manuscript.

A)


95% confidence intervals
delta/epsilon subdivisions
Gammaproteobacteria
Planctomycetacia
Clostridia
Methanomicrobia
Bacilli
Thermococci
Bacillariophyta
Spirochaetes (class)
Bacteroidetes
Dehalococcoidetes
Thermoplasmata
Archaeoglobi
Methanococci
Halobacteria
Betaproteobacteria
Chlorobi
Metazoa
Methanobacteria
Chlorobia
Alphaproteobacteria
Thermotogae (class)
Deinococci
Chloroflexi (class)
Apicomplexa
Mimivirus
Methanopyri
Fungi
Aquificae (class)
Phycodnaviridae
Chroococcales
< 1e-15
< 1e-15
< 1e-15
< 1e-15
< 1e-15
3.13e-13
4.38e-13
3.47e-12
1.86e-10
2.73e-10
1.75e-9
1.05e-8
1.30e-8
1.76e-8
3.06e-8
4.59e-8
1.26e-7
1.66e-7
2.35e-7
3.50e-6
3.50e-6
4.34e-6
1.55e-5
5.31e-5
4.07e-4
6.74e-4
1.15e-3
1.31e-3
4.53e-3
0.010
0.015
q-value (corrected)
5909
0
7480
Sequences
−6
−4
−2
0
2
4
6
Difference between proportions (%)

B)

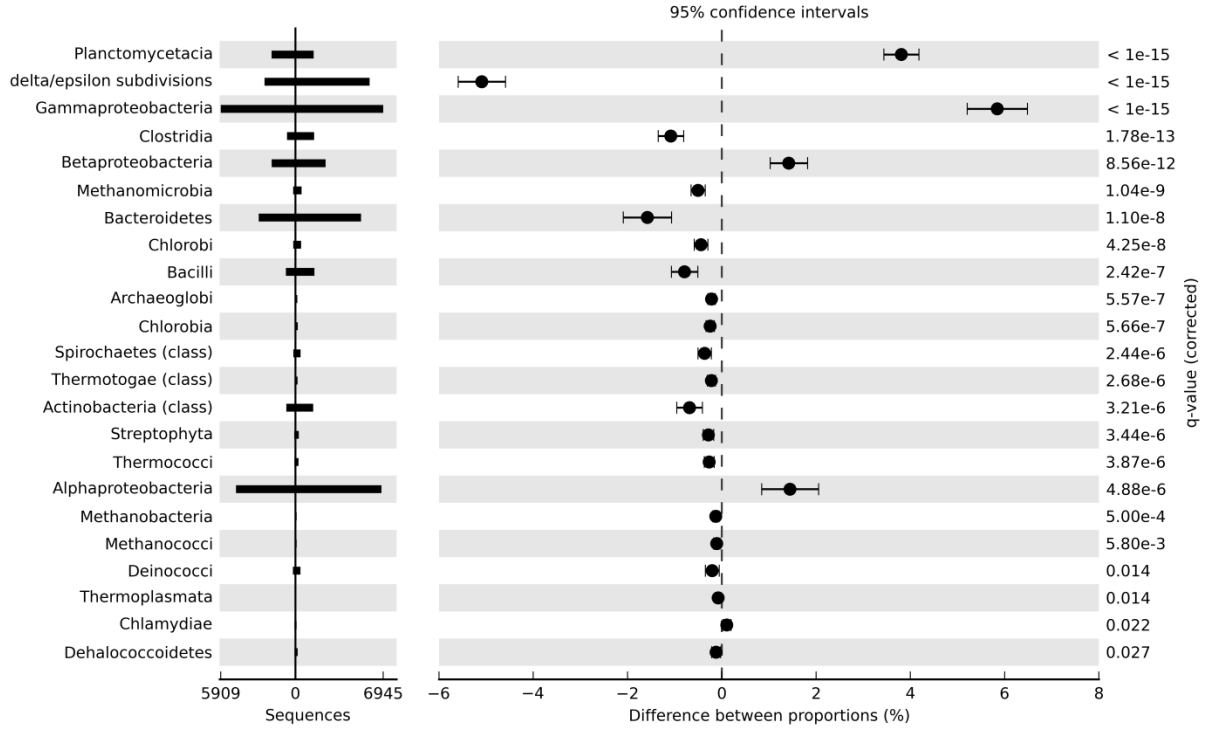


Figure S2 Taxa enriched or depleted in the 37 PSU and A) 109 PSU B) 132 PSU metagenomes. Corrected P-values were calculated using Storey's FDR approach. Taxa enriched in the marine (37 PSU) metagenomes have positive differences between proportions.

\

CHAPTER IV

Substrate type determines metagenomic profiles from diverse chemical habitats

A version of this manuscript is published in *PLoS One* as: Jeffries, T. C., J. R. Seymour, J. A. Gilbert, E. A. Dinsdale, K. Newton, S. C. Leterme, B. Roudnew, R. J. Smith, L. Seuront, and J. G. Mitchell. 2011. Substrate type determines metagenomic profiles from diverse chemical habitats. PLoS One 6: e25173..

Abstract

Environmental parameters drive phenotypic and genotypic frequency variations in microbial communities and thus control the extent and structure of microbial diversity. We tested the extent to which microbial community composition changes are controlled by shifting physiochemical properties within a hypersaline lagoon. We sequenced four sediment metagenomes from the Coorong, South Australia from samples which varied in salinity by 99 Practical Salinity Units (PSU), an order of magnitude in ammonia concentration and two orders of magnitude in microbial abundance. Despite the marked divergence in environmental parameters observed between samples, hierarchical clustering of taxonomic and metabolic profiles of these metagenomes showed striking similarity between the samples (>89%). Comparison of these profiles to those derived from a wide variety of publically available datasets demonstrated that the Coorong sediment metagenomes were similar to other sediment, soil, biofilm and microbial mat samples regardless of salinity (>85% similarity). Overall, clustering of solid substrate and water metagenomes into discrete similarity groups based on functional potential indicated that the dichotomy between water and solid matrices is a fundamental determinant of community microbial metabolism that is not masked by salinity, nutrient concentration or microbial abundance.

Introduction

Microbes numerically dominate the biosphere and play crucial roles in maintaining ecosystem function by driving chemical cycles and primary productivity [1,2]. They represent the largest reservoir of genetic diversity on Earth, with the number of microbial species inhabiting terrestrial and aquatic environments estimated to be at least in the millions [3]. However, the factors determining the spatiotemporal distributions of microbial species and genes in the environment are only vaguely understood, but are likely to include micro-scale to global-scale phenomena with different controlling elements.

Microbial community structure is determined on varying scales by a complex combination of historical factors (e.g. dispersal limitation and past environmental conditions) [4], the overall habitat characteristics [5], the physical structure of the habitat (e.g. fluid or sediment) and by changes in current environmental parameters (e.g. salinity and pH) [6-9]. Understanding the relative importance of these different effectors is central to understanding the role of microbes in ecosystem function, and therefore to predicting how resident microbial communities will adapt to, for example, increasing salinity levels due to localized climate driven evaporation and reduced rainfall [10].

Physicochemical gradients provide natural model systems for investigating the influence of environmental variables on microbial community structure. In aquatic systems, salinity is a core factor influencing microbial distribution [6,11] and has been identified as the primary factor influencing the global spatial distribution of microbial taxa [6]. Salinity gradients occur in estuaries, solar salterns and ocean depth profiles. Evidence exists for increases in abundance and decreases in the diversity of microbial communities spanning salinity gradients [9, 11-14]. This

change is wrought by variance in the halo-tolerance of different taxa and the influence of salinity on nutrient concentrations [15].

We examined the resident microbial communities inhabiting sediment at four points along a continuous natural salinity gradient in the Coorong, a temperate coastal lagoon located at the mouth of the Murray River, South Australia. To determine the relative importance of salinity, nutrient status and microbial abundance in structuring microbial community composition and function, we used shotgun metagenomics to compare the taxonomic and metabolic profiles of our samples to representative metagenomes in public databases. Our results demonstrate that the taxonomic composition and metabolic potential of our metagenomes show a conserved signature, despite the microbes existing in disparate chemical environments. Comparison to other metagenomes indicates that this signature is determined by the substrate type (i.e. sediment) of the samples.

Results

Biogeochemical environment

Dramatic shifts in physiochemical conditions occurred across the Coorong lagoon, with salinity notably varying from 37 to 136 practical salinity units (PSU) and inorganic nutrient levels changing by over an order of magnitude between sampling locations (Table 1). Practical Salinity Units (PSU) are the standard measurement of salinity in oceanography and represent a ratio of the conductivity of a solution relative to a standard, and is approximately convertible to parts per thousand of salt. For context seawater has an average salinity of 35 PSU [16]. Additionally, the abundance of heterotrophic bacteria and viruses, as determined by flow cytometry [17, 18],

increased along the salinity gradient by 31 fold and 28 fold respectively. The microbial community inhabiting this environmental gradient was explored using metagenomics, where microbial DNA was extracted and sequenced from each sampling site using a 454 GS-FLX platform (Roche). The sampling yielded between 16Mbp and 27Mbp of sequence information per library (Table 1). Approximately 30 % of the sequences from each library had significant (BLASTX E-value < 10^{-5}) matches to the SEED non-redundant database [19] as determined using the MetaGenomics Rapid Annotation using Subsystem Technology (MG-RAST) pipeline [20].

Taxonomic and metabolic profiling of metagenomes along an environmental gradient

All metagenomic libraries were dominated by bacteria (94 % of hits to the SEED database) with sequences also matching the archaea (4%), eukarya (1.5%) and viruses (0.2%). The bacterial phylum, *Proteobacteria*, dominated all four metagenomic libraries, representing over 50 % of taxonomic matches for SEED taxonomy (Fig. 1) and over 40% of ribosomal DNA matches (Table S1). Other prominent phyla included the *Bacteroidetes/Chlorobi* group (approx. 8 – 14 %), *Firmicutes* (approx. 6 - 8 %), and *Planctomycetes* (approx. 4 - 7 %). In the metagenome from the 136 PSU environment, *Cyanobacteria* were the second most represented phylum, representing approximately 12 % of the community, in the metagenomic datasets (Fig. 1) but were less prominent in the other samples, representing approximately 4 %. In the ribosomal DNA profiles generated from BLAST matches of metagenome sequences against the Ribosomal Database Project (21) (Table S1), *Cyanobacteria* were the second most abundant classified phylum in both the 132 PSU and 136 PSU metagenomes. At the phylum level, profiles were highly conserved between the four samples (Fig. 1). At level 3 within the MG-RAST

hierarchical classification scheme, which includes orders and classes [20], the most abundant taxa in all four metagenomes were the classes *γ-proteobacteria* and *α-proteobacteria* which represented approximately 20% of sequence matches. *Cyanobacteria* in the 136 PSU metagenome were predominantly represented by the orders *Nostocales* (order) an*d Chroococcales*, which each comprised approximately 40% of cyanobacterial hits (Table S2). Metagenomic profiles remained highly conserved at the genome level, which we used to compare the Coorong metagenomes to each other and to other metagenomes from diverse habitats (Fig. 3).

All Coorong metagenomes were dominated by the core metabolic functions of carbohydrate, amino acid and protein metabolism. Metabolisms indicative of a functionally diverse community were represented with heterotrophic nutrition, photosynthesis, nitrogen metabolism and sulfur metabolism contributing to the profile (Fig. 2). Paralleling the pattern observed for the taxonomic profiles, metabolic profiles were conserved between the four samples in terms of broadly defined metabolic processes, classified at the coarsest level of functional hierarchy within the MG-RAST database (Fig. 2). Metagenomic profiles remained highly conserved at the level of individual cellular processes, termed subsystems, which is the finest level of metabolic hierarchy within the MG-RAST database [20] (Fig. 4).

Comparison to metagenomic profiles from other habitats

We compared the taxonomic and metabolic structures of our metagenomes to those from a wide variety of habitats, including other hypersaline and marine sediment environments (Table 2, Table S3), using high resolution profiles derived at the genome and metabolic subsystem [19] level. For both taxonomic and metabolic profiles (Figs. 3 & 4), Coorong metagenomes showed a

high degree of statistical similarity (Bray-Curtis) to each other, despite the strong habitat gradients from which they were derived. Taxonomically, our metagenomes were all > 89% similar with the 136 PSU sample diverging at 92% similarity from the 109 PSU and 132 PSU profiles which were 94% similar. In terms of metabolic potential, they were > 89.5 % similar with the 136 PSU sample diverging at 93% similarity from the 109 PSU and 132 PSU profiles which were 93.5% similar.

The metagenomes which exhibited the greatest taxonomic similarity to the Coorong samples were from a hypersaline microbial mat, farm soil, hypersaline sediment and a freshwater stromatolite. These samples formed a discrete cluster of >82% similarity in our hierarchical tree (Fig. 3). Those with the greatest metabolic similarity to the Coorong samples were from marine sediment, farm soil, phosphorous removing sludge and a whalefall microbial mat. These samples formed a discrete cluster of >85% similarity in our hierarchical tree (Fig. 4). Notably, these metagenomes were all derived from sediment, soil, biofilm or mat samples (termed 'solid substrate' in this study) and particle rich bioreactor sludge, but varied in salinity from non-saline to hypersaline. Hypersaline water samples from the Coorong lagoon (Newton *et al*, in prep), with similar salinities to our data, did not cluster with the Coorong sediment metagenomes in terms of taxonomy or metabolism, but rather clustered with water samples from a variety of other habitats. Marine sediment samples however, clustered with the Coorong sediment metagenomes for metabolic but not taxonomic profiles. Overall, solid substrate and water metagenomes clustered into discrete metabolic similarity groups with nodes of 85% similarity.

Discussion

Despite the strong environmental heterogeneity along the gradient studied here (Table 1), taxonomic and metabolic profiles were conserved at the phyla and SEED hierarchy 1 level (Figs. 1 & 2). This similarity was even more striking at finer levels of resolution. Coorong metagenome profiles were >89% and 89.5% similar in taxonomic and metabolic composition at the genome and subsystem level respectively (Figs. 3 & 4). This indicates that the four microbial communities had similar structure, despite the intense environmental variability that occurred along the gradient. While the strong similarity between these samples, relative to other samples of comparable salinity, may to some extent be attributable to identical DNA extraction and sequencing procedures, biogeography and a shared environmental history between the samples, the clustering of our metagenomes with other solid substrate metagenomes for both taxonomic and metabolic profiles at >82% and >85% respectively, indicates that the signature of our profiles is largely determined by the substrate type of the samples (i.e. sediment). The metagenomes which show a high degree of similarity to our profiles are derived from a wide range of salinities, indicating that salinity is not the major structuring factor.

Particularly evident is the close metabolic clustering of the four Coorong sediment metagenomes with other examples of marine sediment (Fig. 4) despite these samples coming from a lower salinity than the Coorong sediment samples. This principle is highlighted by the observation that Coorong water samples of a similar salinity and identical geographic location (Table S3) do not cluster with Coorong sediment samples in terms of taxonomy or metabolic potential, but rather cluster with other water samples. We interpret this as an indication that the substrate type (e.g. water vs solid substrate) is an important determinant of microbial functional composition that

supersedes bulk environmental parameters (e.g. salinity) as the dominant structuring factor. This is further supported by the observation that the majority of metagenomes analyzed for metabolic potential cluster into two groups: a water group and a solid substrate group (Fig. 4), regardless of salinity or geographic location. Whilst it has been shown that metagenomic profiles cluster into defined biome groups [5,22], this is the first observation of such a clear dichotomy between water and solid substrate habitats which is not masked by salinity.

Salinity has previously been identified as the primary factor governing the global distribution of prokaryotic 16S rRNA sequences [6,23,24, 25]. Whilst Lozupone & Knight [6] identified substrate type (water vs sediment) as the second most important factor structuring microbial diversity after salinity, Tamames *et al* [24] concluded that salinity is more relevant than substrate type as sediment/soil and water from similar salinities clustered together in their analysis. These findings contradict the patterns apparent in our metabolic profile clustering (Fig. 4) and indicate that the phylogenetic and metabolic aspects of microbial community diversity may be driven by different dominant factors. This also implies that accessing genetic information from the entire length of the genome as opposed to a specific taxonomic marker gene can yield different interpretations. This is potentially due to the influence of lateral gene transfer and a wider representation of taxa in 16S rDNA databases as opposed to genomic databases [26, 27]. Whilst Coorong metagenomes clustered taxonomically with other solid substrate metagenomes (Fig. 3), there was not a clear dichotomy between samples from water and solid substrate types as was observed for the metabolic profiles. This indicates that the substrate type may not be as important a controlling factor for taxonomy as it is for metabolism. That substrate type is a more important determinant of metabolic composition indicates that some genes, important for living in different substrate types, are shared by varying taxa adapted to different salinities.

The samples that did not metabolically cluster within the two larger branches of 'solid substrate' and water (Fig. 4) were typically derived from more extreme hypersaline environments, such as solar salterns [28] and a hypersaline mat [29]. This indicates that in some cases, salinity can be the major factor driving the metabolic profile grouping, probably in instances where salinity reaches a critical level, whereby it selects for less diversity and more dominant taxa. This is consistent with the salinity driven clustering of the saltern metagenomes when ordinated using di-nucleotide signatures [22].

The characteristics of particular substrate types that can select the metabolic content of the microbial community could be related to the differing degree of chemical heterogeneity in fluid and solid substrate habitats. Water is mixed to a higher degree than soil/sediment thus resulting in less physiochemical heterogeneity. Soil, sediment and biofilms are extremely heterogeneous resulting in the high degree of diversity commonly observed in these habitats compared to water substrates [3,6]. This differing division of resources and niches likely explains the dichotomous clustering of water and solid substrate metagenomes observed in our data. Additionally, in aquatic systems, sediment and benthic habitats are generally more anoxic than the overlying water suggesting that reduction and oxidation (REDOX) status is also a potentially important factor driving this split. Indeed, initial investigations indicate that a prevalence of virulence, motility and anaerobic respiration genes in solid substrate habitats drive the water versus solid substrate split (Jeffries *et al*, in prep).

Our interpretation that the matrix from which the sample is derived is more important in determining the functional community structure than bulk physicochemical conditions has important implications for how we predict changes in microbial community function in the

context of climate change driven increases in salinity levels or eutrophication associated with anthropogenic inputs. For example, the Coorong is currently undergoing a period of increasing salinity levels and eutrophication [30], reflected in the gradient examined here. Our results suggest that, whilst small scale changes in gene abundance occur across this salinity gradient (for example regulation/signaling and metabolism of aromatic compounds; Fig. 2), the overall functional potential of the microbial community remains similar between salinities and demonstrates a high degree of similarity to lower salinity marine sediment at the subsystem level (Fig. 4). This indicates that while shifts in the composition of the microbial community may occur following further shifts in salinity, the overall biogeochemical potential of the community may remain relatively unchanged. Of course, extreme increases in salinity will potentially result in the emergence of dominant specialist species, decreasing diversity and potentially influencing function.

There is the potential that the discrete clustering of our samples may be related to technical bias, because of the different strategies for sample collection, sequencing and analysis of metagenomes from other locations. However, when we compared our data with metagenomes generated using different DNA extraction techniques and sequencing platforms, no discernible pattern emerged that can link the relatedness of metagenomes to elements of methodology (Figs 3 & 4). DNA extraction and sequencing techniques have also been shown not to significantly influence metagenomic profile discrimination by habitat [31]. Additionally, marine sediment samples extracted in the same lab using identical techniques did not cluster taxonomically with the Coorong samples (Fig. 3) and Coorong water samples extracted using the same lab and techniques did not cluster with the Coorong sediment samples (Figs 3 & 4), indicating methodology is not obscuring environmental clustering. One caveat that should be considered

when interpreting our data is the use of annotated data to compare metagenomes. Our data is reflective of the genomes and metabolic subsystems present in the MG-RAST database [20] and should be interpreted as patterns observed in the context of this diversity. Metagenomic databases are composed of taxa for which whole genome sequences exist, which represent a biased subsection of microbial diversity heavily skewed towards cultured organisms chosen because of ease of growth or interesting phenotypes [26, 27]. Thus the databases tend to be skewed towards the phyla Proteobacteria, Firmicutes, Actinobacteria and Bacteroidetes [26]. Whilst genome based databases represent a valid reference point for relative comparison of the taxonomic affiliation of subsystems observed in the data, which has been routinely applied for metagenomes [20] a much broader view of the taxonomic variability can be provided by the 16S rDNA gene [26]. Further analysis using clustering algorithms [32] and di-nucleotide frequencies [22] will shed light on how our un-annotated data is similar to other metagenomes.

This study focused on the balance between taxonomic and metabolic identifiers to determine the dominant controlling environmental factor. We found substrate type is the dominant controller of gene abundance. To date, the majority of community scale microbial biogeography studies have considered the presence or absence of particular taxonomic units. In many cases however, microbial biogeography is not binary, with most taxa being present but at a low abundance in the so called 'rare biosphere' [33]. Additionally, functional genes may be passed between different taxa via lateral gene transfer [34,35] indicating that taxonomy alone is not a determinant of community function. More sophisticated approaches which consider complex patterns in the metagenomic structure of communities and the complex interactions between different drivers acting on different scales are necessary to understand the spatial distribution of microbial diversity. High throughput sequencing allows profiling of both taxonomic and metabolic

diversity and when coupled to statistical techniques [5, 36-39] and standardized records of metadata [40] patterns in the composition of microbial metagenomes begin to emerge. One such pattern in our data is the high degree of taxonomic and functional similarity between metagenomes derived across a strong salinity, nutrient and abundance gradient and between metagenomes derived from sediment/soil/mat metagenomes regardless of salinity. Another pattern is the dichotomous clustering of solid substrate metagenomes and water metagenomes into discrete similarity groups which are not masked by differences in salinity. Overall our results suggest that substrate type (water or solid substrate) plays a fundamental role in determining the composition of the metagenome and that, in addition to extant physiochemical parameters, needs to be considered when interpreting patterns in microbial community diversity.

Materials and methods

Site selection and sediment sampling

Sampling was conducted along the 100 km long, shallow temperate coastal lagoon comprising the Coorong, in South Australia (35°33'3.05"S, 138°52'58.80"E), which is characterized by a strong continuous gradient from estuarine to hypersaline salinities. Samples were collected from four sites along the salinity gradient. The sites were characterized by differing salinities and nutrient status (Table 1). Sediment for DNA extraction was sampled using a new 1.5 cm diameter sterile corer at each site, and included the upper 10 cm of sediment. Sample cores were transferred to a sterile 50mL centrifuge tube, stored and transported on ice in the dark following collection, and DNA extraction was undertaken within six hours of sampling.

For each site, nutrient levels in porewater and overlying water were determined using a Lachat QuikChem 8500 nutrient analyzer and pH, dissolved oxygen and salinity were measured using a 90FL-T (TPS) multi-parameter probe. Abundance of heterotrophic bacteria and viruses in sediment porewater was assessed using a Becton Dickinson FACScanto flow cytometer and previously described protocols [17, 18]. In line with previous studies [e.g. 41], porewater microbial abundance was used to compare sediment samples using flow cytometry, potentially representing a lower estimate of the entire sediment abundance [42], which includes particle-attached bacteria and viruses. Sampling was conducted under a Government of South Australia Department of Environment and Heritage Permit to Undertake Scientific Research.

Metagenomic sequencing

Microbial community DNA was extracted from c.a.10g of homogenized sediment, using the entire volume of the sediment core, using a bead beating and chemical lysis extraction kit (MoBio, Solano Beach, CA.) and further concentrated using ethanol precipitation. DNA quality and concentration was determined by agarose gel electrophoresis and spectrophotometry and >5µg of high molecular weight DNA was sequenced at the Australian Genome Research Facility. Sequencing was conducted on a GS-FLX pyrosequencing platform (Roche) using a multiplex barcoding approach to distinguish between the four libraries on a single plate. Sequencing yielded between 16Mbp and 27Mbp of sequence information per library, with an average read length of 232.5bp (Table 1).

Bioinformatics and statistical analysis

Unassembled sequences (environmental gene tags) were annotated using the MetaGenomics Rapid Annotation using Subsystem Technology (MG-RAST) pipeline version 2.0 (http://metagenomics.nmpdr.org/) [20], with a BLASTX E-value cut-off of $E<1\times10^{-5}$ and a minimum alignment length of 50bp. The abundance of individual sequences matching a particular SEED subsystem (groups of genes involved in a particular metabolic function) [19] were normalized by sequencing effort and used to generate a metabolic profile of the metagenome. Taxonomic profiles were generated within MG-RAST using the normalized abundance of the phylogenetic identity of sequence matches to the SEED database [19] and Ribosomal Database Project (Table S1) both with a BLAST E-value cut-off of $E<1\times10^{-5}$ and a minimum alignment length of 50bp [21]. The MG-RAST pipeline [20] implements the automated BLASTX annotation of metagenomic sequencing reads against the SEED non-redundant database [19], a manually curated collection of genome project derived genes grouped into specific metabolic processes termed 'subsystems'. The SEED matches of Protein Encoding Genes (PEGs) derived from the sampled metagenome may be reconstructed either in terms of metabolic function or taxonomic identity at varying hierarchical levels of organization. For taxonomy, there are five levels from domain to genome level and for metabolism there are three sequential nested groupings termed level 1, level 2 and subsystem. In our data, metabolic information was derived at the coarsest level of organization, the generalized cellular functions, termed level 1 (Fig. 2), and the finest, individual subsystems (Fig. 4). Taxonomy was profiled at the phylum (Fig. 1) and genome (Fig. 3) level. In order to statistically investigate the similarity of the four Coorong metagenomes, as well as the metagenomic profiles publicly available on the MG-RAST server and in our own database (Table 2, Table S3), we generated a heatmap of the

frequency of MG-RAST hits to each individual taxa (genome level) or subsystem for each metagenome, which had been normalized by dividing by the total number of hits to remove bias in sequencing effort or differences in read length. These hits were identified using an E-value cut-off of E<0.001. Statistical analyses were conducted on square root transformed frequency data using Primer 6 for Windows (Version 6.1.6, Primer-E Ltd. Plymouth) [43]. Hierarchical agglomerative clustering (CLUSTER) [44] was used to display the Bray-Curtis similarity relationships between our profiles and those of the publicly available metagenomes with the results displayed as a group average dendogram. Specific Bray-Curtis similarities for individual clusters were taken from the Primer 6 CLUSTER output , which displays the stepwise construction of the dendogram.

Acknowledgements

We thank the anonymous reviewers for their feedback and suggestions. This work was funded by the Australian Research Council.

References

1. Falkowski PG, Fenchel T, Delong EF (2008) The microbial engines that drive Earth's biogeochemical cycles. Science 320: 1034-1039.

2. Azam F, Malfatti F (2007) Microbial structuring of marine ecosystems. Nat Rev Microbiol 5: 782-791.

3. Torsvik V, Ovreas L, Thingstad TF (2002) Prokaryotic diversity - Magnitude, dynamics, and controlling factors. Science 296: 1064-1066.

4. Martiny JBH, Bohannan BJM, Brown JH, Colwell RK, Fuhrman JA, et al. (2006) Microbial biogeography: putting microorganisms on the map. Nat Rev Microbiol 4: 102-112.

5. Dinsdale EA, Edwards RA, Hall D, Angly F, Breitbart M, et al. (2008) Functional metagenomic profiling of nine biomes. Nature 452: 629-U628.

6. Lozupone CA, Knight R (2007) Global patterns in bacterial diversity. Proc Natl Acad Sci USA104: 11436-11440.

7. Hewson I, Paerl RW, Tripp HJ, Zehr JP, Karl DM (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.

8. Ramette A, Tiedje JM (2007) Multiscale responses of microbial life to spatial distance and environmental heterogeneity in a patchy ecosystem. Proc Natl Acad Sci USA 104: 2761-2766.

9. Hollister EB, Engledow AS, Hammett AJM, Provin TL, Wilkinson HH, et al. (2010) Shifts in microbial community structure along an ecological gradient of hypersaline soils and sediments. ISME J 4: 829-838.

10. Hughes L (2003) Climate change and Australia: Trends, projections and impacts. Austral

Ecology 28: 423-443.

11. Schapira M, Buscot MJ, Leterme SC, Pollet T, Chapperon C, et al. (2009) Distribution of heterotrophic bacteria and virus-like particles along a salinity gradient in a hypersaline coastal lagoon. Aquatic Microbial Ecology 54: 171-183.
12. Benlloch S, Lopez-Lopez A, Casamayor EO, Ovreas L, Goddard V, et al. (2002) Prokaryotic genetic diversity throughout the salinity gradient of a coastal solar saltern. Environ Microbiol 4: 349-360.
13. Gasol JM, Casamayor EO, Joint I, Garde K, Gustavson K, et al. (2004) Control of heterotrophic prokaryotic abundance and growth rate in hypersaline planktonic environments. Aquatic Microbial Ecology 34: 193-206.
14. Schapira M, Buscot MJ, Pollet T, Leterme S, Seuront L (2010) Distribution of picophytoplankton communities from brackish to hypersaline waters in a South Australian coastal lagoon. Saline Systems 6: 2. Doi 10.3354/Ame01262
15. Javor B (1989) Hypersaline environments : microbiology and biogeochemistry. Berlin ; New York: Springer-Verlag. viii, 328 p. p.
16. Segar DA (1998) Introduction to ocean sciences. Belmont, CA: Wadsworth Pub. xxiii, 497, 489 p. p.
17. Seymour JR, Patten N, Bourne DG, Mitchell JG (2005) Spatial dynamics of virus-like particles and heterotrophic bacteria within a shallow coral reef system. Marine Ecology-Progress Series 288: 1-8.
18. Marie D, Brussaard CPD, Thyrhaug R, Bratbak G, Vaulot D (1999) Enumeration of marine viruses in culture and natural samples by flow cytometry. Appl Environ Microbiol 65: 45-52.

19. Overbeek R, Begley T, Butler RM, Choudhuri JV, Chuang HY, et al. (2005) The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. Nucleic Acids Res 33: 5691-5702.

20. Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, et al. (2008) The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC Bioinformatics 9: Doi 10.1186/1471-2105-9-386

21. Cole JR, Chai B, Farris RJ, Wang Q, Kulam-Syed-Mohideen AS, et al. (2007) The ribosomal database project (RDP-II): introducing myRDP space and quality controlled public data. Nucleic Acids Research 35: D169-D172.

22. Willner D, Thurber RV, Rohwer F (2009) Metagenomic signatures of 86 microbial and viral metagenomes. Environ Microbiol 11: 1752-1766.

23. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, et al. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proc Natl Acad Sci USA 10.1073/pnas.1000080107

24. Tamames J, Abellan JJ, Pignatelli M, Camacho A, Moya A (2010) Environmental distribution of prokaryotic taxa. BMC Microbiology 10: Doi 10.1186/1471-2180-10-85

25. Auguet, J.C., Barberan, A., & Casamayor, E. O. (2010). Global ecological patterns in uncultured Archaea. ISME J, 4: 182-190.

26. Hugenholtz, P. (2002) Exploring prokaryotic diversity in the genomic era. Genome Biol 3: REVIEWS0003.

27. Wu, D., Hugenholtz, P., Mavromatis, K., Pukall, R., Dalin, E., Ivanova, N. N., Kunin, V., et al. (2009). A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea. Nature, 462, 1056-1060.

28. Rodriguez-Brito B, Li L, Wegley L, Furlan M, Angly F, et al. (2010) Viral and microbial community dynamics in four aquatic environments. ISME J 4: 739-751.

29. Kunin V, Raes J, Harris JK, Spear JR, Walker JJ, et al. (2008) Millimeter-scale genetic gradients and community-level molecular convergence in a hypersaline microbial mat. Mol Syst Biol 4: 198.

30. Lester RE, Fairweather PG (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.

31. Delmont, T. O., Malandain, C., Prestat, E., Larose, C., Monier, J.-M., Simonet, P., & Vogel, T. M. (2011). Metagenomic mining for microbiologists. ISME J, http://dx.doi.org/10.1038/ismej.2011.61

32. Li WZ (2009) Analysis and comparison of very large metagenomes with fast clustering and functional annotation. BMC Bioinformatics 10: Doi 10.1186/1471-2105-10-359

33. Sogin ML, Morrison HG, Huber JA, Mark Welch D, Huse SM, et al. (2006) Microbial diversity in the deep sea and the underexplored "rare biosphere". Proc Natl Acad Sci USA 103: 12115-12120.

34. Doolittle WF, Zhaxybayeva O (2010) Metagenomics and the Units of Biological Organization. Bioscience 60: 102-112.

35. Boucher Y, Douady CJ, Papke RT, Walsh DA, Boudreau MER, et al. (2003) Lateral gene transfer and the origins of prokaryotic groups. Annual Review of Genetics 37: 283-328.

36. Rodriguez-Brito B, Rohwer F, Edwards RA (2006) An application of statistics to comparative metagenomics. BMC Bioinformatics 7: Doi 10.1186/1471-2105-7-162

37. Parks DH, Beiko RG (2010) Identifying biologically relevant differences between metagenomic communities. Bioinformatics 26: 715-721.

38. Kristiansson E, Hugenholtz P, Dalevi D (2009) ShotgunFunctionalizeR: an R-package for functional comparison of metagenomes. Bioinformatics 25: 2737-2738.

39. Mitra S, Gilbert JA, Field D, Huson DH (2010) Comparison of multiple metagenomes using phylogenetic networks based on ecological indices. ISME J 4: 1236-1242.

40. Field D, Garrity G, Gray T, Morrison N, Selengut J, et al. (2008) The minimum information about a genome sequence (MIGS) specification. Nature Biotechnology 26: 541-547.

41. Drake LA, Choi KH, Haskell AGE, Dobbs FC (1998) Vertical profiles of virus-like particles and bacteria in the water column and sediments of Chesapeake Bay, USA. Aquatic Microbial Ecology 16: 17-25.

42. Helton RR, Liu L, Wommack KE (2006) Assessment of factors influencing direct enumeration of viruses within estuarine sediments. Applied and Environmental Microbiology 72: 4767-4774.

43. Clarke A, Gorley R (2006) PRIMER v6: User Manual/Tutorial. Plymouth, UK: PRIMER-E.

44. Clarke KR (1993) nonparametric multivariate analyses of changes in community structure.Australian Journal of Ecology 18: 117-143.

45. Rusch DB, Halpern AL, Sutton G, Heidelberg KB, Williamson S, et al. (2007) The Sorcerer II Global Ocean Sampling expedition: Northwest Atlantic through Eastern Tropical Pacific. PLoS Biology 5: 398-431.

46. Dinsdale EA, Pantos O, Smriga S, Edwards RA, Angly F, et al. (2008) Microbial Ecology of Four Coral Atolls in the Northern Line Islands. PLoS One 3: Doi 10.1371/Journal.Pone.0001584

47. Gilbert JA, Field D, Huang Y, Edwards R, Li WZ, et al. (2008) Detection of Large Numbers of Novel Sequences in the Metatranscriptomes of Complex Marine Microbial Communities. PLoS One 3: Doi 10.1371/Journal.Pone.0003042

48. Edwards RA, Rodriguez-Brito B, Wegley L, Haynes M, Breitbart M, et al. (2006) Using pyrosequencing to shed light on deep mine microbial ecology. BMC Genomics 7: Doi 10.1186/1471-2164-7-57

49. Tringe SG, von Mering C, Kobayashi A, Salamov AA, Chen K, et al. (2005) Comparative metagenomics of microbial communities. Science 308: 554-557.

50. Breitbart M, Hoare A, Nitti A, Siefert J, Haynes M, et al. (2009) Metagenomic and stable isotopic analyses of modern freshwater microbialites in Cuatro CiEnegas, Mexico. Environ Microbiol 11: 16-34.

51. Martin HG, Ivanova N, Kunin V, Warnecke F, Barry KW, et al. (2006) Metagenomic analysis of two enhanced biological phosphorus removal (EBPR) sludge communities. Nature Biotechnology 24: 1263-1269.

Figure Legends

Figure 1. Taxonomic composition (Phyla level) of four metagenomic libraries derived from Coorong lagoon sediment. Relative representation in the metagenome was calculated by dividing the number of hits to each category by the total number of hits to all categories, thus normalizing by sequencing effort. Hits were generated by BLASTing sequences to the SEED database with an E-value cut-off of 1×10^{-5} and a minimum alignment of 50bp.

Figure 2. Metabolic composition of four metagenomic libraries derived from Coorong lagoon sediment. Relative representation in the metagenome was calculated by dividing the number of hits to each category by the total number of hits to all categories, thus normalizing by sequencing effort. Hits were generated by BLASTing sequences to the SEED database with an E-value cut-off of 1×10^{-5} and a minimum alignment of 50bp.

Figure 3. Comparison of taxonomic profiles derived from selected metagenomes publicly available on the MG-RAST database. The hierarchical agglomerative cluster plot (group average) is derived from a Bray-Curtis similarity matrix calculated from the square root transformed abundance of DNA fragments matching taxa in the SEED database (BLASTX E-value < 0.001, genome level taxonomy).

Figure 4. Comparison of metabolic profiles derived from selected metagenomes publicly available on the MG-RAST database. The hierarchical agglomerative cluster plot (group average) is derived from a Bray-Curtis similarity matrix calculated from the square root transformed abundance of DNA fragments matching subsystems in the SEED database (BLASTX E-value < 0.001).

Supporting information legends

Table S1. Percentage of Ribosomal DNA matches to bacterial phyla. Relative representation in the metagenome was calculated by dividing the number of hits to each category by the total number of hits to all categories. Hits were generated by BLASTing sequences to the Ribosomal Database Project [21], via MG-RAST [20], with an E-value cut-off of 1×10^{-5} and a minimum alignment of 50bp. Due to inconsistencies in 16S rDNA copy number, these relative abundances represent estimates of overall ribosomal DNA composition at phyla level only.

Table S2. Relative proportion of matches to the SEED taxonomic hierarchy. Relative representation in the metagenome was calculated by dividing the number of hits to each category by the total number of hits to all categories. Hits were generated by BLASTing sequences to the SEED database with an E-value cut-off of 1×10^{-5} and a minimum alignment of 50bp.

Table S3. Detailed summary of metagenomes used in this study. All metagenomes are publicly available on the MG-RAST server (http://metagenomics.nmpdr.org/) [20]. Number of database hits (BLASTX) are determined using an E-value cut-off of 0.001. References are provided in Table 2 of the manuscript. Bold = this study.

Table 1. Sequencing data and environmental metadata for metagenomic sampling sites.

Sampling Site	37 PSU	109 PSU	132 PSU	136 PSU
Number of reads	68888	101003	114335	108257
Average read length (bp)	232	234	232	232
% Sequences matching SEED subsystems	27	30	26	29
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)	4.8×10^6 ($\pm 6.3 \times 10^5$)	7.4×10^7 ($\pm 8.4 \times 10^6$)	7.2×10^7 ($\pm 4.2 \times 10^6$)	1.5×10^8 ($\pm 1.4 \times 10^7$)
Porewater virus concentration (per mL)	1.5×10^7 ($\pm 5.8 \times 10^6$)	2.3×10^8 ($\pm 3.1 \times 10^7$)	1.8×10^8 ($\pm 1.5 \times 10^7$)	4.2×10^8 ($\pm 3.1 \times 10^7$)
Turbidity of water column (NTU)	7	16	16	10
Dissolved Oxygen in water column (%)	93	140	134	89

Percentage of sequences matching SEED subsystems were determined with an E-value cutoff of $E<1\times10^{-5}$. All metadata 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.

Table 2. Summary of metagenomes used in this study.

MG-RAST ID	Description/Reference	MG-RAST ID	Description/Reference
4440984.3	**Coorong sediment (37 PSU)**	4440971.3	Hypersaline mat (22-34mm) [30]
4441020.3	**Coorong sediment (109 PSU)**	4441584.3	GS012 (Estuary) [45]
4441021.3	**Coorong sediment (132 PSU)**	4441590.3	GS020 (freshwater) [5]
4441022.3	**Coorong sediment (136 PSU)**	4441595.3	GS027 (Marine) [45]
4446406.3	Coorong water 1	4441598.3	GS032 (mangrove) [45]
4446412.3	Coorong water 2	4441599.3	GS033 (hypersaline) [45]
4446411.3	Coorong water 3	4441606.3	GS108a (marine) [45]
4446341.3	Marine sediment 1	4441610.3	GS113 (marine) [45]
4446342.3	Marine sediment 2	4441613.3	GS117a (marine) [45]
4440329.3	Hypersaline sediment	4443688.3	Botany Bay (marine)
4440324.3	Saltern 1 (low) [5,24]	4443689.3	Botany Bay 2 (marine)
4440435.3	Saltern 2 (medium) [5,29]	4440041.3	Line Islands (marine) [46]
4440438.3	Saltern 3 (high) [5,29]	4440212.3	Arctic (marine) [47]
4440437.3	Saltern 4 (low) [5,29]	4440440.3	Aquaculture pond [5]
4440426.3	Saltern 5 (low) [5,29]	4440281.3	Soudan mine [48]
4440429.3	Saltern 6 (high) [5,29]	4441656.4	Whalefall mat [49]
4440067.3	Stromatolite 1 [50]	4441093.3	EBPR (USA) [51]
4440060.4	Stromatolite 2 [50]	4441092.3	EBPR (Australia) [51]
4440061.3	Stromatolite 3 [5]	4441091.3	Farm soil [49]
4440964.3	Hypersaline mat (0-1mm) [30]		

All metagenomes are publicly available on the MG-RAST server (http://metagenomics.nmpdr.org/) [20]. Number of database hits (BLASTX) are determined using an E-value cut-off of 0.001. A more detailed table is provided in supporting information Table S3.

Bold = this study.

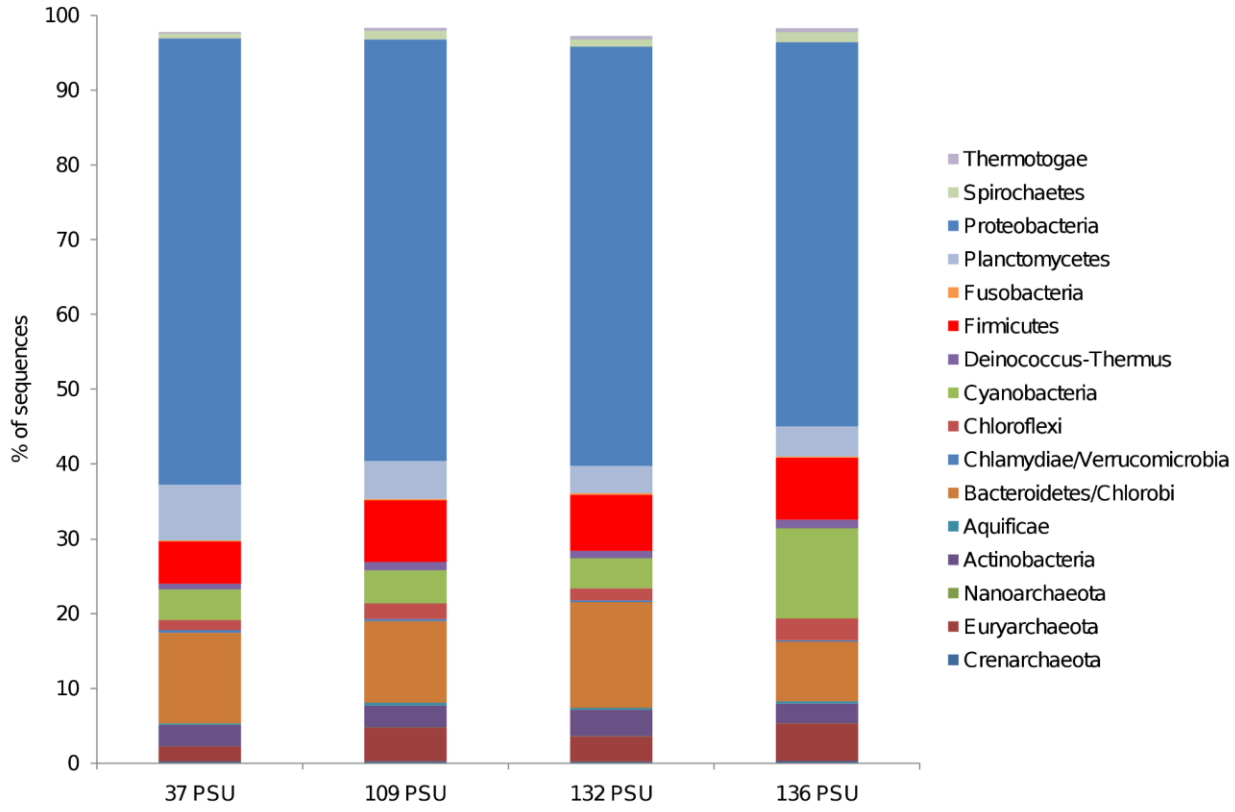

100
90
80
70
60
50
40
30
20
10
0
% of sequences
37 PSU
109 PSU
132 PSU
136 PSU
Thermotogae
Spirochaetes
Proteobacteria
Planctomycetes
Fusobacteria
Firmicutes
Deinococcus-Thermus
Cyanobacteria
Chloroflexi
Chlamydiae/Verrucomicrobia
Bacteroidetes/Chlorobi
Aquificae
Actinobacteria
Nanoarchaeota
Euryarchaeota
Crenarchaeota

Figure 1.

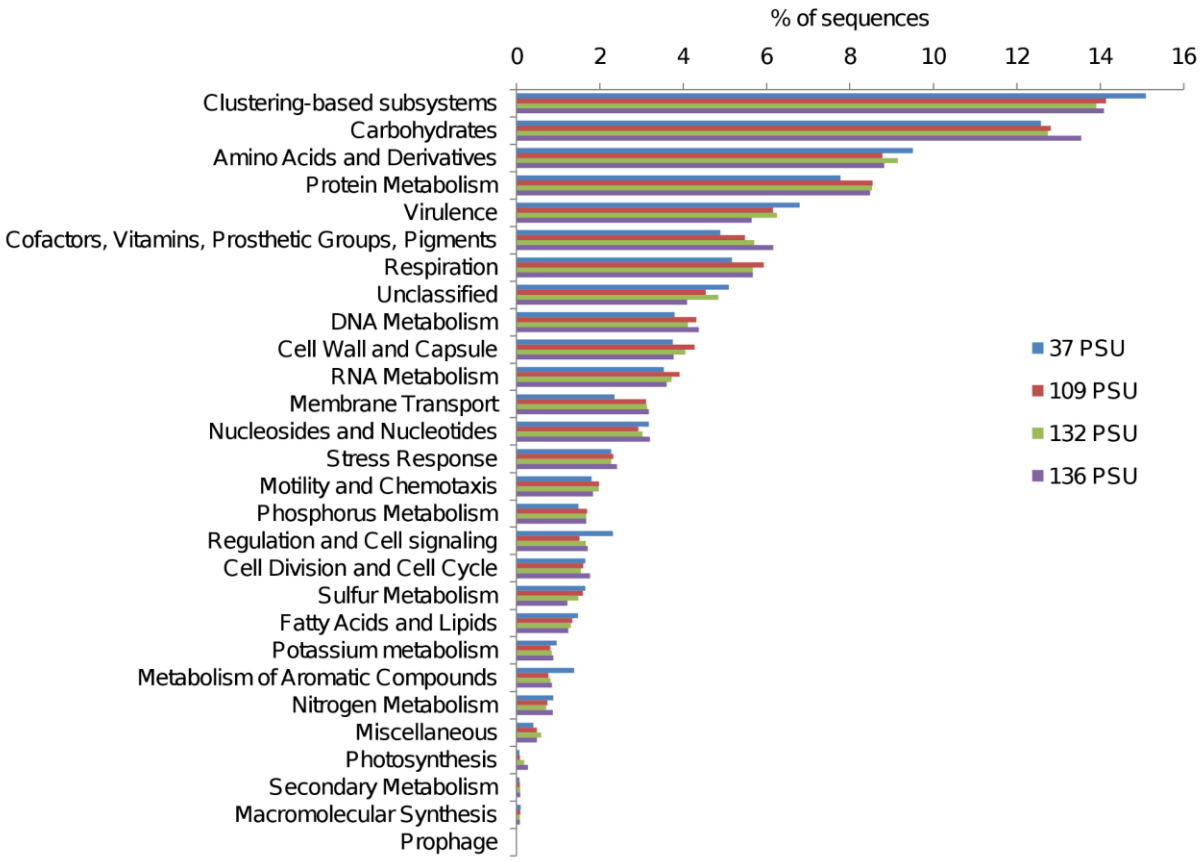

% of sequences
0
2
4
6
8
10
12
14
16
Clustering-based subsystems
Carbohydrates
Amino Acids and Derivatives
Protein Metabolism
Virulence
Cofactors, Vitamins, Prosthetic Groups, Pigments
Respiration
Unclassified
DNA Metabolism
Cell Wall and Capsule
RNA Metabolism
Membrane Transport
Nucleosides and Nucleotides
Stress Response
Motility and Chemotaxis
Phosphorus Metabolism
Regulation and Cell signaling
Cell Division and Cell Cycle
Sulfur Metabolism
Fatty Acids and Lipids
Potassium metabolism
Metabolism of Aromatic Compounds
Nitrogen Metabolism
Miscellaneous
Photosynthesis
Secondary Metabolism
Macromolecular Synthesis
Prophage
37 PSU
109 PSU
132 PSU
136 PSU

Figure 2

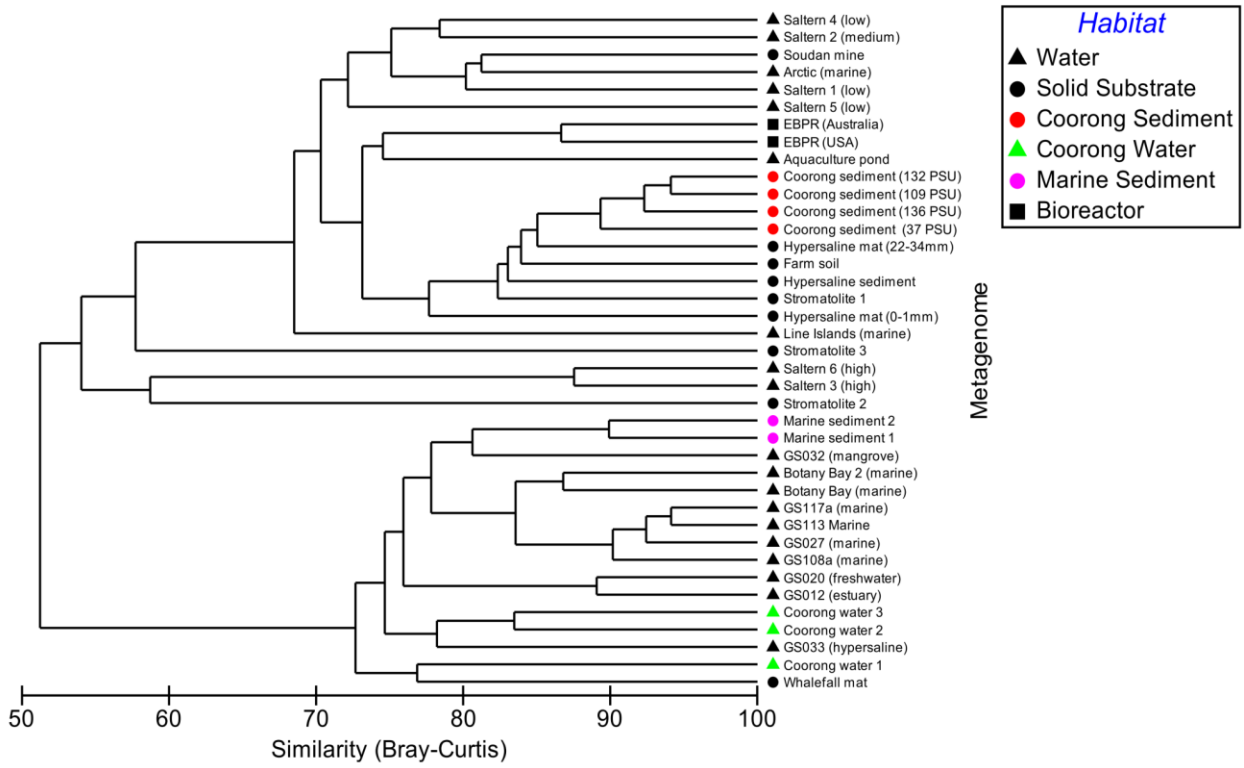

Habitat
Water
Solid Substrate
Coorong Sediment
Coorong Water
Marine Sediment
Bioreactor
Saltern 4 (low)
Saltern 2 (medium)
Soudan mine
Arctic (marine)
Saltern 1 (low)
Saltern 5 (low)
EBPR (Australia)
EBPR (USA)
Aquaculture pond
Coorong sediment (132 PSU)
Coorong sediment (109 PSU)
Coorong sediment (136 PSU)
Coorong sediment (37 PSU)
Hypersaline mat (22-34mm)
Farm soil
Hypersaline sediment
Stromatolite 1
Hypersaline mat (0-1mm)
Line Islands (marine)
Stromatolite 3
Saltern 6 (high)
Saltern 3 (high)
Stromatolite 2
Marine sediment 2
Marine sediment 1
GS032 (mangrove)
Botany Bay 2 (marine)
Botany Bay (marine)
GS117a (marine)
GS113 Marine
GS027 (marine)
GS108a (marine)
GS020 (freshwater)
GS012 (estuary)
Coorong water 3
Coorong water 2
GS033 (hypersaline)
Coorong water 1
Whalefall mat
Metagenome
50
60
70
80
90
100
Similarity (Bray-Curtis)

Figure 3.

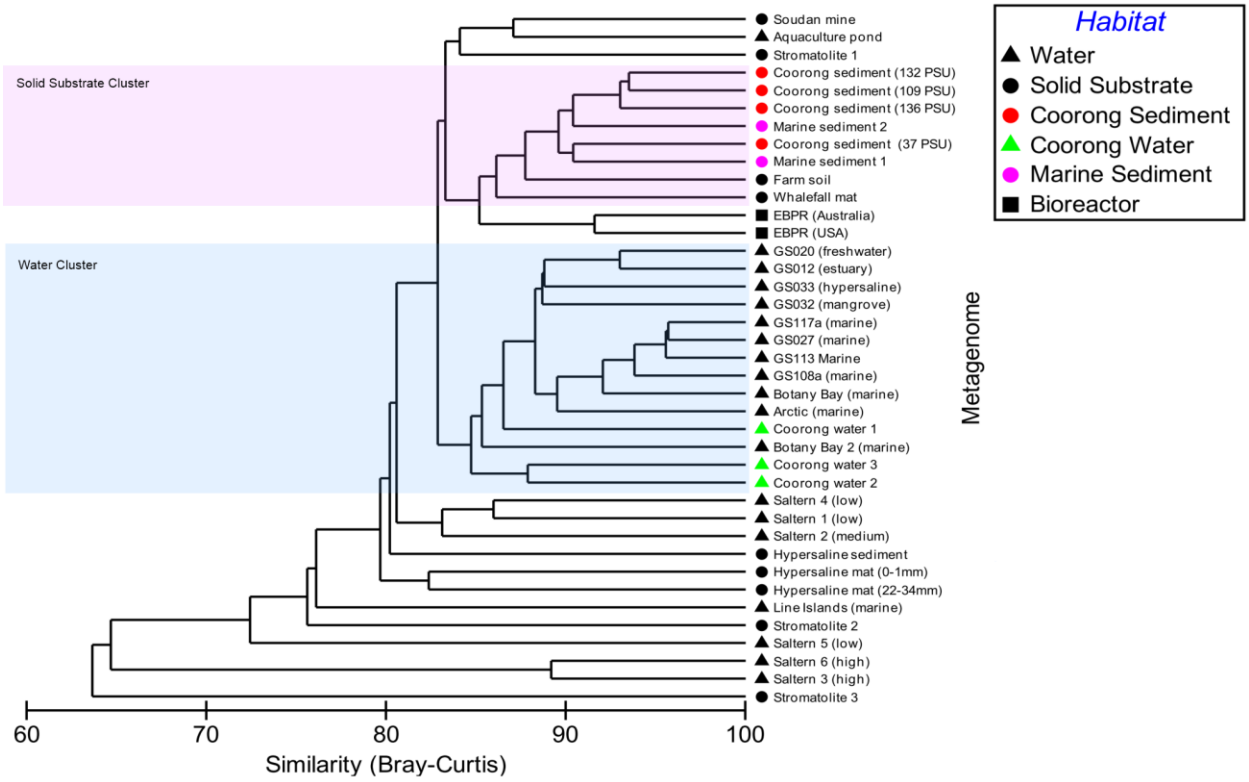

Habitat
Water
Solid Substrate
Coorong Sediment
Coorong Water
Marine Sediment
Bioreactor
Solid Substrate Cluster
Water Cluster
Soudan mine
Aquaculture pond
Stromatolite 1
Coorong sediment (132 PSU)
Coorong sediment (109 PSU)
Coorong sediment (136 PSU)
Marine sediment 2
Coorong sediment (37 PSU)
Marine sediment 1
Farm soil
Whalefall mat
EBPR (Australia)
EBPR (USA)
GS020 (freshwater)
GS012 (estuary)
GS033 (hypersaline)
GS032 (mangrove)
GS117a (marine)
GS027 (marine)
GS113 Marine
GS108a (marine)
Botany Bay (marine)
Arctic (marine)
Coorong water 1
Botany Bay 2 (marine)
Coorong water 3
Coorong water 2
Saltern 4 (low)
Saltern 1 (low)
Saltern 2 (medium)
Hypersaline sediment
Hypersaline mat (0-1mm)
Hypersaline mat (22-34mm)
Line Islands (marine)
Stromatolite 2
Saltern 5 (low)
Saltern 6 (high)
Saltern 3 (high)
Stromatolite 3
Metagenome
60
70
80
90
100
Similarity (Bray-Curtis)

Figure 4.

Table S1

PHYLUM	37 PSU	109 PSU	132 PSU	136 PSU
Actinobacteria	11.43	8.11	2.78	4.82
Aquificae	0	0	0	4.82
Bacteroidetes	8.57	12.16	12.5	6.02
Cyanobacteria	5.71	0	16.67	12.05
Firmicutes	0	5.41	4.17	3.61
Genera_incertae_sedis_WS3	0	4.05	0	0
Planctomycetes	0	2.7	1.39	0
Proteobacteria	57.14	47.3	40.28	45.78
Spirochaetes	0	5.41	8.33	4.82
unclassified_Bacteria	25.71	24.32	27.78	32.53

Table S2

Proportion of matches to the SEED Database (Taxonomy)

Domain	MG-RAST Level 2 (Phyla)	MG-RAST Level 3	37 PSU	109 PSU	132 PSU	136 PSU
Bacteria	Proteobacteria	Gammaproteobacteria	0.2326	0.1839	0.1742	0.1584
Bacteria	Proteobacteria	Alphaproteobacteria	0.1852	0.1707	0.1707	0.1756
Bacteria	Proteobacteria	delta/epsilon subdivisions	0.0963	0.1373	0.1472	0.1103
Bacteria	Bacteroidetes/Chlorobi group	Bacteroidetes	0.1144	0.0983	0.1302	0.0693
Bacteria	Proteobacteria	Betaproteobacteria	0.0741	0.0628	0.0599	0.0612
Bacteria	Planctomycetes	Planctomycetacia	0.0743	0.0506	0.0362	0.0402
Bacteria	Firmicutes	Clostridia	0.0263	0.041	0.0371	0.0415
Bacteria	Firmicutes	Bacilli	0.0297	0.0409	0.0376	0.0409
Bacteria	Actinobacteria	Actinobacteria (class)	0.0284	0.0294	0.0352	0.0265
Bacteria	Cyanobacteria	Nostocales	0.0144	0.0162	0.0135	0.0489
Bacteria	Cyanobacteria	Chroococcales	0.0136	0.0161	0.0152	0.0466
Bacteria	Bacteroidetes/Chlorobi group	Chlorobi	0.009	0.0153	0.0159	0.0149
Bacteria	Chloroflexi	Chloroflexi (class)	0.0105	0.0142	0.0117	0.0216
Archaea	Euryarchaeota	Methanomicrobia	0.0071	0.0139	0.0122	0.015
Bacteria	Spirochaetes	Spirochaetes (class)	0.0065	0.0116	0.0102	0.013
Bacteria	Deinococcus-Thermus	Deinococci	0.0079	0.0114	0.0099	0.0116
Bacteria	Proteobacteria	unclassified Proteobacteria	0.0089	0.009	0.0087	0.0087
Archaea	Euryarchaeota	Thermococci	0.0034	0.0079	0.0061	0.009
Archaea	Euryarchaeota	Halobacteria	0.0039	0.0073	0.0041	0.009
Bacteria	Chloroflexi	Dehalococcoidetes	0.0032	0.0068	0.0044	0.0076
Bacteria	Cyanobacteria	Gloeobacteria	0.0068	0.0058	0.0056	0.0088
Eukaryota	Fungi/Metazoa group	Metazoa	0.008	0.0046	0.0087	0.0053

Bacteria	Cyanobacteria	Oscillatoriales	0.0037	0.0042	0.0038	0.0141
Bacteria	Thermotogae	Thermotogae (class)	0.002	0.0041	0.0042	0.0053
Archaea	Euryarchaeota	Archaeoglobi	0.0015	0.004	0.0037	0.0051
Archaea	Euryarchaeota	Methanococci	0.0014	0.0037	0.0025	0.0042
Bacteria	Aquificae	Aquificae (class)	0.0022	0.0035	0.0025	0.0028
Eukaryota	Viridiplantae	Streptophyta	0.0039	0.0034	0.0067	0.0033
Archaea	Euryarchaeota	Methanobacteria	0.0012	0.0032	0.0025	0.0036
Archaea	Euryarchaeota	Thermoplasmata	0.0009	0.0029	0.0017	0.0024
Bacteria	Chlamydiae/Verrucomicrobia group	Chlamydiae	0.0034	0.0028	0.0023	0.0019
Archaea	Crenarchaeota	Thermoprotei	0.0025	0.0026	0.0023	0.0029
Bacteria	Fusobacteria	Fusobacteria (class)	0.0015	0.0021	0.0021	0.0016
Eukaryota	Fungi/Metazoa group	Fungi	0.0033	0.002	0.0031	0.002
Archaea	Euryarchaeota	Methanopyri	0.0009	0.0019	0.0012	0.0019
Bacteria	Cyanobacteria	Prochlorales	0.002	0.0017	0.0022	0.0023
Viruses	dsDNA viruses, no RNA stage	Caudovirales	0.0018	0.0017	0.0014	0.0011
Bacteria	Firmicutes	Mollicutes	0.0004	0.0004	0.0003	0.0004
Archaea	Nanoarchaeota	Nanoarchaeum	0.0001	0.0003	0.0003	0.0004
Eukaryota	Alveolata	Apicomplexa	0.0006	0.0001	0.0005	0.0002
Eukaryota	Rhodophyta	Bangiophyceae	0.0001	0.0001	0	0
broad host range plasmids	Plasmid pIPO2T.		0.0002	0.0001	0.0002	0.0002
Eukaryota	Acanthamoebidae	Acanthamoeba	0	0	0	0
Eukaryota	Alveolata	Ciliophora	0	0	0.0001	0
Eukaryota	Cryptophyta	Cryptomonadaceae	0.0001	0	0	0
Eukaryota	Euglenozoa	Kinetoplastida	0	0	0.0001	0
Eukaryota	Glaucocystophyceae	Cyanophoraceae	0	0	0	0
Eukaryota	Heterolobosea	Schizopyrenida	0	0	0	0
Eukaryota	Jakobidae	Reclinomonas	0	0	0	0

Eukaryota	Malawimonadidae	Malawimonas	0	0	0	0
Eukaryota	Rhodophyta	Florideophyceae	0	0	0	0
Eukaryota	Viridiplantae	Chlorophyta	0	0	0.0001	0
Eukaryota	stramenopiles	Bacillariophyta	0.0014	0	0.0021	0.0001
Eukaryota	stramenopiles	Oomycetes	0	0	0	0
Eukaryota	stramenopiles	Synurophyceae	0	0	0	0
Viruses	Retro-transcribing viruses	Caulimoviridae	0	0	0	0
Viruses	dsDNA viruses, no RNA stage	Asfarviridae	0	0	0	0
Viruses	dsDNA viruses, no RNA stage	Baculoviridae	0	0	0	0
Viruses	dsDNA viruses, no RNA stage	Herpesviridae	0.0001	0	0	0
Viruses	dsDNA viruses, no RNA stage	Mimivirus	0.0003	0	0.0001	0
Viruses	dsDNA viruses, no RNA stage	Nimaviridae	0	0	0	0
Viruses	dsDNA viruses, no RNA stage	Phycodnaviridae	0.0002	0	0.0001	0
Viruses	dsDNA viruses, no RNA stage	Poxviridae	0.0001	0	0	0
Viruses	ssDNA viruses	Microviridae	0	0	0	0
Viruses	unclassified viruses	Haloviruses	0	0	0	0

Table S3

Description	Latitude	Longitude	Salinity	Number of matches (taxonomy)	Number of matches (metabolism)
Australian Coorong sediment (37 PSU)	-35.5508	138.883	37 PSU	34456	**22538**
Australian Coorong sediment (109 PSU)	-35.7973	139.3174	105 PSU	54697	**36700**
Australian Coorong sediment (132 PSU)	-35.9376	139.4882	132 PSU	54135	**35642**
Australian Coorong sediment (136 PSU)	-36.1664	139.6509	136 PSU	**57571**	**37737**
Australian Coorong water MDA S1	-35° 32' 974	138° 52' 977	37 PSU	50570	37731
Australian Coorong water S2	-36° 09' 806	139° 38' 963	106 PSU	14360	10758
Australian Coorong water S2 MDA	-36° 09' 807	139° 38' 964	106 PSU	28939	20698
Australian Marine Sediment (Port Flinders)	33°06'09"	138°01'26"	Marine	42807	27918
Australian Marine Sediment (Yorke Peninsula)	34°56'59"	137°21'17"	Marine	32588	21049
Anoxic Hypersaline sediment	33.338864	-115.84164	high salinity	19875	13657
Solar saltern (low salinity)	32.599198	-117.11906	low salinity	8503	6090
Solar saltern (medium salinity)	32.599198	-117.11906	medium salinity	11157	7713
Solar saltern (high salinity)	32.60397	-117.0983	high salinity	98138	61342
Solar saltern (low salinity)	33.599198	-116.11906	low salinity	58190	41980
Solar saltern (low salinity)	34.599198	-115.11906	low salinity	4176	3139
Solar saltern (high salinity)	32.60397	-117.0983	high salinity	10383	6539
Pozas Azule microbialite	27.985876	-101.06639	freshwater	40098	26206
Rio Mesquites microbialite	26.985876	-102.06639	freshwater	24024	11379
Highbourne Cay stromatolite	24.709876	-76.822521	Marine	6607	1352
Guerro Negro hypersaline mat (0-1mm)	27.688889	-113.91695	90 PSU	9756	6512
Guerro Negro hypersaline mat (22-33mm)	27.688889	-113.91695	91 PSU	8814	6083
GS012 Estuary - North American East Coast	38°56'49N”	76°25'2W”	3.5 PSU	112589	92897
GS020 Fresh Water - Panama Canal	9°9'52N”	79°25'2W"	0.1 PSU	254328	197680
GS027 Coastal - Galapagos Islands	1°12'58S”	90°25'22W”	34.9 PSU	203974	171123

GS032 Mangrove - Galapagos Islands	0°35'38S”	91°4'10W"	-	128761	103263
GS033 Hypersaline	1°13'42S”	90°25'45W”	63.4 PSU	631985	507684
GS108a Lagoon Reef	12°5'33S”	96°52'54E”	32.4 PSU	47184	38903
GS113 Open Ocean - Indian Ocean	7°0'27S”	76°19'52E”	33.3 PSU	101787	85348
GS117a Coastal - Indian Ocean	4°36'49S”	55°30'31E”	35.5 PSU	310097	253950
Botany Bay	33° 59'27S”	151° 13'55E”	Marine	65306	54452
Botany Bay	33° 59'27S”	151° 13'55E”	Marine	12322	9484
Christmas Island coral reef	1.992317	-157.48407	Marine	10709	5083
Coastal Arctic mesocosm	60.269	5.2222	Marine	117631	86167
Aquaculture pond	33.537594	-116.09775	freshwater	60411	37297
Soudan mine (red sample)	47.819307	-92.242595	-	57647	36947
Whalefall microbial mat	33.3	-1919.22	Marine	33714	25209
Phosphorous removing EBPR sludge - Australia	-27.466667	153.016667	-	88073	87709
Phosphorous removing EBPR sludge - USA	43.06666667	-89.4	-	117407	66082
Waseca farm soil	44.025	-93.586667	-	112760	82771

CHAPTER V

Functional and taxonomic drivers of metagenome partitioning by substrate type

Formatted for *Applied and Environmental Microbiology*

Abstract

Recent analyses of 16S rDNA and metagenomic libraries have shown that the substrate type, whether a community lives in a sediment or water habitat, is a fundamental determinant of microbial community structure and metabolism. In this study we have conducted the first survey of the genes and taxa driving functional partitioning between water and solid substrate metagenomes. Each habitat type had a distinctive core of more abundant taxa and processes which defined life in each substrate type. Metabolisms over-represented in solid substrate habitats were largely involved in resistance to toxic compounds and pili mediated motility over surfaces. This is in contrast to water metagenomes which showed a higher proportion of processes related to essential cellular metabolism indicating that less genomic complexity is necessary to survive and reproduce in this substrate type. Solid substrate microbial communities may be more metabolically complex due to the high degree of resource and physiochemical heterogeneity, high microbial abundance and increased rates of lateral gene transfer in soil and sediment habitats. These factors enrich for flexible adaptive metabolisms which converge in microorganisms adapted to a wide range of physiochemical conditions and which define microbial life in solid substrate habitats.

Introduction

Identifying patterns in the spatial distribution of microbial diversity is a central challenge of microbial ecology, that when combined with explanatory environmental variables leads to a mechanistic view of how microbial taxa and gene abundance is structured by habitat .

One variable which potentially controls microbial community composition is the physical substrate of the habitat, i.e. whether a community exists suspended in open water or within a porous substrate. Microbial diversity is orders of magnitude higher (41) in soil and sediment than in water per unit volume. This diversity is thought to be a result of the high degree of spatial heterogeneity of resources and niche partitioning. This view is consistent with a global meta-analysis of the conserved 16S rDNA gene taxonomic marker (24) which has shown that whether the substrate is sediment or water is an important determinant of taxonomic patterns, indicating that similar community structures exist in soils and sediment in different habitats and biomes. Whilst these studies have relied purely on taxonomy, this similarity is supported by recent metagenomic data (17) that demonstrates a conserved functional signature between soils, sediments and microbial mats from a variety of habitats, which to a degree is also reflected in the taxonomic clustering of these samples. The conserved functional signature is also fundamentally different from the conserved metagenomic signature of water habitats from a variety of biomes. This dichotomy between solid substrate and water metagenomes indicates that different functional processes define life in these habitats and that the physical structure of a community's surroundings has a profound influence on genomic composition.

Recently, multidimensional statistical tools have been applied to elucidate the environmental variables determining ordination patterns of globally distributed 16S rDNA sequences (24) and

metagenomic profiles (13, 30). and to determine which functional processes drive the biome clustering of metagenomes (9). Here, we have employed statistical analyses of metagenomic profiles to elucidate the functional processes and taxonomic groups which contribute to the functional dichotomy between solid substrate habitats and water habitats.

Materials and Methods

Metagenomic sequencing of Coorong sediment samples

Moderately hypersaline (~ 130 Practical Salinity Units) sediment (c.a. 10g) and water (c.a. 3L) were collected from the Coorong, South Australia (35°33'3.05"S, 138°52'58.80"E) as part of previously described sampling expeditions (18). DNA was extracted directly from the sediment using bead beating and chemical lysis (PowerMax Soil, MoBio). Water samples were filtered using a 0.2μm filter and DNA was extracted from this filter using bead beating and chemical lysis (Powerwater, MoBio). DNA quality and concentration was determined by agarose gel electrophoresis and spectrophotometry and >5μg of high molecular weight DNA was sequenced at the Australian Genome Research Facility. Sequencing was conducted on a GS-FLX pyrosequencing platform (Roche). Sediment and water metagenomes analyzed contained 114335 and 18532 sequences respectively and had average read lengths of 232 base pairs and 341 base pairs. Unassembled sequences (environmental gene tags) were annotated using the MetaGenomics Rapid Annotation using Subsystem Technology (MG-RAST) pipeline version 2.0 (http://metagenomics.nmpdr.org/) (29), with a BLASTX E-value cut-off of $E<1\times10^{-5}$ and a minimum alignment length of 50bp. The abundance of individual sequences matching a particular SEED subsystem, groups of genes involved in a particular metabolic function (30), were normalized by sequencing effort and used to generate a metabolic profile of the

metagenome. Taxonomic profiles were generated within MG-RAST using the normalized abundance of the phylogenetic identity of sequence matches to the SEED database. These two metagenomes were used as examples of water and sediment habitats for a pairwise statistical comparison.

Pairwise comparison of hypersaline sediment and water metagenomes

Functional and Taxonomic profiles of metagenomes from the Coorong were visualized as scatter plots of the relative abundance of individual metabolic processes and taxa in each metagenome. This allowed visual discrimination of processes and taxa showing the highest degree of variation between the samples, within each metabolic and taxonomic category. These metagenomes were chosen because their signatures are characteristic of the solid substrate and water clusters in Jeffries *et al*, 2011 (17) and many other environmental variables, such as salinity, and geographic location are controlled. To test for statistically significant categories we used the STatistical Analysis of Metagenomic Profiles (STAMP) package (33) to implement Fisher's exact test with a Storey's FDR multiple test correction (1, 40). Confidence intervals were determined using the Newcombe-Wilson method (31). Statistical significance was assessed using a corrected p-value, termed a q-value, resultant from multiple test correction.

Multidimensional determination of dissimilarity drivers between solid substrate and water habitats

To investigate if the functional and taxonomic categories found to differ in abundance between metagenomes in the pair-wise analysis were consistent drivers of the dichotomy between solid substrate and water metagenomes from all habitats we generated metabolic and taxonomic

profiles of publically available metagenomes from the MetaGenomics using Rapid Annotation Subsystems (MG-RAST) database (29) showing matches to the SEED database (32) with an BLASTX e-value cut-off of 0.001. Profiles were generated at all subsystem hierarchy ranks within MG-RAST and at phylum and class level for SEED taxonomic profiles. All profiles were normalized by total sequence abundance to control for different sized datasets. Metagenomes were chosen which reflected a range of different habitat conditions and were designated as belonging to the 'solid substrate' or 'water' clusters in Jeffries *et al* (2011) (17). Details of the metagenomes are provided in Table 1.

To investigate the statistical distance between these metagenomes we conducted Non-Metric Multidimensional Scaling (NM-MDS) (8) of metabolic profiles derived at the subsystem level. Bray-Cutis similarity was used as a distance measure on square root transformed abundance data. We employed a SIMilarity of PERcentages analysis (SIMPER) (8) to determine which categories most contributed to the statistical dissimilarity between the solid substrate and water clusters displayed in the MDS. These analyses were carried out using the PRIMER-E statistical package (7). To determine if these clustering patterns observed using rank-based similarity measures were also supported by a canonical ordination, and to identify which metabolic functions determine the clustering of solid substrate and water habitats in two dimensional space, we conducted a canonical discriminant analysis (CDA) using PASW statistics 18.0 (SPSS Inc). We applied a stepwise analysis using Wilk's lambda ($p=0.05$) to identify which metabolic variables were used to build a discriminative model of group membership. Variables found by the stepwise analysis to contribute to group membership were plotted over the ordination data as vectors based on their canonical discriminant function coefficients with a transformation of $\times 10$ to allow plotting on the same set of axes. These procedures follow those employed for CDA by Dinsdale

et al, 2008 (9). In order to give the CDA more statistical power, we included other metagenomes in addition to those used for the statistical analyses described above (Table 1). These metagenomes were previously found to not cluster with solid substrate and water groupings (15). We designated this group 'other'.

Results

Pairwise comparison of hypersaline sediment and water metagenomes

To explore the relative abundance of functional genes and taxonomic groups present in model metagenomes from solid substrate and water groupings we profiled the metabolic potential and taxonomic affiliation of sequences present in hypersaline Coorong sediment and water metagenomes, visualizing each variable as a function of its relative abundance in each sample (Figs. 1 & 2).

At the coarsest rank of metabolic organization in the MG-RAST database (Fig. 1A), genes encoding virulence, respiration, and motility and chemotaxis were the most divergent from a 1:1 ratio in the sediment metagenome, relative to the water metagenome. Genes central to cellular growth via DNA metabolism and protein metabolism, and genes grouped into miscellaneous metabolic subsystems due to clustering on the genome, were slightly higher in proportion in the water metagenome than the sediment metagenome, but were abundant in each. Protein metabolism and DNA metabolism abundance differences in Fig. 1A were driven by peptide biosynthesis and DNA replication respectively which were higher in proportion in the water metagenome along with genes involved in cell wall and capsule synthesis (Fig. 1B). At the subsystem level (Fig. 1C), the abundance of the protein metabolism gene category was driven by

tRNA aminoacylation and cell wall construction by an increase in peptidoglycan biosynthesis genes. Virulence genes encoding the category of 'resistance to antibiotics and toxic compounds (Fig. 1B) and expression of the extracellular Type IV pili (Fig. 1C) were higher in abundance in sediment as were genes encoding anaerobic respiratory reductase enzymes (Fig. 1C). Generally, processes increased in water represented high abundance categories in each metagenome, whereas processes with higher representation in sediment relative to water were from categories lower in proportion overall. Metagenomes showed greater variation in structure at finer levels of resolution, being more conserved, i.e. closer to a 1:1 ratio, in Figure 1A than in Figure1C.

Taxonomically *Protebacteria* were 13 % higher in abundance in sediment than water but were the most abundant phylum in both metagenomes (Fig. 2A). This was due to a 12% higher abundance of *δ/ε-Proteobacteria* and 2% higher abundance of *Betaproteobacteria* in sediment (Fig. 2B). The phylum *Firmicutes*, whilst only a small proportion of both metagenomes, was also more abundant in sediment. *Actinobacteria* showed the most difference between the substrate types being 28% higher in water than sediment at phylum (Fig. 2A) and class (Fig. 2B) level.

Having explored differences in the context of overall abundance, we tested for statistically significant over-representation of functional and taxonomic categories (Fig. 3 & Fig. 4). Many of the categories which showed the largest difference in abundance in Figure 1 were statistically significant. In particular, virulence, motility and chemotaxis, and respiration were the most over-represented categories in the sediment metagenome (Fig 3A). Virulence was again driven by an over-representation of antibiotic and toxin resistance genes (Fig 3B) and type IV pili synthesis, which was the most overrepresented category in sediment at subsystem level (Fig.

3C). Motility and chemotaxis overrepresentation included genes for flagella and, non-flagella motility (Fig. 3B) and subsystems encoding gliding and chemotaxis (Fig. 2C). The respiration subsystem coding anaerobic reductases was the second most overrepresented category in sediment (Fig. 3C). STAMP Analysis also revealed several other higher level over-represented categories not evident based on raw difference of proportion. In particular, functional categories related to sulphur and nitrogen metabolism were among the most overrepresented categories in sediment (Fig. 3B and 3C) and a protein secretion pathway was the second most significantly overrepresented level 2 pathway in sediment (Fig, 3B). Sequences encoding the core cellular processes of protein and DNA metabolism, cell wall construction and cell division were significantly overrepresented in the water metagenome at all levels (Fig 3).Taxonomically, increased representation of *δ/ε-Proteobacteria* in sediment and *Actinobacteria* in water were the most statistically significant differences in taxon abundance (Fig. 4).

Multivariate comparison of solid substrate and water metagenome clusters

To investigate if the variables over-represented in the pairwise comparison of characteristic sediment and water metagenomes were also more generally responsible for driving the dichotomy between solid substrate and water metabolic clusters collected from a variety of habitats and at different times we applied multivariate statistics to metagenomes from a variety of habitats (Table 1). Two dimensional ordination of the metabolic profiles of these metagenomes showed discrete clustering of solid substrate and water metagenomes (Fig. 5). Metagenomes within each cluster shared >83% similarity to each other and solid substrate and water clusters diverged into separate clusters at 82% similarity. In agreement with the pairwise analysis, virulence, motility and chemotaxis and respiration were all consistent drivers of the dichotomy

between solid substrate and water metagenomes due to a higher abundance in solid substrate metagenomes (Table 2). Consistent with the pairwise analysis, the contribution of virulence to dissimilarity was largely due to antibiotic compound and toxin resistance genes and sequences for the expression of type IV pili (within the type IV ESAT secretion system). Additionally, genes encoding regulation and cell signalling were also higher in the sediment cluster contributing to 8% of the dissimilarity, as were genes involved in nitrogen, potassium and sulphur metabolism.

Also confirming the pairwise analysis, genes encoding core metabolic processes such as protein biosynthesis and DNA replication contributed most to the dissimilarity between substrate types due to a higher abundance in water metagenomes. Taxonomically, *Actinobacteria* contributed most to the disparity between substrate clusters due to a higher abundance in water habitats, confirming the pairwise analysis, and combined with *Alphaproteobacteri*a contributed up to 15% of the overall dissimilarity. Due to a higher abundance in the solid substrate cluster, the *δ/ε* divisions of *Proteobacteria, Betaproteobacteria and Plactomycetacia* were also major drivers of the substrate dichotomy (Table 3).

To further test the overall influence of metabolic variables on the ordination of metagenomic profiles we conducted a Canonical Discriminant Analysis (CDA) of the metagenomes used in Jeffries *et al* 2011 (17). Solid substrate and water metagenomes showed very strong group clustering (Fig. 6) indicating that each substrate has a distinguishing profile. Virulence was a major determinant of solid substrate group membership as was potassium metabolism. Cell wall and capsule synthesis genes were the most important in distinguishing the water cluster. DNA

metabolism processes also drove ordination to the negative side of the x-axis whereas respiration contributed to the clustering of solid substrate metagenomes to the positive axis.

Discussion

Functional processes that define life in solid substrate habitats

Pairwise analysis of characteristic metagenomes, analysis of variables contributing to the discrete clustering of water and solid cluster metagenomes and CDA analysis all showed virulence to be one of the major functional categories driving the dichotomy between solid substrate and water habitats. Whilst the term virulence denotes an increased ability to infect eukaryotic hosts, in this case many genes incorporated under the virulence category may have an environmental function. An increased abundance of genes encoding resistance to antibiotics and toxic compounds, and genes encoding type IV pili in solid substrate metagenomes were the main contributor to the influence of virulence. The presence of antibiotic resistance genes in environmental samples is not surprising given that natural habitats act as a reservoir of this 'resistome'(2, 27, 35). Metagenomic analyses of soil (3, 35) and contaminated river sediments (22) have revealed the prevalence and novel diversity of antibiotic resistance genes in solid substrate habitats. Antibiotic resistance initially evolved in natural populations (27), probably to provide a defence against natural concentrations of microbially produced antibiotics involved in inhibiting the growth of competitors. These naturally occurring resistance profiles are particularly evident in stationary complexes like soil and sediment which are less likely to be influenced by the movement of anthropogenic inputs (2), however in habitats closely associated with medical or agricultural practice could result in natural and artificial resistance (27). The roles of antibiotic compounds and thus their resistance genes in natural settings are different from clinical settings.

For example, low concentrations of antibiotic compounds can signal cellular responses modulating community interaction, and are potentially involved in microbial signal communication (12, 27), quorum sensing and biofilm formation (2, 14), processes which are important in surface rich habitats with a high cell density such as soils and sediments. This is concurrent with the contribution of regulation and cell signaling genes to substrate dissimilarity, due to their higher abundance in solid substrate metagenomes than water metagenomes. Heavy metal contamination has also been shown to select for antibiotic resistance (27) and many enzymes that confer antibiotic resistance, such as efflux pumps, also play a role in general mechanisms of resistance to toxins such as heavy metals (2). Indeed, genes related to Cobalt, Zinc and Cadmium resistance, grouped under the resistance to antibiotics and toxic compounds category, were found to be higher in the solid substrate grouping and to contribute to the dissimilarity between clusters. The influence of heavy metal resistance was potentially greater in solid substrates due to higher accumulation rates of heavy metals in sediments. Our data indicate that resistance to toxic chemicals is generally more important in solid substrate habitats than water habitats potentially due to a higher density of antibiotic producing cells (2), increased importance of these molecules for cell signaling (12, 25), or increased rates of toxin contamination via sinking particles.

Type IV pili over-representation in Coorong sediment and in the solid substrate cluster overall indicate that this is a potentially important cellular appendage in solid substrate habitats. Whilst generally studied within the context of clinical virulence, Type IV pili play a major role in the biofilm colonization of a wide variety of moist surfaces via twitching motility (20, 28) allowing exploration of the surface and movement through viscous extracellular polymeric substances (EPS) (20). Twitching is also referred to as social motility, the genes for which were drivers of

the dissimilarity between clusters and were overrepresented in the Coorong sediment metagenome and solid substrate cluster. Overall, motility and chemotaxis was one of the most overrepresented categories in the Coorong sediment and in the solid substrate being an important driver of the substrate type partitioning indicating that both flagella and non-flagella motility, and chemotaxis, are important in porous substrates.

Motility is an important trait in the nutrient poor ocean, where it provides cells with a competitive edge to exploit micro-nutrient patches (38). Our data indicate that motility could be an even more important process in sediment where there is a lack of resource mixing and where resources are highly spatially partitioned into heterogeneous microniches requiring cellular movement and specialized motility mechanisms, such as type 1V pili, to exploit particle surfaces. Additionally, chemotaxis requires chemical gradients (38) and the porosity of the substrate stabilizes chemical gradients compared to open water where turbulence destroys or limits them.

Functional categories related to sulphur, potassium and nitrogen metabolism also drove the split between solid substrate and water metagenomes, and potassium metabolism was a major driver of solid substrate ordination in the CDA. One possible explanation for this is that all of the samples included in the solid substrate cluster are from nutrient rich habitats such as coastal sediments and farm soil. The potassium metabolism category was composed of genes responsible for efflux pumps, indicating that maintenance of the internal ion concentration of the cell was the determining factor. The contribution of anaerobic respiratory reductases and hydrogenases to the dissimilarity between solid substrate and water metagenomes was not surprising given the anoxic nature of sediment habitats and indicated that anoxic respiration is a fundamental process in solid substrate habitats with low oxygen availability.

Core metabolisms dominate in water metagenomes

Generally processes related to protein metabolism, DNA replication and cell wall synthesis were over-represented in Coorong water relative to sediment and drove the dichotomy between substrate type clusters due to their higher abundance in water habitats. These processes can all be considered as 'housekeeping' metabolic functions that are essential to survival and reproduction in any habitat. Whilst these processes are also obviously central to life in solid substrate, and were indeed still abundant in the example of Coorong sediment, their higher abundance in water metagenomes indicates that they were not obscured by other processes that increase in abundance in solid substrate. Thus we speculate that generally microbial communities in water habitats allocate less of their genomic space to 'flexible processes' such as virulence and motility allowing a functional focus on housekeeping genes. This is reflected in ocean metagenomes which are dominated by a core suite of genes associated with cellular metabolism which remains similar across environmental gradients (16, 37, 43). This lack of requirement for a more flexible genetic repertoire in water metagenomes is apparent in the average genome size of microbes derived from metagenomic datasets, which is significantly lower in the terrestrial biome than the aquatic biome (4) and is larger in soil than in ocean water (34). A large genome is selectively advantageous in complex habitats where diverse resources are available, such as soil and sediment (4, 21, 34), and these larger genomes are depleted in genes encoding protein metabolism, DNA synthesis and cell division relative to smaller genomes where they are enriched (21). This is consistent with our data that find an over-representation of these housekeeping genes in water habitats relative to solid substrates. Within the ocean biome, it has been proposed that the most abundant and ubiquitous surface water plankton have streamlined, inflexible genomes with a reduced capacity for sensory and response mechanisms,

which are adapted to persistence at low nutrient conditions and slow growth (14, 15, 44). In contrast genomes associated with particles or at low abundance in the ocean are adapted to a 'feast or famine' lifestyle and have more genomic variation. Our data support this hypothesis of genome streamlining in the most abundant members of aquatic habitats.

Taxonomic partitioning between solid substrate and water metagenomes

Taxonomic dissimilarity between water and solid substrates was largely driven by a higher representation of *Actinobacteria* and *Alphaproteobacteria* in water habitats and an increase in *δ/ε-Proteobacteria* in the solid substrate cluster. Diverse *Actinobacteria* populate both ocean water (6, 19) and freshwater (39), however, some debate exists as to whether these taxa are adapted to the marine environment or are the consequence of terrestrial runoff due to their commonality in soil (6, 19). Our results indicate that *Actinobacteria* may be an important inhabitant of water habitats, consistent with recent metagenomic evidence of highly adapted actinobacterial photorhodopsin genes in the photic zone of marine (37) and freshwater (39) habitats. An increase in the class *Alphaproteobacteria* in the water cluster is not surprising given the abundance of the SAR11 clade of *Alphaproteobacteria* in marine waters. The group *δ/ε-Proteobacteria* contains the anaerobic sulphate-reducing bacteria and other anaerobic genera such as *Geobacter,* thus their larger representation in solid substrate metagenomes is consistent with an increase in sulphur metabolism and anaerobic respiration in the solid substrate cluster. The predominance of proteobacterial antibiotic resistance (2) and the restriction of type 1V pili mediated motility to the *Proteobacteri*a (28) indicate that the metabolic partitioning between substrate types is reflected in the taxonomic differences between the clusters.

Concluding remarks

In this study we have conducted the first survey of the genes and taxa driving functional partitioning between water and solid substrate metagenomes. Each habitat type had a distinctive core of more abundant taxa and processes which defined life in each substrate type. Metabolisms over-represented in the solid substrate habitat painted a genomic portrait of an anaerobic, heterogenous habitat where pili mediated motility over surfaces and resistance to toxic compounds is fundamental to life. This is contrasted by water metagenomes which showed a higher proportion of processes related to essential cellular metabolism indicating that less genomic complexity is necessary to survive and reproduce in this substrate type. Solid substrate microbial communities may be more metabolically complex due to the high degree of resource and physiochemical heterogeneity, high microbial abundance and increased rates of lateral gene transfer (41) in soil and sediment habitats. These factors enrich for flexible adaptive metabolisms which converge in microorganisms adapted to a wide range of physiochemical conditions (17) and which define microbial life in solid substrate habitats.

References

1. **Agresti, A.** 1990. Categorical data analysis. Wiley, New York.
2. **Allen, H. K., J. Donato, H. H. Wang, K. A. Cloud-Hansen, J. Davies, and J. Handelsman.** 2010. Call of the wild: antibiotic resistance genes in natural environments. Nat Rev Micro **8:**251-259.
3. **Allen, H. K., L. A. Moe, J. Rodbumrer, A. Gaarder, and J. Handelsman.** 2008. Functional metagenomics reveals diverse [beta]-lactamases in a remote Alaskan soil. ISME J **3:**243-251.
4. **Angly, F. E., D. Willner, A. Prieto-Davo, R. A. Edwards, R. Schmieder, R. Vega-Thurber, D. A. Antonopoulos, K. Barott, M. T. Cottrell, C. Desnues, E. A. Dinsdale, M. Furlan, M. Haynes, M. R. Henn, Y. F. Hu, D. L. Kirchman, T. McDole, J. D. McPherson, F. Meyer, R. M. Miller, E. Mundt, R. K. Naviaux, B. Rodriguez-Mueller, R. Stevens, L. Wegley, L. X. Zhang, B. L. Zhu, and F. Rohwer.** 2009. The GAAS Metagenomic Tool and Its Estimations of Viral and Microbial Average Genome Size in Four Major Biomes. PLoS Computational Biology **5:** e1000593.
5. **Breitbart, M., A. Hoare, A. Nitti, J. Siefert, M. Haynes, E. Dinsdale, R. Edwards, V. Souza, F. Rohwer, and D. Hollander.** 2009. Metagenomic and stable isotopic analyses of modern freshwater microbialites in Cuatro CiEnegas, Mexico. Environmental Microbiology **11:**16-34.
6. **Bull, A., J. Stach, A. Ward, and M. Goodfellow.** 2005. Marine actinobacteria: perspectives, challenges, future directions. Antonie van Leeuwenhoek **87:**65-79.
7. **Clarke, A., and R. Gorley.** 2006. PRIMER v6: User Manual/Tutorial. PRIMER-E, Plymouth, UK.

8. **Clarke, K. R.** 1993. nonparametric multivariate analyses of changes in community structure. Australian Journal of Ecology **18:**117-143.

9. **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.

10. **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.

11. **Edwards, R. A., B. Rodriguez-Brito, L. Wegley, M. Haynes, M. Breitbart, D. M. Peterson, M. O. Saar, S. Alexander, E. C. Alexander, and F. Rohwer.** 2006. Using pyrosequencing to shed light on deep mine microbial ecology. BMC Genomics **7:** 57.

12. **Fajardo, A., and J. L. Martínez.** 2008. Antibiotics as signals that trigger specific bacterial responses. Current Opinion in Microbiology **11:**161-167.

13. **Gilbert, J. A., D. Field, Y. Huang, R. Edwards, W. Z. Li, P. Gilna, and I. Joint.** 2008. Detection of Large Numbers of Novel Sequences in the Metatranscriptomes of Complex Marine Microbial Communities. PLoS One **3:** e3042.

14. **Giovannoni, S. J., and U. Stingl.** 2005. Molecular diversity and ecology of microbial plankton. Nature **437:**343-348.

15. **Giovannoni, S. J., H. J. Tripp, S. Givan, M. Podar, K. L. Vergin, D. Baptista, L. Bibbs, J. Eads, T. H. Richardson, M. Noordewier, M. S. Rappe, J. M. Short, J. C.**

Carrington, and E. J. Mathur. 2005. Genome streamlining in a cosmopolitan oceanic bacterium. Science **309:**1242-1245.

16. **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.
17. **Jeffries, T. C., J. R. Seymour, J. A. Gilbert, E. A. Dinsdale, K. Newton, S. C. Leterme, B. Roudnew, R. J. Smith, L. Seuront, and J. G. Mitchell.** 2011. Substrate type determines metagenomic profiles from diverse chemical habitats. PLoS One **6:** e25173..
18. **Jeffries, T. C., J. R. Seymour, K. Newton, R. J. Smith, L. Seuront, and J. G. Mitchell.** 2012. Increases in the abundance of microbial genes encoding halotolerance and photosynthesis along a sediment salinity gradient. Biogeosciences **9**: 815-825..
19. **Jensen, P., and F. Lauro.** 2008. An assessment of actinobacterial diversity in the marine environment. Antonie van Leeuwenhoek **94:**51-62.
20. **Jin, F., J. C. Conrad, M. L. Gibiansky, and G. C. L. Wong.** 2011. Bacteria use type-IV pili to slingshot on surfaces. Proceedings of the National Academy of Sciences **108:**12617-12622.
21. **Konstantinidis, K. T., and J. M. Tiedje.** 2004. Trends between gene content and genome size in prokaryotic species with larger genomes. Proceedings of the National Academy of Sciences of the United States of America **101:**3160-3165.
22. **Kristiansson, E., J. Fick, A. Janzon, R. Grabic, C. Rutgersson, B. Weijdegård, H. Söderström, and D. G. J. Larsson.** 2011. Pyrosequencing of Antibiotic-Contaminated

River Sediments Reveals High Levels of Resistance and Gene Transfer Elements. PLoS One **6:**e17038.

23. **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.

24. **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.

25. **Marie, D., C. P. D. Brussaard, R. Thyrhaug, G. Bratbak, and D. Vaulot.** 1999. Enumeration of marine viruses in culture and natural samples by flow cytometry. Applied and Environmental Microbiology **65:**45-52.

26. **Martin, H. G., N. Ivanova, V. Kunin, F. Warnecke, K. W. Barry, A. C. McHardy, C. Yeates, S. He, A. A. Salamov, E. Szeto, E. Dalin, N. H. Putnam, H. J. Shapiro, J. L. Pangilinan, I. Rigoutsos, N. C. Kyrpides, L. L. Blackall, K. D. McMahon, and P. Hugenholtz.** 2006. Metagenomic analysis of two enhanced biological phosphorus removal (EBPR) sludge communities. Nat Biotech **24:**1263-1269.

27. **Martínez, J. L.** 2008. Antibiotics and Antibiotic Resistance Genes in Natural Environments. Science **321:**365-367.

28. **Mattick, J. S.** 2002. Type iv pili and twitching motility. Annual Review of Microbiology **56:**289-314.

29. **Meyer, F., D. Paarmann, M. D'Souza, R. Olson, E. M. Glass, M. Kubal, T. Paczian, A. Rodriguez, R. Stevens, A. Wilke, J. Wilkening, and R. A. Edwards.** 2008. The

metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC Bioinformatics **9:** 386.

30. **Mitra, S., J. A. Gilbert, D. Field, and D. H. Huson.** 2010. Comparison of multiple metagenomes using phylogenetic networks based on ecological indices. ISME J **4:**1236-1242.

31. **Newcombe, R. G.** 1998. Interval estimation for the difference between independent proportions: comparison of eleven methods. Statistics in Medicine **17:**873-890.

32. **Overbeek, R., T. Begley, R. M. Butler, J. V. Choudhuri, H. Y. Chuang, M. Cohoon, V. de Crecy-Lagard, N. Diaz, T. Disz, R. Edwards, M. Fonstein, E. D. Frank, S. Gerdes, E. M. Glass, A. Goesmann, A. Hanson, D. Iwata-Reuyl, R. Jensen, N. Jamshidi, L. Krause, M. Kubal, N. Larsen, B. Linke, A. C. McHardy, F. Meyer, H. Neuweger, G. Olsen, R. Olson, A. Osterman, V. Portnoy, G. D. Pusch, D. A. Rodionov, C. Ruckert, J. Steiner, R. Stevens, I. Thiele, O. Vassieva, Y. Ye, O. Zagnitko, and V. Vonstein.** 2005. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. Nucleic Acids Research **33:**5691-5702.

33. **Parks, D. H., and R. G. Beiko.** 2010. Identifying biologically relevant differences between metagenomic communities. Bioinformatics **26:**715-721.

34. **Raes, J., J. Korbel, M. Lercher, C. von Mering, and P. Bork.** 2007. Prediction of effective genome size in metagenomic samples. Genome Biology **8:**R10.

35. **Riesenfeld, C. S., R. M. Goodman, and J. Handelsman.** 2004. Uncultured soil bacteria are a reservoir of new antibiotic resistance genes. Environmental Microbiology **6:**981-989.

36. **Rodriguez-Brito, B., L. Li, L. Wegley, M. Furlan, F. Angly, M. Breitbart, J. Buchanan, C. Desnues, E. Dinsdale, R. Edwards, B. Felts, M. Haynes, H. Liu, D. Lipson, J. Mahaffy, A. B. Martin-Cuadrado, A. Mira, J. Nulton, L. Pasic, S. Rayhawk, J. Rodriguez-Mueller, F. Rodriguez-Valera, P. Salamon, S. Srinagesh, T. F. Thingstad, T. Tran, R. V. Thurber, D. Willner, M. Youle, and F. Rohwer.** 2010. Viral and microbial community dynamics in four aquatic environments. ISME J **4:**739-751.

37. **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.

38. **Seymour, J. R., Marcos, and R. Stocker.** 2009. Resource Patch Formation and Exploitation throughout the Marine Microbial Food Web. American Naturalist **173:**E15-E29.

39. **Sharma, A. K., K. Sommerfeld, G. S. Bullerjahn, A. R. Matteson, S. W. Wilhelm, J. Jezbera, U. Brandt, W. F. Doolittle, and M. W. Hahn.** 2009. Actinorhodopsin genes discovered in diverse freshwater habitats and among cultivated freshwater Actinobacteria. ISME J **3:**726-737.

40. **Storey, J. D., and R. Tibshirani.** 2003. Statistical significance for genomewide studies. Proceedings of the National Academy of Sciences **100:**9440-9445.

41. **Torsvik, V., L. Ovreas, and T. F. Thingstad.** 2002. Prokaryotic diversity - Magnitude, dynamics, and controlling factors. Science **296:**1064-1066.

42. **Tringe, S. G., C. von Mering, A. Kobayashi, A. A. Salamov, K. Chen, H. W. Chang, M. Podar, J. M. Short, E. J. Mathur, J. C. Detter, P. Bork, P. Hugenholtz, and E. M. Rubin.** 2005. Comparative metagenomics of microbial communities. Science **308:**554-557.

43. **Venter, J. C., K. Remington, J. F. Heidelberg, A. L. Halpern, D. Rusch, J. A. Eisen, D. Wu, I. Paulsen, K. E. Nelson, W. Nelson, D. E. Fouts, S. Levy, A. H. Knap, M. W. Lomas, K. Nealson, O. White, J. Peterson, J. Hoffman, R. Parsons, H. Baden-Tillson, C. Pfannkoch, Y.-H. Rogers, and H. O. Smith.** 2004. Environmental Genome Shotgun Sequencing of the Sargasso Sea. Science **304:**66-74.

44. **Yooseph, S., K. H. Nealson, D. B. Rusch, J. P. McCrow, C. L. Dupont, M. Kim, J. Johnson, R. Montgomery, S. Ferriera, K. Beeson, S. J. Williamson, A. Tovchigrechko, A. E. Allen, L. A. Zeigler, G. Sutton, E. Eisenstadt, Y.-H. Rogers, R. Friedman, M. Frazier, and J. C. Venter.** 2010. Genomic and functional adaptation in surface ocean planktonic prokaryotes. Nature **468:**60-67..

Figure legends

Figure 1 Relative proportion of metabolic processes at MG-RAST hierarchy levels (A) Top-level (overall metabolic processes) (B) Second level and (C) Third level (subsystems). Red line indicates 1:1 ratio of abundance. Top five most variable categories are labeled.

Figure 2 Relative proportion of taxa at A) phyla level & B) class level. Red line indicates 1:1 ratio of abundance. Top three most variable categories are labelled.

Figure 3 Metabolic processes significantly different in abundance between hypersaline sediment and hypersaline water at MG-RAST hierarchy levels A) one (overall metabolic processes) B) two & C) three (subsystems, filtered at $q<0.01$). Differences were determined using Black = over-represented in sediment, Red = water.

Figure 4 Taxa significantly different in abundance between hypersaline sediment and hypersaline water at A) phyla and B) class level. Black = over-represented in sediment, Red = water. Only the most significant categories (q-value $<10^{-15}$) are displayed.

Figure 5 Non-Metric Multidimensional Scaling Plot (NM-MDS) of the Bray-Curtis distance between metabolic profiles of metagenomes from solid-substrate and water habitats. Contour line represents 83% similarity. Metagenomes are publically available on MG-RAST (details Table 1).

Figure 6 CDA of metabolic profiles derived from selected metagenomes publicly available on the MG-RAST database. The plot is based on canonical discriminate functions 1 and 2 with vectors representing the structural matrix for metabolic processes identified as influencing separation of the profiles using a stepwise procedure.

Table 1 Summary of metagenomes used in this study

MG-RAST ID	Description/Reference	Grouping	MG-RAST ID	Description/Reference	Grouping
4440984.3	**Coorong sediment (37 PSU)**	Solid	4440971.3	Hypersaline mat (22-34mm) (21)	Other
4441020.3	**Coorong sediment (109 PSU)**	Solid	4441584.3	GS012 (Estuary) (35)	Water
4441021.3	**Coorong sediment (132 PSU)**	Solid	4441590.3	GS020 (freshwater) (35)	Water
4441022.3	**Coorong sediment (136 PSU)**	Solid	4441595.3	GS027 (Marine) (35)	Water
4446406.3	Coorong water 1	Water	4441598.3	GS032 (mangrove) (35)	Water
4446412.3	Coorong water 2	Water	4441599.3	GS033 (hypersaline) (35)	Water
4446411.3	Coorong water 3	Water	4441606.3	GS108a (marine) (35)	Water
4446341.3	Marine sediment 1	Solid	4441610.3	GS113 (marine) (35)	Water
4446342.3	Marine sediment 2	Solid	4441613.3	GS117a (marine) (35)	Water
4440329.3	Hypersaline sediment	Other	4443688.3	Botany Bay (marine)	Water
4440324.3	Saltern 1 (low) (9, 34)	Other	4443689.3	Botany Bay 2 (marine)	Water
4440435.3	Saltern 2 (medium) (9, 34)	Other	4440041.3	Line Islands (marine) (10)	Other
4440438.3	Saltern 3 (high) (9, 34)	Other	4440212.3	Arctic (marine) (13)	Water
4440437.3	Saltern 4 (low) (9, 34)	Other	4440440.3	Aquaculture pond (9)	Other
4440426.3	Saltern 5 (low) (9, 34)	Other	4440281.3	Soudan mine (11)	Other
4440429.3	Saltern 6 (high) (9, 34)	Other	4441656.4	Whalefall mat (40)	Solid
4440067.3	Stromatolite 1 (5)	Other	4441093.3	EBPR (USA) (24)	Solid
4440060.4	Stromatolite 2 (5)	Other	4441092.3	EBPR (Australia) (24)	Solid
4440061.3	Stromatolite 3 (9)	Other	4441091.3	Farm soil (40)	Solid
4440964.3	Hypersaline mat (0-1mm) (21)	Other			

Table 2 SIMPER analysis of metabolic dissimilarity between solid substrate and water metagenome clusters

	Function	Average abundance in water cluster (%)	Average abundance in solid substrate cluster (%)	% contribution to dissimilarity	Cumulative % contribution to dissimilarity
	Level 1				
	Virulence	4.84	7.29	9.13	9.13
	Regulation and Cell signaling	1.21	2.56	8.03	17.16
	Protein Metabolism	9	6.76	5.95	23.11
	Potassium metabolism	0.49	1.21	5.8	28.91
	Motility and Chemotaxis	1.69	2.56	5.71	34.62
	Sulfur Metabolism	1.44	1.96	4.57	39.19
	Nucleosides and Nucleotides	3.61	2.56	4.48	43.67
	Respiration	4	4.84	4.22	47.89
	Stress Response	1.96	2.89	4.18	52.08
Parent category	**Level 2**				
Regulation and Cell signaling	Unclassified (regulation and cell signalling)	0.11	0.16	2.08	2.08
Virulence	Type III, Type IV, ESAT secretion systems	0.05	0.1	2.08	4.16
Virulence	Resistance to antibiotics and toxic	0.14	0.18	1.93	6.09

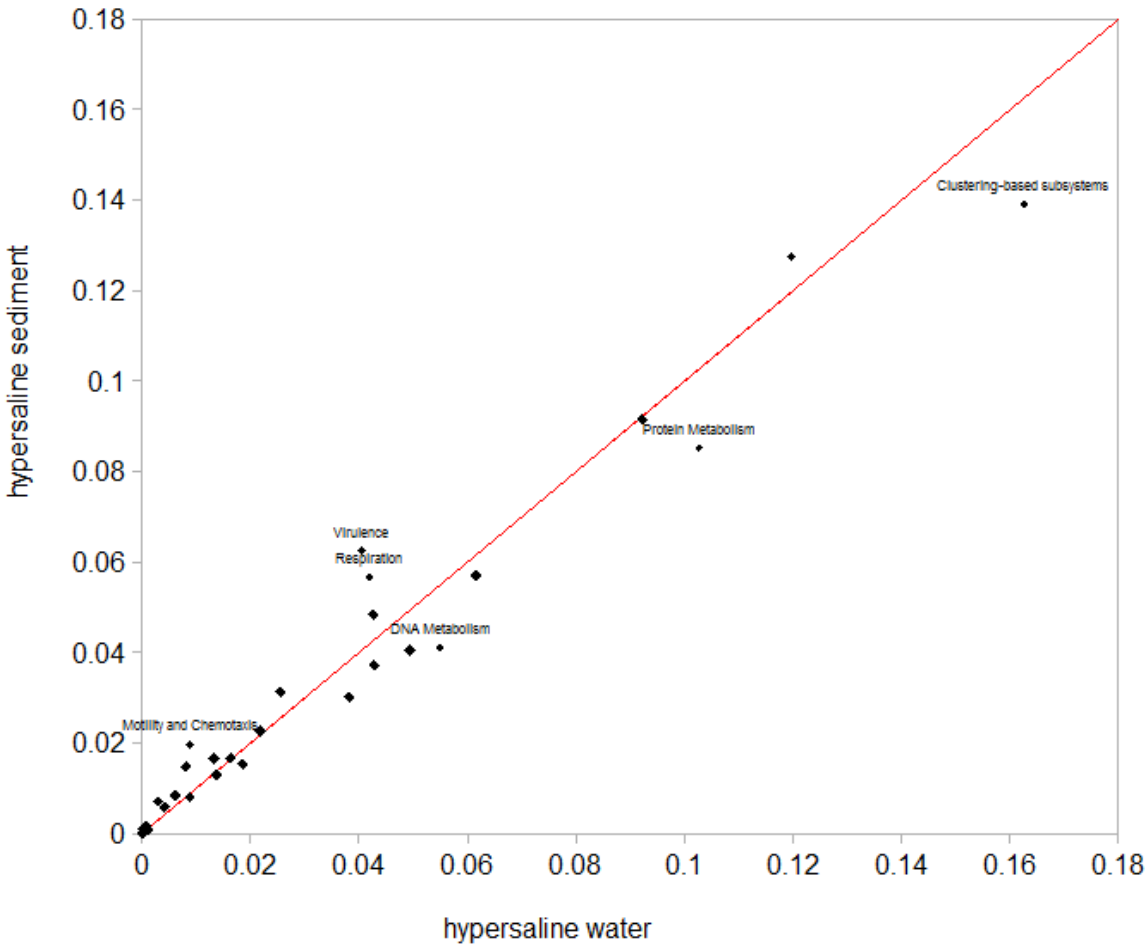
	compounds				
Protein Metabolism	Protein biosynthesis	0.24	0.2	1.86	7.95
Motility and Chemotaxis	Unclassified (motillity & chemotaxis)	0.05	0.09	1.76	9.71
Carbohydrates	Monosaccharides	0.15	0.13	1.58	11.28
Potassium metabolism	Unclassified (potassium metabolism)	0.07	0.11	1.55	12.83
Sulfur Metabolism	Unclassified (sulfur metabolism)	0.09	0.11	1.37	14.2
DNA metabolism	DNA replication	0.14	0.11	1.31	15.51
Clustering-based subsystems	Ribosomal Protein L28P	0.02	0.05	1.21	16.72
Unclassified	Unclassified	0.06	0.08	1.2	17.93
Clustering-based subsystems	Sarcosine oxidase	0.06	0.03	1.09	19.01
RNA metabolism	Transcription	0.12	0.1	1.08	20.09
Motility and Chemotaxis	Social motility and nonflagellar swimming in bacteria	0.06	0.08	1.03	21.13
	Level 3				
Cell Division and Cell Cycle	Cyanobacterial_Circadian_Clock	0.09	0.64	0.67	0.67
Regulation and Cell signaling	cAMP_signaling_in_bacteria	0.49	1.44	0.67	1.34
Motility and Chemotaxis	Bacterial_Chemotaxis	0.25	0.81	0.56	1.9
Sulfur Metabolism	Galactosylceramide_and_Sulfatide_ metabolism	0.36	0.81	0.53	2.43
Virulence	Type_IV_pilus	0.09	0.49	0.51	2.94
Respiration	NiFe_hydrogenase_maturation	0.01	0.16	0.5	3.44
Respiration	Hydrogenases	0.04	0.25	0.5	3.94

Potassium metabolism	Glutathione-regulated_potassium-efflux_system	0.49	1.21	0.49	4.44
Protein Metabolism	tRNA_aminoacylation	2.25	1.44	0.46	4.9
Carbohydrates	Pyruvate:ferredoxin_oxidoreductase	0	0.16	0.46	5.36
DNA Metabolism	Restriction-Modification_System	0.16	0.49	0.44	5.81
Virulence	Cobalt-zinc-cadmium_resistance	0.36	0.81	0.43	6.24
Nitrogen Metabolism	Nitrogen_fixation	0	0.09	0.42	6.66
Virulence	Pyoverdine_biosynthesis_new	0.09	0.36	0.42	7.07
Carbohydrates	Trehalose_Biosynthesis	0.09	0.25	0.41	7.48
Protein Metabolism	General_Secretion_Pathway	0.09	0.36	0.41	7.89
Virulence	Type_4_secretion_and_conjugative_transfer	0.04	0.16	0.41	8.3
Carbohydrates	Deoxyribose_and_Deoxynucleoside_Catabolism	0.49	0.16	0.4	8.71
Regulation and Cell signaling	Phytochromes	0.04	0.16	0.4	9.11
Carbohydrates	D-ribose_utilization	0.64	0.25	0.4	9.5
Respiration	Anaerobic_respiratory_reductases	0.25	0.64	0.39	9.89

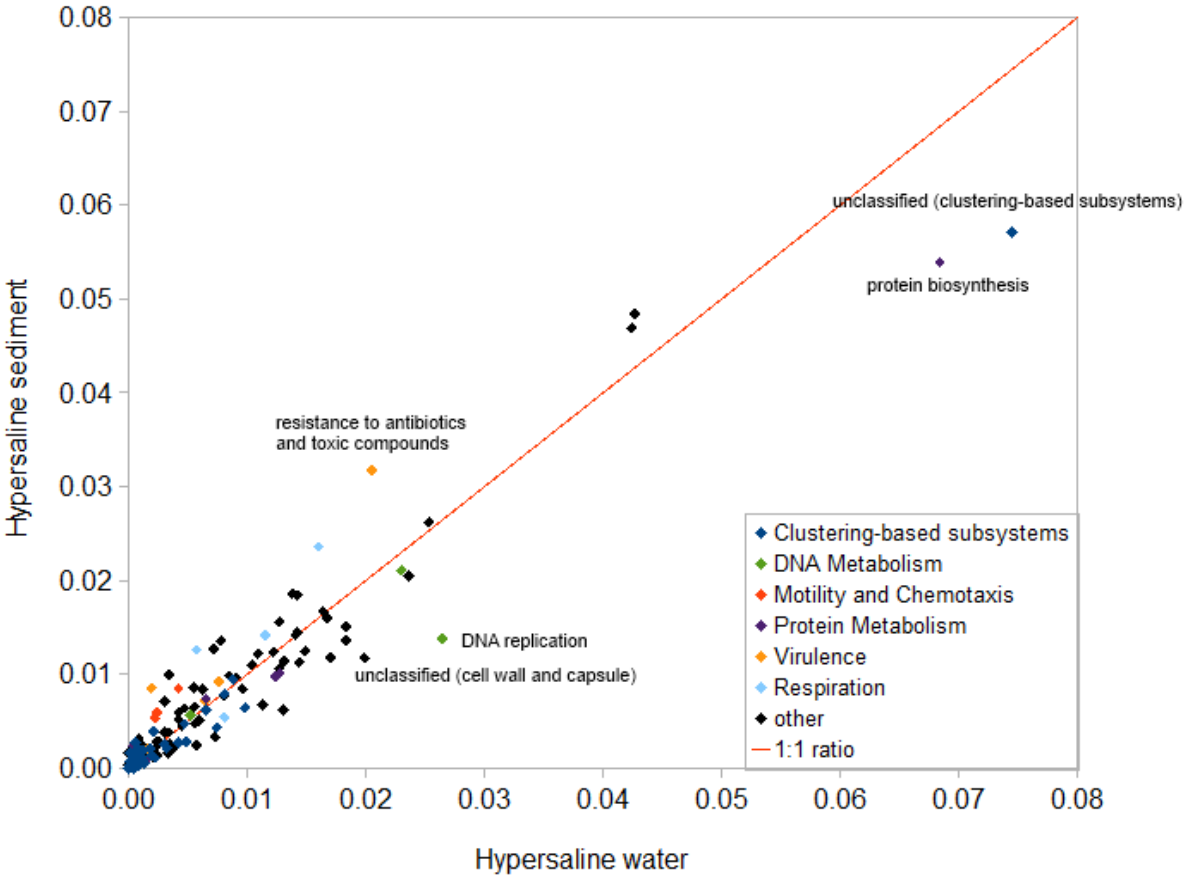
Table 3 SIMPER analysis of taxonomic dissimilarity between solid substrate and water metagenome clusters (Class level).

Taxon	Average abundance in water cluster (%)	Average abundance in solid substrate cluster (%)	% contribution to dissimilarity	Cumulative % contribution to dissimilarity
Alphaproteobacteria	0.55	0.42	8.09	8.09
Actinobacteria (class)	0.29	0.18	6.98	15.06
Betaproteobacteria	0.2	0.33	6.6	21.66
delta/epsilon subdivisions	0.17	0.3	5.71	27.37
Planctomycetacia	0.09	0.18	4.19	31.57
Caudovirales	0.13	0.04	3.92	35.49
Bacteroidetes	0.33	0.31	3.55	39.04
Gammaproteobacteria	0.39	0.43	3.39	42.44
Nostocales	0.06	0.14	3.39	45.83
Prochlorales	0.11	0.04	3.04	48.86
Acidobacteria	0.07	0.04	2.99	51.85

A)


hypersaline sediment
hypersaline water
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
0
0.02
0.04
0.06
0.08
0.1
0.12
0.14
0.16
0.18
Clustering-based subsystems
Protein Metabolism
Virulence
Respiration
DNA Metabolism
Motility and Chemotaxis

B)


0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
Hypersaline sediment
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0.08
Hypersaline water
unclassified (clustering-based subsystems)
protein biosynthesis
resistance to antibiotics
and toxic compounds
DNA replication
unclassified (cell wall and capsule)
Clustering-based subsystems
DNA Metabolism
Motility and Chemotaxis
Protein Metabolism
Virulence
Respiration
other
1:1 ratio

C)

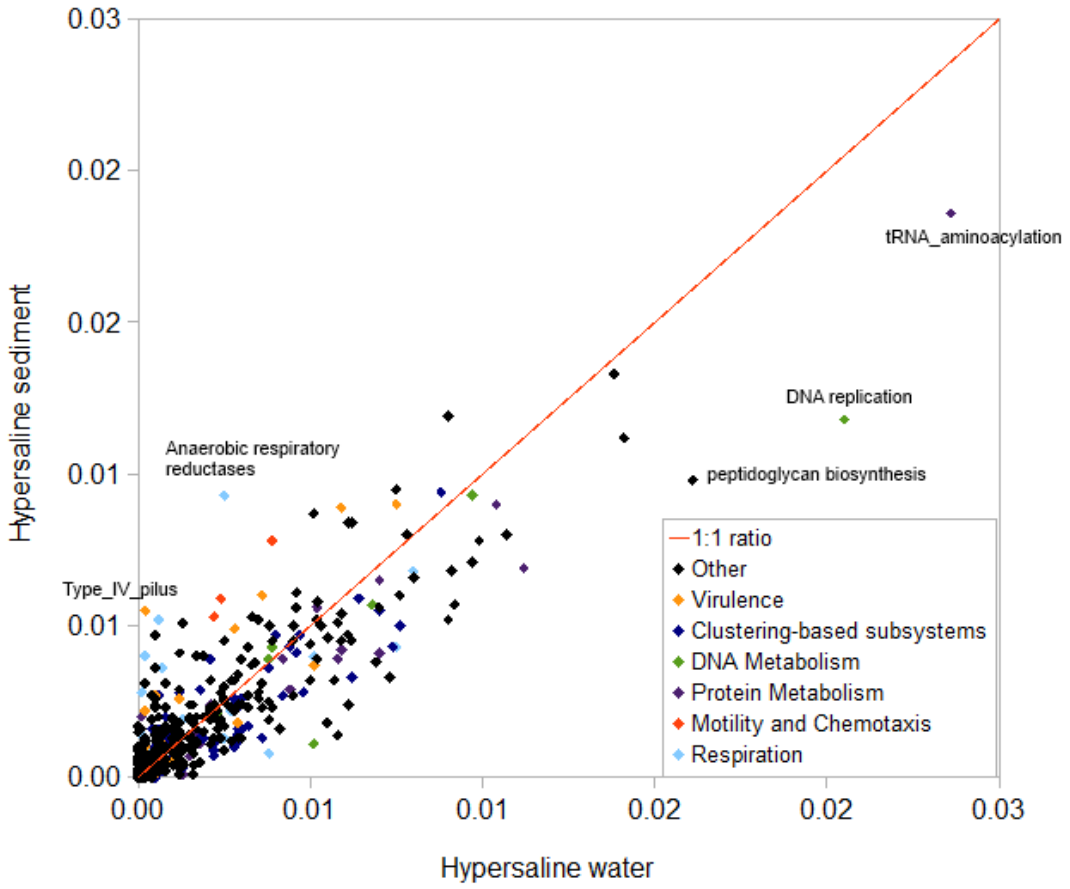
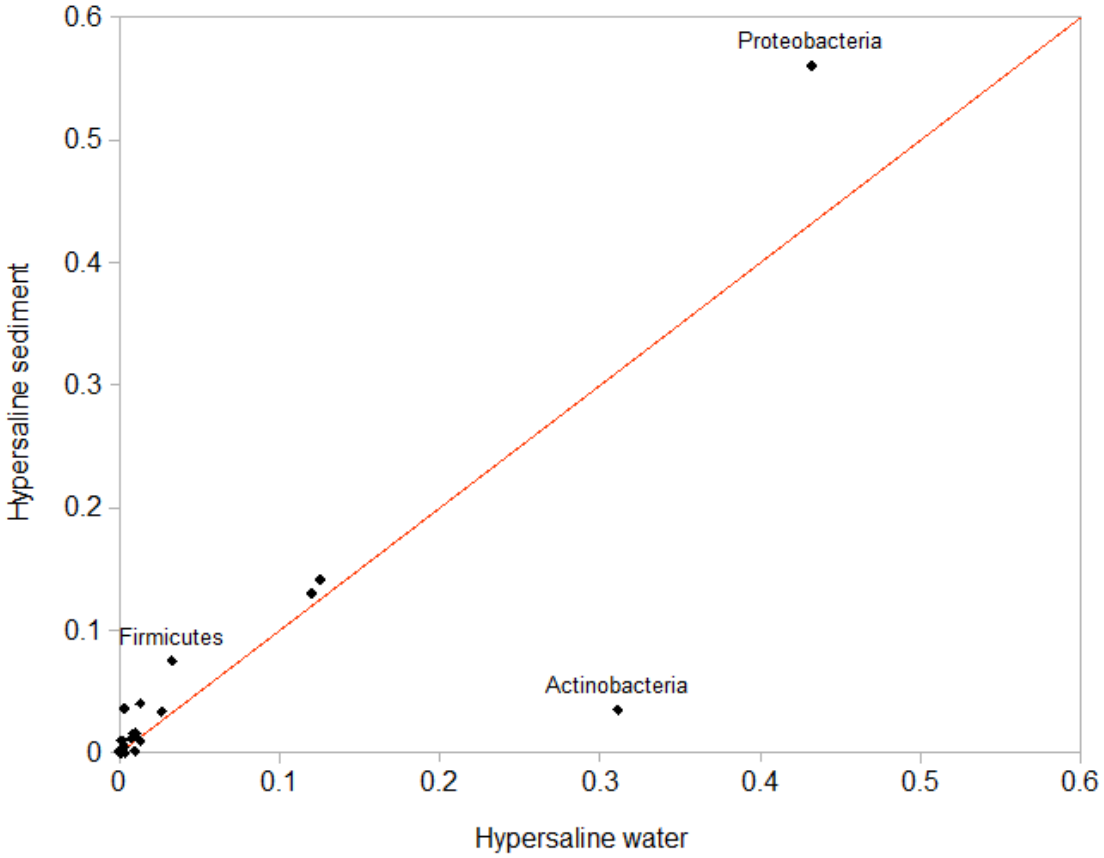

0.03
0.02
0.02
0.01
0.01
0.00
0.00
0.01
0.01
0.02
0.02
0.03
Hypersaline sediment
Hypersaline water
tRNA_aminoacylation
DNA replication
peptidoglycan biosynthesis
Anaerobic respiratory reductases
Type_IV_pilus
1:1 ratio
Other
Virulence
Clustering-based subsystems
DNA Metabolism
Protein Metabolism
Motility and Chemotaxis
Respiration

Figure 1

A)


Proteobacteria
Firmicutes
Actinobacteria
Hypersaline sediment
Hypersaline water
0
0.1
0.2
0.3
0.4
0.5
0.6

B)

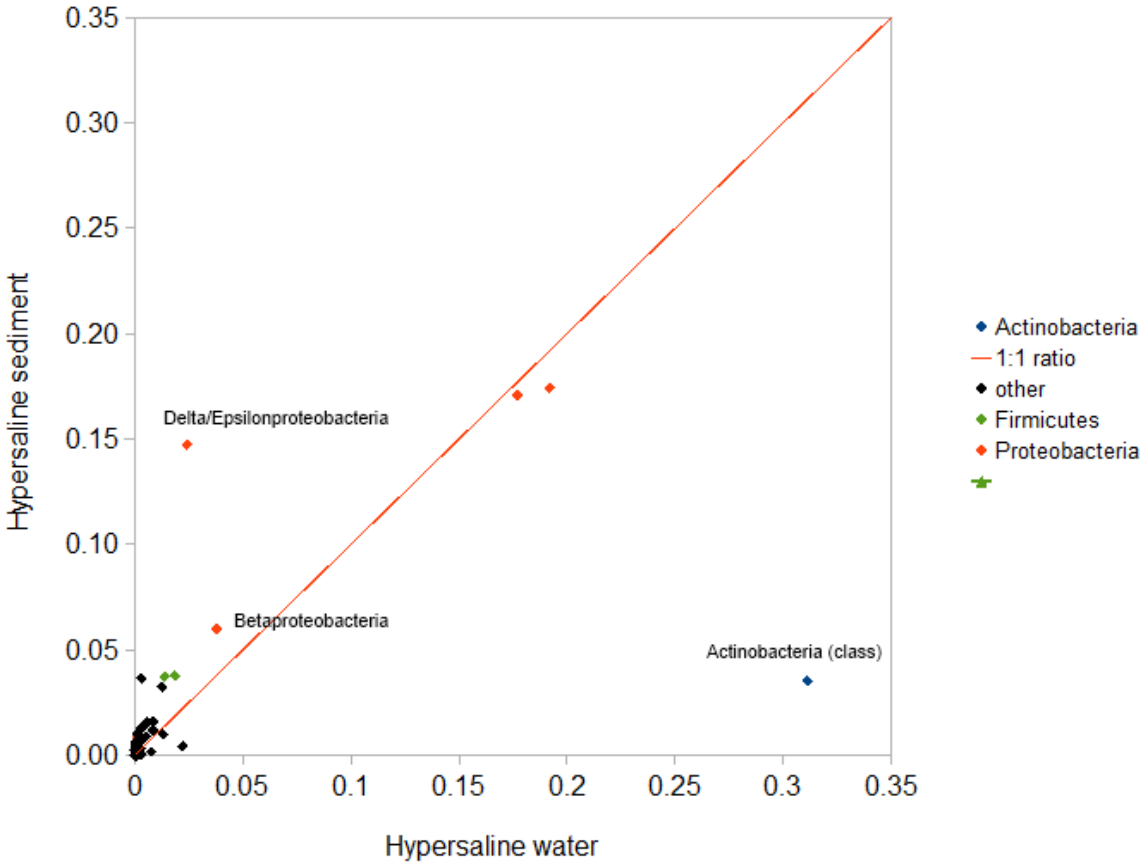
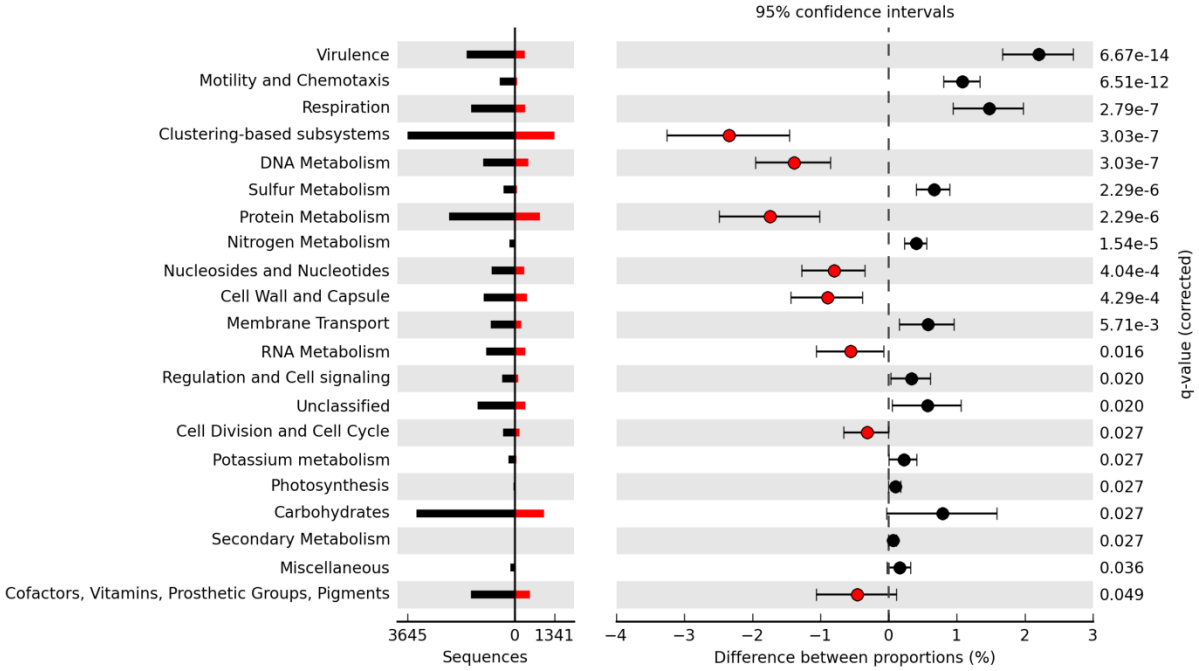
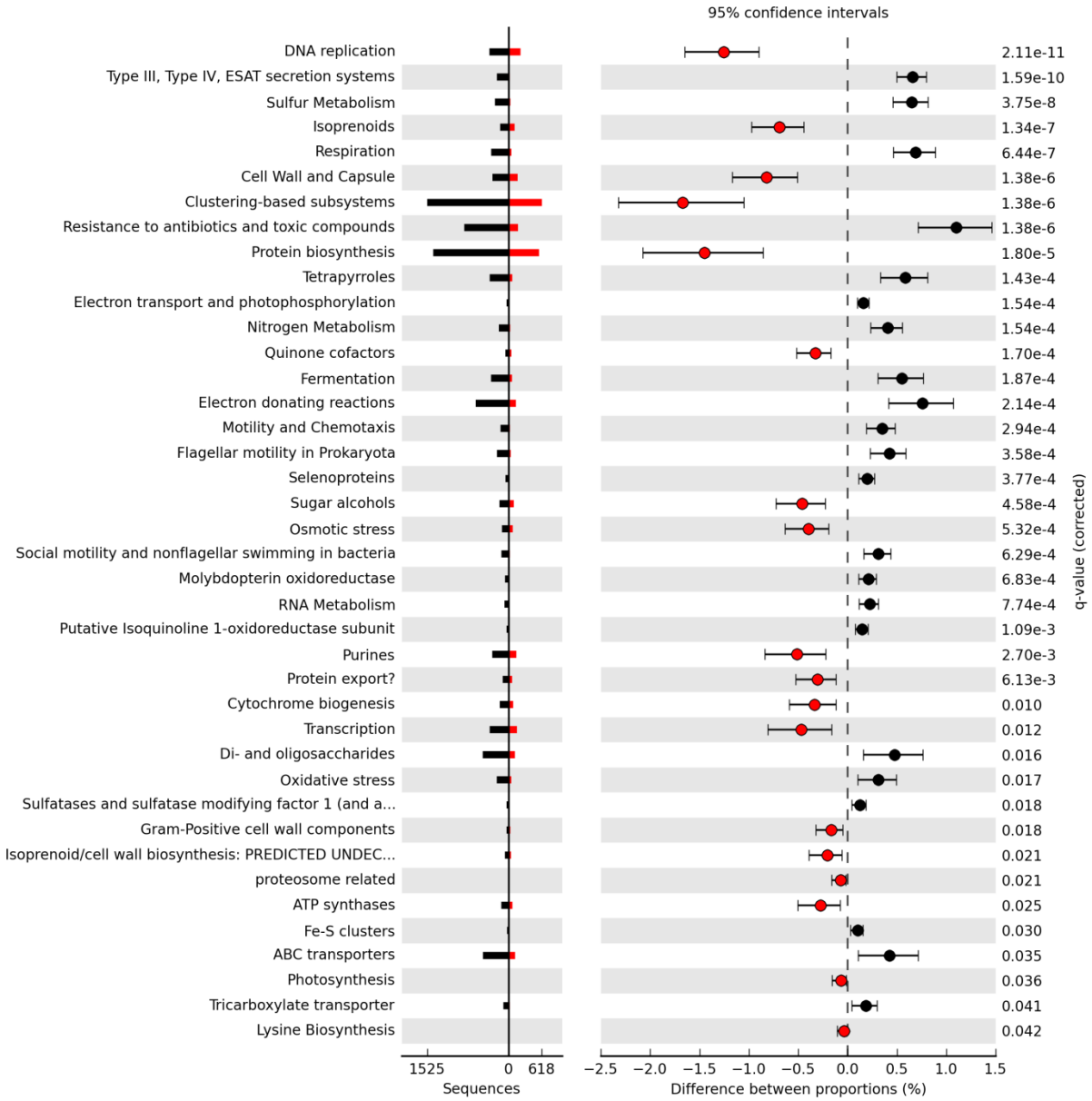

0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
Hypersaline sediment
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
Hypersaline water
Delta/Epsilonproteobacteria
Betaproteobacteria
Actinobacteria (class)
Actinobacteria
1:1 ratio
other
Firmicutes
Proteobacteria

Figure 2

A)


95% confidence intervals
Virulence
Motility and Chemotaxis
Respiration
Clustering-based subsystems
DNA Metabolism
Sulfur Metabolism
Protein Metabolism
Nitrogen Metabolism
Nucleosides and Nucleotides
Cell Wall and Capsule
Membrane Transport
RNA Metabolism
Regulation and Cell signaling
Unclassified
Cell Division and Cell Cycle
Potassium metabolism
Photosynthesis
Carbohydrates
Secondary Metabolism
Miscellaneous
Cofactors, Vitamins, Prosthetic Groups, Pigments
6.67e-14
6.51e-12
2.79e-7
3.03e-7
3.03e-7
2.29e-6
2.29e-6
1.54e-5
4.04e-4
4.29e-4
5.71e-3
0.016
0.020
0.020
0.027
0.027
0.027
0.027
0.027
0.036
0.049
q-value (corrected)
3645
0
1341
Sequences
−4
−3
−2
−1
0
1
2
3
Difference between proportions (%)

B)

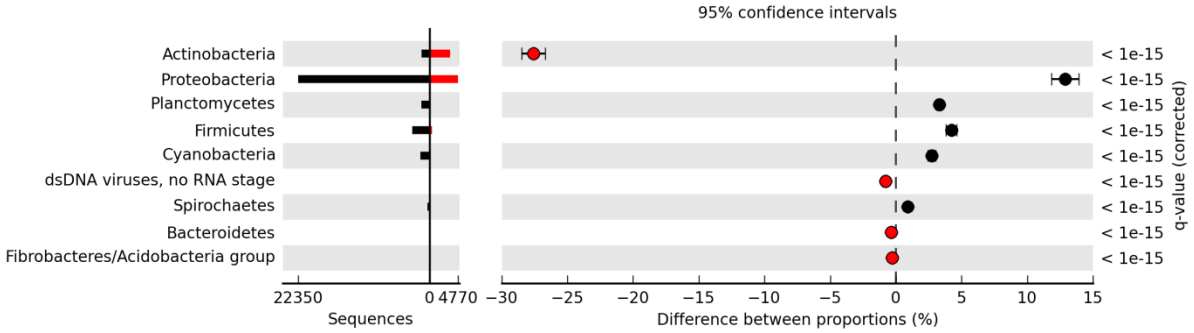

95% confidence intervals
DNA replication
Type III, Type IV, ESAT secretion systems
Sulfur Metabolism
Isoprenoids
Respiration
Cell Wall and Capsule
Clustering-based subsystems
Resistance to antibiotics and toxic compounds
Protein biosynthesis
Tetrapyrroles
Electron transport and photophosphorylation
Nitrogen Metabolism
Quinone cofactors
Fermentation
Electron donating reactions
Motility and Chemotaxis
Flagellar motility in Prokaryota
Selenoproteins
Sugar alcohols
Osmotic stress
Social motility and nonflagellar swimming in bacteria
Molybdopterin oxidoreductase
RNA Metabolism
Putative Isoquinoline 1-oxidoreductase subunit
Purines
Protein export?
Cytochrome biogenesis
Transcription
Di- and oligosaccharides
Oxidative stress
Sulfatases and sulfatase modifying factor 1 (and a...
Gram-Positive cell wall components
Isoprenoid/cell wall biosynthesis: PREDICTED UNDEC...
proteosome related
ATP synthases
Fe-S clusters
ABC transporters
Photosynthesis
Tricarboxylate transporter
Lysine Biosynthesis
2.11e-11
1.59e-10
3.75e-8
1.34e-7
6.44e-7
1.38e-6
1.38e-6
1.38e-6
1.80e-5
1.43e-4
1.54e-4
1.54e-4
1.70e-4
1.87e-4
2.14e-4
2.94e-4
3.58e-4
3.77e-4
4.58e-4
5.32e-4
6.29e-4
6.83e-4
7.74e-4
1.09e-3
2.70e-3
6.13e-3
0.010
0.012
0.016
0.017
0.018
0.018
0.021
0.021
0.025
0.030
0.035
0.036
0.041
0.042
q-value (corrected)
1525
0
618
Sequences
−2.5
−2.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
Difference between proportions (%)

C)


95% confidence intervals
Type_IV_pilus
Anaerobic_respiratory_reductases
Formate_hydrogenase
Galactosylceramide_and_Sulfatide_metabolism
Glycerol_and_Glycerol-3-phosphate_Uptake_and_Utili...
Hydrogenases
DNA_topoisomerases,_Type_II,_ATP-dependent
DNA-replication
NiFe_hydrogenase_maturation
Ubiquinone_Menaquinone-cytochrome_c_reductase_comp...
Pyruvate:ferredoxin_oxidoreductase
Chlorophyll_Biosynthesis
Bacterial_Cell_Division
RNA_polymerase_II
Cyanobacterial_Circadian_Clock
Carotenoids
Na(+)-translocating_NADH-quinone_oxidoreductase_an...
Peptidoglycan_Biosynthesis
Choline_and_Betaine_Uptake_and_Betaine_Biosynthesis
Histidine_Degradation
Nitrogen_fixation
Selenocysteine_metabolism
Coenzyme_B12_biosynthesis
Zinc_resistance
Pyoverdine_biosynthesis_new
Proteolysis_in_bacteria,_ATP-dependent
Sulfate_reduction-associated_complexes
Choline_Transport
Bacterial_Chemotaxis
Menaquinone_and_Phylloquinone_Biosynthesis
Benzoate_degradation
Flagellum
CBSS-235.1.peg.567
Bacterial_motility:Gliding
CBSS-269799.3.peg.2220
Group_II_intron-associated_genes
ABC_transporter_tungstate_(TC_3.A.1.6.2)
Aromatic_amino_acid_interconversions_with_aryl_acids
CBSS-314267.3.peg.390
De_Novo_Purine_Biosynthesis
Ribosome_LSU_bacterial
CBSS-196164.1.peg.461
Glycogen_metabolism
Stringent_Response,_(p)ppGpp_metabolism
Isoprenoid_Biosynthesis
YjeE
CBSS-335283.3.peg.454
Translation_elongation_factor_G_family
Putative_hemin_transporter
F0F1-type_ATP_synthase
cAMP_signaling_in_bacteria
Carbon_monoxide_induced_hydrogenase
5.19e-12
2.86e-9
7.12e-9
1.75e-8
2.58e-8
2.86e-8
4.10e-8
1.77e-6
2.77e-6
4.94e-6
4.94e-6
5.26e-6
8.46e-6
2.41e-5
2.57e-5
7.41e-5
1.15e-4
1.82e-4
2.05e-4
2.14e-4
2.88e-4
4.03e-4
5.34e-4
5.48e-4
6.18e-4
6.18e-4
8.53e-4
1.05e-3
1.13e-3
1.46e-3
1.72e-3
1.72e-3
1.98e-3
1.98e-3
2.19e-3
2.52e-3
2.60e-3
3.43e-3
3.43e-3
3.61e-3
3.91e-3
4.24e-3
4.91e-3
5.83e-3
6.80e-3
7.95e-3
8.19e-3
8.19e-3
8.53e-3
9.07e-3
9.33e-3
9.48e-3
q-value (corrected)
310
0
169
Sequences
−1.5
−1.0
−0.5
0.0
0.5
1.0
Difference between proportions (%)

Figure 3

A)


95% confidence intervals
Actinobacteria
Proteobacteria
Planctomycetes
Firmicutes
Cyanobacteria
dsDNA viruses, no RNA stage
Spirochaetes
Bacteroidetes
Fibrobacteres/Acidobacteria group
< 1e-15
q-value (corrected)
22350
0 4770
Sequences
−30 −25 −20 −15 −10 −5 0 5 10 15
Difference between proportions (%)

B)

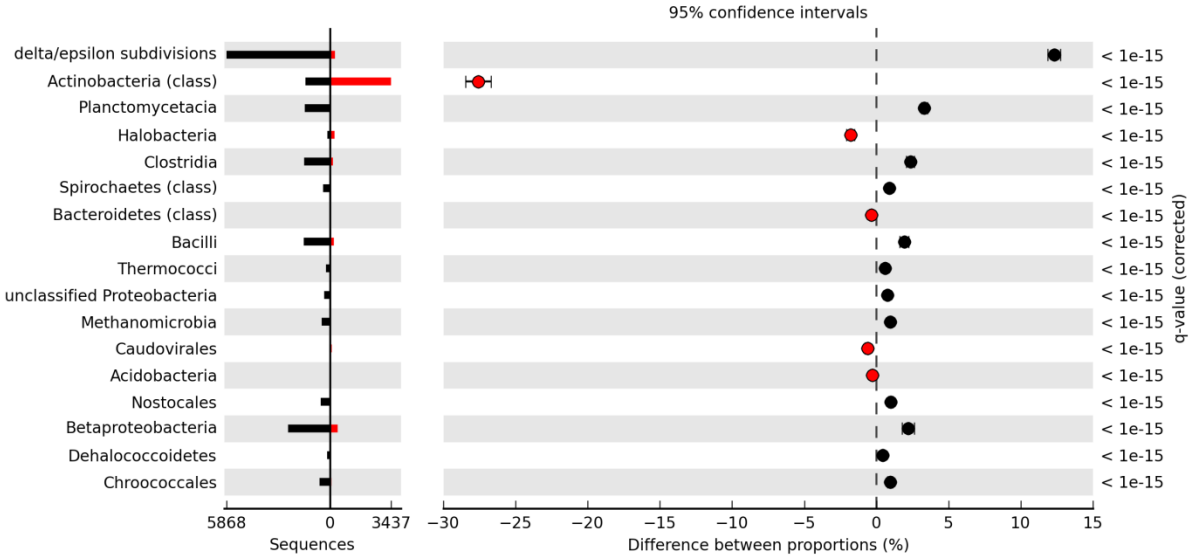

95% confidence intervals
delta/epsilon subdivisions
Actinobacteria (class)
Planctomycetacia
Halobacteria
Clostridia
Spirochaetes (class)
Bacteroidetes (class)
Bacilli
Thermococci
unclassified Proteobacteria
Methanomicrobia
Caudovirales
Acidobacteria
Nostocales
Betaproteobacteria
Dehalococcoidetes
Chroococcales
< 1e-15
q-value (corrected)
5868
0
3437
Sequences
−30 −25 −20 −15 −10 −5 0 5 10 15
Difference between proportions (%)

Figure 4

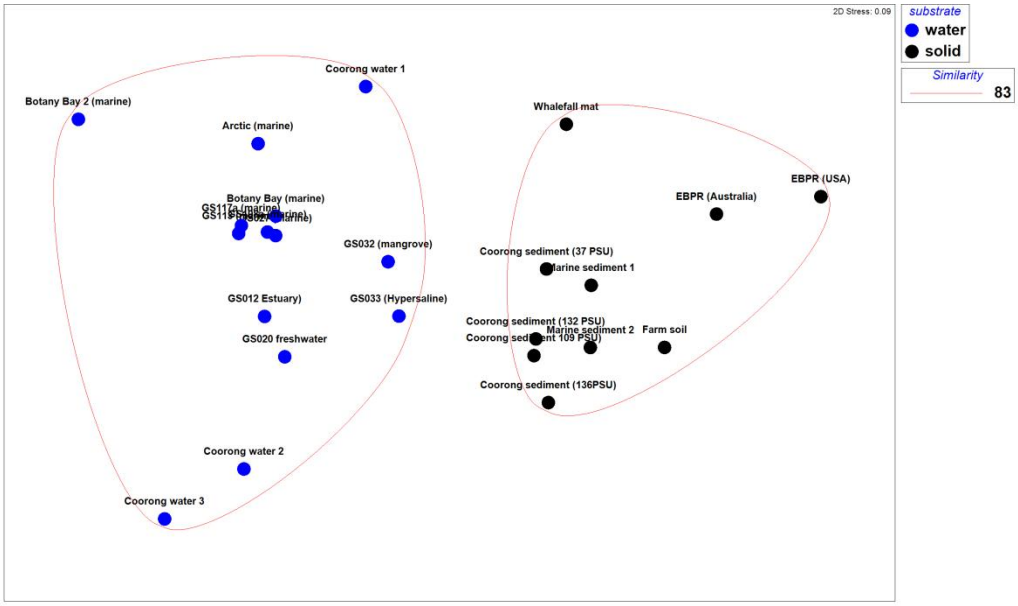

2D Stress: 0.09
substrate
water
solid
Similarity
83
Coorong water 1
Botany Bay 2 (marine)
Arctic (marine)
Botany Bay (marine)
GS032 (mangrove)
GS012 Estuary)
GS033 (Hypersaline)
GS020 freshwater
Coorong water 2
Coorong water 3
Whalefall mat
EBPR (USA)
EBPR (Australia)
Coorong sediment (37 PSU)
Coorong sediment (132 PSU)
Farm soil
Coorong sediment (136PSU)

Figure 5

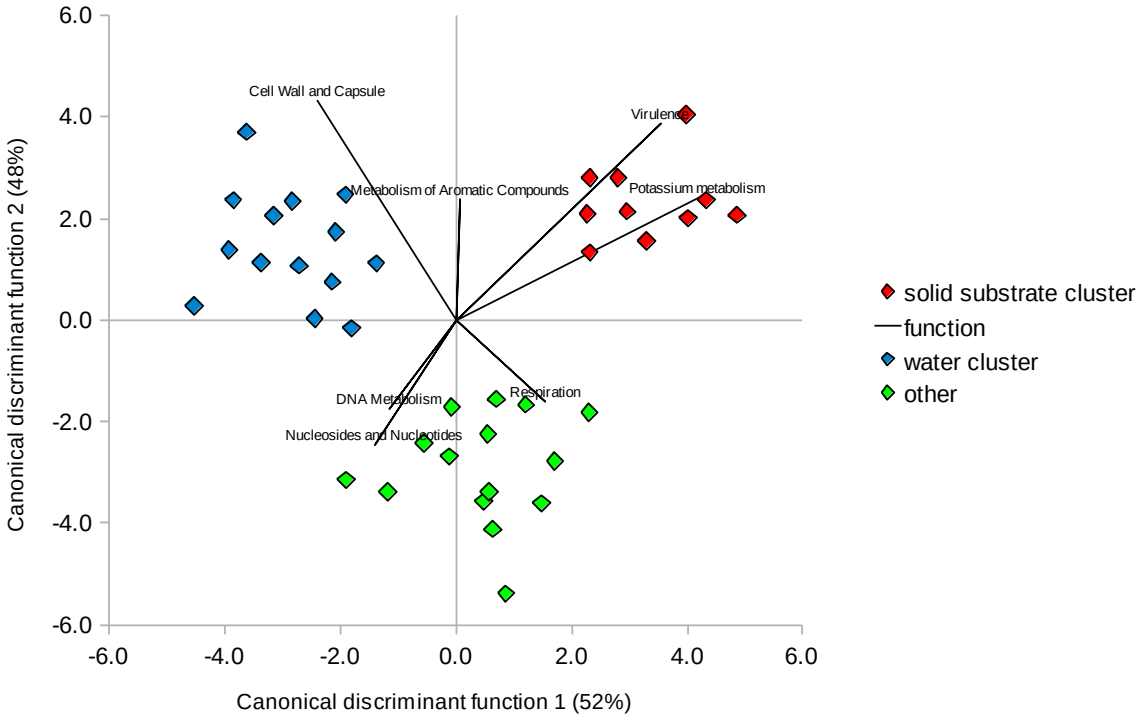

6.0
4.0
2.0
0.0
-2.0
-4.0
-6.0
-6.0
-4.0
-2.0
0.0
2.0
4.0
6.0
Canonical discriminant function 2 (48%)
Canonical discriminant function 1 (52%)
Cell Wall and Capsule
Virulence
Metabolism of Aromatic Compounds
Potassium metabolism
DNA Metabolism
Respiration
Nucleosides and Nucleotides
solid substrate cluster
function
water cluster
other

Figure 6

GENERAL DISCUSSION

Summary of the thesis

This thesis used DNA extracted from four stations along the Coorong lagoon, Australia, to construct 16S rDNA and metagenomic sequence libraries. This sequence data was used to determine the influence of salinity and nutrient variability on community structure and to compare the Coorong to other habitats at the genetic level. Each chapter of the thesis provided novel information regarding the spatial variability of microbial communities, and can be summarized as follows:

Chapter one employed Tag Encoded FLX Amplicon Pyrosequencing (TEFAP) of the bacterial 16S rDNA gene to determine shifts in microbial community structure along the Coorong salinity and nutrient gradient. The overall community structure showed variation along the gradient using measures of both ecological and evolutionary similarity, which was concordant with changes in salinity and nutrient concentration. The proteobacterial genera *Roseobacter* and *Roseovarius* in particular peaked in abundance at intermediate the salinity points and the cyanobacterial genus *Euhalothece* dominated the community at the most saline and nutrient rich site. Overall, the Coorong was ecologically and evolutionarily distinct when compared to other habitats, indicating that the overall nature of the habitat had a role in determining community structure in addition to salinity and nutrient concentrations. This was the first application of high throughput pyrotag sequencing to microbial communities along a continuous sediment salinity gradient from marine to hypersaline salinities.

Chapter two employed Tag Encoded FLX Amplicon Pyrosequencing (TEFAP) targeting the 16S rDNA gene of Archaea to determine shifts in archaeal community structure along the Coorong gradient. Archaeal communities showed strong transitions from 70% *Crenarchaeota*

at marine salinity to 70% *Euryarchaeota* of the class *Thermoplasmata* at the most hypersaline site. At intermediate salinities rare archaeal groups rose in abundance 18 to 25 times indicating that they are a highly dynamic part of the community. Such a clear and dramatic relationship between *Crenarchaeota and Thermoplasmata* sequence abundance and salinity has not been observed previously. The observed sharp changes in community structure and dominance of key groups demonstrate the importance of salinity in structuring archaeal communities.

Combined, chapters one and two addressed the first aim of the thesis by demonstrating how community composition changed with physiochemical parameters along a continuous natural gradient of salinity and nutrients concentration, and identified which taxonomic groups demonstrated the largest shifts.

Chapter three employed metagenomic sequencing to determine what the shifts were in the abundance of functional gene categories along the Coorong gradient. The most significant shifts were in the genetic potential for halotolerance and photosynthesis, which were more highly represented in hypersaline samples. At these sites, halotolerance was achieved by an increase in genes responsible for the acquisition of compatible solutes – organic chemicals that influence the carbon, nitrogen and methane cycles of sediment. Increases in photosynthetic genes were coupled to an increase in genes matching *Cyanobacteria*. These results provide a model for how environmental gradients can drive shifts in biogeochemically important metabolic processes and taxonomic groups.

Chapter three addressed the second aim of the thesis by demonstrating that the community composition was functionally driven by underlying shifts in the abundance of metabolic gene categories. We identified these categories as belonging to salinity tolerance functions providing novel insight into localized adaptation to habitat variability at the genetic level, addressing the third aim of the thesis.

Combined, chapters one, two and three all demonstrated that microbial community composition is influenced by local habitat variability, in this case salinity and nutrient status, and that this variability influences the abundance of individual taxonomic groups and genes encoding specific functions relevant to the gradient. One theme that was consistent between the 16S rDNA and metagenomic datasets was an increase in cyanobacterial ribotypes and metagenomic sequences at the most saline site, which was concordant with an increase in genes encoding photosynthesis. This indicates that one of the most ecologically significant groups of bacteria are strongly influenced by local gradients.

Chapter four compared the metagenomic profiles from the Coorong gradient to other habitats using publically available metagenomes. Despite the habitat variability between the Coorong samples, these metagenomes were more similar to each other than to other habitats including those from comparable salinities. Those that were most similar to the Coorong metabolically were all from solid substrate habitats and clustered closely despite marked differences in local habitat parameters. Overall, clustering of solid substrate and water metagenomes into discrete similarity groups based on functional potential indicated for the first time that the dichotomy between water and solid matrices is a fundamental determinant of overall community metabolism that is not masked by salinity, nutrient concentration or microbial abundance. This

indicates that taxonomic groups adapted to different salinities share a core functional repertoire that is determined by the substrate type and that this signature should be consistent across gradients and in different biomes.

Chapter five used the above metagenomic datasets to determine which functional processes and taxonomic groups drove the substrate partitioning of metagenomic profiles, first using Coorong sediment and water metagenomes as examples to generate candidate categories on a local scale, and then testing whether these were consistent contributors to substrate dissimilarity on a global scale in a variety of biomes. Each habitat type had a distinctive core of more abundant taxa and processes which defined life in each substrate type. Specifically, functions related to toxin resistance, motility and anaerobic life were more abundant in solid substrate habitats, whereas core metabolic processes central to survival were more abundant in water habitats indicating less variable genomic content in these environments. This was the first identification of processes which consistently vary between solid substrate and water habitats, irrespective of local conditions.

Combined chapters four and five provided novel insight into the relative influence of hierarchical variables simultaneously controlling different aspects of community composition. As substrate type influenced metagenomic community structure regardless of physiochemical variability, this added to our *a priori* hypotheses regarding the influence of local physiochemical gradients on community composition.

Summary: Demonstration of hierarchical controllers of microbial community structure

Overall, this thesis conveys a scaled view of microbial ecology in that it identifies several contrasting factors that simultaneously control microbial community composition on different scales. On the local scale, variation in chemical parameters determined the abundance of specific taxonomic groups and metabolic processes. For metabolism many of these processes were directly related to salinity tolerance and were not abundant core processes related to central metabolism. Core processes showed little variation across the gradient. Thus, this conserved overall signature, evidenced by the discrete clustering of Coorong samples in both 16S rDNA and metagenomic datasets, was determined by a 'higher' hierarchical structuring variable which our evidence identifies as the substrate type in which the community exists.

Metagenomes may indeed be viewed as discrete units of biological organization (6) which demonstrate patterns not necessarily masked by small scale fluctuations in the abundance of genes and taxa within them. Our data supports this view in that discrete metagenomic signatures, largely determined by abundant core genomic features, emerge in solid substrate habitats regardless of the local environmental conditions and that this signature may then fluctuate based on the local habitat specific selection of individual flexible genomic categories, such as salinity tolerance and *Cyanobacteria* abundance that vary along physiochemical gradients.

Hierarchical control of microbial community composition fits with the emerging view of macro-organism ecology that suggests ecology should not be viewed as a question of which variables are most important in controlling community structure, but which is most important at particular scales (13). Thus, it may be that on one scale variation in factors such as salinity (3, 9, 11, 16),

nutrient concentration (8, 15) and temperature (14) determine community structure but at another scale factors such as the overall biome classification (5, 17), water mass classification (7), environmental history (12), habitat complexity (1, 10) and substrate type determine community structure or the 'seed bank potential' of microbial communities. These variables act on different spatial scales and may also influence the abundance of rarer taxa and flexible genes or dominant taxa and core genes thus influencing the overall signature of the metagenome to varying degrees.

Our thesis provides a novel analysis of how salinity and nutrient gradients influence the abundance of taxa and genes on a local scale and for the first time demonstrates how the overall metabolic function of a community may be determined by the physical structure of the habitat rather than local physiochemical conditions, indicating the presence of a shared global sediment and water metabolic signature. This provides a clear example of hierarchical controlling factors which act simultaneously to determine microbial community structure and highlights the complexity of the interaction between the habitat and microbiota. The complex interplay between different variables acting on genes, genomes and metagenomes, all on different scales, results in the complexity, diversity and flexibility of microbial communities that have allowed microbes to exist for three and a half billion years, occupy the extremes of the biosphere, and drive the ecology of the planet.

References

1. **Angly, F. E., D. Willner, A. Prieto-Davo, R. A. Edwards, R. Schmieder, R. Vega-Thurber, D. A. Antonopoulos, K. Barott, M. T. Cottrell, C. Desnues, E. A. Dinsdale, M. Furlan, M. Haynes, M. R. Henn, Y. F. Hu, D. L. Kirchman, T. McDole, J. D. McPherson, F. Meyer, R. M. Miller, E. Mundt, R. K. Naviaux, B. Rodriguez-Mueller, R. Stevens, L. Wegley, L. X. Zhang, B. L. Zhu, and F. Rohwer.** 2009. The GAAS Metagenomic Tool and Its Estimations of Viral and Microbial Average Genome Size in Four Major Biomes. PLoS Computational Biology **5:** e1000593 .
2. **Baas Becking, L. G. M.** 1934. Geobiologie of inleiding tot de milieukunde.
3. **Caporaso, J. G., C. L. Lauber, W. A. Walters, D. Berg-Lyons, C. A. Lozupone, P. J. Turnbaugh, N. Fierer, and R. Knight.** Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proceedings of the National Academy of Sciences. **108** Supplement 1 4516-4522.
4. **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.
5. **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.
6. **Doolittle, W. F., and O. Zhaxybayeva.** 2010. Metagenomics and the Units of Biological Organization. Bioscience **60:**102-112.

7. **Lovejoy, C., P. Galand, and D. Kirchman.** 2011. Picoplankton diversity in the Arctic Ocean and surrounding seas. Marine Biodiversity 41:5-12

8. **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.

9. **Jeffries, T. C., J. R. Seymour, K. Newton, R. J. Smith, L. Seuront, and J. G. Mitchell.** 2011. Increases in the abundance of microbial genes encoding halotolerance and photosynthesis along a sediment salinity gradient. Biogeosciences **9: 815-825.**.

10. **Konstantinidis, K. T., and J. M. Tiedje.** 2004. Trends between gene content and genome size in prokaryotic species with larger genomes. Proceedings of the National Academy of Sciences of the United States of America **101:**3160-3165.

11. **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.

12. **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.

13. **McGill, B. J.** 2010. Matters of Scale. Science **328:**575-576.

14. **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. Mol Syst Biol **7**: 47.3

15. **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.

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

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