environmental microbiology reports

Environmental Microbiology Reports (2012) doi:10.1111/j.1758-2229.2012.00362.x

Contrasting microbial assemblages in adjacent water masses associated with the East Australian Current

Justin R. Seymour,[1]* Martina A. Doblin,[1] Thomas C. Jeffries,[2] Mark V. Brown,[3] Kelly Newton,[2] Peter J. Ralph,[1] Mark Baird[1] and James G. Mitchell[2]

[1]Plant Functional Biology & Climate Change Cluster, University of Technology, PO Box 123, Broadway, Sydney, NSW 2007, Australia.

[2]School of Biological Sciences, Flinders University, PO Box 2100, Adelaide, SA 5001, Australia.

[3]School of Biotechnology and Biomolecular Science, University of New South Wales, Kensington, NSW 2052, Australia.

Summary

Different oceanographic provinces host discrete microbial assemblages that are adapted to local physicochemical conditions. We sequenced and compared the metagenomes of two microbial communities inhabiting adjacent water masses in the Tasman Sea, where the recent strengthening of the East Australian Current (EAC) has altered the ecology of coastal environments. Despite the comparable latitude of the samples, significant phylogenetic differences were apparent, including shifts in the relative frequency of matches to *Cyanobacteria*, *Crenarchaeota* and *Euryarchaeota*. Fine-scale variability in the structure of *SAR11*, *Prochlorococcus* and *Synechococcus* populations, with more matches to 'warm-water' ecotypes observed in the EAC, indicates the EAC may drive an intrusion of tropical microbes into temperate regions of the Tasman Sea. Furthermore, significant shifts in the relative importance of 17 metabolic categories indicate that the EAC prokaryotic community has different physiological properties than surrounding waters.

Introduction

The oceanography of eastern Australia is dominated by the East Australian Current (EAC), a western boundary current, which redistributes warm tropical waters from the Coral Sea into the temperate Tasman Sea (Fig. 1) (Baird *et al*., 2008). In recent years, increases in wind stress across the South Pacific Ocean, driven by climate change, have strengthened the subtropical gyre and resulted in increased southerly intrusions of the EAC's warm oligotrophic waters (Cai *et al*., 2005; Ridgway, 2007; Hill *et al*., 2008; Wu *et al*., 2012). As a consequence, the Tasman Sea is amongst the most rapidly warming ocean provinces in the Southern Hemisphere (Cai *et al*., 2005) and marked poleward range expansions of marine macrofauna have been observed in eastern Australian waters (Ling *et al*., 2009; Last *et al*., 2011).

There is also evidence that the strengthening EAC has altered phytoplankton diversity in this region during the last 60 years (Thompson *et al*., 2009), with further range expansions of tropical species into southerly waters predicted (Hallegraeff *et al*., 2009). Given that discrete oceanographic features can demarcate microbial community structure (Johnson *et al*., 2006; Hewson *et al*., 2009) and climate change processes are predicted to affect marine microbial diversity and function (Ducklow *et al*., 2010), similar shifts may be expected for prokaryotic communities in this region. Here, we used metagenomics to assess the potential role of the EAC in shifting the structure and function of microbial assemblages in the Tasman Sea by comparing the prokaryote communities within the southern EAC and an adjacent Tasman Sea water mass (TSW) (Fig. 1).

Results and discussion

Consistent with its tropical origin, surface temperatures in the EAC (23.0°C) were higher than in the adjacent TSW site (18.7°C). Nutrient concentrations were also lower within the EAC sample (Table 1). Water samples were collected from the chlorophyll maximum at each location, corresponding to 25 and 47 m in the EAC and TSW respectively. Metagenomes were sequenced using a 454 GS-FLX (Roche) pyrosequencer, and the resultant 89 047 sequences were compared with the SEED (2011 version) protein non-redundant database in the *Meta Genome Rapid Annotation using Subsystems Technology* (MG-RAST; Version 2.0) pipeline (Overbeek *et al*., 2005; Meyer *et al*., 2008), using BLASTX across all alignment lengths, and matches with an *e*-value of $< 10^{-5}$ were considered significant (Vega Thurber *et al*., 2009; Edwards *et al*., 2010; Pfister *et al*., 2010; Poroyko *et al*., 2010; Delmont *et al*., 2011; Jeffries *et al*. 2011; Smith *et al*. 2012).

Received 28 September, 2011; accepted 30 May, 2012. *For correspondence. E-mail justin.seymour@uts.edu.au; Tel. (+61) 2 9514 4092; Fax (+61) 2 9514 4079.

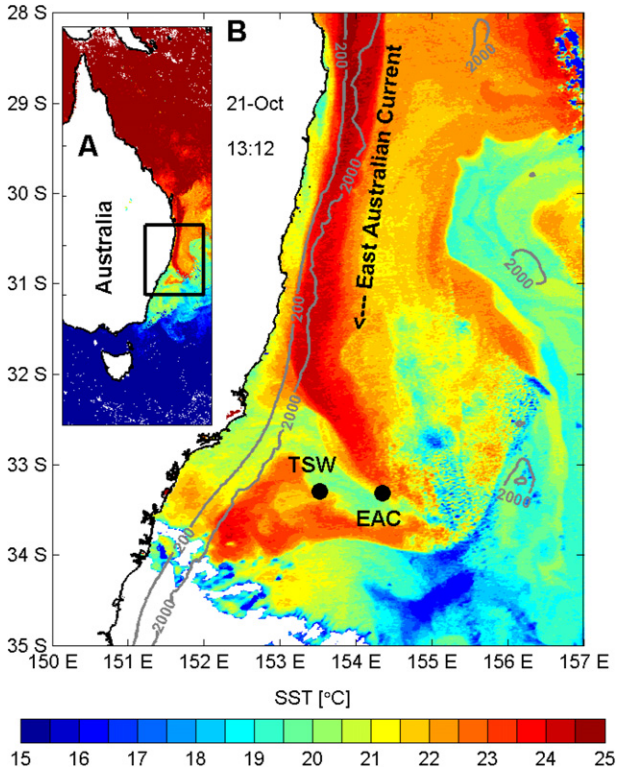


Fig. 1. Sample locations and sea surface temperature (SST) at the time of sampling, indicating the position of the EAC.A. Legacy Bureau of Meteorology SST (LBoMSST) product from 21 October 2009; a 0.01° resolution weighted combination of AVHRR data estimating the sub-skin (~ 5 m depth) ocean temperature.B. MODIS Aqua 1 km resolution SST from 13:12 AEST 21 October 2009. Grey lines show the 200 m and 2000 m isobaths (a detailed description of the oceanography of the region at the time of sampling and sample analysis protocols is provided as Supporting Information).

Using SEED, a total of 50% and 54% of sequences could be matched to known phylogenies in the EAC and TSW samples respectively, which is a level of phylogenetic assignment consistent with previous ocean metagenomes (Hewson *et al.*, 2009). Both samples were dominated by *Bacteria* (93% and 91% of SEED matches), in particular the *Proteobacteria*, which made up 70% and 76% of bacterial matches in the EAC and TSW respectively. In both samples, the bulk of genus level matches were to SAR11 (Fig. 2), which is consistent with the dominance of this organism in most oceanic environments (Morris *et al.*, 2002). Similar genus level patterns were observed when we annotated our sequences against the NCBI NR database in the Community Infrastructure for Advanced Microbial Ecology Research and Analysis (CAMERA) system (Sun *et al.*, 2011) and MEGAN software package (Huson *et al.*, 2007) (Fig. S1).

Given the observed predominance of SAR11-affiliated sequences in both samples, we next compared the metagenomes to two boutique SAR11 sequence databases. The first comprised full-length 16S rRNA gene sequences from the SILVA database (Pruesse *et al.*, 2007) representative of previously defined SAR11 subgroups (Carlson *et al.*, 2009). The second consisted of an internal transcribed spacer (ITS) database derived from SAR11 16S-ITS clones, which further differentiates several of the 16S-based subgroups into different phylotypes (Brown and Fuhrman, 2005). The 16S-based analysis identified SAR11 subgroup S1a (Carlson *et al.*, 2009) as the major SAR11 subgroup in both the EAC and TSW samples (Fig. 3A and B). This subgroup is generally the most abundant in marine surface waters and has several cultivated and sequenced representatives including *Candidatus* Pelagibacter ubique HTCC1062, HTCC1002 (Rappé *et al.*, 2002) and HTCC7211 (Stingl *et al.*, 2007). Within this S1a subgroup, P1a.3, of which the isolate HTCC7211 isolated from the Sargasso Sea is a member, (Stingl *et al.*, 2007) was the dominant phylotype (Fig. 3C and D). P1a.3 has previously been shown to occur in predominantly temperate and tropical waters (Phylotype previously named S2 in Brown and Fuhrman, 2005). The SAR11 16S Subgroup S2 was also identified in both samples using 16S analysis (Fig. 3A and B); however, ITS sequences from this subgroup (belonging to phylotype P2.1) were only observed in the TSW sample. It has been suggested the abundance of this subgroup may be related to nutrient-rich upwelling events (Carlson *et al.*, 2009), consistent with the cooler nutrient-rich waters of the Tasman Sea sample. Finally, subgroup S1b was identified only in the EAC sample. This subgroup (and associated Phylotype S1b) generally occurs only in tropical warmer waters (Brown and Fuhrman, 2005) and its appearance here is consistent with the tropical origin of the EAC. The relative dominance of SAR11 was also

Table 1. Characteristics of EAC and TSW water masses at the depth of chlorophyll max where samples were obtained.

	EAC	TSW
Depth (m)	25	47
Temperature (°C)	22.0	18.7
Dissolved nitrate (μM)	BD	2.18
Dissolved silicate (μM)	BD	0.42
Dissolved phosphate (μM)	0.08	0.24
PAR (μmol photon m^{-2} s^{-1})	87	24
Chlorophyll-a (μg l^{-1})	0.47	0.49
Divinyl chlorophyll-a and b (μg l^{-1})	0.015	BD
Zeaxanthin (μg l^{-1})	0.051	0.016
Bacterial abundance (ml^{-1})	2.2×10^6	2.0×10^6
Synechococcus abundance (ml^{-1})	8.5×10^4	1.8×10^4
Prochlorococcus abundance (ml^{-1})	1.0×10^4	1.2×10^3
Number of sequences	53 165	35 882
Total sequence size (bp)	16 104 563	11 261 883
Average sequence length (bp)	303	314

Sample analysis methodology is described in Supplementary Information.
BD, below detection.

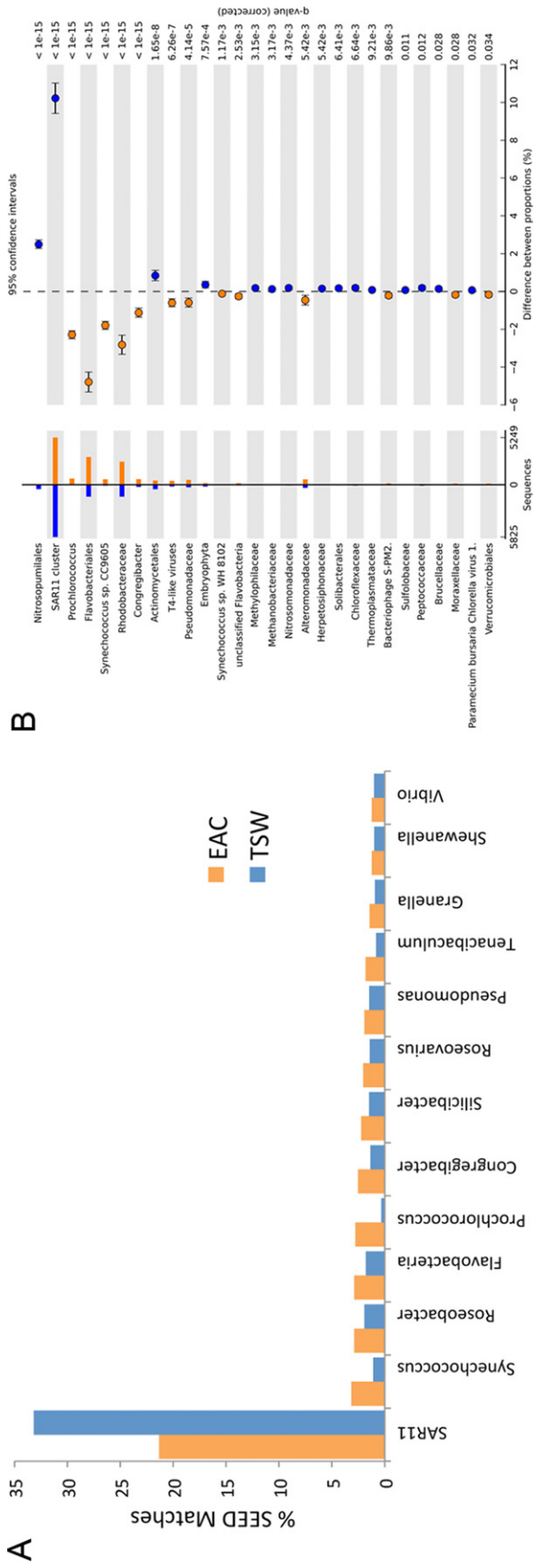


Fig. 2. A. Frequency distribution (relative % of bacterial SEED matches) of top 20 bacterial genera in EAC and TSW metegenomes (ordered according to frequency of occurrence in the EAC sample). Recovered unassembled sequences were annotated using the Meta Genome Rapid Annotation using Subsystems Technology (MG-RAST) pipeline Version 2.0 (Meyer *et al.*, 2008). Phylogeny was assigned by matching sequences to the SEED database (Overbeek *et al.*, 2005) using BLASTX across all alignment lengths and matches with an *e*-value of $< 10^{-5}$ were considered significant.B. The STAMP (STatistical Analysis of Metagenomic Profiles) package (Parks and Beiko, 2010) was used to statistically test for significant differences in the relative abundance of phylogenetic groupings. Fisher's exact test was used to determine significance with a Storey's FDR multiple test correction applied. Hence, all quoted *P*-values represent corrected values (equating to *q*), with values < 0.05 considered significant (Parks and Beiko, 2010). Confidence intervals (95%) were determined using the Newcombe–Wilson method. Groups over-represented in the EAC community (orange) correspond to negative differences between proportions. Groups over-represented in the TSW community (blue) correspond to positive differences between proportions. The metagenomes can be accesses through MG-Rast (http://metagenomics.anl.gov/) under the project numbers 4446457.3 and 4446407.3.

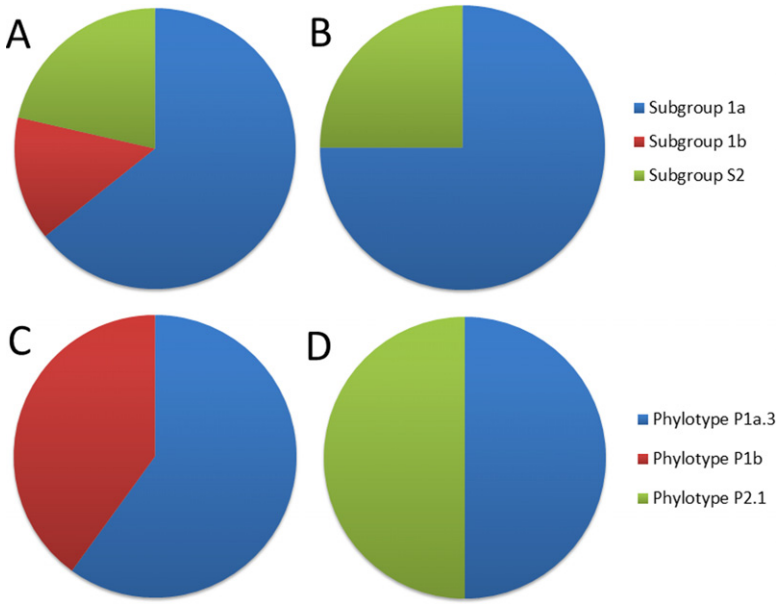


Fig. 3. Frequency of matches to SAR11 16S rRNA subgroups (as defined by Carlson *et al*., 2009) in the (A) EAC and (B) TSW, and matches to ITS phylotypes (as defined by Brown and Fuhrman, 2005) in the (C) EAC and (D) TSW samples. Metagenomic datasets were interrogated using the annotated 16S or ITS reference databases using BLASTN $e = 0.0001$. The subgroup identity of 16S sequence fragments covering at least 100 bp and having a sequence identity of > 97%, and phylotype identity ITS sequence fragments covering at least 80 bp and having a sequence identity of > 90%, were identified from high scoring pairs.

found to be lower in the EAC sample, where there was a significantly ($P < 10^{-15}$) higher frequency of SEED matches to the *Flavobacteriales*, *Rhodobacteriaceae* and the major oceanic cyanobacteria genera *Synechococcus* and *Prochlorococcus*, as determined using Fisher's exact test in the STAMP (STatistical Analysis of Metagenomic Profiles) package (Parks and Beiko, 2010) (Fig. 2, Fig. S1).

The relative number of SEED matches to cyanobacteria was almost 3 times higher in the EAC sample than the TSW sample, but most striking was a shift in the distribution of cyanobacterial genera (Figs 2 and 4). Matches to *Synechococcus* were the dominant feature of both samples, which is typical for sub-tropical and temperate marine ecosystems (Partensky *et al*., 1999a). However, there were significantly ($P < 10^{-15}$) more matches to *Prochlorococcus*, the most abundant photoautotroph in tropical oligotrophic waters (Partensky *et al*., 1999b), in the EAC sample (42% of cyanobacteria) than in the TSW sample (14%). This pattern is in agreement with pigment analyses, where divinyl chlorophyll *a* and *b* were observed in the EAC, but not the TSW, flow cytometry counts, where *Prochlorococcus* abundances were almost an order of magnitude higher in the EAC sample (Table 1), and comparison of our data to a 16S-ITS database that resulted in no hits to *Prochlorococcus* in the TSW sample.

Superimposed upon the broad-scale shifts in the relative importance of *Prochlorococcus* were differences in the distribution of *Prochlorococcus* ecotypes. Within the EAC sample, 81% of matches to *Prochlorococcus* proteins, derived from sequenced representatives in the SEED data-base, were associated with MIT9312, the dominant *Prochlorococcus* ecotype in tropical oceans (Johnson *et al*., 2006). However, MIT9312 accounted for only 22% of matches in the TSW sample, where the largest number of matches were to CCMP1986 (also known as MED4), an ecotype that dominates the *Prochlorococcus* community in cooler, high latitude waters (Fig 4C) (Johnson *et al*., 2006). Similar patterns occurred when we compared our data with a 16S-ITS database to identify *Prochlorococcus* ecotypes in our samples. Consistent with the very low proportion of SEED hits to *Prochlorococcus* in the TSW sample, no *Prochlorococcus* 16S-ITS sequences were observed in the TSW sample. On the other hand, in the EAC sample, only sequences matching MIT9312 and MIT9301, a related strain that has previously been shown to exhibit distribution patterns that closely mirror MIT9312 (Hewson *et al*., 2009), were observed. In both samples the *Prochlorococcus* community was dominated by high-light adapted strains (Malmstrom *et al*., 2010), indicating that the patterns observed here were more compatible with latitudinal shifts in *Prochlorococcus* community composition (Johnson *et al*., 2006) than depth-related shifts (Malmstrom *et al*., 2010).

Similar patterns occurred amongst *Synechococcus* ecotypes, whereby 78% of matches to *Synechococcus* proteins in the EAC sample were associated with CC9605 (*Synechococcus* Clade II), an ecotype most frequently found in tropical/subtropical oligotrophic samples (Toledo and Palenik, 2003; Zwirglmaier *et al*., 2008). While this ecotype still dominated the TSW sample, it represented a significantly ($P < 10^{-15}$) smaller proportion of the *Synechococcus* community, with a more even distribution of ecotypes, including a higher proportion of matches to temperate-zone ecotypes, occurring in this sample

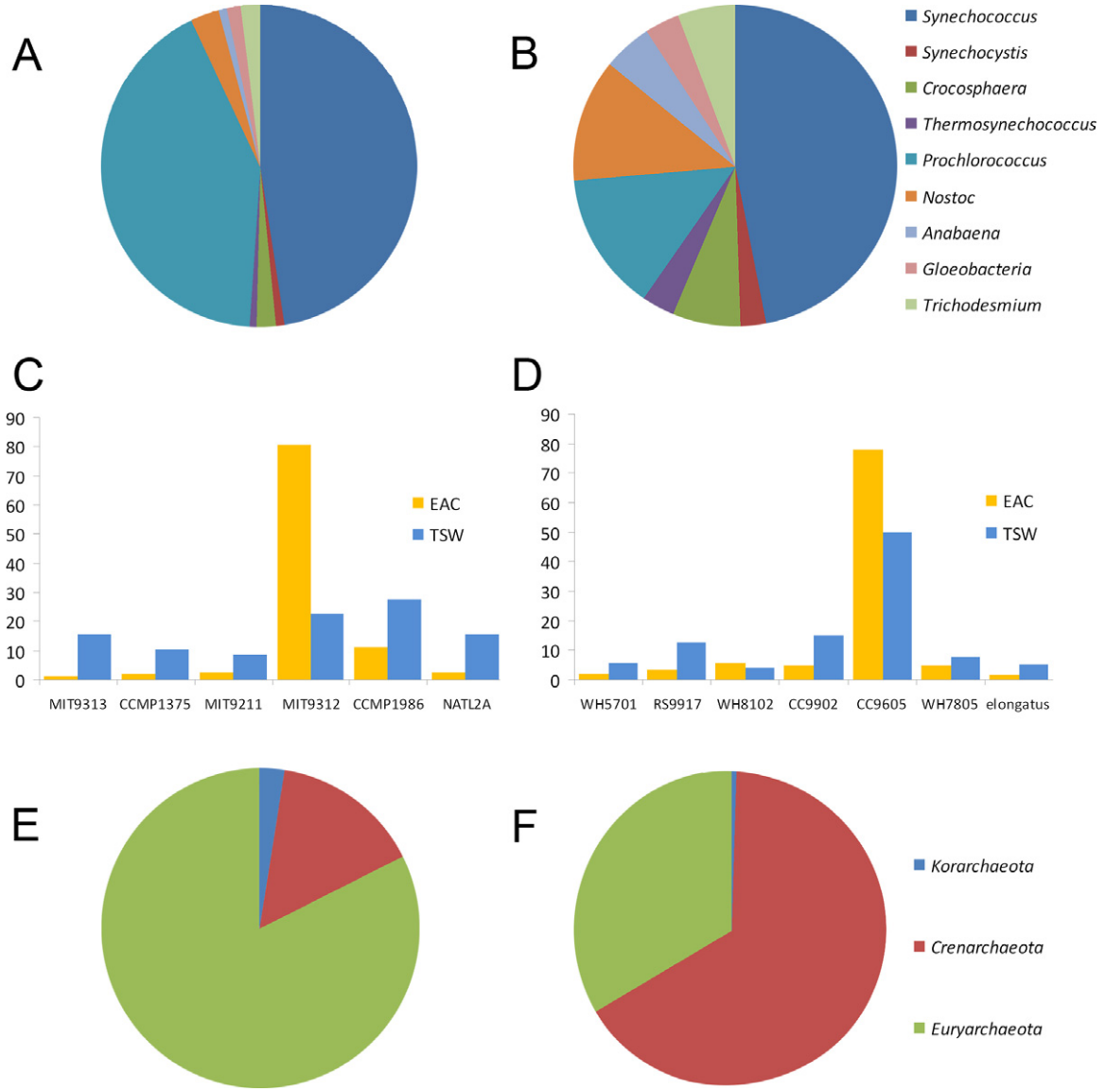


Fig. 4. Distribution of SEED matches ($e < 10^{-5}$) to; cyanobacterial genera in the (A) EAC and (B) TSW samples, major (C) *Prochlorococcus* and (D) *Synechococcus* ecotypes in EAC and TSW samples, and *Archaea* groups in the (E) EAC and (F) TSW samples.

(Fig 4D). These patterns in the distribution of *Synechococcus* and *Prochlorococcus* ecotypes indicate that the EAC may advect tropical microbial assemblages into the temperate latitudes of the Tasman Sea.

Shifts in the *Archaea* community paralleled those observed amongst the cyanobacteria. In the EAC sample, only 1.1% of sequences were matched to *Archaea*, compared with 4.2% in the TSW sample. The smaller EAC *Archaea* community was dominated by matches to the *Euryarchaeota* (82%), while the TSW was instead dominated by the *Crenarchaeota* (Fig. 2; Fig. 4E and F). The increase in matches to *Crenarchaeota* in the TSW sample was primarily driven by a significant ($P < 10^{-15}$) increase in matches to *Nitrosopumilus maritimus*, which made up 61% of total archaeal matches. Conversely, *N. maritimus* accounted for only 5% of *Archaea* matches in the EAC sample. *Nitrosopumilus maritimus* is an important ammonia oxidizing, carbon fixing archeon, which is widely distributed in surface ocean waters and plays a critical role in the ocean's nitrogen and carbon cycles (Könneke *et al.*, 2005; Walker *et al.*, 2010). However, it is important to note that matches to SEED are dependent upon comparisons to sequenced representative organisms, and *N. maritimus* is currently one of very few sequenced pelagic archaea. Consequently, the high percentage of matches to this organism may in part be an artefact of its dominance of the SEED database, rather than its numerical dominance of the *Crenarchaeota* community. None-

theless, the strong divergence in the relative importance of *Crenarchaeota* and *Euryarchaeota* between the EAC and TSW samples provides further evidence for fundamental differences between the two water masses.

Sequence reads were assigned to metabolic subsystem pathways using MG-RAST and a BLASTX *e*-value cut-off of 10^{-5}. In comparison to the phylogenetic shifts observed among key bacterial and archaeal groups, smaller differences in the distribution of matches to metabolic function categories occurred between the samples (Fig. S2). At the coarsest level of metabolic classification, 17 of 27 categories exhibited significant ($P < 0.05$) differences between samples, with amino acid synthesis and metabolism, motility and chemotaxis, clustering based subsystems and photosynthesis exhibiting the greatest divergence between samples. However, at the finer resolution, subsystem pathway level there was little quantitative difference between the samples, with only two categories (Electron transport and photophosphorylation, which was over-represented in the EAC, and Sarcosine oxidase, which was over-represented in the TSW sample), showing significant ($P < 0.05$) differences between samples. This apparent low degree of overall metabolic divergence is a recurrent phenomenon in marine metagenomes and has been explained by the dominance of a core suite of metabolic functions that remain relatively invariable due to their essential role in microbial physiology in ocean environments (Hewson *et al*., 2009). Despite differences in the nutrient status of the two water masses (Table 1), no significant differences in the relative importance of nitrogen, phosphorous or sulfur metabolism were observed between samples at any level. However, proteins involved in photosynthesis were significantly ($P < 0.05$) more prevalent in the EAC sample, where the relative proportion of matches to Phycoblisome and Photosystem I and II proteins were 3.6 times higher than in the TSW sample. This is consistent with the higher proportion of cyanobacteria in the EAC sample.

The microbial communities inhabiting the EAC and TSW water masses displayed several disparate features, indicative of marked ecological and biogeochemical divergence. These included fundamental shifts in the population structure and relative abundance of important ocean microbes, including SAR11, *Prochlorococcus* and *Crenarchaeota,* and is consistent with the premise that discrete oceanographic provinces host specific microbial assemblages (Hewson *et al*., 2009). Significantly, patterns in the distributions of key organisms indicate that the EAC microbial community is characterized by a somewhat 'tropical' signature, despite being sampled at virtually the same latitude (33.3°S) to the TSW community. We tested this hypothesis by comparing our data to a set of publically available metagenomic data (Fig. S3). While there was no clear demarcation of the data into two discrete tropical and temperate groups, it is notable that for both taxonomy and metabolism, the EAC sample was always most similar to metagenomes from the equatorial Pacific (latitudes: 7.5 and 8.6°N), while the TSW sample was generally most similar to samples obtained from comparable temperate latitudes [i.e. Botany Bay (33.9°S) and Monterey Bay (36.6°N)]. These trends, along with the observed patterns in cyanobacterial community organization, provide evidence for a poleward incursion of tropical microbes within the EAC, which mirrors the southerly spread of macrofauna and phytoplankton in this region (Ling *et al*., 2009; Thompson *et al*., 2009; Last *et al*., 2011). Increases in the southerly flow of the EAC have occurred at a rate of approximately 6 km $year^{-1}$ over the last 60 years, and are predicted to increase by a further 20% by 2100 (Ridgway and Hill, 2009). The different characteristics of our two metagenomes indicate that further increases in the influence of the EAC may substantially alter microbial community structure and biogeochemical dynamics in the Tasman Sea. A more extensive analysis of the microbial oceanography within this 'climate change hotspot' (Last *et al*., 2011) and in other regions where shifting circulation patterns are occurring (Bryden *et al*., 2005; Wu *et al*., 2012) may provide a preview of future patterns in ocean function.

Acknowledgements

We thank the captain and crew of the *RV Southern Surveyor* and the chief scientist on the voyage, I. Suthers. G. Tyson provided valuable advice on DNA recovery. This work was supported by Australian Research Council grants DP0772186 to J.R.S., DP1092892 to M.A.D., DP0988002 to M.V.B., DP0880078 to M.B. and I.S., and DP0988818 to I.S. SST data in Fig. 1A was provided by the Australian Bureau of Meteorology and the Integrated Marine Observing System (IMOS) – an initiative of the Australian Commonwealth Government's National Collaborative Research Infrastructure Strategy. We thank the MODIS software development and support team for processing MODIS Aqua Level-2 SST data used in Fig. 1B, and three anonymous reviewers for their helpful comments and suggestions on an earlier version of the manuscript.

References

Baird, M.E., Timko, P.G., Suthers, I.M., Middleton, J.H., Mullaney, T.J., and Cox, D.R. (2008) Biological properties across the Tasman Front off southeast Australia. *Deep Sea Res I* **55:** 1438–1455.

Brown, M.V., and Fuhrman, J.A. (2005) Marine bacterial microdiversity as revealed by internal transcribed spacer analysis. *Aquat Microb Ecol* **41:** 15–23.

Bryden, H.L., Longworth, H.R., and Cunningham, S.A. (2005) Slowing of the Atlantic meridional overturning circulation at 25°N. *Nature* **438:** 655–657.

Cai, W., Shi, G., Cowan, T., Bi, D., and Ribbe, J. (2005) The response of the southern annular mode, the East Australian Current and the southern mid-latitude ocean circulation to global warming. *Geophys Res Lett* **32:** L23706. doi: 10.1029/2005GL024701.

Carlson, C.A., Morris, R., Parsons, R., Treusch, A.H., Giovannoni, S.J., and Vergin, K. (2009) Seasonal dynamics of SAR11 populations in the euphotic and mesopelagic zones of the northwestern Sargasso Sea. *ISME J* **3:** 283–295.

Delmont, T.O., Malandain, C., Prestat, E., Larose, C., Monier, J.-M., Simonet, P., and Vogel, T. (2011) Metagenomic mining for microbiologists. *ISME J* **5:** 1837–1843.

Ducklow, H.W., Moran, X.A.G., and Murray, A.E. (2010) Bacteria in the greenhouse: marine microbes and climate change. In *Environmental Microbiology*, 2nd edn, Mitchell, R., and Gu, J.-D. (eds). Hoboken, NJ, USA: John Wiley & Sons, Inc., pp. 1–32.

Edwards, J.L., Smith, D.L., Conolly, J., McDonald, J.E., Cox, M.J., Joint, I., *et al*. (2010) Identification of carbohydrate metabolism genes in the metagenome of a marine biofilm community shown to be dominated by *Gammaproteobacteria* and *Bacteroidetes*. *Genes* **1:** 371–384.

Hallegraeff, G., Beardall, J., Brett, S., Doblin, M., Hosja, W., de Salas, M., and Thompson, P. (2009) Phytoplankton. In *A Marine Climate Change Impacts and Adaptation Report Card for Australia 2009*. Poloczanska, E.S., Hobday, A.J., and Richardson, A.J. (eds). Australia: National Climate Change Adaptation Research Facility, pp. 1–10. National Climate Change Adaptation Research Facility, Australia Publication 05/09, ISBN 978-1-921609-03-9.

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

Hill, K.L., Rintoul, S.R., Coleman, R., and Ridgway, K.R. (2008) Wind-forced low frequency variability of the East Australian Current. *Geophys Res Lett* **35:** L08602. doi: 10.1029/2007GL032912.

Huson, D.H., Auch, A.F., Qi, J., and Schuster, S.C. (2007) MEGAN analysis of metagenomic data. *Genome Res* **17:** 377–386.

Jeffries, T., Seymour, J., Dinsdale, E., Newton, K., Leterme, S., Gilbert, J., *et al*. (2011) Substrate type determines metagenomic profiles from diverse chemical habitats. *PLoS ONE* **6:** e25173.

Johnson, Z.I., Zinser, E.R., Coe, A., McNulty, N.P., Woodward, E.M.S., and Chisholm, S.W. (2006) Niche partitioning among *Prochlorococcus* ecotypes along ocean-scale environmental gradients. *Science* **311:** 1737–1740.

Könneke, M., Bernhard, A.E., Torre, J.R., Walker, C.B., Waterbury, J.B., and Stahl, D.A. (2005) Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature* **437:** 543–546.

Last, P.R., White, W.T., Gledhill, D.C., Hobday, A.J., Brown, R., Edgar, G.J., and Pecl, G. (2011) Long-term shifts in abundance and distribution of a temperate fish fauna: a response to climate change and fishing practices. *Glob Ecol Biogeogr* **20:** 58–72.

Ling, S.D., Johnson, C.R., Frusher, S.D., and Ridgway, K.R. (2009) Overfishing reduces resilience of kelp beds to climate-driven catastrophic phase shift. *Proc Natl Acad Sci USA* **106:** 22341–22345.

Malmstrom, R.R., Coe, A., Kettler, G.G., Martiny, A.C., Frias-lopez, J., Zinser, E.R., and Chisholm, S.W. (2010) Temporal dynamics of *Prochlorococcus* ecotypes in the Atlantic and Pacific Oceans. *ISME J* **4:** 1252–1264.

Meyer, D., Paarmann, D., D'Souza, M., Olson, R., Glass, E.M., Kubal, M., *et al*. (2008) The Metagenomics RAST server – a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinform* **9:** 386.

Morris, R.M., Rappe, M.S., Connon, S.A., Vergin, K.L., Siebold, W.A., Carlson, C.A., and Giovannoni, S.J. (2002) SAR11 clade dominates ocean surface bacterioplankton communities. *Nature* **420:** 806–810.

Overbeek, R., Begley, T., Butler, R.M., Choudhuri, J.V., Chuang, H.Y., Cohoon, M., *et al*. (2005) The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acid Res* **33:** 5691–5702.

Parks, D.H., and Beiko, R.G. (2010) Identifying biologically relevant differences between metagenomic communities. *Bioinform* **26:** 715–721.

Partensky, F., Blanchot, J., and Vaulot, D. (1999a) Differential distribution and ecology of *Prochlorococcus* and *Synechococcus* in oceanic waters: a review. In *Marine Cyanobacteria*. No. spécial 19. Charpy, L., and Larkum, A.W.D. (eds). Monaco: Bulletin de l'Institut oceanographique, pp. 457–475.

Partensky, F., Hess, W.R., and Vaulot, D. (1999b) *Prochlorococcus*, a marine photosynthetic prokaryote of global significance. *Microbiol Mol Biol Rev* **63:** 106–127.

Pfister, C.A., Meyer, F., and Antonopoulos, D.A. (2010) Metagenomic profiling of a microbial assemblage associated with the California mussel: a node in networks of carbon and nitrogen cycling. *PLoS ONE* **5:** e10518.

Poroyko, V., White, J.R., Wang, M., Donovan, S., Alverdy, J., Liu, D.C., and Morowitz, M.J. (2010) Gut microbial gene expression in mother-fed and formula-fed piglets. *PLoS ONE* **5:** e12459.

Pruesse, E., Quast, C., Knittel, K., Fuchs, B., Ludwig, W., Peplies, J., and Glöckner, F.O. (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res* **35:** 7188–7196.

Rappé, M.S., Connon, S.A., Vergin, K.L.l, and Giovannoni, S.J. (2002) Cultivation of the ubiquitous SARII marine bacterioplankton clade. *Nature* **418:** 630–633.

Ridgway, K.R. (2007) Long-term trend and decadal variability of the southward penetration of the East Australian current. *Geophys Res Lett* **34:** L13612. doi: 10.1029/2007GL030393.

Ridgway, K.R., and Hill, K. (2009) The East Australian Current. In *A Marine Climate Change Impacts and Adaptation Report Card for Australia 2009*. Poloczanska, E.S., Hobday, A.J., and Richardson, A.J. (eds). Australia: National Climate Change Adaptation Research Facility, pp. 1–16. National Climate Change Adaptation Research Facility, Australia Publication 05/09, ISBN 978-1-921609-03-9.

Smith, R.J., Jeffries, T.C., Roudnew, B., Fitch, A.J., Seymour, J.R., Delpin, M.W., *et al*. (2012) Metagenomic comparison

of microbial communities inhabiting confined and unconfined aquifer ecosystems. *Environ Microbiol* **14:** 240–253.
Stingl, U., Tripp, H.J., and Giovannoni, S.J. (2007) Improvements of high-throughput culturing yielded novel SAR11 strains and other abundant marine bacteria from the Oregon coast and the Bermuda Atlantic Time Series study site. *ISME J* **1:** 361–371.
Sun, S., Chen, J., Li, W., Altintas, I., Lin, A., Peltier, S., *et al*. (2011) Community cyberinfrastructure for Advanced Microbial Ecology Research and Analysis: the CAMERA resource. *Nucleic Acid Res* **39:** D546–D551. doi: 10.1093/nar/gkq1102.
Thompson, P.A., Baird, M.E., Ingleton, T., and Doblin, M.A. (2009) Long-term changes in temperate Australian coastal waters: implications for phytoplankton. *Mar Ecol Prog Ser* **394:** 1–19.
Toledo, G., and Palenik, B. (2003) A *Synechococcus* serotype is found preferentially in surface marine waters. *Limnol Oceanogr* **48:** 1744–1755.
Vega Thurber, R., Willner-Hall, D., Rodriquez-Mueller, B., Desnues, C., Edwards, R.A., Angly, F., *et al*. (2009) Metagenomic analysis of stressed coral holobionts. *Environ Microbiol* **11:** 2148–2163.
Walker, C.B., de la Torre, J.R., Klotz, M.G., Urakaw, H., Pinel, N., Arp, D.J., *et al*. (2010) *Nitrosopumilus maritimus* genome reveals unique mechanisms for nitrification and autotrophy in globally distributed marine crenarchaea. *Proc Natl Acad Sci USA* **107:** 8818–8823.
Wu, L., Cai, W., Zhang, L., Nakamura, H., Timmermann, A., Joyce, T., *et al*. (2012) Enhanced warming over the global subtropical western boundary currents. *Nat Clim Change* **2:** 161–166.
Zwirglmaier, K., Jardillier, L., Ostrowski, M., Mazard, S., Garczarek, L., Vaulot, D., *et al*. (2008) Global phylogeography of marine *Synechococcus* and *Prochlorococcus* reveals distinct partitioning of lineages among oceanic biomes. *Environ Microbiol* **10:** 147–161.

Supporting information

Additional Supporting Information may be found in the online version of this article:

Fig. S1. Top 10 identified bacterial genus-level matches determined using a BLASTX search ($E < 10^{-5}$) against the NCBI NR database using CAMERA (Sun *et al*., 2011) and mapped against the NCBI taxonomy within the MEGAN software package (Huson *et al*., 2007) (ordered according to frequency of occurrence in the EAC sample).

Fig. S2. STAMP analysis showing relative importance of broad metabolic categories (SEED, Level 1) in TSW and EAC samples. Corrected *P*-values (*q*-values) were calculated using Storey's FDR approach (Parks and Beiko, 2010). Groups over-represented in the EAC community (orange) correspond to negative differences between proportions. Groups over-represented in the TSW community (blue) correspond to positive differences between proportions.

Fig. S3. Hierarchical clustering of metagenomic profiles at (A) class, (B) genus and (C) functional level. Dendrograms represent group average clustering of the Bray–Curtis similarity between profiles. Abundance profiles were generated using the SEED database (Overbeek *et al*., 2005, $E < 10^{-5}$) in MG-RAST (Meyer *et al*., 2008) and normalized abundance data was exported into PRIMER-E (Clarke and Gorley, 2006) for analysis using the CLUSTER algorithm (Clarke, 1993). Normalization was based on a log transformation and data centring as per the MG-RAST standard protocol (http://blog.metagenomics.anl.gov/howto/mg-rast-analysis-tools/). Metagenomes representative of tropical and temperate ocean surface habitats (red and blue symbols respectively) were chosen for the analysis based on their geographic location (Tropical = < 23° latitude) and consisted of greater than 1000 hits. Datasets were as follows: GOS Temperate ('Global Ocean Survey', Rusch *et al*. 2007; MG-RAST ID 4441143.3, 4441144.3, 4441570.3, 4441573.3,4441574.3, 4441575.3, 4441578.3, 4441579.3, 4441583.3, 4441585.3, 4441659.3), GOS Tropical ('Global Ocean Survey', Rusch *et al*. 2007; 4441145.3, 4441146.3, 4441594.3, 4441603.3, 4441605.3), Equatorial Pacific ('Marine Bacterioplankton Metagenomes'; 4443766.3, 4443695.3, 4443697.3, 4443698.3, 4443699.3, 4443700.3, 4443701.3), Study ('EAC/TSW'; 4446407.3, 4446457.3), Monterey Bay ('Monterey Bay Microbial Study'; 4443713.3, 4443712.3, 4443714.3, 4443715.3, 4443716.3, 4443717.3), Botany Bay ('Botany Bay Metagenomes'; 4443688.3, 4443689.3), HOT/ALOHA ('Microbial Community Genomics at the HOT/ALOHA' De Long *et al*. 2006; 4441051.3, 4441057.4), All data are publically available on MG-RAST (Meyer *et al*., 2008; http://metagenomics.anl.gov/metagenomics.cgi?page=Home; accessed 21/3/12)

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

Biogeosciences, 9, 815–825, 2012
www.biogeosciences.net/9/815/2012/
doi:10.5194/bg-9-815-2012

Biogeosciences

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

T. C. Jeffries[1,2], J. R. Seymour[2], K. Newton[1], R. J. Smith[1], L. Seuront[1,3,4], and J. G. Mitchell[1]

[1]School of Biological Sciences, Flinders University, Adelaide, South Australia 5001, Australia
[2]Plant Functional Biology and Climate Change Cluster, University of Technology Sydney, Australia
[3]Aquatic Sciences, South Australian Research and Development Institute, Henley Beach 5022, Australia
[4]Centre National de la Recherche Scientifique, Paris cedex 16, France

Correspondence to: T. C. Jeffries (jeffries.thomas@gmail.com)

Received: 14 July 2011 – Published in Biogeosciences Discuss.: 28 July 2011
Revised: 30 January 2012 – Accepted: 31 January 2012 – Published: 20 February 2012

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 salinity (Practical Salinity Scale) 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 150 km long, 2 km 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, 2010; Pollet 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 across stations, expressed using the Practical Salinity Scale (Lewis, 1980). The sites were named by their salinity and defined by their GPS coordinates, which were as follows: 37 ($-35.551°$ S, 138.883° E), 109 ($-35.797°$ S, 139.317° E), 132 (-35.938S, 139.488° E) & 136 ($-36.166°$ S, 139.651° E). Ammonia concentrations at these sites ranged between 0.21 ($\pm$0.09) and 3.10 ($\pm$0.84) mgN l^{-1}, phosphate concentrations ranged between 0.05 ($\pm$0.01) and 0.27 ($\pm$0.09) mgP l^{-1} (Supplement Fig. S1). Heterotrophic bacteria and virus like particles in porewater, as determined by flow cytometry (Marie et al., 1995; Seymour et al., 2005), increased from 4.8×10^6 ($\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 (Supplement Fig. S1).

At each site, 10 g of sediment, submerged in approximately 2 m 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 7 cm of the core was dark grey and black mud overlaid by approximately 3 cm 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 h of collection.

2.2 DNA extraction and sequencing

DNA was extracted from 10 g 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 68 888 DNA sequences in the 37 salinity metagenome, 101 003 sequences in the 109 salinity metagenome, 114 335 sequences in the 132 salinity metagenome and 108 257 sequences in the 136 salinity metagenome, with an average read length of 232 bp.

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 50 bp 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. Fisher's exact test uses a hypergeometric distribution of sequences drawn without replacement from a pair of metagenomic samples to generate a statistical significance value in a computationally efficient manner (Parks and Beiko, 2010) and is thus ideal for the pairwise comparison of metagenomes. Fisher's exact test has been routinely applied to observe statistically significant differences between single metagenomic profiles (e.g. Lamendella et al., 2011; McCarthy et al., 2011; Biddle et al., 2011). 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 lower salinity (37) 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 marine (37) 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 and c) 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 salinity metagenome.

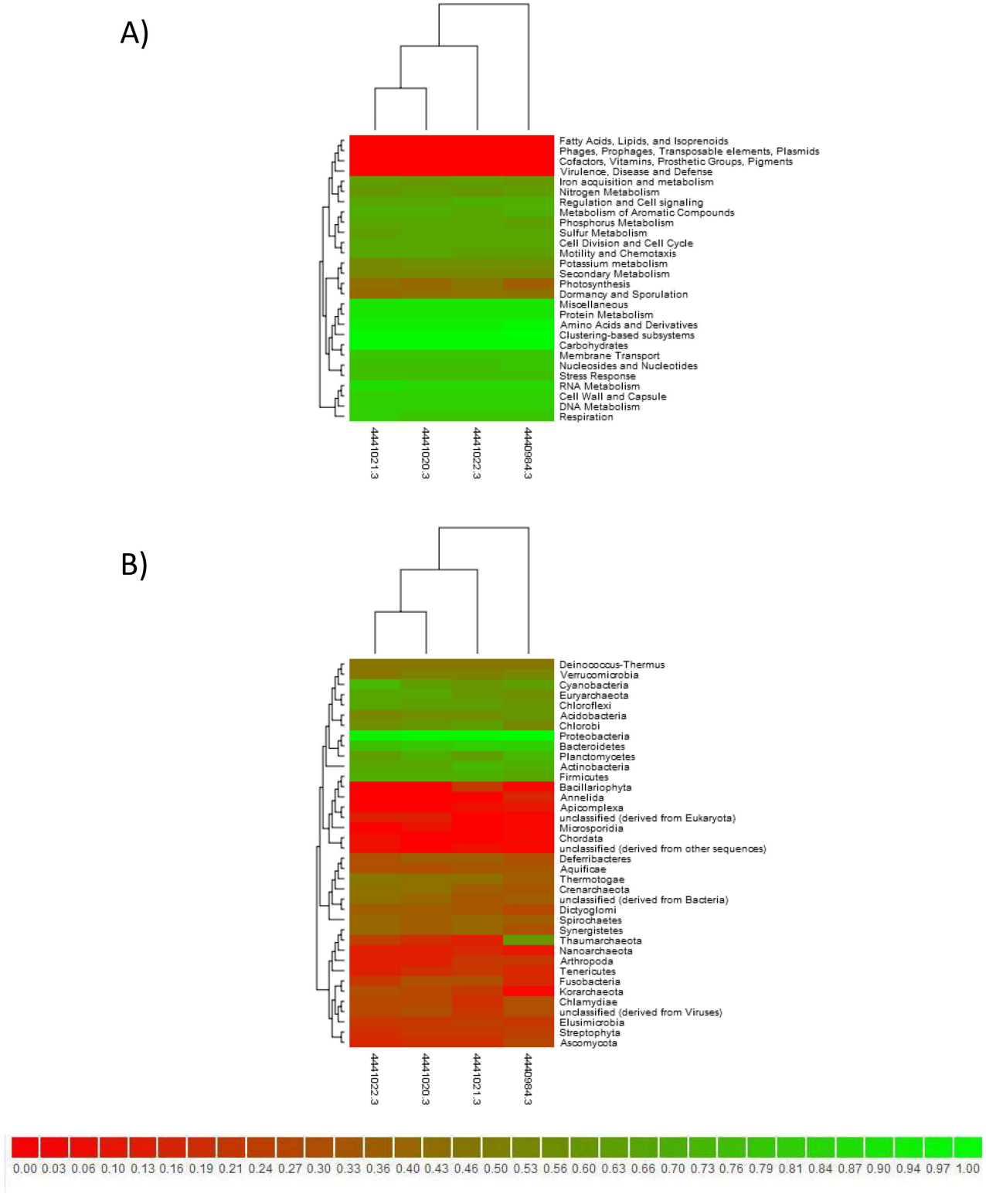


Fig. 1. Disimilarity between metagenomic profiles. **(A)** Functional potential **(B)** Taxonomic composition. 4440984.3 = 37, 4441020.3 = 109, 4441021.3 = 132, 4441022.3 = 136. Colour gradient represents proportion of sequences.

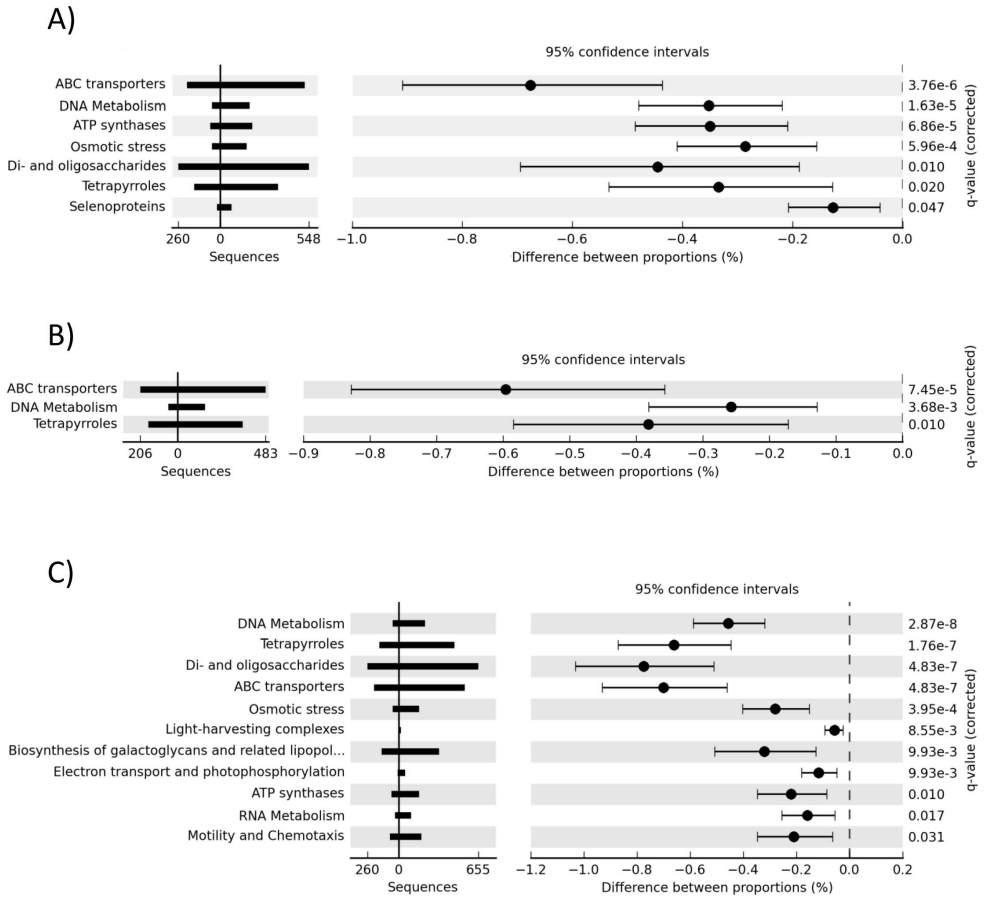


Fig. 2. Metabolic processes over-represented in hypersaline metagenomes relative to the 37 metagenome. **(A)** 109 **(B)** 132 **(C)** 136. Corrected P-values were calculated using Storey's FDR approach.

Sequences related to photosynthesis and pigment synthesis were over-represented in all hypersaline metagenomes relative to the 37 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 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 metagenome and the hypersaline metagenomes (Fig. 4, Supplement Fig. S2). The cyanobacterial classes Nostocales, Oscillatoriales and Chroococcales were found to be over-represented in the most hypersaline metagenome (136) relative to the lower salinity 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, 132 and 136 metagenomes relative to the 37 sample. Of these, the class Methanomicrobia was the most over-represented in all cases. The halophilic class Halobacteria were over-represented in the 136 and 109 metagenomes showing the highest increase in proportion in the most hypersaline metagenome (136) (Fig. 4; Supplement Fig. S2).

We also observed shifts in the structure of the Proteobacteria. The class Delta/Epsilon-Proteobacteria were over-represented in hypersaline metagenomes, while the relative abundance of Gammaproteobacteria, Betaproteobactera and Alphaproteobacteria were significantly higher in the lower salinity metagenome. The classes Bacteroidetes and Plactomycetacia were also strongly over-represented in the lower salinity 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

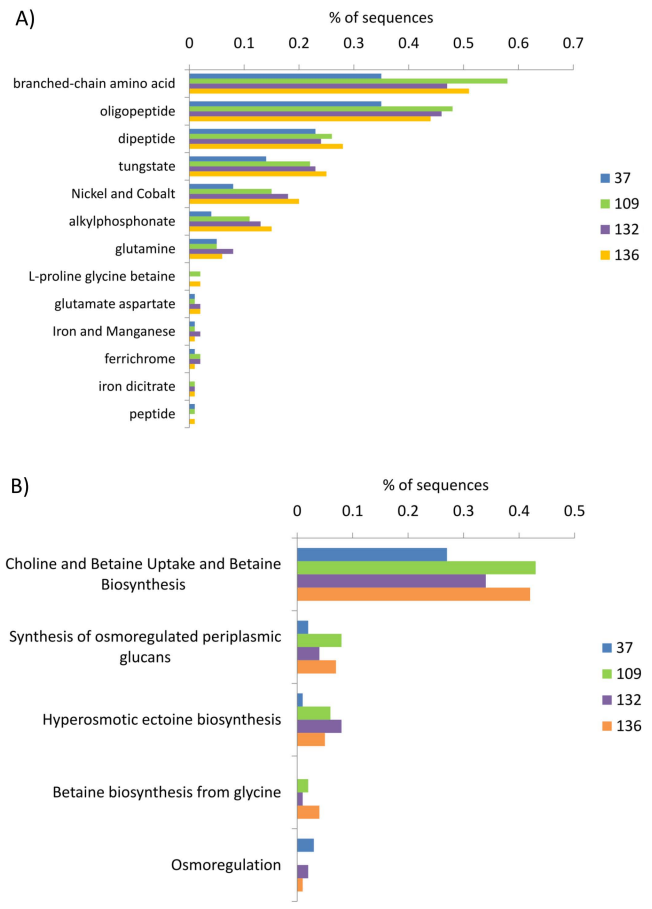


Fig. 3. Breakdown of subsystem contribution to **(A)** ABC transporter and **(B)** osmotic stress categories. Subsystems are the third level of organization within the MG-RAST hierarchy.

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 salinity metagenome relative to the 37 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).

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

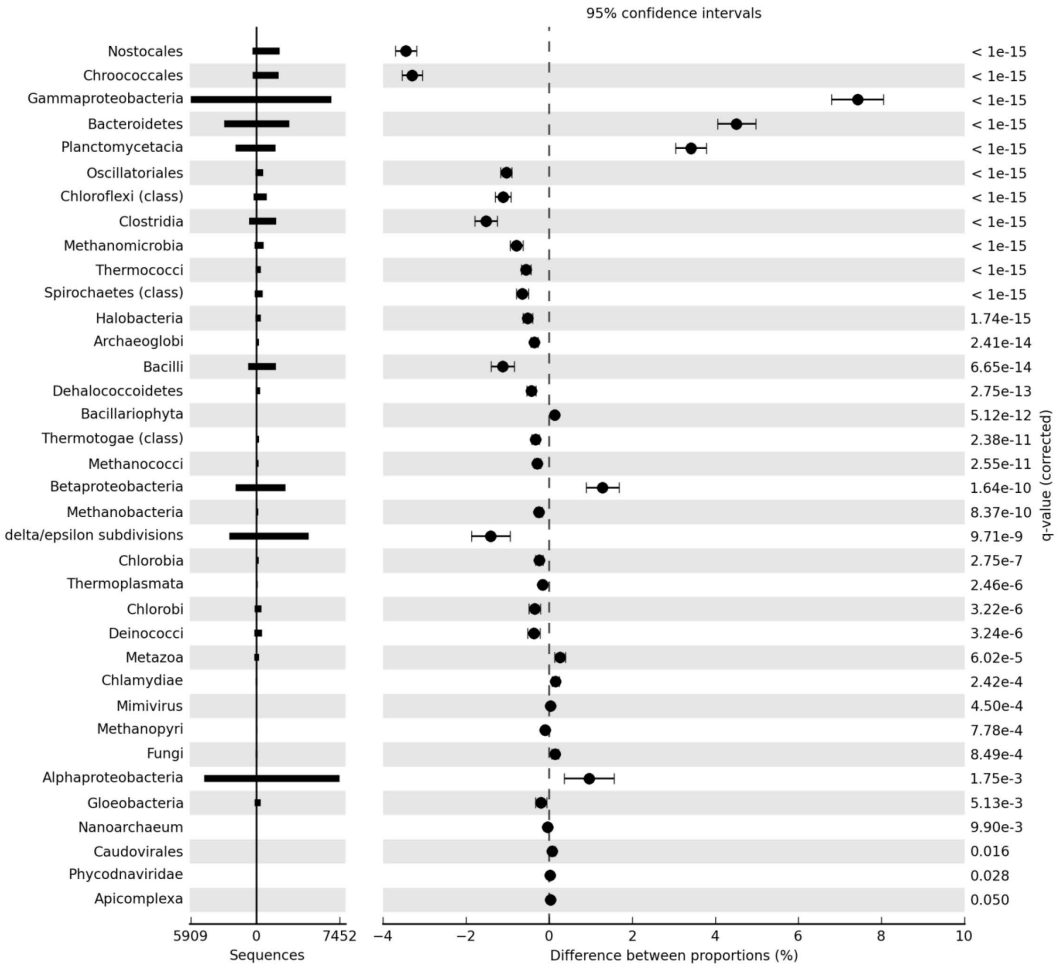


Fig. 4. Taxa enriched or depleted in the 37 and 136 metagenomes. Corrected P-values were calculated using Storey's FDR approach. Taxa enriched in the marine (37) metagenomes have positive differences between proportions.

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). Members of the class Methanosarcinales, which is over-represented in the most hypersaline metagenome, contain the order Methanosarcinales which metabolize methylated amines, formed from the

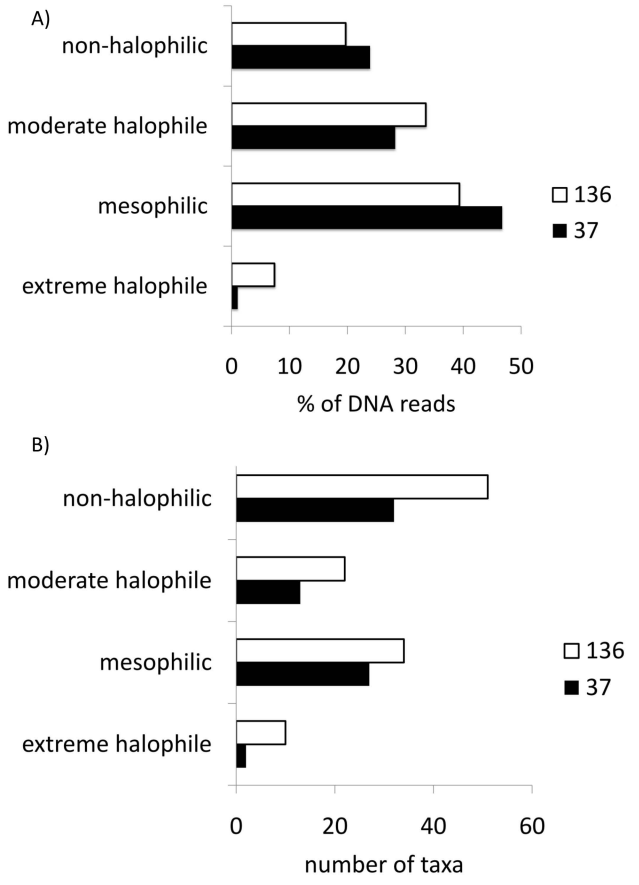


Fig. 5. Representation of halophilic taxa in the 37 and 136 metagenomes. **(A)** % DNA reads matching taxa with a defined salinity tolerance **(B)** number of taxa with a defined salinity tolerance.

breakdown of the osmoprotectant glycine betaine (Kendall and Boone, 2006), potentially influencing the rates of methane flux in the sediment. 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.

The presence of sequences matching the Halobacteriacea further indicates an adapted halophilic community as these Archaea tend to be found at the highest salinities and generally use a "high-salt in" strategy (Oren, 2008), which suggests that this mode of salt adaptation is also present in our samples. Sequences and isolates matching the Halobacteriales have been observed in the hypersaline microbial mats of Shark Bay, Australia (Goh et al., 2006; Burns et al., 2004; Allen et al., 2008) indicating that they are an important constituent of benthic microbial communities that exist at more moderate hypersaline conditions as well as at the extremes of salt saturation.

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 most saline 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 $g\,l^{-1}$ salinity (Oren, 2009). The Cyanobacteria over-represented in our most hypersaline metagenome represent filamentous Cyanobacteria. Many taxa comprising mats are filamentous (Oren, 2002, 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. The occurrence of abundant Nostocales sequences in our metagenomes is unexpected as while Nostocales have also been observed in saline Antarctic lakes (Jungblut and Neilan, 2010) and in the hypersaline stromatolites of Shark Bay, Australia (Burns et al., 2004), there are indeed no reports that we could find of high abundance Nostocales in samples as hypersaline as ours. Our data could either describe a novel group of halophilic heteroscystous Nostocales, which would require further microscopic analysis and detailed molecular taxonomic classification to confirm, or potentially our Nostocales sequences are derived from other filamentous cyanobacterial taxa which are closely related to Nostocales, for which there are no examples in the SEED database used for taxonomic classification.

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 (Supplement 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., 2012) 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). Another interesting trend in our data is the higher level of ammonium in hypersaline sediments compared to the lower salinity (37) sediment (Supplement Fig. S1). This is potentially related to the potential negative influence of high salt concentrations on nitrifying bacteria, such as the genera Nitrosomonas and Nitrosococcus, which are both present in our data further suggesting that salinity may be influencing other taxonomic groups relevant to the nitrogen cycle.

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).

5 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.

Supplementary material related to this article is available online at: http://www.biogeosciences.net/9/815/2012/bg-9-815-2012-supplement.pdf.

Acknowledgements. This work was supported by the Australian Research Council. We thank Aharon Oren and an anonymous reviewer for their comments on this manuscript during open discussion.

Edited by: G. Herndl

References

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

Allen, M. A., Goh, F., Leuko, S., Echigo, A., Mizuki, T., Usami, R., Kamekura, M., Neilan, B. A., and Burns, B. P.: Haloferax elongans sp. nov. and Haloferax mucosum sp. nov., isolated from microbial mats from Hamelin Pool, Shark Bay, Australia, International Journal of Systematic and Evolutionary Microbiology, 58, 798–802, 2008.

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, doi:10.1111/j.1462-2920.2005.00808.x, 2005.

Biddle, J. F., White, J. R., Teske, A. P., and House, C. H.: Metagenomics of the subsurface Brazos-Trinity Basin (IODP site 1320): comparison with other sediment and pyrosequenced metagenomes, ISME J., 5, 1038–1047, 2011.

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.

Burns, B. P., Goh, F., Allen, M., and Neilan, B. A.: Microbial diversity of extant stromatolites in the hypersaline marine environment of Shark Bay, Australia, Environmental Microbiology, 6, 1096–1101, 2004.

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, doi: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, doi:10.1371/journal.pone.0001584, 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, doi: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.

Goh, F., Leuko, S., Allen, M. A., Bowman, J. P., Kamekura, M., Neilan, B. A., and Burns, B. P.: Halococcus hamelinensis sp. nov., a novel halophilic archaeon isolated from stromatolites in Shark Bay, Australia, International Journal of Systematic and Evolutionary Microbiology, 56, 1323–1329, doi:10.1099/ijs.0.64180-0, 2006.

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, doi: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, doi: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, doi: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], doi: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 pp., 1989.

Jungblut, A. D. and Neilan, B. A.: Cyanobacterial Mats of the Meltwater Ponds on the McMurdo Ice Shelf (Antarctica) Microbial Mats, in: Cellular Origin, Life in Extreme Habitats and Astrobiology, edited by: Seckbach, J. and Oren, A., Springer Netherlands, 499–514, 2010.

Kendall, M. and Boone, D.: The Order Methanosarcinales, in: The Prokaryotes, edited by: Dworkin, M., Falkow, S., Rosenberg, E., Schleifer, K.-H., and Stackebrandt, E., Springer New York, 244–256, 2006.

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, doi: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, doi:10.1038/msb.2008.35, 2008.

Lamendella, R., Santo Domingo, J., Ghosh, S., Martinson, J., and Oerther, D.: Comparative fecal metagenomics unveils unique functional capacity of the swine gut, BMC Microbiology, 11, 103, doi:10.1186/1471-2180-11-103, 2011.

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, doi:10.1016/j.ecss.2009.04.018, 2009.

Lewis, E.: The practical salinity scale 1978 and its antecedents, Oceanic Engineering, IEEE Journal of, 5, 3–8, 1980.

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

McCarthy, C. B., Diambra, L. A., and Rivera Pomar, R. V.: Metagenomic Analysis of Taxa Associated with Lutzomyia longipalpis, Vector of Visceral Leishmaniasis, Using an Unbiased High-Throughput Approach, PLoS Negl. Trop. Diseases, 5, e1304, 2011.

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, 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, available at: 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, 2000.

Oren, A.: Microbial life at high salt concentrations: phylogenetic and metabolic diversity, Saline Systems, 4, 2, doi: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, doi: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 Ecol., 31, 225–247, 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, doi: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, doi: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, doi:10.1186/1746-1448-6-5, 2010.

Roberts, M. F.: Organic compatible solutes of halotolerant and halophilic microorganisms, Saline Systems, 1, doi: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, available at: 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, 0398–0431, 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, doi: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, Mar. Ecol.-Prog. Ser., 251, 181–189, doi: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, doi: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, J. Biol. Chem., 286, 10169–10176, doi:10.1074/jbc.M110.208892, 2011.

Storey, J. D. and Tibshirani, R.: Statistical significance for genomewide studies, P. Natl. Acad. Sci. USA, 100, 9440–9445, doi: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, Nucl. Acids Res., 39, D546–D551, doi: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, doi: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, doi:10.1111/j.1574-6976.2000.tb00542.x, 2000.

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

Yale, J. and Bohnert, H. J.: Transcript Expression in Saccharomyces cerevisiae at High Salinity, J. Biol. Chem., 276, 15996–16007, doi:10.1074/jbc.M008209200, 2001.

environmental microbiology

Environmental Microbiology (2012) **14**(1), 240–253 doi:10.1111/j.1462-2920.2011.02614.x

Metagenomic comparison of microbial communities inhabiting confined and unconfined aquifer ecosystems

Renee J. Smith,[1]* Thomas C. Jeffries,[1] Ben Roudnew,[1] Alison J. Fitch,[1] Justin R. Seymour,[2] Marina W. Delpin,[1] Kelly Newton,[1] Melissa H. Brown[1] and James G. Mitchell[1]

[1]*School of Biological Sciences, Flinders University, Adelaide, SA, 5001, Australia.*

[2]*Plant Functional Biology and Climate Change Cluster, University of Technology Sydney, Sydney, NSW 2007, Australia.*

Summary

A metagenomic analysis of two aquifer systems located under a dairy farming region was performed to examine to what extent the composition and function of microbial communities varies between confined and surface-influenced unconfined groundwater ecosystems. A fundamental shift in taxa was seen with an overrepresentation of *Rhodospirillales*, *Rhodocyclales*, *Chlorobia* and *Circovirus* in the unconfined aquifer, while *Deltaproteobacteria* and *Clostridiales* were overrepresented in the confined aquifer. A relative overrepresentation of metabolic processes including antibiotic resistance (β-lactamase genes), lactose and glucose utilization and DNA replication were observed in the unconfined aquifer, while flagella production, phosphate metabolism and starch uptake pathways were all overrepresented in the confined aquifer. These differences were likely driven by differences in the nutrient status and extent of exposure to contaminants of the two groundwater systems. However, when compared with freshwater, ocean, sediment and animal gut metagenomes, the unconfined and confined aquifers were taxonomically and metabolically more similar to each other than to any other environment. This suggests that intrinsic features of groundwater ecosystems, including low oxygen levels and a lack of sunlight, have provided specific niches for evolution to create unique microbial communities. Obtaining a broader understanding of the structure and function of microbial communities inhabiting different groundwater systems is particularly important given the increased need for managing groundwater reserves of potable water.

Received 6 June, 2011; revised 17 July, 2011; accepted 12 September, 2011. *For correspondence. E-mail renee.smith@flinders.edu.au; Tel. (+61) 8 8201 3490; Fax (+61) 8 8201 3015.

Introduction

Terrestrial subsurface environments, including groundwater, accommodate the largest reservoir of microbes in the biosphere, with estimates of bacterial abundances reaching 3.8–6.0 $\times$ 10^{30} cells (Whitman *et al.*, 1998). Due to the lack of sunlight and input of nutrients and energy from external sources, these microbial communities are largely responsible for the turnover of energy and matter, forming the basis of subterranean food webs (Sherr and Sherr, 1991). These communities also influence the purity of groundwater and subsequent availability of potable drinking water (Danielopol *et al.*, 2003).

Holding more than 97% of the world's freshwater reserves, aquifers are a largely untapped resource of potable drinking water, but also harbour a high diversity of microbes (Gibert and Deharveng, 2002). These reserves are becoming increasingly important (Bond *et al.*, 2008) in countries such as Australia, which are susceptible to drought events (Mpelasoka *et al.*, 2008). However, the nature of the microbial communities inhabiting aquifers remains largely unexplored. To effectively understand and maintain groundwater reserves it is important to investigate the identity and biogeochemical function of the microbes within aquifer systems.

Aquifer systems, defined by a permeable zone below the earth's surface through which groundwater moves (Hamblin and Christiansen, 2004), are generally classified into two major types; unconfined and confined aquifers. 'Unconfined aquifers' are connected to the surface via open pore space and thus can receive external input from the surrounding area. They are sensitive to precipitation via seepage through the soil, and are directly affected by human impact (Al-Zabet, 2002). 'Confined aquifers' occur at greater depth and lie below an impermeable strata layer. The thick confining strata layer ensures that there is no input from the overlaying surface environment. Input to confined aquifers occurs only from distant recharge sources and due to slow flow rates, can be isolated for hundreds to thousands of years (Gibert and Deharveng, 2002). Microbes inhabiting these systems must be capable of surviving with limited resources, as external

inputs of nutrients and oxygen are not readily available (Pedersen, 2000; Griebler and Lueders, 2009). Survival strategies to cope in this environment include increased affinity to limiting nutrients and reduced metabolic rates and growth (Teixeira de Mattos and Neijssel, 1997; Brune *et al*., 2000).

Sporadic changes in limiting resources in these groundwater systems, driven by external input, can lead to major shifts in the taxonomy and the metabolism of microbial communities (Hemme *et al*., 2010). The sensitivity of microbes to environmental change allows them to be used as bioindicators (Avidano *et al*., 2005; Steube *et al*., 2009). A major goal in the study of groundwater microbiology is to determine what the effects of these shifts in microbial ecology have on water quality (Langworthy *et al*., 1998; Hemme *et al*., 2010).

The concentration of chemical contaminants and pathogens in groundwater systems is influenced by the biogeochemical and ecological dynamics of subterranean microbial communities (Hemme *et al*., 2010). Shifts in microbial taxonomy resulting from pollution in groundwater have been investigated (Männistö *et al*., 1999; Chang *et al*., 2001) but the effects of introduced contaminants on the metabolic potential of groundwater microbes are only vaguely understood. Previous groundwater studies have shown that microbes respond to external contaminants at both the phenotypic and genotypic level, with changes in microbial community structure, as well as an increase in the number of genes responsible for the degradation of introduced contaminants (Langworthy *et al*., 1998). Furthermore, Hemme and colleagues (2010) showed that introduced contaminants into groundwater systems can decrease species and allelic diversity and eliminate some metabolic pathways. Evolutionary analysis of a microbial community in groundwater contaminated with heavy metals has shown that lateral gene transfer could play a key role in the rapid response and adaptation to environmental contamination (Hemme *et al*., 2010). Hence, to obtain a complete description of the effect of external influences on groundwater systems, both the taxonomy and the metabolic potential of microbial communities need to be studied.

The effect of agricultural modification on groundwater is less well characterized; however, it has been shown that introduced manure from a livestock farm caused the microbial composition of previously uncontaminated groundwater to taxonomically resemble livestock wastewater (Cho and Kim, 2000). This study used 16S rDNA technology that is limited to prokaryote taxonomy and discounts viruses and eukaryotes. Advances in metagenomic studies have allowed for the direct sequencing of whole environmental microbial genomes (Kennedy *et al*., 2010) and have greatly increased our knowledge of gene function, metabolic processes, community structure and ecosystems response to environmental change. Previous metagenomic studies have revealed clear shifts in the structure of microbial assemblages related to human impact (Dinsdale *et al*., 2008a).

With this in mind, the aim of the present study is to compare an unconfined and a confined groundwater system using metagenomics approaches, and provide insight into the endemic taxonomy and metabolic processes of the resident microbial communities, and how these may be affected by introduced contaminants.

Results

Overview of the biogeochemical environment and microbial enumeration

The unconfined and confined aquifers were characterized by low oxygen levels of 0.2 mg l^{-1} and 0.26 mg l^{-1} respectively. Iron, sulfur and total organic carbon were all significantly higher ($P < 0.05$) in the unconfined aquifer than the confined aquifer. All other nutrients were not statistically different between samples. Salinity and pH were higher in the unconfined aquifer, while temperature was lower. Microbial cell counts were similar in the unconfined and confined aquifers (Table 1).

Taxonomic and metabolic profiling of groundwater metagenomes

A total of 64 506 and 409 743 sequences with an average read length of 386 and 387 bases were obtained from the unconfined and confined aquifer samples respectively. Both metagenomic libraries were dominated by bacteria (82% of hits to SEED) (http://metagenomics.theseed.org/) (Overbeek *et al*., 2005) with sequences also matching viruses (9%), archaea (6%) and eukaryota (2%). Proteobacteria represented the highest percentage of matches to the SEED database for both the unconfined and confined aquifers with 18% and 13% of all sequences respectively (Fig. 1A). Within this, the delta/epsilon subdivision contributed to 5% and 7% of the total sequences in the unconfined and confined aquifer respectively. Viruses (ssDNA) were also major contributors with 3–4% of sequences matching the SEED database (Table S1). A total of 278 organisms and 3683 novel sequences could not be assigned to known sequences in the database.

When aquifers were compared using the Statistical Analysis of Metagenomic Profiles (STAMP) software package (Parks and Beiko, 2010), there was an overrepresentation of crenarchaeota, proteobacteria, actinobacteria, chloroflexi, ssDNA viruses, bacteroidetes/chlorobi

Table 1. Geophysical and microbial enumeration data.

Parameter	Unconfined aquifer (mean ± SD)[a]	Confined aquifer (mean ± SD)[a]	*P*-value
Iron (mg l^{-1})	3.041 ± 0.184	1.232 ± 0.003	0.000[c]
Sulfur (mg l^{-1})	76.3 ± 4.747	57.5 ± 0.173	0.002[c]
Ammonia (mg l^{-1})	0.025 ± 0.001	0.023 ± 0.004	0.330
Nitrate (mg l^{-1})	0.012 ± 0.001	0.012 ± 0.011	0.959
Nitrite (mg l^{-1})	0[b]	0[b]	–
Phosphorus (mg l^{-1})	0.015 ± 0.001	0.02 ± 0.019	0.718
TOC (mg l^{-1})	2.033 ± 0.208	0.9 ± 0.173	0.002[c]
Sulfide (mg l^{-1})	0[b]	0[b]	–
pH	7.56	7.16	–
Temperature (°C)	16.5	17.54	–
Salinity (ppm)	1.65	1.27	–
Oxygen (mg l^{-1})	0.2	0.26	–
Total bacterial and viral cell count (cell ml^{-1})	0.15E + 05 ± 1.43E + 04	1.12E + 05 ± 1.08E + 04	0.775

a. Variance is denoted by Standard Deviation.
b. A value of zero indicates the nutrient is below the detectable limit of the machine. In the case of Nitrite and sulfide this is 0.003 and 0.1 mg l^{-1} respectively.
c. Denotes statistically significant values.

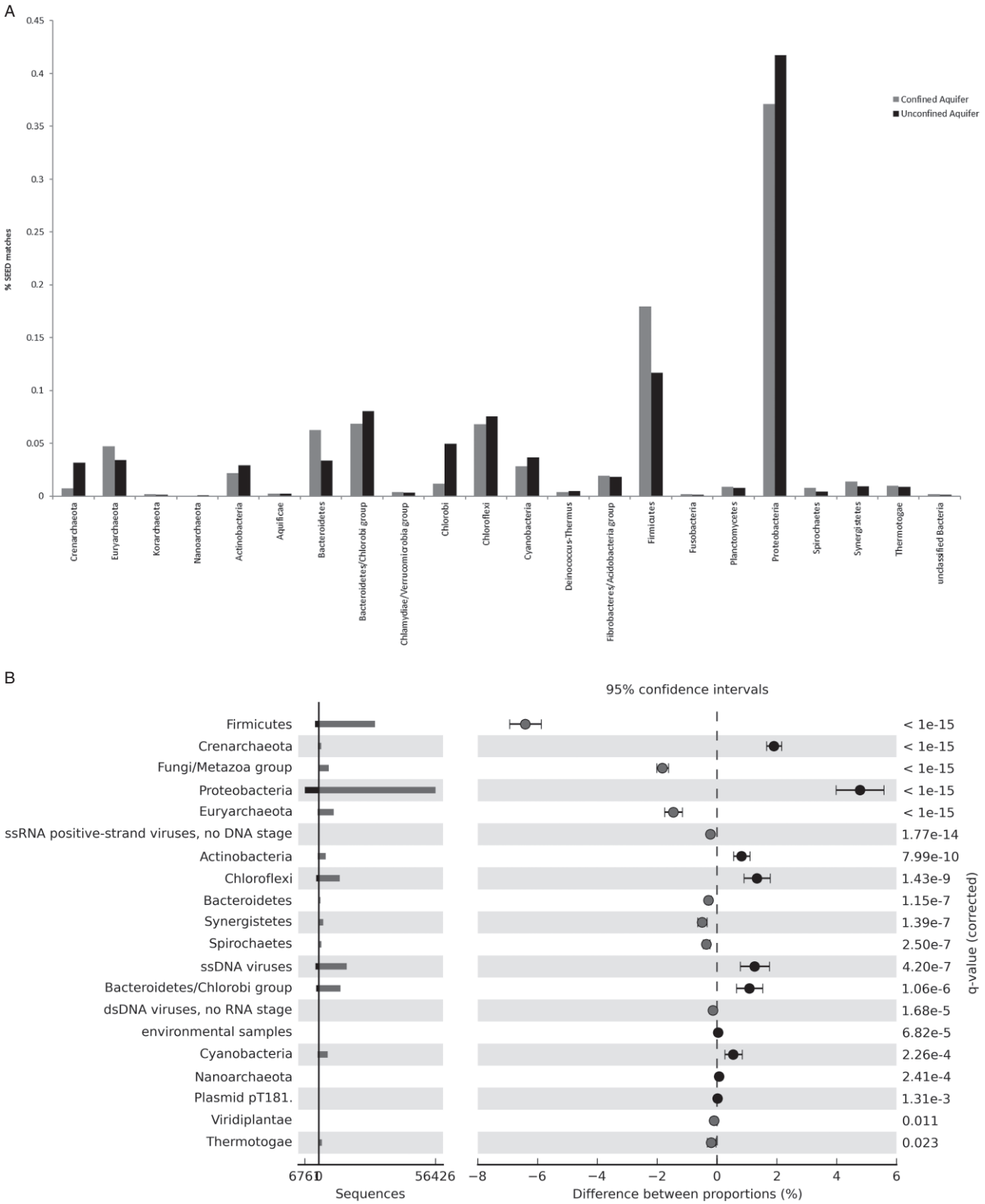
group and cyanobacteria in the unconfined aquifer (q-value < 1.06e^{-5}). Conversely, there was an overrepresentation of firmicutes, the fungi/metazoa group and euryarchaeota in the confined aquifer (q-value < 1e^{-15}) (Fig. 1B). SIMilarity PERcentage (SIMPER) analysis (Clarke, 1993) revealed the main contributors to the dissimilarity between the unconfined and the confined aquifer at phyla level were crenarchaeota and firmicutes, which contributed to 13% and 11% of the dissimilarity respectively (Table S2). At finer levels of taxonomic resolution (order level), *Deltaproteobacteria* represented the highest percentage of matches to the SEED database for both unconfined and confined aquifers with 5% and 7% of all sequences respectively (Fig. 2A). STAMP comparisons revealed an overrepresentation of *Rhodospirillales*, *Rhodocyclales*, *Chlorobia* and *Circovirus* occurred in the unconfined aquifer, whereas an overrepresentation of *Deltaproteobacteria* and *Clostridiales* occurred in the confined aquifer (Fig. 2B).

In both aquifer samples the core metabolic functions comprising DNA and protein metabolism were most prevalent, while a high level of phosphorous metabolism occurred in the confined aquifer (Table S3). Comparisons of the metabolic profiles of the unconfined and confined aquifer using STAMP, revealed an overrepresentation of DNA metabolism in the unconfined aquifer and an overrepresentation of motility and chemotaxis in the confined aquifer (Fig. 3A). SIMPER analysis revealed that overall DNA metabolism contributed to 15% of the dissimilarity between the unconfined and confined aquifers, while stress response and motility and chemotaxis contributed approximately 7.5% of the dissimilarity (Table S4). Finer levels (subsystem level) of resolution indicated that the unconfined aquifer had an overrepresentation of lactose and galactose uptake and utilization, beta-lactamase resistance and DNA replication. The confined aquifer had an overrepresentation of sequences matching sigmaB stress response regulation, flagellum, cobalt-zinc-cadmium resistance, phosphate metabolism and cellulosome degradation (i.e. starch uptake) (Fig. 3B).

Comparison of metabolic and taxonomic profiles from other habitats

In order to determine the overall effect the groundwater environment has on the inhabitant microbial assemblages, we compared our groundwater metagenomes with 37 publicly available metagenomes on the MetaGenomics Rapid Annotation using Subsystem Technology (MG-RAST) pipeline version 2.0 (Meyer *et al.*, 2008), covering a wide variety of habitats including other freshwater and low oxygen environments (Table S5). The highest metabolism (subsystem) and taxonomy (organism) resolution available was used to create cluster profiles that revealed the unconfined and the confined aquifers were more similar to each other than to any other metagenome (85% and 90% similarity, respectively). When the microbial taxonomy of these samples was compared with metagenomes from other

Fig. 1. Comparison of aquifer taxonomic profiles at phyla level.
A. Frequency distribution (relative % of bacterial SEED matches) of bacterial phyla in the unconfined and the confined aquifer.
B. STAMP analysis of taxonomy enriched or depleted between the confined and unconfined aquifers, using approach describes in Parks and Beiko (2010). Groups overrepresented in the unconfined aquifer (black) correspond to positive differences between proportions and groups overrepresented in the confined aquifer (grey) correspond to negative differences between proportions. Corrected *P*-values (q-values) were calculated using Storey's FDR approach.


A
% SEED matches
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
Confined Aquifer
Unconfined Aquifer
Crenarchaeota
Euryarchaeota
Korarchaeota
Nanoarchaeota
Actinobacteria
Aquificae
Bacteroidetes
Bacteroidetes/Chlorobi group
Chlamydiae/Verrucomicrobia group
Chlorobi
Chloroflexi
Cyanobacteria
Deinococcus-Thermus
Fibrobacteres/Acidobacteria group
Firmicutes
Fusobacteria
Planctomycetes
Proteobacteria
Spirochaetes
Synergistetes
Thermotogae
unclassified Bacteria
B
95% confidence intervals
Firmicutes
Crenarchaeota
Fungi/Metazoa group
Proteobacteria
Euryarchaeota
ssRNA positive-strand viruses, no DNA stage
Actinobacteria
Chloroflexi
Bacteroidetes
Synergistetes
Spirochaetes
ssDNA viruses
Bacteroidetes/Chlorobi group
dsDNA viruses, no RNA stage
environmental samples
Cyanobacteria
Nanoarchaeota
Plasmid pT181.
Viridiplantae
Thermotogae
< 1e-15
< 1e-15
< 1e-15
< 1e-15
< 1e-15
1.77e-14
7.99e-10
1.43e-9
1.15e-7
1.39e-7
2.50e-7
4.20e-7
1.06e-6
1.68e-5
6.82e-5
2.26e-4
2.41e-4
1.31e-3
0.011
0.023
q-value (corrected)
6761
56426
Sequences
−8
−6
−4
−2
0
2
4
6
Difference between proportions (%)

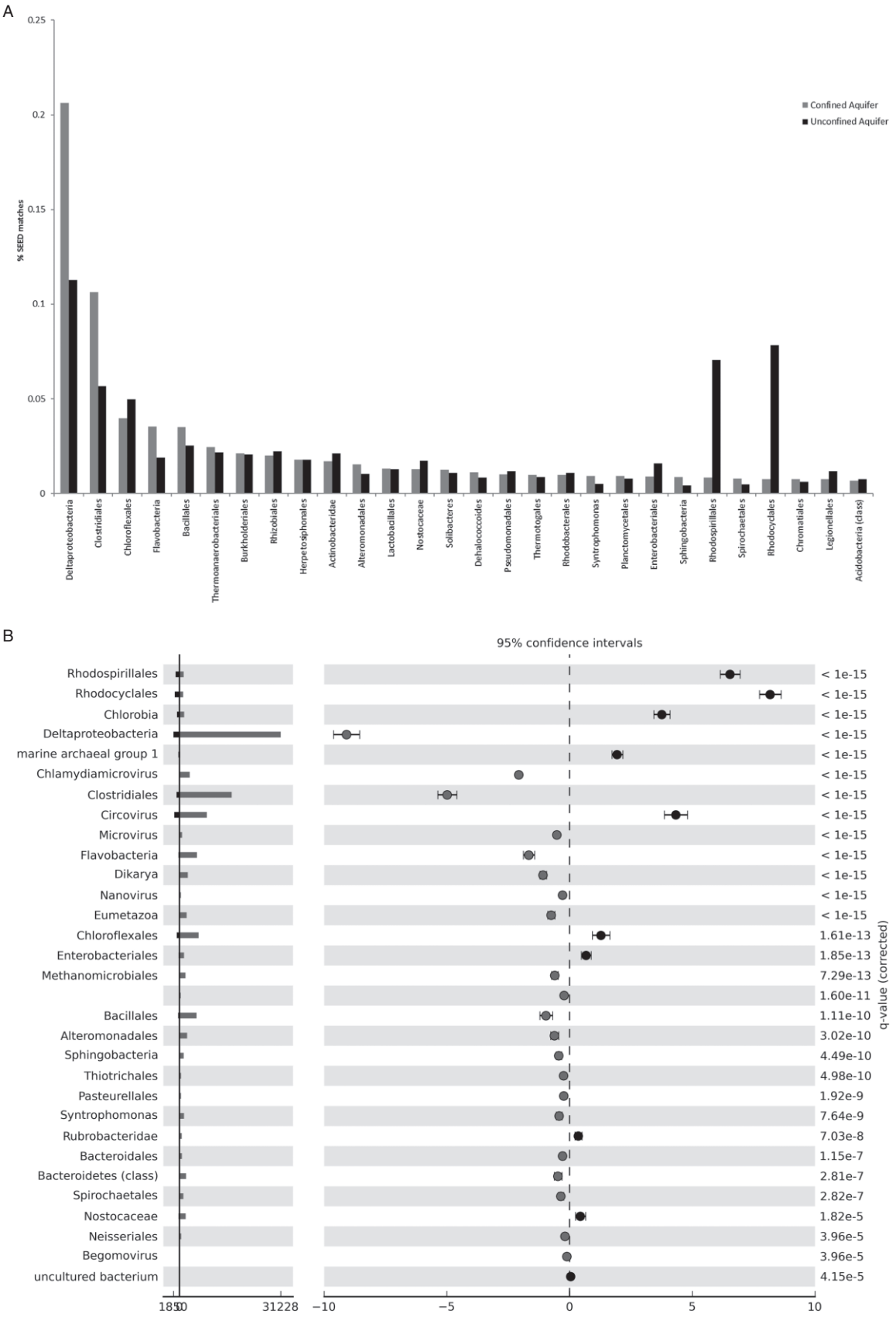

A
% SEED matches
0.25
0.2
0.15
0.1
0.05
0
Confined Aquifer
Unconfined Aquifer
Deltaproteobacteria
Clostridiales
Chloroflexales
Flavobacteria
Bacillales
Thermoanaerobacterales
Burkholderiales
Rhizobiales
Herpetosiphonales
Actinobacteridae
Alteromonadales
Lactobacillales
Nostocaceae
Solibacteres
Dehalococcoides
Pseudomonadales
Thermotogales
Rhodobacterales
Syntrophomonas
Planctomycetales
Enterobacteriales
Sphingobacteria
Rhodospirillales
Spirochaetales
Rhodocyclales
Chromatiales
Legionellales
Acidobacteria (class)
B
95% confidence intervals
Rhodospirillales
Rhodocyclales
Chlorobia
Deltaproteobacteria
marine archaeal group 1
Chlamydiamicrovirus
Clostridiales
Circovirus
Microvirus
Flavobacteria
Dikarya
Nanovirus
Eumetazoa
Chloroflexales
Enterobacteriales
Methanomicrobiales
Bacillales
Alteromonadales
Sphingobacteria
Thiotrichales
Pasteurellales
Syntrophomonas
Rubrobacteridae
Bacteroidales
Bacteroidetes (class)
Spirochaetales
Nostocaceae
Neisseriales
Begomovirus
uncultured bacterium
< 1e-15
< 1e-15
< 1e-15
< 1e-15
< 1e-15
< 1e-15
< 1e-15
< 1e-15
< 1e-15
< 1e-15
< 1e-15
< 1e-15
< 1e-15
1.61e-13
1.85e-13
7.29e-13
1.60e-11
1.11e-10
3.02e-10
4.49e-10
4.98e-10
1.92e-9
7.64e-9
7.03e-8
1.15e-7
2.81e-7
2.82e-7
1.82e-5
3.96e-5
3.96e-5
4.15e-5
q-value (corrected)
1851
31228
Sequences
−10
−5
0
5
10
Difference between proportions (%)

Fig. 2. Comparison of aquifer taxonomic profiles at order level taxonomy.
A. Frequency distribution (relative % of bacterial SEED matches) of taxonomy in the unconfined and the confined aquifer.
B. STAMP analysis of taxonomy enriched or depleted between the confined and unconfined aquifers. Groups overrepresented in the unconfined aquifer (black) correspond to positive differences between proportions and groups overrepresented in the confined aquifer (grey) correspond to negative differences between proportions. Corrected *P*-values (*q*-values) were calculated using Storey's FDR approach.

environments, the groundwater samples were most similar to termite gut and cow rumen metagenomes with a cluster node at 75% similarity (Fig. 4). When the metabolic potential of these samples was compared with metagenomes from other environments, groundwater samples were most similar to whale fall, phosphorous removing sludge, marine sediment samples and farm soil with a cluster node at 85% similarity (Fig. 5).

Discussion

Aquifer systems

Aquifer systems are considered to be extreme environments due to a lack of easily accessible organic carbon and low levels of inorganic nutrient input, low oxygen levels and a lack of sunlight (Danielopol *et al*., 2000). Consequently, microbial communities inhabiting these

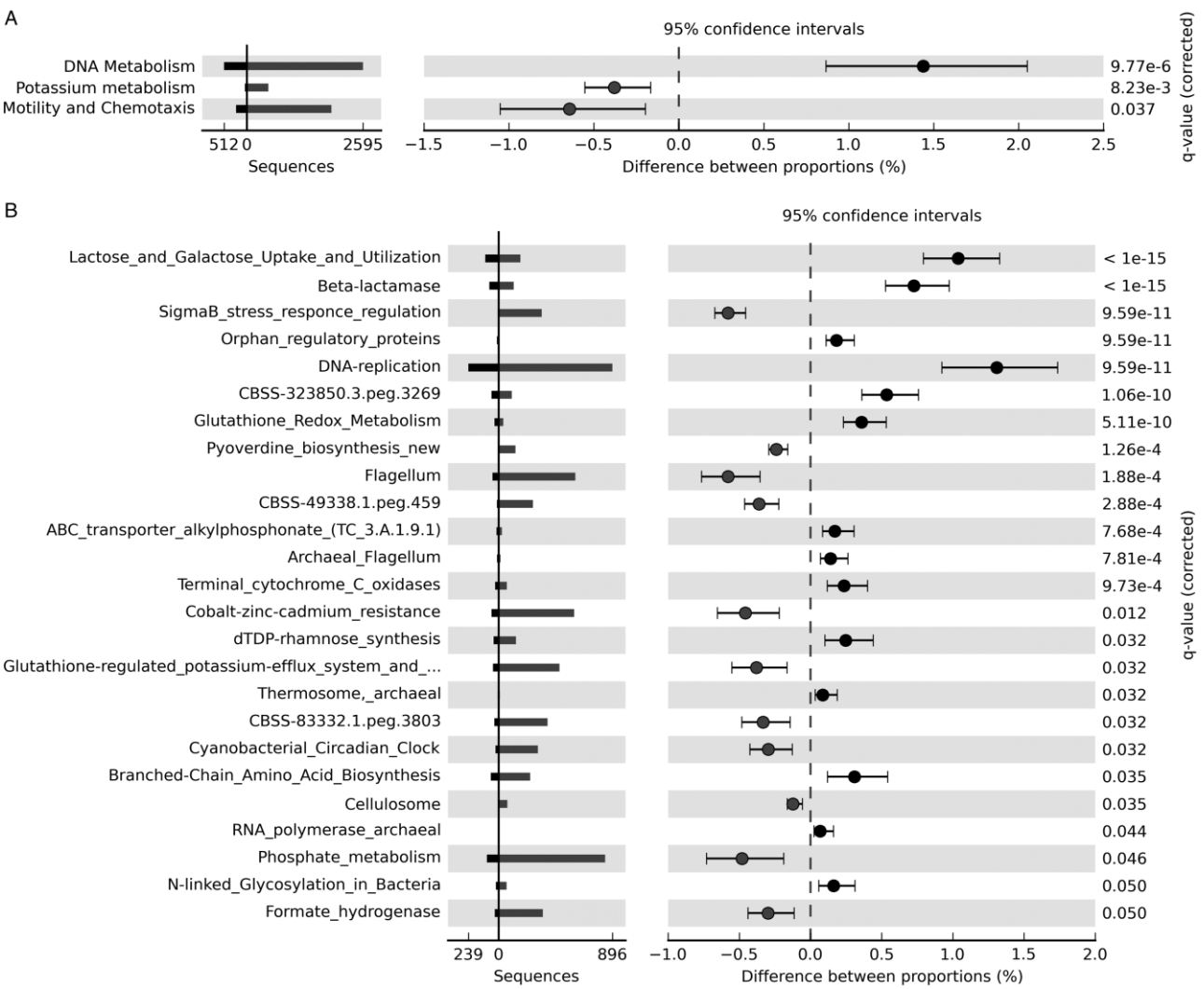


Fig. 3. Comparison of aquifer metabolism profiles.
A. STAMP analysis of hierarchy 1 enriched or depleted between the confined and unconfined aquifers. Groups overrepresented in the unconfined aquifer (black) correspond to positive differences between proportions and groups overrepresented in the confined aquifer (grey) correspond to negative differences between proportions. Corrected *P*-values (*q*-values) were calculated using Storey's FDR approach.
B. STAMP analysis of subsystems enriched or depleted between the confined and unconfined aquifers. Groups overrepresented in the unconfined aquifer (black) correspond to positive differences between proportions and groups overrepresented in the confined aquifer (grey) correspond to negative differences between proportions.

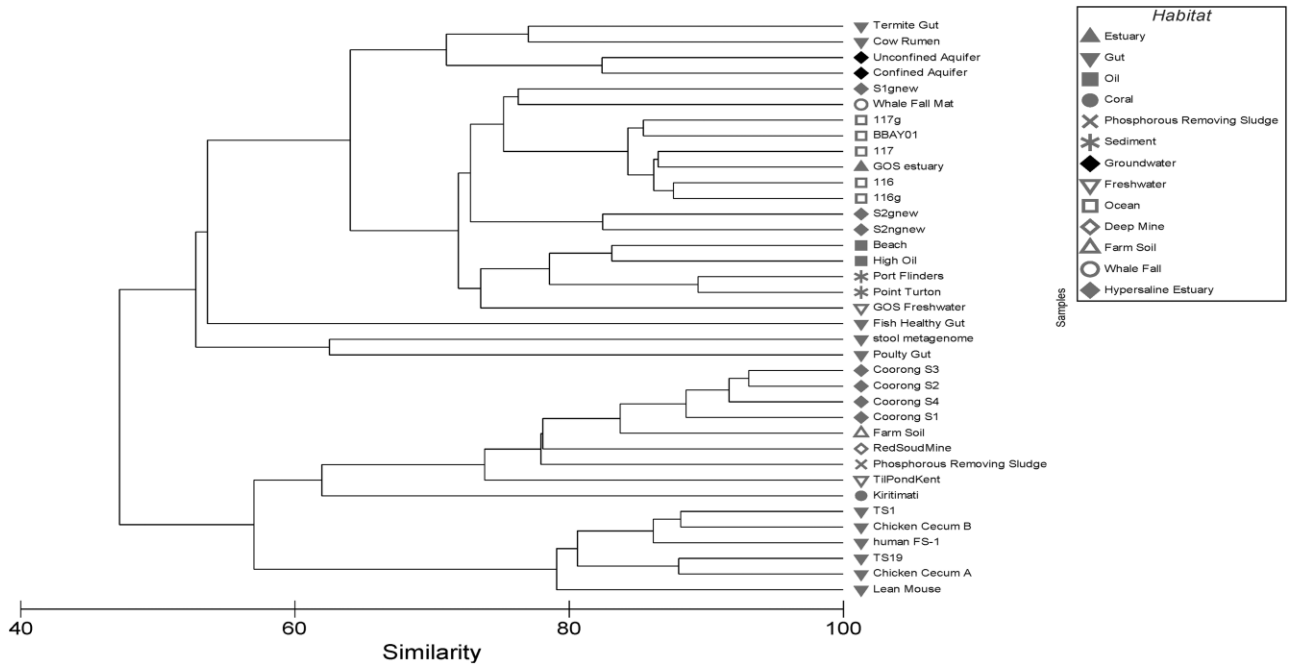


Fig. 4. Comparison of aquifer taxonomic profiles along with publicly available profiles available on the MG-RAST database. Cluster plot is derived from a Bray-Curtis similarity matrix calculated from the square-root transformed abundance of DNA fragments matching genome level taxonomy in the SEED database (BLASTX *E*-value < 0.001). Details of metagenomes in Table S5.

environments consist of microbes adapted to surviving in nutrient poor groundwater environments (Pedersen, 2000). In addition, strong environmental changes driven by anthropogenic influences present a consistent challenge for these communities (Griebler and Lueders, 2009). To determine the effects of anthropogenic influences on groundwater microbes, the microbial ecology of pristine aquifer systems needs to be compared with unconfined aquifers to determine how external factors influence microbial taxonomy and metabolism.

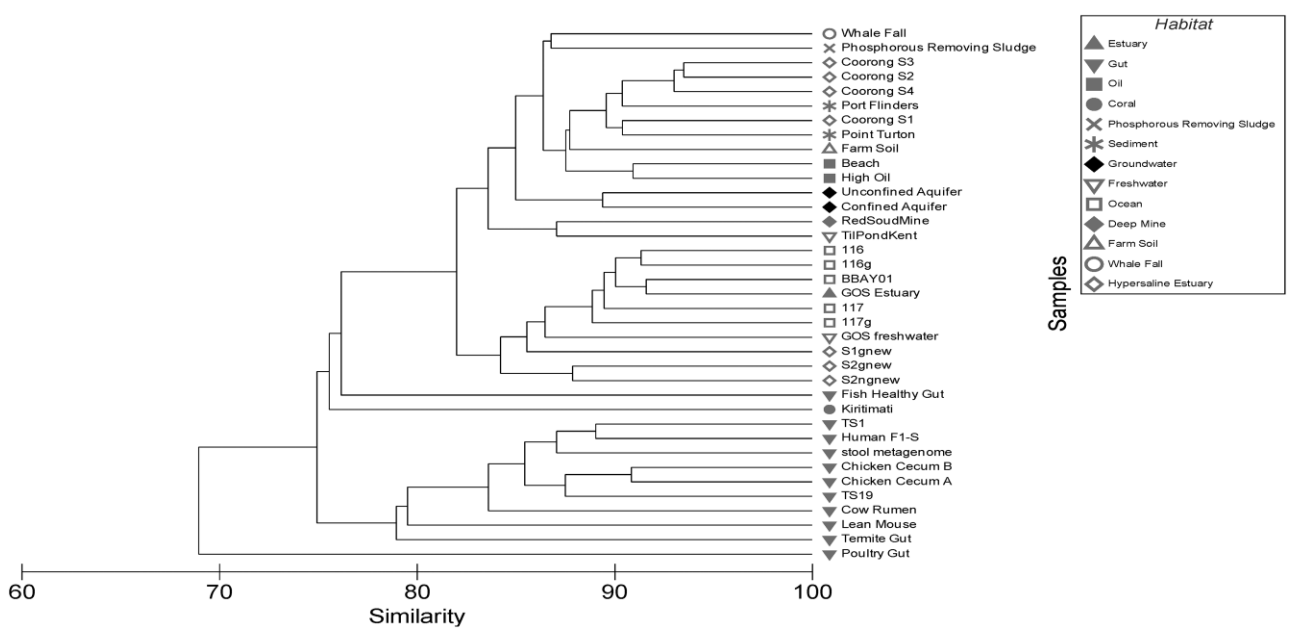


Fig. 5. Comparison of aquifer metabolic profiles along with publicly available profiles available on the MG-RAST database. Cluster plot is derived from a Bray-Curtis similarity matrix calculated from the square-root transformed abundance of DNA fragments matching subsystems in the SEED database (BLASTX *E*-value < 0.001). Details of metagenomes in Table S5.

We assessed the chemical properties and the microbial communities within an unconfined aquifer, which has been exposed to external input from a dairy farm, and an adjacent confined aquifer, which has had no external input for approximately 1500 years (Banks *et al*., 2006), to determine the effect of anthropogenic inputs on groundwater ecosystems. Nutrient analysis comparing these two systems showed that the confined aquifer had significantly lower sulfur, iron and total organic carbon (TOC) concentrations than the unconfined aquifer. In groundwater, the amount of suspended microbes is largely dependent on the availability of dissolved organic carbon (DOC) and nutrients (Griebler and Lueders, 2009). Typically phosphorous and iron are limiting factors in groundwater systems (Bennett *et al*., 2001). Those microbes able to increase the bioavailability of such critical nutrients can increase the viability of the native population (Rogers and Bennett, 2004). Flow cytometry counts showed that total bacterial and viral abundances were relatively similar between the unconfined and confined aquifer with mean values of $1.15 \times 10^5 \pm 1.43 \times 10^4$ and $1.12 \times 10^5 \pm 1.08 \times 10^4$ cells ml^{-1} respectively (Table 1). This is consistent with commonly reported microbial cell counts of 10^3–10^8 cells ml^{-1} in groundwater regardless of contamination (Pedersen, 1993; 2000; Griebler and Lueders, 2009).

Taxonomic profiling of groundwater

A shift in dominant taxa was observed between the unconfined and the confined aquifer, with fundamentally different communities inhabiting each environment. In the unconfined aquifer there was an overrepresentation of *Rhodospirillales*, *Rhodocyclales*, *Chlorobia* and *Circovirus* (Fig. 2). The dominance of these taxa in the unconfined aquifer differs from a recent metagenomic study in which uranium contaminated aquifers where dominated by *Rhodanobacter*-like gammaproteobacterial and *Burkholderia*-like betaproteobacterial species (Hemme *et al*., 2010). However, *Rhodocyclales* are commonly found in wastewater treatment systems (Hesselsoe *et al*., 2009) and are noted for their ability to degrade and transform pollutants such as nitrogen, phosphorous and aromatic compounds (Loy *et al*., 2005). This suggests that the microbial communities in the unconfined aquifer are responding to the influx of nutrients similar to those seen in wastewater. Furthermore, *Chlorobia* are green sulfur bacteria that are typically found in deep anoxic aquatic environments where low light intensity and sulfide concentrations favour their growth (Guerrero *et al*., 2002; Madigan *et al*., 2003). This suggests the increased sulfur concentration in the unconfined aquifer could be responsible for the overrepresented *Chlorobia*. Taken together, these patterns indicate that different types of contamination can drive markedly different community profiles within aquifer systems.

The overrepresentation of circovirus in the unconfined aquifer is also notable, due to its known vertebrate pathogenicity (Rosario *et al*., 2009a). Circoviridae has been linked to a number to livestock related diseases including infections of dairy cattle (Nayar *et al*., 1999) and has previously been found in reclaimed water, suggesting it is resistant to wastewater treatment (Rosario *et al*., 2009b). The occurrence of circoviridae in the unconfined aquifer could indicate contamination from nearby farmland and is consistent with a study by Dinsdale and colleagues (2008a) who found increased numbers of pathogens in a human impacted versus non-human impacted marine environments.

In the confined aquifer there was an overrepresentation of *Deltaproteobacteria* and *Clostridiales* (Fig. 2). *Clostridiales* are obligate anaerobes and have the ability to form endospores when growing cells are subjected to nutritional deficiencies (Paredes-Sabja *et al*., 2011). *Clostridiales* have not been widely reported in aquifer systems; however, their survival strategies make them well adapted to survive in low nutrient conditions, such as subsurface environments like those observed in the confined aquifer (Leclerc and Moreau, 2002).

Metabolic profiling of groundwater

Generally, the rate of metabolism in subsurface communities is slower in comparison with other aquatic or sediment environments (Swindoll *et al*., 1988). Within groundwater systems, previous studies have shown metabolic rates were higher in a shallow sandy aquifer compared with a confined clayey aquifer (Chapelle and Lovley, 1990). The authors suggested this lower metabolism could be due to the reduced interconnectivity, and thus, a reduction in microbial and nutrient mobility. The core metabolic function in each of our aquifer systems was DNA metabolism; however, an overrepresentation of DNA replication was seen in the unconfined aquifer compared with the confined (Fig. 3). This indicates that the reduced nutrient levels in the confined aquifer may have led to reduced reproduction.

When nutrient levels are low, it is advantageous for microbes to attach themselves to sediment particles, detritus, rock surfaces and biofilms (Griebler and Lueders, 2009). This attachment mode is successful as nutrient availability is higher at surfaces (Hall-Stoodley *et al*., 2004). Thus, microbes dominating groundwater systems are more commonly found attached to surfaces than in suspension (Griebler and Lueders, 2009). Repulsive forces of the substratum require microbial cells to

produce flagella for the early stages of attachment (Donlan, 2002). Overrepresentation of flagella in the confined aquifer community (Fig. 3) could be indicative of a greater need to attach to surfaces in the low nutrient confined aquifer.

Our data also indicate that β-lactamase genes were overrepresented in the unconfined aquifer (Fig. 3). This antibiotic resistance gene is widely seen in Gram-negative bacteria and has been shown to be a product of the extensive use of β-lactams in dairy farms to prevent bacterial infections (Berghash *et al*., 1983; Gianneechini *et al*., 2002; Sawant *et al*., 2005; Liebana *et al*., 2006). Within livestock, the majority of antibiotics are excreted unchanged by the animal, where they subsequently enter water sources via leaching and run-off (Zhang *et al*., 2009). This has caused concern about the potential impacts that antibacterial resistance in waterways can have on humans and animal health (Kemper, 2008). The overrepresentation of β-lactamase in the unconfined aquifer suggests that external input, potentially in the form of farm affected input, may introduce new cellular processes that would not normally be required by endemic groundwater microbes. This is consistent with a study that investigated the use of antibiotics in farm animals and illustrated that antibiotic resistance can be spread into the surrounding environment through the use of antimicrobial drugs (Ghosh and LaPara, 2007). Further, microbes able to utilize lactose have previously been linked to dairy farms (Klijn *et al*., 1995) and thus, the overrepresentation of lactose and glucose utilization found in the unconfined aquifer (Fig. 3) could be linked to external input from the overlaying dairy farms.

Comparison with other microbial communities

To determine how the unique features of the groundwater environment influence the structure of microbial communities, we compared the metagenomes from our aquifer systems with metagenomes from different environments (Table S5). The unconfined and confined aquifer metagenomes were more similar to each other than to any other community, both in terms of taxonomy and metabolism (Figs 4 and 5). This suggests the features of subterranean aquatic environments, including low oxygen concentrations, coupled with a lack of sunlight and low external inputs of nutrients have led to a unique niche for microbial communities to evolve. In a recent study, four sediment metagenomes from a naturally occurring salinity gradient were compared and it was found that despite differences in salinity and nutrient levels, these four samples clustered more closely to each other and other sediment samples, than to other similar hypersaline environments (Jeffries *et al*., 2011). It was found that the substrate type, i.e. sediment or water, rather than salinity drove the similarity. Willner and colleagues (2009) also found that microbiomes and viromes have distinct sequence-based signatures, which are driven by environmental selection. This is further supported by Dinsdale and colleagues (2008b), who compared metagenomic sequences from 45 distinct microbiomes and 42 distinct viromes to show there was a strong discrimatory profile across different environments. Our data similarity suggest that the unique features of the subterranean aquatic environment act to structure microbial assemblages that retain a high level of similarity between different aquifers.

The taxonomy of the aquifer metagenomes was most similar to cow rumen and termite gut metagenomes (Fig. 4). A common feature among these environments is the incidence of anaerobic fungi that is overrepresented in the confined aquifer (Fry *et al*., 1997; Ramšak *et al*., 2000; Ekendahl *et al*., 2003; Warnecke *et al*., 2007). A primary role of anaerobic fungi in gut systems is the large-scale breakdown of plant material, including cellulose (Ramšak *et al*., 2000; Warnecke *et al*., 2007). The breakdown of cellulose in groundwater is also known to occur in shallow aquifers (Vreeland *et al*., 1998), which, along with the overrepresentation in cellulosome genes in the confined aquifer (Fig. 3), suggests that cellulose is present and possibly an important food source for the overrepresented fungi/metazoa group (Fig. 1). Furthermore, the cellulosome gene is similarly represented in the groundwater, termite gut and cow rumen, suggesting cellulose is a major factor linking the three environmental metagenomes.

The metabolism of the aquifer metagenomes was most similar to other sediment metagenomes (85% similar) rather than freshwater environments (80% similar) (Fig. 5). Common features to groundwater and sediment environments are low oxygen concentrations, a lack of sunlight and large surfaces for biofilm formation (Griebler and Lueders, 2009). As previously discussed, due to low nutrient levels in groundwater environments, a common survival strategy is for the microbes to attach to sediment particles or form biofilms (Hall-Stoodley *et al*., 2004; Griebler and Lueders, 2009). This suggests, the attachment mode of life coupled with the low oxygen concentrations and a lack of sunlight, are the main factors driving the similarity between these metagenomes.

Caveats

Due to the low microbial biomass in groundwater systems, we used multiple displacement amplification (MDA) before 454 pyrosequencing. This method has been used widely to amplify DNA before sequencing (Binga *et al*.,

2008; Dinsdale *et al*., 2008a; Neufeld *et al*., 2008; Palenik *et al*., 2009), but its suitability for use in quantitative metagenomic analysis has been debated (Yilmaz *et al*., 2010) because of the GC bias introduced. However, in our study, as GenomiPhi was used on both aquifer samples compared here, any bias in the process is applied to both aquifers. Furthermore, we are concerned with differences between aquifer groups rather than absolute changes in particular genes. Edwards and colleagues (2006) used GenomiPhi to amplify microbial DNA from a Soudan Mine and found that the whole genome amplification bias was minimal and was found preferentially towards the ends of linear DNA. The authors concluded that as these biases were applied equally to both libraries, this bias would have been negated during the comparative study when assessing differences in the community structure (Edwards *et al*., 2006).

There is a possibility that the clustering of our samples may be due to the way in which the samples were collected, sequenced and analysed, which may be different to the metagenomes from other environments. However, there is no evidence of clustering based on collection, DNA extraction, MDA or sequencing protocols (Figs 4 and 5), and thus a technical bias is not evident.

Conclusion

Our data indicate that aquifer ecosystems host unique microbial assemblages that have different phylogenetic and metabolic properties to other environments. We suggest this pattern is driven by the unique physiochemical properties of subterranean aquatic environments, and that groundwater ecosystems represent a specific microbial niche. Our data also revealed that the unconfined aquifer examined in this study has significantly different features to the more pristine confined aquifer, which in some cases appear to have been influenced by external input from a surrounding dairy farm. Increased nutrient concentrations, the overrepresentation of DNA replication as well as lactose and galactose utilization and β-lactamase genes are all consistent with inputs of nutrients and contaminants from dairy farm practises. Preservation of groundwater is of increasing importance due to its use as potable water sources and as water sources for global industrial and agricultural production. This study provides important insights and suggests further investigation into the differences between unconfined and confined aquifers. Further to this, the subterranean dispersal of agricultural contaminants is needed in order to fully determine the effects of anthropogenic processes on groundwater.

Experimental procedures

Site selection

Samples were collected from two depths in the Ashbourne aquifer system, situated within the Finniss River Catchment, South Australia (35°18′S 138°46′E) in June 2010. The Ashbourne aquifer system is two aquifer ecosystems with separate recharge processes that have distinct water sources. The confined aquifer has been isolated from external input for approximately 1500 years (Banks *et al*., 2006), and thus provides a baseline for which the unconfined aquifer can be compared.

Sampling groundwater

Unconfined and confined aquifer samples were collected from a nested set of piezometers. Each piezometer consisted of a 10 mm diameter PVC casing, with slotted PVC screens that provide discrete sampling points at specific depths. The unconfined aquifer was sampled from a piezometer at 13–19 m and the confined aquifer at 79–84 m. To ensure that only aquifer water was sampled, bores were purged by pumping out 3 bore volumes using a 12 V, 36 m monsoon pump (EnviroEquip) before sampling. Based on microbial abundances at each depth determined previously using flow cytometry, 20 l and 200 l of water was collected from the unconfined and confined aquifers respectively, to ensure sufficient biomass for microbial DNA recovery.

From each sampling location, triplicate 600 ml water samples for inorganic and organic chemistry analysis were collected and stored on ice. Nutrient analysis for ammonia, nitrite, nitrate and filterable reactive phosphorous were conducted using a flow injection analyser. TOC was analysed using OI analytical 1010 & 1030 low level TOC analysers, iron and sulfur were determined by the ICP-006 and ICP-004 elemental analysis using an ICP-mass spectrometer, and sulfide (S^{2-}) concentrations were determined using the colorimetric method (APHA 1995). All analysis was conducted at the Australian Water Quality Centre (Adelaide). For enumeration of microbes at each site, triplicate 1 ml samples were fixed with gluteraldehyde (2% final concentration), quick frozen in liquid nitrogen and stored at –80°C prior to flow cytometric analysis (Brussaard, 2004). Physical parameters, including temperature, salinity, pH and oxygen concentration, were recorded at each sampling point with the use of a MS5 water quality sonde (Hach Hydrolab).

Microbial enumeration

Bacteria and viruses were enumerated using a FACSCanto flow cytometer (Becton-Dickson). Before analysis, triplicate samples were quick thawed and diluted 1:10 with 0.2 μm filtered TE buffer (10 mM Tris, 1 mM EDTA pH 7.5). Samples were then stained with SYBR-I Green solution (1:20000 dilution; Molecular Probes, Eugene, OR) and incubated in the dark for 10 min at 80°C (Brussaard, 2004). As an internal size standard fluorescent 1 μm diameter beads (Molecular Probes, Eugene, OR) were added to each sample at a final concentration of approximately 10^5 beads ml^{-1} (Gasol and Del Giorgio, 2000). Forward scatter (FSC), side scatter (SSC)

and green (SYBRGreen-I) fluorescence were acquired for each sample. WinMDI 2.9 (Joseph Trotter) software was used to identify and enumerate microbes according to variations in green fluorescence and side scatter (Marie *et al*., 1997; 1999; Brussaard, 2004).

Sample filtration, microbial community DNA extraction and sequencing

Following collection, samples for metagenomic analysis were filtered through 5 μm membranes to remove sediment particles before being concentrated by 2000-fold using a 100 kDa tangential flow filtration (TFF) filter (MilliporeTM). Microbial community DNA was extracted using a bead beating and chemical lysis extraction protocol (PowerWater DNA Isolation Kit; MoBio laboratories). Due to the low microbial biomass in the aquifer samples, DNA was then amplified using the multiple strand displacement Phi29 DNA polymerase (GenomiPhi V2 Kit; GE Healthcare Life Sciences) and cleaned up using a PCR clean-up kit (UltraClean PCR Clean-Up Kit; MoBio laboratories). DNA quality and concentration were determined by 1.5% TBE agarose gel electrophoresis (Bioline) and a Qubit fluorometer (Quant-iT dsDNA HS Assay Kit; Invitrogen). Approximately 500 ng of high molecular weight DNA was then sequenced by the Ramaciotti Centre for Gene Function Analysis, Sydney, Australia. Sequencing was conducted on the GS-FLX pyrosequencing platform using Titanium series reagents (Roche).

Data analysis

To determine if the nutrient data were statistically different between the unconfined and the confined aquifer, *P*-values were determined by an Independent *t*-test. All analysis was performed using PASW version 18 statistical software.

Unassembled DNA sequences were annotated with the MetaGenomics Rapid Annotation using Subsystem Technology (MG-RAST) pipeline version 2.0 (Meyer *et al*., 2008). BLASTX was used with a minimum alignment length of 50 bp and an *E*-value cut-off of $E < 1 \times 10^{-5}$ as described by Dinsdale and colleagues (2008b). Taxonomic profiles were generated using the normalized abundance of sequence matches to the SEED database (Overbeek *et al*., 2005), while the normalized abundance of sequence matches to a given subsystem were used to generate metabolic profiles.

To determine statistically significant differences between the two aquifer samples, the Statistical Analysis of Metagenomic Profiles (STAMP) software package was used (Parks and Beiko, 2010). First, a table of the frequency of hits to each individual taxa or subsystem for each metagenome was generated, which had been normalized by dividing by the total number of hits to remove bias in difference in read lengths and sequencing effort. An *E*-value cut-off of $E < 1 \times 10^{-5}$ was used to identify hits. The highest level of resolution available on MG-RAST was used for metabolism (subsystem) and taxonomy (genome). *P*-values were calculated in STAMP using the two sided Fisher's Exact test (Fisher, 1958), while the confidence intervals were calculated using the Newcombe-Wilson method (Newcombe, 1998). False discovery rate was corrected for using the Storey's FDR method (Storey and Tibshirani, 2003).

We next compared the metagenomes of our groundwater samples with 37 publicly available metagenomes from a variety of environments on MG-RAST (Table S5), to statistically investigate the similarities between the two groundwater samples as well as other environments. Heatmaps were generated and normalized, as described above; however, as groundwater samples were compared with datasets with a variety of different read lengths, a lower *E*-value cut-off of $E < 0.001$ was used. Statistical analyses were conducted on square root transformed data using the statistical package Primer 6 for Windows (Version 6.1.6, Primer-E, Plymouth) (Clarke and Gorley, 2006). Metagenomes were then analysed using hierarchial agglomerative clustering (CLUSTER) (Clarke, 1993) analyses of the Bray-Curtis similarities. The main taxa or subsystems contributing to the differences were identified using similarity percentage (SIMPER) analysis (Clarke, 1993).

Acknowledgements

The authors gratefully acknowledge Eugene Ng from the Flow Cytometry Unit of the Flinders University Medical Centre for providing technical support during the flow cytometry work. Funding was provided by ARC linkage Grant LP0776478. R.S. is the recipient of a Flinders University Research Scholarship (FURS).

References

Al-Zabet, T. (2002) Evaluation of aquifer vulnerability to contamination potential using the DRASTIC method. *Environ Geol* **43:** 203–208.

APHA (2005) *Standard Methods for the Examination of Water and Wastewater*, 21st edn. Washington, DC, USA: American Public Health Association.

Avidano, L., Gamalero, E., Cossa, G.P., and Carraro, E. (2005) Characterization of soil health in an Italian polluted site by using microorganisms as bioindicators. *Appl Soil Ecol* **30:** 21–33.

Banks, E.W., Wilson, T., Green, G., and Love, A. (2006) Groundwater recharge investigations in the Eastern Mount Lofty Ranges, South Australia. Report DWLBC 2007/20, Government of South Australia, through Department of Water, Land and Biodiversity Conservation, Adelaide.

Bennett, P.C., Rogers, J.R., and Choi, W.J. (2001) Silicates, silicate weathering, and microbial ecology. *Geomicrobiol J* **18:** 3–19.

Berghash, S.R., Davidson, J.N., Armstrong, J.C., and Dunny, G.M. (1983) Effects of antibiotic treatment of nonlactating dairy cows on antibiotic resistance patterns of bovine mastitis pathogens. *Antimicrob Agents Chemother* **24:** 771–776.

Binga, E.K., Lasken, R.S., and Neufeld, J.D. (2008) Something from (almost) nothing: the impact of multiple displacement amplification on microbial ecology. *ISME J* **2:** 233–241.

Bond, N.R., Lake, P.S., and Arthington, A.H. (2008) The impacts of drought on freshwater ecosystems: an Australian perspective. *Hydrobiologia* **600:** 3–16.

Brune, A., Frenzel, P., and Cypionka, H. (2000) Life at the oxic-anoxic interface: microbial activities and adaptations. *FEMS Microbiol Rev* **24:** 691–710.

Brussaard, C.P.D. (2004) Optimization of procedures for counting viruses by flow cytometry. *Appl Environ Microbiol* **2004:** 3.

Chang, Y., Peacock, A.D., Long, P.E., Stephen, J.R., McKinley, J.P., Macnaughton, S.J., *et al*. (2001) Diversity and characterization of sulfate-reducing bacteria in groundwater at a Uranium Mill Tailings Site. *Appl Environ Microbiol* **67:** 3149–3160.

Chapelle, F.H., and Lovley, D.R. (1990) Rates of microbial metabolism in deep coastal plain aquifers. *Appl Environ Microbiol* **56:** 1865–1874.

Cho, J., and Kim, S. (2000) Increase in bacterial community diversity in subsurface aquifers receiving livestock wastewater input. *Appl Environ Microbiol* **66:** 956–965.

Clarke, A., and Gorley, R. (2006) *PRIMER V6: User Manula/Tutorial*. Plymouth, UK: PRIMER-E.

Clarke, K.R. (1993) Nonparametric multivariate analysis of changes in community structure. *Aust J Ecol* **18:** 117–143.

Danielopol, D.L., Pospisil, P., and Rouch, R. (2000) Biodiversity in groundwater: a large-scale view. *Trends Ecol Evol* **15:** 223–224.

Danielopol, D.L., Griebler, C., Gunatilaka, A., and Notenboom, J. (2003) Present state and future prospects for groundwater ecosystems. *Environ Conserv* **30:** 104–130.

Dinsdale, E.A., Pantos, O., Smriga, S., Edwards, R.A., Angly, F., Wegley, L., *et al*. (2008a) Microbial ecology of four coral atolls in the Northern Line Islands. *PLoS ONE* **3:** 1–17.

Dinsdale, E.A., Edwards, R.A., Hall, D., Angly, F., Breitbart, M., Brulc, J.M., *et al*. (2008b) Functional metagenomic profiling of nine biomes. *Nature* **452:** 629–633.

Donlan, R.M. (2002) Biofilms: microbial life on surfaces. *Emerg Infect Dis* **8:** 881–890.

Edwards, R.A., Rodriguez-Brito, B., Wegley, L., Haynes, M., Breitbart, M., Peterson, D.M., *et al*. (2006) Using pyrosequencing to shed light on deep mine microbial ecology. *BMC Genomics* **7:** 1–13.

Ekendahl, S., O'Neill, A.H., Thomsson, E., and Pedersen, K. (2003) Characterisation of yeasts isolated from deep igneous rock aquifers of the Fennoscandian Shield. *Microb Ecol* **46:** 416–428.

Fisher, W.D. (1958) On grouping for maximum homogeneity. *J Am Stat Assoc* **53:** 789–798.

Fry, N.K., Fredrickson, J.K., Fishbain, S., Wagner, M., and Stahl, D.A. (1997) Population structure of microbial communities associated with two deep, anaerobic, alkaline aquifers. *Appl Environ Microbiol* **63:** 1498–1504.

Gasol, J.M., and Del Giorgio, P.A. (2000) Using flow cytometry for counting natural planktonic bacteria and understanding the structure of planktonic bacterial communities. *Sci Mar* **64:** 197–224.

Ghosh, S., and LaPara, T.M. (2007) The effects of subtherapeutic antibiotic use in farm animals on the proliferation and persistence of antibiotic resistance among soil bacteria. *ISME J* **1:** 191–203.

Gianneechini, R.E., Concha, C., and Franklin, A. (2002) Antimicrobial susceptibility of udder pathogens isolated from dairy herds in the West Littoral Region of Uruguay. *Acta Vet Scand* **43:** 31–41.

Gibert, J., and Deharveng, L. (2002) Subterranean ecosystems: a truncated functional biodiversity. *Bioscience* **52:** 473–481.

Griebler, C., and Lueders, T. (2009) Microbial biodiversity in groundwater ecosystems. *Freshw Biol* **54:** 649–677.

Guerrero, R., Montesinos, E., Pedrós-Alió, C., Esteve, I., Mas, J., Gemerden, H., *et al*. (2002) Phototrophic sulfur bacteria in two Spanish lakes: vertical distribution and limiting factors. *Limnol Oceanogr* **30:** 919–931.

Hall-Stoodley, L., Costerton, W.J., and Stoodley, P. (2004) Bacterial biofilms: from the natural environment to infectious diseases. *Nat Rev Microbiol* **2:** 95–108.

Hamblin, K.W., and Christiansen, E.H. (2004) *Earth's Dynamic Systems*, 10th edn. New Jersey, USA: Pearson Education.

Hemme, C.L., Deng, Y., Gentry, T.J., Fields, M.W., Wu, L., Barua, S., *et al*. (2010) Metagenomic insights into evolution of a heavy metal-contaminated groundwater microbial community. *ISME J* **4:** 660–672.

Hesselsoe, M., Füreder, S., Schloter, M., Bodrossy, L., Iversen, N., Roslev, P., *et al*. (2009) Isotope array analysis of Rhodocyclales uncovers functional redundancy and versatility in an activated sludge. *ISME J* **3:** 1349–1364.

Jeffries, T.C., Seymour, J.R., Gilbert, J.A., Dinsdale, E.A., Newton, K., Leterme, S.S.C., *et al*. (2011) Substrate type determines metagenomic profiles from diverse chemical habitats. *PLoS ONE* (in press).

Kemper, N. (2008) Veterinary antibiotics in the aquatic and terrestrial environment. *Ecol Indic* **8:** 1–13.

Kennedy, J., Flemer, B., Jackson, S.A., Lejon, D.P.H., Morrissey, J.P., O'Gara, F., and Dobson, A.D.W. (2010) Marine metagenomics: new tools for the study and exploitation of marine microbial metabolism. *Mar Drugs* **8:** 608–628.

Klijn, N., Weerkamp, A.H., and Nizo, W.M.D. (1995) Detection and characterization of lactose-utilizing *Lactococcus* spp. in natural ecosystems. *Appl Environ Microbiol* **61:** 788–792.

Langworthy, D.E., Stapleton, R.D., Sayler, G.S., and Findlay, R.H. (1998) Genotypic and phenotypic responses of a riverine microbial community to polycyclic aromatic hydrocarbon contamination. *Appl Environ Microbiol* **64:** 3422–3428.

Leclerc, H., and Moreau, A. (2002) Microbiological safety of natural mineral water. *FEMS Microbiol Rev* **26:** 207–222.

Liebana, E., Batchelor, M., Hopkins, K.L., Clifton-Hadley, F.A., Teale, C.J., Foster, A., *et al*. (2006) Longitudinal farm study of extended-spectrum beta-lactamase-mediated resistance. *J Clin Microbiol* **44:** 1630–1634.

Loy, A., Schulz, C., Lücker, S., Schöpfer-Wendels, A., Stoecker, K., Baranyi, C., *et al*. (2005) 16S rRNA gene-based oligonucleotide microarray for environmental monitoring of the betaproteobacterial order 'Rhodocyclales'. *Appl Environ Microbiol* **71:** 1373–1386.

Madigan, M.T., Martinko, J.M., Dunlap, P.V., and Clark, D.P. (2003) *Brock Biology of Microorganisms*, 10th edn. New Jersey, USA: Pearson Education.

Männistö, M.K., Tiirola, M.A., Salkinoja-Salonen, M.S., Kulomaa, M.S., and Puhakka, J.A. (1999) Diversity of chlorophenol-degrading bacteria isolated from contaminated boreal groundwater. *Arch Microbiol* **171:** 189–197.

Marie, D., Partensky, F., Jacquet, S., and Vaulot, D. (1997) Enumeration and cell cycle analysis of natural populations of marine picoplankton by flow cytometry using the nucleic acid stain SYBR Green I. *Appl Environ Microbiol* **63:** 186–193.

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

Meyer, M., Stenzel, U., and Hofreiter, M. (2008) Parallel tagged sequencing on the 454 platform. *Nat Protoc* **3:** 267–278.

Mpelasoka, F., Hennessy, K., Jones, R., and Bates, B. (2008) Comparison of suitable drought indices for climate change impacts assessment over Australia towards resource management. *Int J Climatol* **28:** 1283–1292.

Nayar, G.P.S., Hamel, A.L., Lin, L., Sachvie, C., Grudeski, C., and Spearman, G. (1999) Evidence for circovirus in cattle with respiratory disease and from aborted bovine fetuses. *Can Vet J* **40:** 277–278.

Neufeld, J.D., Chen, Y., Dumont, M.G., and Murrell, J.C. (2008) Marine methylotrophs revealed by stable-isotope probing, multiple displacement amplification and metagenomics. *Environ Microbiol* **10:** 1526–1535.

Newcombe, R.G. (1998) Improved confidence intervals for the difference between binomial proportions based on paired data. *Stat Med* **17:** 2635–2650.

Overbeek, R., Begley, T., Butler, R.M., Choudhuri, J.V., Chuang, H., Cohoon, M., *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.

Palenik, B., Ren, Q., Tai, V., and Paulsen, I.T. (2009) Coastal *Synechococcus metagenome* reveals major roles for horizontal gene transfer and plasmids in population diversity. *Environ Microbiol* **11:** 349–359.

Paredes-Sabja, D., Setlow, P., and Sarker, M.R. (2011) Germination of spores of *Bacillales* and *Clostridiales* species: mechanisms and proteins involved. *Trends Microbiol* **19:** 85–94.

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

Pedersen, K. (1993) The deep subterranean biosphere. *Earth-Sci Rev* **34:** 243–260.

Pedersen, K. (2000) Exploration of deep intraterrestrial microbial life: current perspectives. *FEMS Microbiol Lett* **185:** 9–16.

Ramšak, A., Peterka, M., Tajima, K., Martin, J.C., Wood, J., Johnston, M.E.A., *et al*. (2000) Unravelling the genetic diversity of ruminal bacteria belonging to the CFB phylum. *FEMS Microbiol Ecol* **33:** 69–79.

Rogers, J.R., and Bennett, P.C. (2004) Mineral stimulation of subsurface microorganisms: release of limiting nutrients from silicates. *Chem Geol* **203:** 91–108.

Rosario, K., Duffy, S., and Breitbart, M. (2009a) Diverse circovirus-like genome architectures revealed by environmental metagenomics. *J Gen Virol* **90:** 2418–2424.

Rosario, K., Nilsson, C., Lim, Y.W., Ruan, Y., and Breitbart, M. (2009b) Metagenomic analysis of viruses in reclaimed water. *Environ Microbiol* **11:** 2806–2820.

Sawant, A.A., Sordillo, L.M., and Jayarao, B.M. (2005) A survey on antibiotic usage in dairy herds in Pennsylvania. *J Dairy Sci* **88:** 2991–2999.

Sherr, E.B., and Sherr, B.F. (1991) Planktonic microbes: tiny ceils at the base of the ocean's food webs. *Trends Ecol Evol* **6:** 50–54.

Steube, C., Richter, S., and Griebler, C. (2009) First attempts towards an integrative concept for the ecological assessment of groundwater ecosystems. *Hydrogeol J* **17:** 23–35.

Storey, J.D., and Tibshirani, R. (2003) SAM thresholding and false discovery rates for detecting differential gene expression in DNA microarrays. In *The Analysis of Gene Expression Data: Methods and Software*. Parmigiani, G., Garrett, E.S., Irizarry, R.A., and Zeger, S.L. (eds). New York, USA: Springer, pp. 272–290.

Swindoll, M.C., Aelion, M.C., and Pfaender, F.K. (1988) Influence of inorganic and organic nutrients on aerobic biodegradation and on the adaptation response of subsurface microbial communities. *Appl Environ Microbiol* **54:** 212–217.

Teixeira de Mattos, M.J., and Neijssel, O.M. (1997) Bioenergetic consequences of microbial adaptation to low-nutrient environments. *J Biotechnol* **59:** 117–126.

Vreeland, R.H., Piselli, A.F., Jr, McDonnough, S., and Meyers, S.S. (1998) Distribution and diversity of halophilic bacteria in a subsurface salt formation. *Extremophiles* **2:** 321–331.

Warnecke, F., Luginbühl, P., Ivanova, N., Ghassemian, M., Richardson, T.H., Stege, J.T., *et al*. (2007) Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature* **450:** 560–569.

Whitman, W.B., Coleman, D.C., and Wiebe, W.J. (1998) Prokaryotes: the unseen majority. *Proc Natl Acad Sci USA* **95:** 6578–6583.

Willner, D., Thurber, R.V., and Rohwer, F. (2009) Metagenomic signatures of 86 microbial and viral metagenomes. *Environ Microbiol* **11:** 1752–1766.

Yilmaz, S., Allgaier, M., and Hugenholtz, P. (2010) Multiple displacement amplification compromises quantitative analysis of metagenomes. *Nat Methods* **7:** 943–944.

Zhang, X., Zhang, T., and Fang, H.H.P. (2009) Antibiotic resistance genes in water environment. *Appl Microbiol Biotechnol* **82:** 397–414.

Supporting information

Additional Supporting Information may be found in the online version of this article:

Table S1. Relative proportion of matches to the SEED database taxonomic hierarchy.

Table S2. Contribution of phyla level taxonomy to the dissimilarity of confined and unconfined aquifer metagenomes.
Table S3. Relative proportion of matches to a given subsystem hierarchy 1.
Table S4. Contribution of metabolic hierarchical 1 system to the dissimilarity of confined and unconfined aquifer metagenomes.
Table S5. Summary of publicly available metagenomes used in this study.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

OPEN ACCESS Freely available online

PLoS one

Substrate Type Determines Metagenomic Profiles from Diverse Chemical Habitats

Thomas C. Jeffries[1,2]*, Justin R. Seymour[2], Jack A. Gilbert[3,4,5], Elizabeth A. Dinsdale[6], Kelly Newton[1], Sophie S. C. Leterme[1,7], Ben Roudnew[1], Renee J. Smith[1], Laurent Seuront[1,7,8], James G. Mitchell[1]

1 School of Biological Sciences, Flinders University, Adelaide, South Australia, Australia, **2** Plant Functional Biology and Climate Change Cluster, University of Technology Sydney, Sydney, Australia, **3** Plymouth Marine Laboratory, Plymouth, United Kingdom, **4** Institute of Genomic and Systems Biology and Department of Biosciences, Argonne National Laboratory, Argonne, Illinois, United States of America, **5** Department of Ecology and Evolution, University of Chicago, Chicago, Illinois, United States of America, **6** Department of Biology, San Diego State University, San Diego, California, United States of America, **7** Aquatic Sciences, South Australian Research and Development Institute, Henley Beach, South Australia, Australia, **8** Centre National de la Recherche Scientifique, Paris, France

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.

Citation: 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(9): e25173. doi:10.1371/journal.pone.0025173

Editor: Purificación López-García, Université Paris Sud, France

Received January 17, 2011; **Accepted** August 29, 2011; **Published** September 23, 2011

Funding: This work was supported by an Australian Research Council (ARC) Discovery Grant (http://www.arc.gov.au/). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: jeffries.thomas@gmail.com

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 16 Mbp and 27 Mbp 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) and *Chroococcales*, which each comprised approximately 40% of cyanobacterial hits (Table S2).

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 genome level, which we used to compare the Coorong metagenomes to each other and to other metagenomes from diverse habitats (Fig. 3), and 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

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 ($\pm6.3\times10^5$)	7.4×10^7 ($\pm8.4\times10^6$)	7.2×10^7 ($\pm4.2\times10^6$)	1.5×10^8 ($\pm1.4\times10^7$)
Porewater virus concentration (per mL)	1.5×10^7 ($\pm5.8\times10^6$)	2.3×10^8 ($\pm3.1\times10^7$)	1.8×10^8 ($\pm1.5\times10^7$)	4.2×10^8 ($\pm3.1\times10^7$)
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.
doi:10.1371/journal.pone.0025173.t001

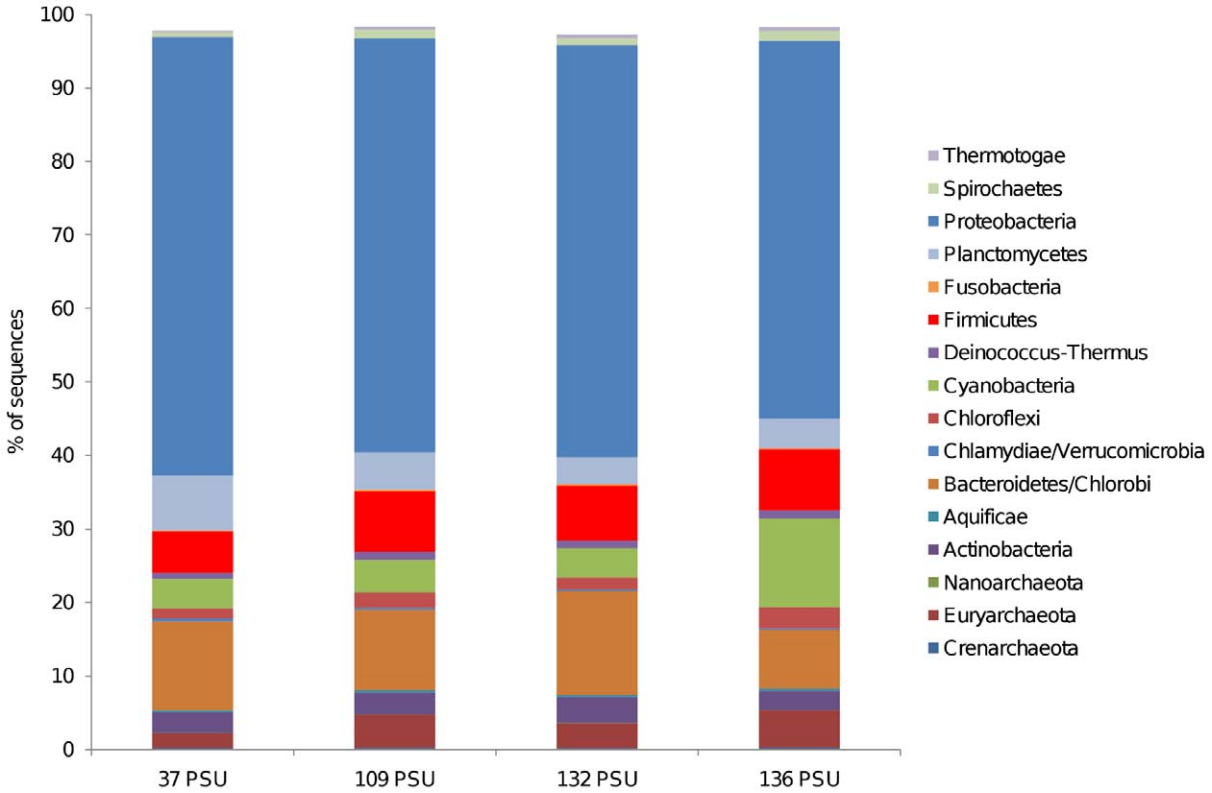


Figure 1. Taxonomic composition (Phyla level) of four metagenomic libraries derived from Coorong lagoon sediment. Relative representation in the metagenome was calculated by dividing the number of hits to each category by the total number of hits to all categories, thus normalizing by sequencing effort. 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 50 bp.
doi:10.1371/journal.pone.0025173.g001

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

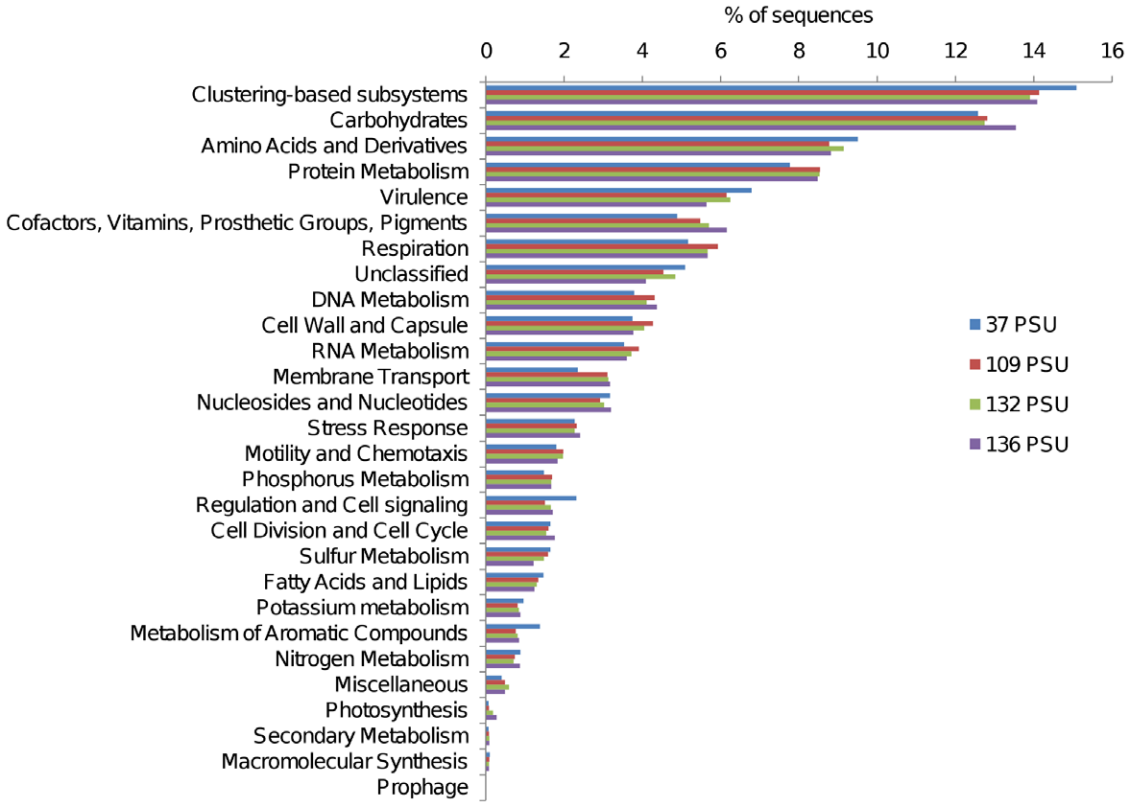


Figure 2. Metabolic composition of four metagenomic libraries derived from Coorong lagoon sediment. Relative representation in the metagenome was calculated by dividing the number of hits to each category by the total number of hits to all categories, thus normalizing by sequencing effort. Hits were generated by BLASTing sequences to the SEED database with an E-value cut-off of 1×10^{-5} and a minimum alignment of 50 bp.
doi:10.1371/journal.pone.0025173.g002

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

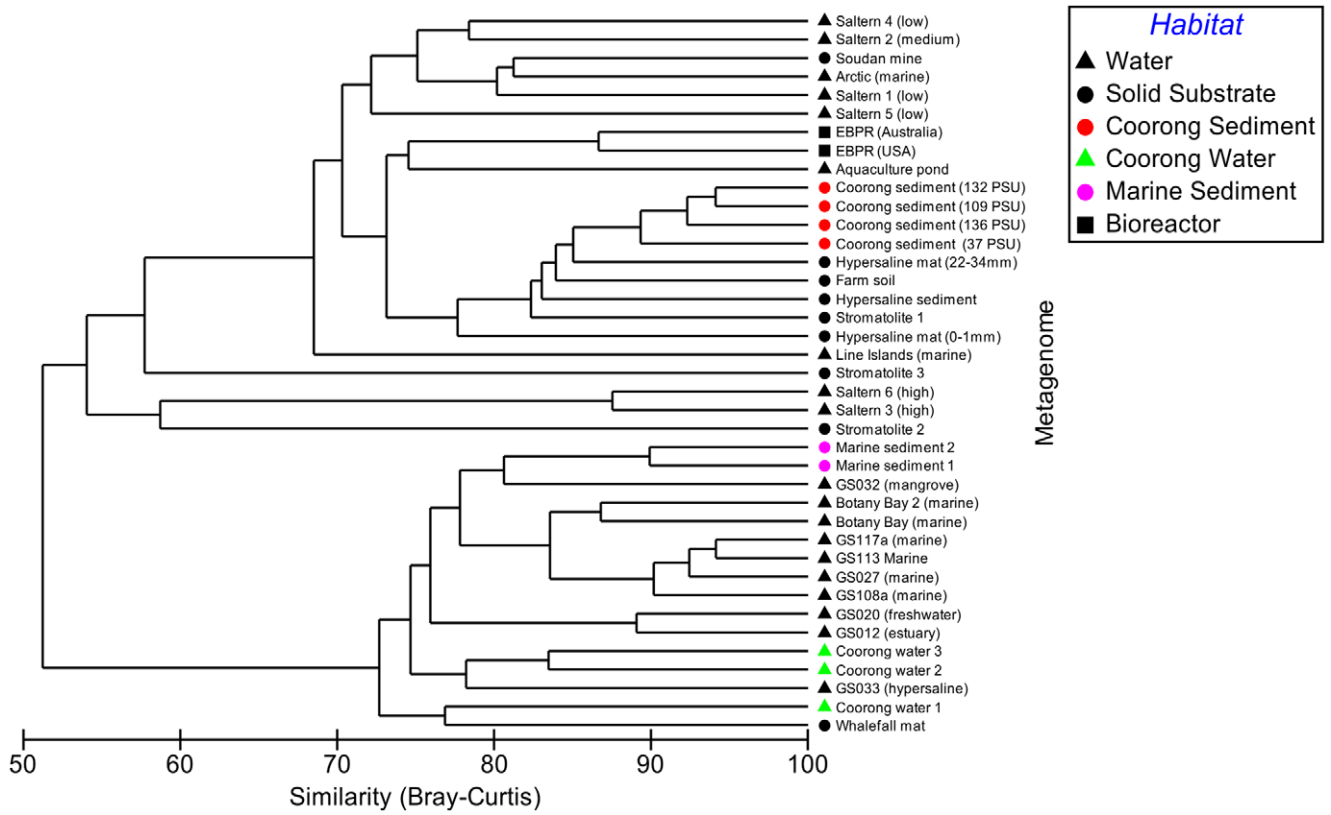


Figure 3. Comparison of taxonomic profiles derived from selected metagenomes publicly available on the MG-RAST database. The hierarchical agglomerative cluster plot (group average) is derived from a Bray-Curtis similarity matrix calculated from the square root transformed abundance of DNA fragments matching taxa in the SEED database (BLASTX E-value<0.001, genome level taxonomy).
doi:10.1371/journal.pone.0025173.g003

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

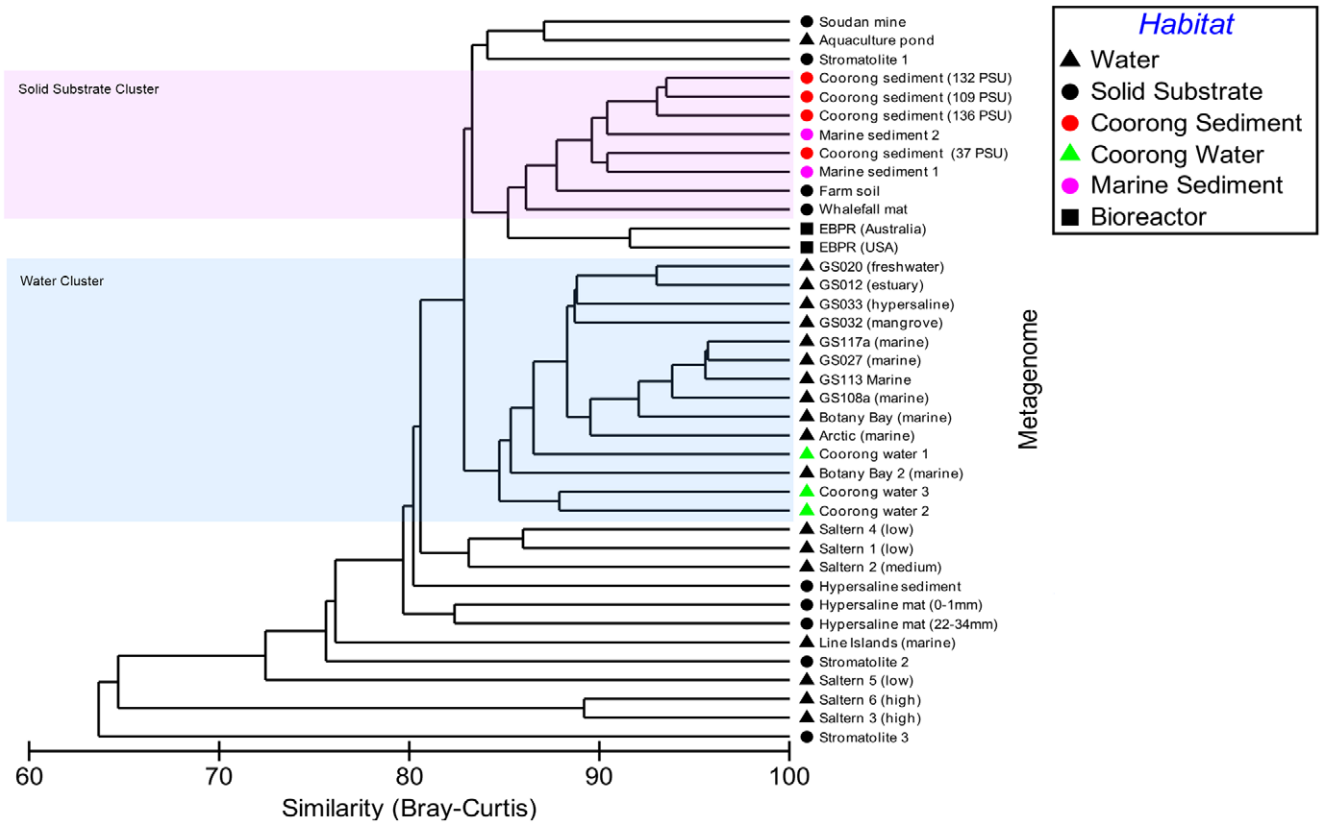


Figure 4. Comparison of metabolic profiles derived from selected metagenomes publicly available on the MG-RAST database. The hierarchical agglomerative cluster plot (group average) is derived from a Bray-Curtis similarity matrix calculated from the square root transformed abundance of DNA fragments matching subsystems in the SEED database (BLASTX E-value<0.001).
doi:10.1371/journal.pone.0025173.g004

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 50 mL 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

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 (22–34 mm) [29]
4441020.3	**Coorong sediment (109 PSU)**	4441584.3	GS012 (Estuary) [45]
4441021.3	**Coorong sediment (132 PSU)**	4441590.3	GS020 (freshwater) [45]
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) [5,28]	4443689.3	Botany Bay 2 (marine)
4440435.3	Saltern 2 (medium) [5,28]	4440041.3	Line Islands (marine) [46]
4440438.3	Saltern 3 (high) [5,28]	4440212.3	Arctic (marine) [47]
4440437.3	Saltern 4 (low) [5,28]	4440440.3	Aquaculture pond [5]
4440426.3	Saltern 5 (low) [5,28]	4440281.3	Soudan mine [48]
4440429.3	Saltern 6 (high) [5,28]	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 (0–1 mm) [29]		

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.

doi:10.1371/journal.pone.0025173.t002

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.10 g 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 >5 μg 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 16 Mbp and 27 Mbp of sequence information per library, with an average read length of 232.5 bp (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 50 bp. 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 50 bp [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.

Supporting Information

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 50 bp. Due to inconsistencies in 16S rDNA copy number, these relative abundances represent estimates of overall ribosomal DNA composition at phyla level only.
(DOC)

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 50 bp.
(XLS)

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.
(XLS)

Acknowledgments

We thank the anonymous reviewers for their feedback and suggestions.

Author Contributions

Conceived and designed the experiments: TCJ LS JGM. Performed the experiments: TCJ JRS KN SSCL BR RJS. Analyzed the data: TCJ JRS JAG EAD. Wrote the paper: TCJ JRS JGM.

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 USA 104: 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.
16. Segar DA (1998) Introduction to ocean sciences. BelmontCA: Wadsworth Pub. xxiii, 497, 489 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. (2010) 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 JC, Barberan A, Casamayor EO (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, 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 TO, Malandain C, Prestat E, Larose C, Monier J-M, et al. (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.

Biogeochemistry (2007) 83:173–188
DOI 10.1007/s10533-007-9097-z

ORIGINAL PAPER

The influence of *Phaeocystis globosa* on microscale spatial patterns of chlorophyll *a* and bulk-phase seawater viscosity

L. Seuront · C. Lacheze · M. J. Doubell · J. R. Seymour · V. Van Dongen-Vogels · K. Newton · A. C. Alderkamp · J. G. Mitchell

Received: 17 October 2005 / Accepted: 8 May 2006 / Published online: 27 April 2007

Abstract A two-dimensional microscale (5 cm resolution) sampler was used over the course of a phytoplankton spring bloom dominated by *Phaeocystis globosa* to investigate the structural properties of chlorophyll *a* and seawater excess viscosity distributions. The microscale distribution patterns of chlorophyll *a* and excess viscosity were never uniform nor random. Instead they exhibited different types and levels of aggregated spatial patterns that were related to the dynamics of the bloom. The chlorophyll *a* and seawater viscosity correlation patterns were also controlled by the dynamics of the bloom with positive and negative correlations before and after the formation of foam in the turbulent surf zone. The ecological relevance and implications of the observed patchiness and biologically induced increase in seawater viscosity are discussed and the combination of the enlarged colonial form and mucus secretion is suggested as a competitive advantage of *P. globosa* in highly turbulent environments where this species flourishes.

Keywords Eastern English Channel · Patchiness · *Phaeocystic globosa* · Plankton rheology · Turbulence

L. Seuront (✉) · C. Lacheze
CNRS FRE 2816 ELICO, Station Marine de Wimereux, University of Sciences and Technologies of Lille-Lille 1, 28 avenue Foch, 62930 Wimereux, France
e-mail: Laurent.Seuront@univ-lille1.fr

L. Seuront · M. J. Doubell · K. Newton · J. G. Mitchell
School of Biological Sciences, Flinders University, GPO Box 2100, Adelaide, SA 5001, Australia

J. R. Seymour
Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, 77 Massachussets Avenue, Cambridge, MA 02139-4307, USA

V. Van Dongen-Vogels
Marine Biology Laboratory, Catholic University of Louvain, Batiment Kellner, 3 Place Croix du Sud, Louvain-La-Neuve 1348, Belgium

A. C. Alderkamp
Department of Marine Biology, University of Groningen, PO Box 14, Haren 9750AA, The Netherlands

Introduction

Plankton patchiness is widely acknowledged as a ubiquitous and key feature of marine ecosystems (Martin 2003). Many organisms have been shown to exploit patches of food (e.g., Tiselius 1992), and patch formation may be important in the foraging success of many marine invertebrates (Seuront et al. 2001) and vertebrates (Cartamil and Lowe 2004), as well as for the sexual encounters among individuals of relatively rare species (Buskey 1998). While the quantification of the spatial and temporal structure of phytoplankton distributions has for the most part focussed on empirical observations at scales greater

than 10 m to several kilometres (Martin 2003), there is a growing evidence for the existence of both vertical and horizontal phytoplankton patchiness at scales smaller than one metre (e.g., Dekshenieks et al. 2001). Because the analysis of large-scale patterns must integrate processes occurring at much smaller scales (Levin 1992), investigating the physical and biological dynamics controlling microscale plankton patchiness may be an absolute prerequisite to improve our understanding of the scales, intensities, durations and ecological relevance of these patterns.

The cosmopolitan genus *Phaeocystis* is a key organism in driving global geochemical cycles, climate regulation and fisheries yield (Schoemann et al. 2005), and has recently been suggested as a potential source of microscale phytoplankton patchiness (Seuront 2005). Of particular importance is the formation of large, gel-like colonies (from several mm to several cm; Verity and Medlin 2003) which form thick brown jelly layers (Al-Hasan et al. 1990) that modify the rheological properties of seawater, produce foam that resembles whipped egg white (Dreyfuss 1962), and clog plankton and fish nets (Peperzak 2002). Large quantities of foam are generated in the turbulent surf zone of beaches along the North Sea and the Eastern English Channel (Lancelot et al. 1987; Seuront et al. 2006). Quantitative descriptions of the bulk-phase properties of seawater during phytoplankton blooms have estimated the changes in the viscous properties of seawater that are induced by phytoplankton mucus secretion (Jenkinson 1986, 1993; Jenkinson and Biddanda 1995; Seuront et al. 2006), and found a positive correlation between seawater viscosity and chlorophyll *a* concentration during *Phaeocystis* sp. blooms in the German Bight and the North Sea (Jenkinson 1993; Jenkinson and Biddanda 1995). More recent work (Seuront et al. 2006) showed a significant increase in seawater viscosity with the development of a *Phaeocystis globosa* bloom in the Eastern English Channel, and a positive correlation between chlorophyll concentration and seawater viscosity before foam formation as well as a negative correlation after foam formation. These results suggest that seawater viscosity is influenced by extracellular materials associated with colony formation and colony maintenance rather than by cell composition and standing stock.

In this context, the aim of this study is to use a microscale (5 cm resolution) two-dimensional sampler to investigate: (i) the heterogeneity of the microscale distributions of phytoplankton biomass and bulk-phase seawater viscosity, (ii) the potential changes in these distributions over the course of a *P. globosa* spring bloom and (iii) the nature of the correlation between seawater viscosity and phytoplankton before, during and after the formation of foam in the turbulent surf zone.

Materials and methods

Study site

Samples were taken from the end of a coastal pier situated in the intertidal waters of the French coast of the Eastern English Channel, approximately 20 m from the shoreline (50°45′896″ N, 1°36′364″ E), near Boulogne-sur-Mer (France). The Eastern English Channel is characterised by 3 to 9 m tides. These tides are characterised by a residual translation parallel to the coast, with nearshore coastal waters drifting from the English Channel into the North Sea. Coastal waters are influenced by freshwater run-off from the Seine estuary to the Straits of Dover. This coastal flow (Brylinski et al. 1991) is separated from offshore waters by a tidally maintained frontal area. The coastal flow water mass is characterised by its low salinity, high turbidity, high phytoplankton richness and high productivity compared to the oceanic offshore waters (Seuront 2005). This sampling site was chosen because the physical and hydrological properties are representative of the inshore water masses of the Eastern English Channel (Seuront 2005). Sampling was conducted at high tide, before, during and after the *Phaeocystis globosa* bloom in the Eastern English Channel on 26 September 2003, 5 March 2004, 18 March 2004, 1 April 2004, 20 April 2004, 10 May 2004, 25 May 2004 and 20 June 2004.

Microscale sampling device

A purpose-designed two-dimensional (2D) sampling device was developed to investigate the microscale distributions of chlorophyll *a* and seawater viscosity.

The design of the device was based on sampling systems previously employed to measure the microscale spatial distributions of bacteria and phytoplankton (Seymour et al. 2000; Waters et al. 2003) and consisted of a 10 × 10 array of 50 ml syringes, separated by 5 cm and with a 3 mm wide opening. The pneumatically operated device was attached to a series of electric pumps allowing for the simultaneous collection of 100 samples across an area of 0.2 m^2. On retrieval, a polyvinyl chloride (PVC) rack holding 60 ml vials was inserted below the syringe array on stainless-steel rails, and the samples were transferred into the vials, which were immediately brought back to the laboratory in a cooler, stored in the dark in a room at the in situ temperature; all subsequent analysis were conducted within 2 h to avoid potential biases related to exudation during temperature changes and in high light. Samples were taken from a depth of 1 m, and the temperature and salinity of each water sample was measured using a Hydrolab probe at each of the sampling dates. The presence and the size of *P. globosa* colonies were investigated from bulk seawater samples taken simultaneously to the microscale samples.

Chlorophyll *a* analysis

Chlorophyll concentrations were estimated from 40 ml subsamples following Suzuki and Ishimaru (1990). Samples were filtered on Whatman GF/F glass-fibre filters (pore size 0.45 µm). Chlorophyllous pigments were extracted by direct immersion of the filters in 5 ml of N,N-dimethylformamide, and actual extractions were made in the dark at −20°C. Concentrations of chlorophyll *a* in the extracts were determined following Strickland and Parsons (1972) using a Turner 450 fluorometer previously calibrated with chlorophyll *a* extracted from *Anacystis nidulans* (Sigma Chemicals, St Louis).

Bulk-phase seawater viscosity measurements

Viscosity measurements were conducted using a portable ViscoLab400 viscometer (Cambridge Applied Systems Inc., Boston) from 10 ml subsamples following Seuront et al. (2006). Viscosity was measured in triplicate from 3 ml water subsamples poured into a small chamber, where a low-mass stainless-steel piston is magnetically forced back and forth, with a 230 µm piston-cylinder gap size. The force driving the piston is constant, and the time required for the piston to move back and forth into the measurement chamber is proportional to the viscosity of the fluid. The more viscous the fluid the longer it takes the piston to move through the chamber, and vice versa. As viscosity is influenced by temperature and salinity (Miyake and Koizumi 1948), the measured viscosity η_m (cP) can be thought of as the sum of a physically controlled viscosity component $\eta_{T,S}$ (cP) and a biologically controlled viscosity component η_{Bio} (cP) (Jenkinson 1986):

$$\eta_m = \eta_{T,S} + \eta_{\mathrm{Bio}} \tag{1}$$

The physically controlled component $\eta_{T,S}$ was estimated from viscosity measurements conducted on subsamples after passing through 0.20 µm pore size filters. The biologically induced excess viscosity η_{Bio} (cP) was subsequently defined from each water sample as:

$$\eta_{\mathrm{Bio}} = \eta_m - \eta_{T,S} \tag{2}$$

The relative excess viscosity η (%) is finally given as (Seuront et al. 2006):

$$\eta = \left(\eta_m - \eta_{T,S}\right)/\eta_{T,S} \tag{3}$$

Before each viscosity measurement, temperature and salinity of the water sample were simultaneously measured using a Hydrolab probe, and did not exhibit any significant difference with the theoretical viscosity estimated from temperature and salinity (Wilcoxon–Mann–Whitney *U*-test, $P > 0.05$). This ensures the relevance of our $\eta_{T,S}$ estimates. Between each viscosity measurement, the viscometer chamber was carefully rinsed first with deionised water and then with bulk-phase seawater filtered through 0.2 µm pore size filters to avoid any potential dilution of the next sample.

Potential biases and limitations

Due to the difficulty in handling *Phaeocystis* sp. colonies without disrupting them, our viscosity measurements could have been biased through the disruption of colonies: (i) during the sampling process and/or (ii) the viscosity measurement in the

chamber of the viscometer. This would result in creating additional mucus particles and releasing additional macromolecules into solution, and thus a biased increase in seawater viscosity. These two potential issues have been carefully addressed to ensure the relevance of our viscosity measurements.

Any *P. globosa* colony larger than the opening of the syringes (i.e. 3 mm) would be destroyed during the sampling. This is unlikely to have occurred during our sampling for two reasons. First, the size of *P. globosa* colonies in bulk samples taken simultaneously to our 2D sampling, observed by inverted microscopy and measured using an ocular micrometer, ranged in size from 0.1 to 1.3 mm at concentrations ranging from 500 to 5,000 per litre (Table 1). Control measurements were carried out at each of our sampling dates, during which we compared the viscosity measured from water samples taken by the syringes of the 2D sampler and from bulk seawater samples. As no significant differences were found between the two estimates (Wilcoxon–Mann–Whitney U-test, $P > 0.05$, $N = 100$), the syringe sampler has not been considered as a potential source of bias in subsequent viscosity measurements.

Fragile colonies, and more specifically the colonies larger than 230 µm (as the piston-cylinder gap is only 230 µm; see above) would still be destroyed in the measurement chamber, and would result in a biased increase in seawater viscosity. Therefore all visible *P. globosa* colonies were systematically removed from the 3 ml samples with a Pasteur pipette under a dissecting microscope (magnification × 100), under temperature controlled conditions before viscosity measurements were conducted. Considering the relatively large colonies at relatively low concentrations (500 to 5000 per litre; Table 1) we can assume that most though maybe not all colonies were removed. It could also be argued here that this could have been achieved through a careful screening of the water samples through a 200 µm mesh before doing the measurements. However, a 200 µm screening would still not remove the colonies smaller than 200 µm that are still likely to be destroyed by the high shear occurring in the measurement chamber and to create additional mucus particles and release additional macromolecules into solution. The two methods are then both potentially biased, but in similar ways, as preliminary testing conducted during the 2002 and 2003 spring blooms did not exhibit any significant difference (Fig. 1). Finally, we have already shown a distinct increase in excess seawater viscosity following the mechanical disruption of *P. globosa* colonies (Seuront et al. 2006). This would not have been possible if just the act of making a measurement increased viscosity.

Table 1 Mean abundance (colonies l^{-1}) and size (µm) of *P. globosa* colonies observed at each sampling date

Date	*Phaeocystis globosa* colonies			
	Size (µm)	N_1	Abundance (col l^{-1})	N_2
26/09/03	–	–	0	–
05/03/04	–	–	0	–
10/03/04	–	–	0	–
01/04/04	180 (100–220)	100	1400(1200–1600)	5
20/04/04	270(230–660)	100	1200(990–1350)	4
10/05/04	500 (300–1300)	100	5000(4820–5250)	4
25/05/04	390 (300–645)	100	500(290–560)	4
20/06/04	–	–	0	–

The values given in parenthesis are the minimum and maximal values. N_1 is the number of colonies considered for size measurements and N_2 is the number of bulk samples used to estimate colony abundance

Data analyses

Quantifying microscale spatial variability. A simple quantification of the degree of microscale variability observed for a given 2D microscale distribution was expressed as the ratio between maximum and minimum values of the distributions, considered as an estimate of the maximum variability (Seuront and Spilmont 2002), and the coefficient of variation CV (CV = SD/$\bar{x}$, where$\bar{x}$ and SD are the mean and the standard deviation, respectively) estimated for that range.

Identifying spatial structure

Spatial autocorrelation analysis was used to quantify patterns in two-dimensional microscale distribution of chlorophyll *a* concentrations (µgChl-a l^{-1}) and relative seawater excess viscosity η (%). The objective is to link these parameters to a common patch size to quantify their spatial linkage.

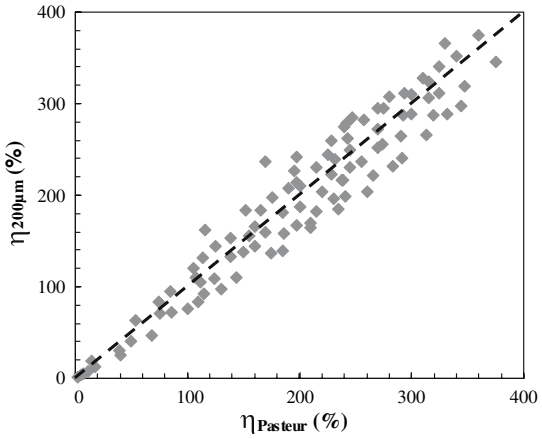


Fig. 1 Comparisons of the excess viscosity measurements η (%) obtained from bulk phase seawater samples after removing all the visible colonies from the water samples with a Pasteur pipette under a dissecting microscope (magnification × 100), η_{Pasteur}, and after carefully screening the water samples through a 200 µm mesh ($\eta_{200\mu m}$) before doing the measurements. A modified *t*-test (Zar 1996) did not exhibit a significant difference between our measurements and the theoretical expectation $\eta_{\text{Pasteur}} = \eta_{200\mu m}$ ($P > 0.05$, $N = 100$)

Generally speaking, positive spatial autocorrelation indicates that abundances in adjacent localities are similar (aggregation) while negative-autocorrelation signifies that adjacent localities alternate between high and low values, i.e. peaks and valleys, and hot spots and cold spots. The theory and use of spatial autocorrelation procedures to analyse spatial associations has been extensively discussed (Sokal and Oden 1978a, b), and the reader is referred to these papers for more-detailed descriptions.

Moran's I and Geary's c spatial autocorrelation statistics (Moran 1950; Geary 1954) are defined as:

$$I(d) = \frac{\frac{1}{W}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(y_i - \bar{y})(y_j - \bar{y})}{\frac{1}{n}\sum_{i=1}^{n}(y_j - \bar{y})^2} \text{ for } i \neq j \qquad (4)$$

and

$$c(d) = \frac{\frac{1}{2W}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(y_i - y_j)^2}{\frac{1}{(n-1)}\sum_{j=1}^{n}(y_j - \bar{y})^2} \text{ for } i \neq j \qquad (5)$$

where y_i and y_j are the values of the observed variable at sites i and j, $\bar{y}$ is their mean and W is defined as $W = \sum_i^n \sum_j^n w_{ij}$ where $w_{ij} = d_{ij}^{-2}$ when sites i and j are at a distance d and $w_{ij} = 0$ otherwise. Only the pairs of site (i, j) within the stated distance class (d) are taken into account.

Moran's I Eq. 4 and Geary's c Eq. 5 are two related, but conceptually different, autocorrelation coefficients. Moran's I, which ranges from +1 to −1, is sensitive to extreme abundances in nearby localities while Geary's c, bounded between 0 and an unspecified value larger than 1, indicates whether nearby localities exhibit similar abundances. A positive autocorrelation (i.e. positive values of I) indicates dependence between samples that are spatially close. In contrast, a negative autocorrelation (i.e. negative values of I) indicates dependence between distant samples. The no-correlation values is thus $I = 0$. A Geary's c correlogram varies as the reverse of a Moran's I correlogram: strong autocorrelation produces high values of I and low values of c. Positive autocorrelation translates into values of c between 0 and 1 whereas negative autocorrelation produces values larger than 1. Hence, the reference no-correlation value is $c = 1$.

Under the hypothesis of a random spatial distribution, the expected values of the Moran's I and Geary's c are $I = -(n-1)^{-1}$ and $c = 1$, respectively. When combined with the Fisher's dispersion index ($S^2/\bar{x}$ or variance-to-mean ratio), Moran's I and Geary's c were used to characterise and classify in an objective way the spatial pattern of a particular descriptor (see Table 2). In addition, patch sizes were determined from spatial correlograms, i.e. Moran's I plotted against the separation distance d. As a minimum number of 30 pairs are necessary to estimate autocorrelation statistics with confidence (Legendre and Legendre 1998), nine distance classes were used, with distances ranging from 5 to 45 cm. The significance of spatial correlograms was tested using the Bonferroni-corrected significance procedure (Legendre and Legendre 1998). For each 2D data set, the distance at which I changes sign was used as an estimate of the average patch radius (Sokal and Oden 1978a).

Statistical analyses

For each date, the chlorophyll *a* distribution and seawater viscosity data were significantly non-normally distributed (Kolmogorov–Smirnov test,

Table 2 Types of distribution patterns based on values of Moran's *I*, Geary's *c* and variance-to-mean ratio. S indicates that the value is significantly different from the expected value, while NS denotes no significant difference

Type	*I*	*c*	$S^2/\bar{x}$	Distribution
A	S	S	–	Transition from high to low abundance occurs within the sampling area, i.e. gradient in abundance
B	S−	S−	S−	Uniform
C	S	NS	–	Aggregative, based on a few extreme peaks in abundance
D	NS	S	–	Aggregative, based on several similar peaks in abundance
E	NS	NS	S+	Aggregative, due to high abundance in a single, isolated location
F	NS	NS	NS	Random

The + and − signs indicate whether the value is higher or lower than the expected value, respectively. Conclusions are based on discussions contained in Sokal and Oden (1978a), and Decho and Fleeger (1988)

$P < 0.01$), so nonparametric statistics were used throughout this work. Simple comparisons between two parameters were conducted using the Wilcoxon–Mann–Whitney (WMW) test. Correlation between variables was investigated using Kendall's coefficient of rank correlation, τ. Kendall's coefficient of correlation was used in preference to Spearman's coefficient of correlation ρ because Spearman's ρ gives greater weight to pairs of ranks that are further apart, while Kendall's τ weights each disagreement in rank equally (Sokal and Rohlf 1995).

Results

Environmental conditions

Salinity did not exhibit any characteristic pattern and fluctuated between 35.1 and 35.3 PSU (35.20 ± 0.02 PSU; $\bar{x} \pm$ SD). Temperature exhibited a clear seasonal cycle, ranging from 6.1°C on March 5 to 17.2°C on September 26.

Microscale spatial variability

The temporal patterns of chlorophyll *a* concentration and seawater excess viscosity are summarised in Table 3. Chlorophyll concentration and seawater excess viscosity were below 3.0 µg Chl-*a* l^{-1} on September 26 and March 5 (Table 3). Chlorophyll concentration then exhibited a sharp increase, corresponding to the initiation and the development of the spring bloom, to reach a maximum value of 57.4 ± 6.5 µg Chl-*a* l^{-1} ($\bar{x} \pm$ SD) on May 10, followed by a 3.5-fold decrease between May 10 and May 25 that coincided with the formation of foam in the turbulent surf zone. In contrast, seawater excess viscosity reached its peak (28 ± 25%;$\bar{x} \pm$ SD) on May 25 after the beginning of foam formation, and finally sharply decreased down to 5 ± 3% on June 20.

The values of the coefficient of variation (CV) and the ratios between maximum and minimum values (max./min.) demonstrate the high variability of microscale 2D distributions of chlorophyll *a* biomass and seawater excess viscosity (Table 3). The variability of chlorophyll and excess viscosity 2D distributions were minimum during the formation of foam in the turbulent surf zone. As further illustration of this variability, the 5 cm resolution, two-dimensional distributions of chlorophyll *a* concentration and seawater excess viscosity are shown for the pre-bloom condition (Fig. 2a), during the spring bloom, before (Fig. 2b) and after (Fig. 2c) the formation of foam in the turbulent surf zone, and after the bloom (Fig. 2d). Chlorophyll concentration and excess viscosity show alternating high- and low-density areas separated by sharp gradients and characterised by localised hot spots and cold spots (Fig. 2). The differential effect of foam formation on the two-dimensional distributions of chlorophyll *a* and excess viscosity is also clear from the transition observed between the pre- and post-foam distributions (Fig. 2b, c). Before the formation of foam, chlorophyll *a* and excess viscosity distributions are dominated by a high-density background and a few cold spots (Fig. 2b), while after the formation of foam, the chlorophyll *a* distributions were dominated by a

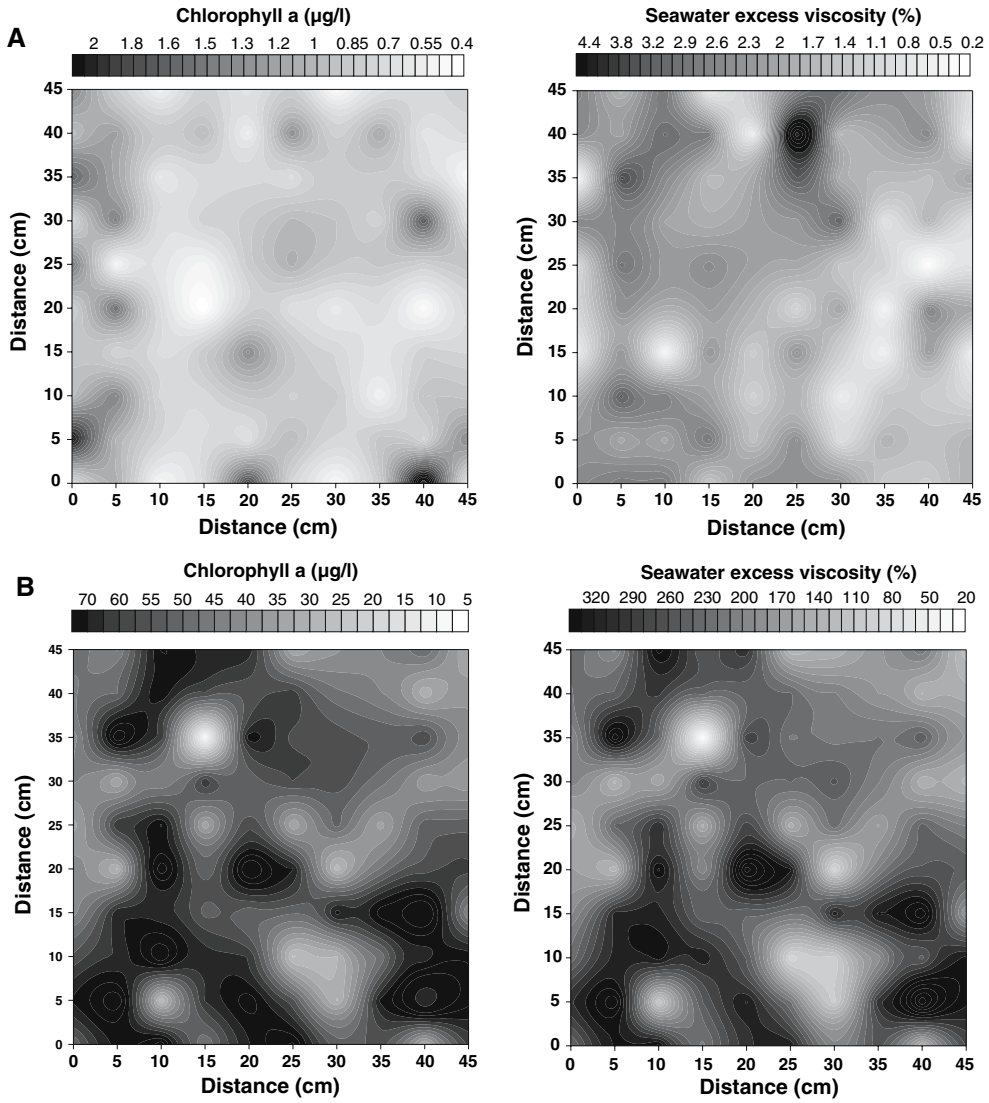


Fig. 2 Two-dimensional distributions of chlorophyll *a* concentrations (µg l^{-1}) and seawater excess viscosity (%) before the spring phytoplankton bloom (**A**) and during the spring bloom before foam formation in the turbulent surf zone (**B**). Two-dimensional distributions of chlorophyll *a* concentrations (µg l^{-1}) and seawater excess viscosity (%) during the spring phytoplankton bloom after foam formation in the turbulent surf zone (**C**), and after the bloom (**D**)

low-density background and a few hot spots (Fig. 2c). Finally, the chlorophyll biomass and excess viscosity do not show any clear correlation pattern in pre- and post-bloom conditions (Fig. 2a, d). In contrast, they exhibit positive and negative correlation before and after the formation of foam, respectively (Fig. 2b, c).

Microscale spatial correlation

To identify any possible correlation between chlorophyll concentration and excess viscosity, excess viscosity was plotted against chlorophyll *a* concentration (Fig. 3). In pre-bloom conditions (September 26), no relationship is visible between chlorophyll concentration and excess viscosity. The development of the spring bloom is subsequently clearly visible, with simultaneous increases in chlorophyll *a* concentration and excess viscosity from March 5 to April 20. After the formation of foam, a progressive decoupling between chlorophyll concentration and excess viscosity occurred. It was initiated on May 10, when positive and negative correlations were observed

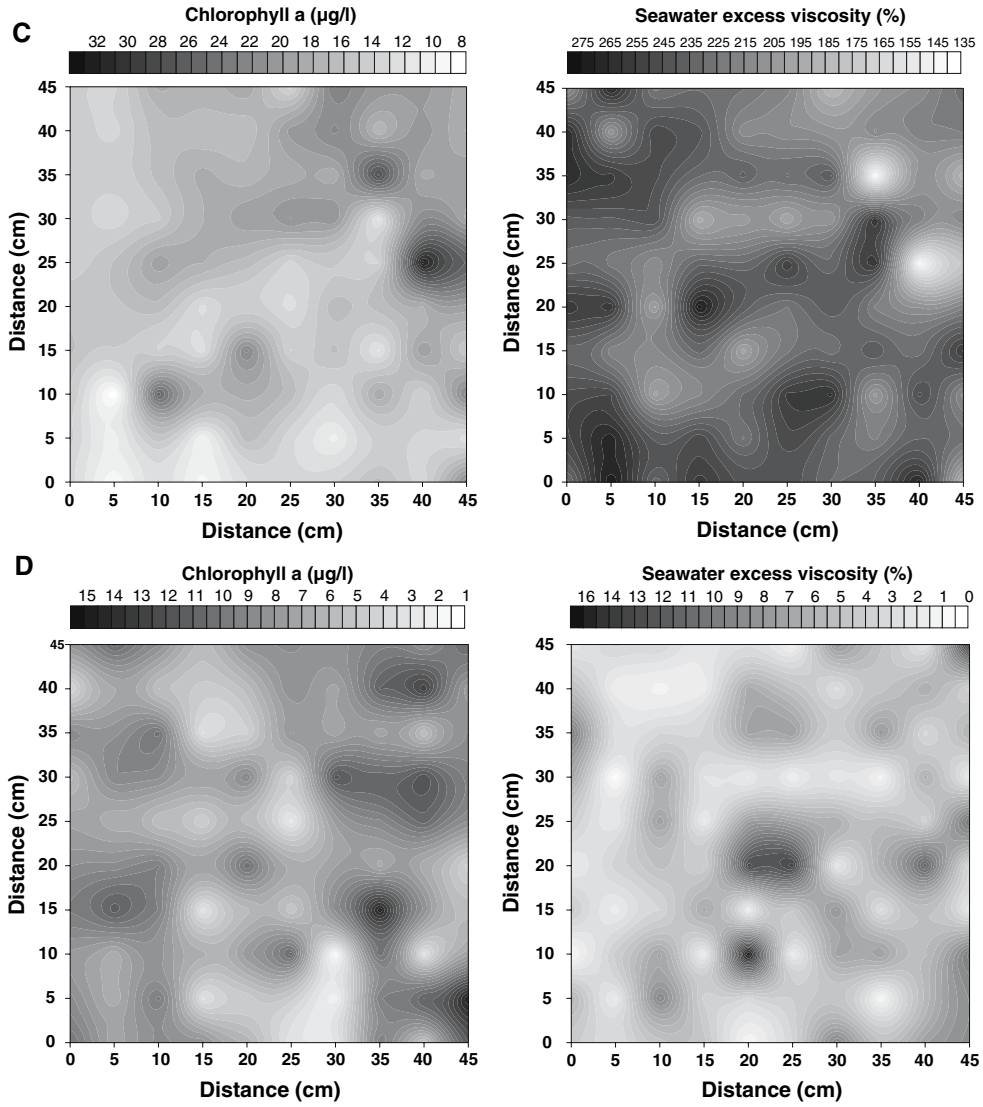


Fig. 2 continued

between chlorophyll concentration and excess viscosity for chlorophyll concentrations higher and lower than 55 µg Chl-*a* l^{-1}, respectively. On May 25, accompanying a sharp decrease in Chl-*a* concentration, excess viscosity was negatively related to chlorophyll concentration. Finally, in post-bloom conditions (June 20), chlorophyll concentration and excess viscosity appeared independently distributed. These observations are specified by the values of Kendall rank correlation coefficients (Fig. 4) estimated at each date between the two distributions. No significant correlation was observed in pre- and post-bloom conditions, while significant positive and negative correlations were observed during the spring bloom before and after the formation of foam, respectively (Fig. 4).

Microscale spatial structure

The results of spatial autocorrelation analysis showed that none of the investigated patterns were uniform nor random (Table 4), indicating the existence of structural complexity in 2D microscale patterns of chlorophyll *a* concentration and seawater viscosity. Except on June 20, consistent spatial patterns were found for chlorophyll and excess

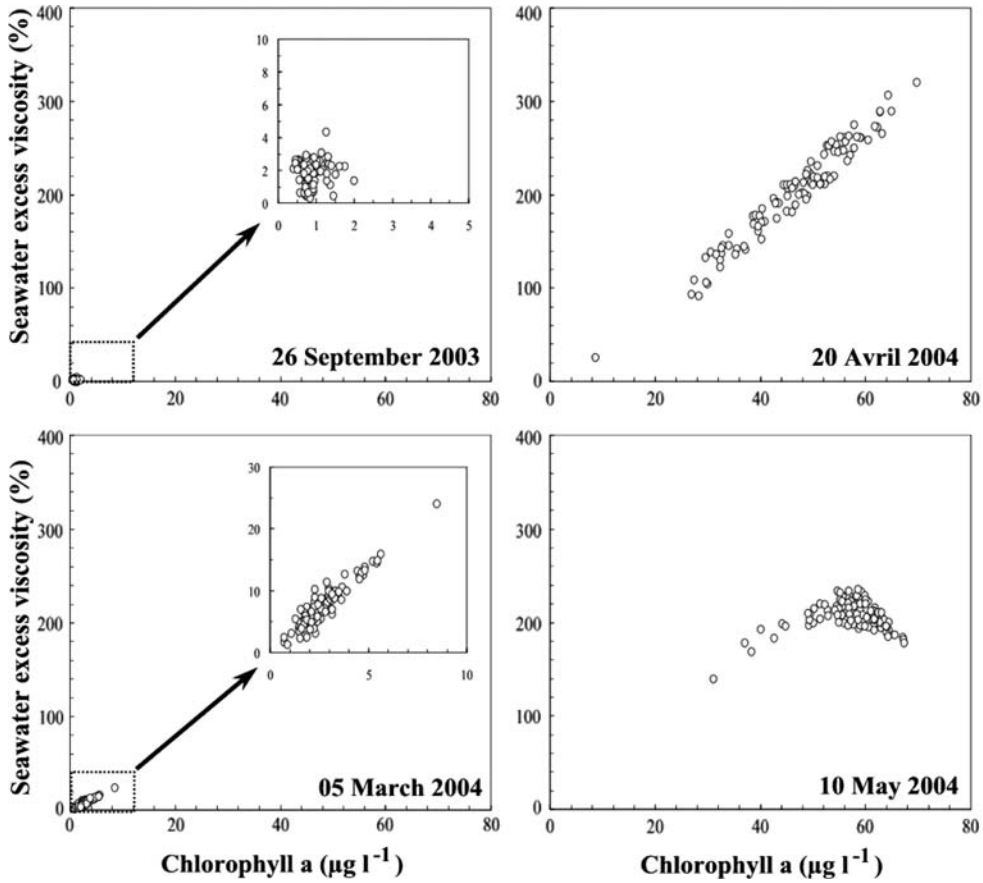


Fig. 3 Scatterplots of seawater excess viscosity η (%) as a function of chlorophyll *a* concentration (µg l^{-1}), showing the two different regimes observed over the course of our survey. The sampling conducted on 26 September has been used as a separate reference sample, characteristic of non-bloom conditions. Scatterplots of seawater excess viscosity η (%) as a function of chlorophyll *a* concentration (µg l^{-1}), showing the two different regimes observed over the course of our survey

viscosity. More specifically, pre- and post-bloom conditions were mainly characterised by three types of aggregative distributions (Table 4). In contrast, the spatial patterns observed during the bloom were either aggregative (i.e. dominated by the presence of a few extreme peaks) or characterised by a variety of transitions from high to low values, the latter being only observed during the formation of foam. Patch sizes were finally estimated from correlograms constructed for each parameter at each date to determine spatial autocorrelation as a function of increasing distance between samples. Because the distance intervals were based on the distance intervals between the centre points of samples, patch sizes were expressed in terms of the distance interval in which the autocorrelation value changed sign (positive to negative autocorrelation or vice versa). Patch sizes for chlorophyll concentration and excess viscosity were between 5 and 10 cm in pre- and post-bloom conditions (Table 4). Larger patch sizes were found for chlorophyll concentration and excess seawater viscosity in bloom conditions. During the formation of foam viscous patches were larger than chlorophyll patches (Table 4).

Discussion

Microscale spatial patterns and Phaeocystis globosa bloom dynamics

Chlorophyll, viscosity and foam formation. The shift from significant positive correlations to significant

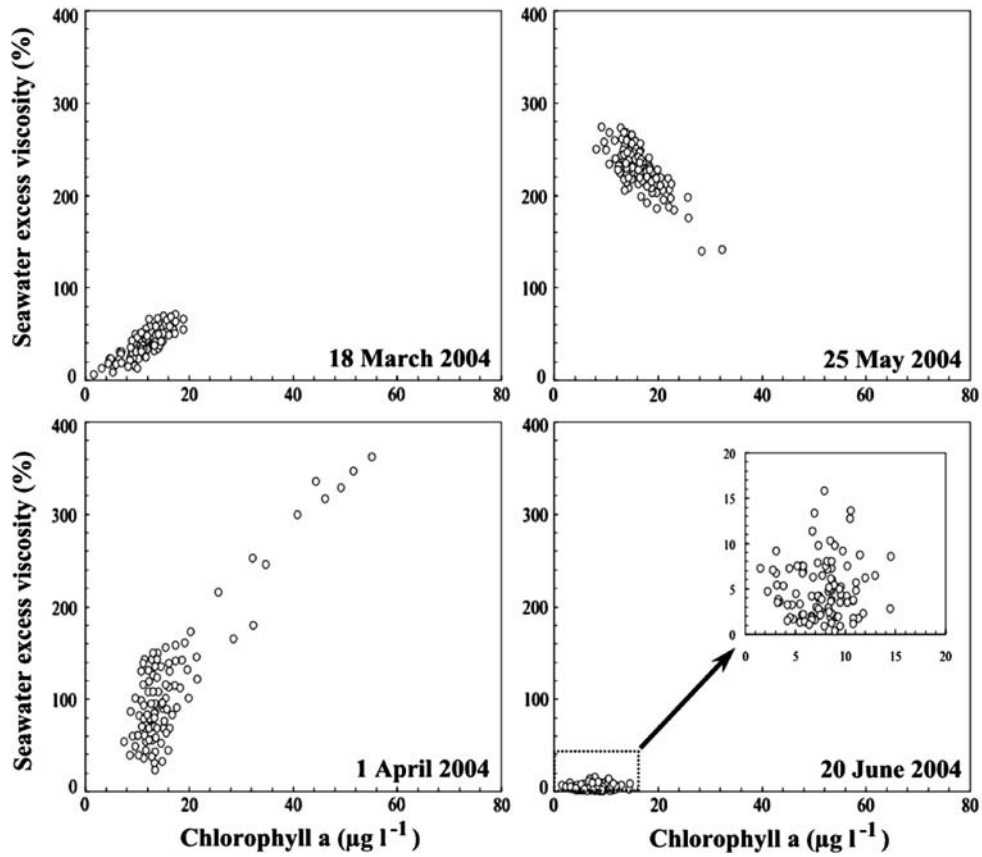


Fig. 3 continued

negative correlations before and after the foam formation is congruent with a recent mechanistic explanation (Seuront et al. 2006), suggesting that the disruption of the mucilaginous colonial matrix by turbulent mixing in the surf zone leads: (i) to the formation of foam and to the transformation of colonial cells into flagellated ones (Peperzak 2002), (ii) to a decrease in chlorophyll *a* concentration (Fig. 5) as a significant proportion of cells are entrained within the foam during the emulsion process (Seuront et al. 2006), and finally (iii) to the decoupling between the viscous (i.e. colonial polymeric materials) and nonviscous (flagellated cells) contribution of *P. globosa* to bulk-phase seawater properties in intertidal (Fig. 5a, b) and inshore (Fig. 5c, d) water masses, located respectively 20 m and 2 nautical miles from the shoreline. However, the relationship observed between chlorophyll concentration and excess viscosity on May 10 through our microscale sampling strategy (Fig. 3) indicates that the aforementioned decoupling process could be more complex than initially thought. As the negative correlation occurs for the chlorophyll *a* concentrations higher than 55 µg l^{-1}, the high-density chlorophyll patches, likely to be the largest and/or the oldest ones, may be more fragile and thus the first ones to be destroyed by turbulent mixing. This hypothesis is consistent with the strong decrease in variability and the changes in the distribution patterns observed in both chlorophyll and excess viscosity spatial patterns between April 20 and May 10 (see Tables 3 and 4, Fig. 5a, b).

The mechanistic explanation proposed above for the dynamics of chlorophyll *a* concentration and seawater excess viscosity is also consistent with recent work conducted on the dynamics of transparent exopolymeric particles (TEP) produced by *Phaeocystis globosa* (Mari et al. 2005). Two phases for the dynamics of TEP were then identified: (i) a production phase during the growth phase of *P. globosa* where TEP and chlorophyll *a* concentration were

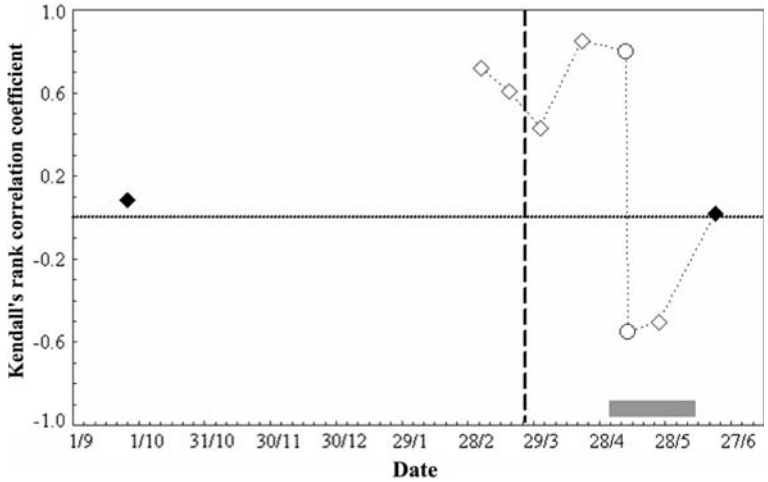


Fig. 4 The time course of the Kendall's correlation coefficient τ, estimated between the 100 simultaneous measurements of chlorophyll concentrations and seawater viscosity at each sampling date. The open and black diamonds represent significant ($P < 0.05$) and nonsignificant Kendall's τ, and the open dots the two correlation patterns observed on May 10 (see Fig. 2). The grey bar indicates the period of foam formation (May 3 to June 10). The sampling conducted on 26 September has been used as a separate reference sample, characteristic of non-bloom conditions

positively correlated, and (ii) a release of large TEP from the mucilaginous matrix of *P. globosa* colonies subsequent to colony disruption (caused by nutrient depletion) where TEP and chlorophyll *a* concentration were negatively correlated. The causes of colony disruption observed by Mari et al. (2005) and the area described in the present work might diverge. The similarity between the dynamics of chlorophyll *a* concentration-excess viscosity shown in the present work and chlorophyll *a* concentration-TEP concen-

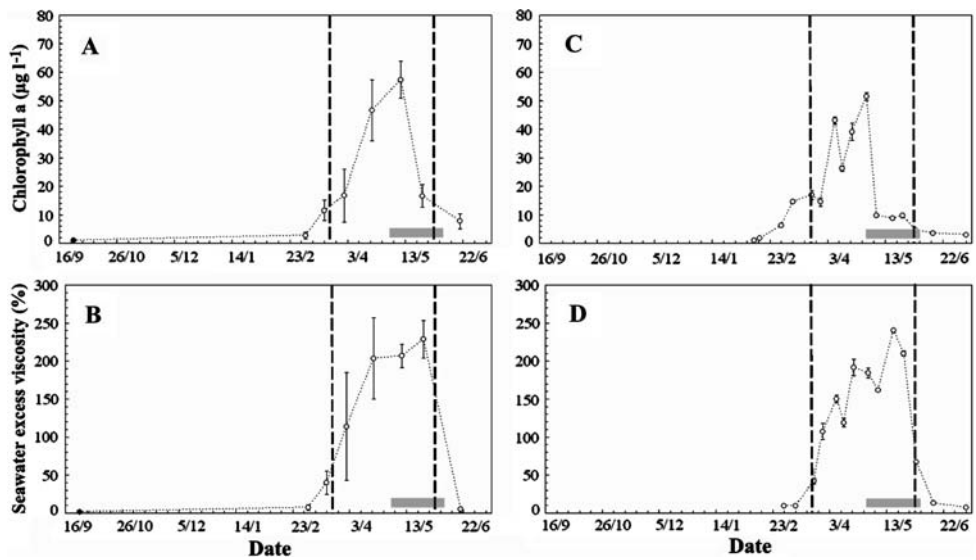


Fig. 5 Time course of chlorophyll *a* concentration (µg l^{-1}; **A**, **C**) and seawater excess viscosity η (%; **B**, **D**), in the intertidal (**A**, **B**; present work) and coastal (**C**, **D**; modified from Seuront et al. 2006) waters of the Eastern English Channel. The grey bar indicates the period of foam formation (May 3 to June 10), and the dashed vertical lines the appearance and disappearance of *P. globosa* in the phytoplankton assemblage. The sampling conducted on 26 September (black dot; **A**, **B**) has been used as a separate reference sample, characteristic of non-bloom conditions

Table 3 Descriptive statistics of the two-dimensional microscale distributions of chlorophyll *a* concentration (μg l^{-1}) and seawater excess viscosity (%) for each sampling date

Date	Mean	SD	CV	Max./min.
Chlorophyll concentration				
26/09/2003	0.9	0.3	31.4	5.0
05/03/2004	2.7	1.2	44.8	11.9
18/03/2004	11.5	3.6	31.2	11.4
01/04/2004	16.7	9.3	55.8	7.4
20/04/2004	46.6	10.6	22.8	8.1
10/05/2004	57.4	6.5	11.3	2.2
25/05/2004	16.6	4.0	24.0	4.0
20/06/2004	7.7	2.6	33.9	9.7
Excess viscosity				
26/09/2003	1.8	0.7	39.1	16.1
05/03/2004	7.5	3.6	47.6	18.5
18/03/2004	40.1	15.5	38.6	12.1
01/04/2004	113.7	70.9	62.4	15.6
20/04/2004	203.5	53.4	26.3	12.4
10/05/2004	206.8	15.4	7.5	1.7
25/05/2004	228.4	24.9	10.9	2.0
20/06/2004	4.9	3.1	63.6	47.5

SD—standard deviation; CV—coefficient of variation; and max./min.—ratio between the maximum and minimum values of a distribution

Table 4 Results of spatial autocorrelation analyses of chlorophyll *a* and seawater excess viscosity two-dimensional microscale distributions

Date	Moran's *I*	Geary's *c*	S^2/Mean	Distribution	Patch size (cm)
Chlorophyll concentration					
26/09/2003	NS	S+	NS	D	5–10
05/03/2004	NS	NS	S+	E	5–10
18/03/2004	S+	NS	S+	C	5–10
01/04/2004	S+	NS	S+	C	5–10
20/04/2004	S+	S−	S+	A	10–15
10/05/2004	S+	S−	S+	A	10–15
25/05/2004	S+	S−	S+	A	10–15
20/06/2004	NS	S+	S+	D	5–10
Escess viscosity					
26/09/2003	NS	S+	NS	D	5–10
05/03/2004	NS	NS	S+	E	5–10
18/03/2004	S+	NS	S+	C	5–10
01/04/2004	S+	NS	S+	C	5–10
20/04/2004	S+	S−	S+	A	15–20
10/05/2004	S+	S−	S+	A	15–20
25/05/2004	S+	S−	S+	A	15–20
20/06/2004	NS	S+	S+	D	5–10

tration recently investigated, strongly suggests, however, that TEP could play a critical role in the increase in seawater viscosity. The variety of dissolved particulate carbohydrates that are differentially produced during a *Phaeocystis globosa* bloom (see Alderkamp et al. 2005), contribute (for some of them) to the formation of microscopic polymer gels in seawater (Chin et al. 1998) and large sedimenting particles (e.g. TEP; Logan et al. 1995). They may also form an intermediate in the formation of the foam observed after *P. globosa* blooms. The issue of relating quantitative seawater viscosity measurements to the quality and the quantity of dissolved and particulate carbohydrates and transparent exopolymeric particles should then be carefully addressed in the future.

Type of spatial patterns and patch sizes

Spatial autocorrelation and correlograms were used to describe the microscale spatial patterns of chlorophyll *a* concentration and seawater excess viscosity. Identical distribution patterns may or may not produce identical correlograms, while different spatial patterns result in different correlograms (Sokal & Oden 1978b). Both distributions exhibited positive spatial autocorrelation at all the sampling dates. This type of spatial pattern is characterised by an association between extreme and similar per sample values and a transition from high to low values within the sampling area (0.2 m^2). Before and after the spring *P. globosa* bloom the observed spatial patterns (Fig. 2a, d) exclusively exhibited aggregative distributions characterised by the alternation of hot spots and cold spots. The similar increases in patch sizes suggest that the aggregative character of the spatial patterns observed before and after the bloom were mainly due to localised hot spots that may be related to local phytoplankton aggregates (e.g. Koike et al. 1990). In contrast, during the *P. globosa* bloom, and mainly after the formation of foam, these distributions were characterised by a variety of smooth transitions from high to low values. The smoother, more continuous

and structured distributions observed during the bloom and characterised by larger patch sizes may be thought of as: (i) the result of the strong biomodification of the fluid properties mainly by *P. globosa*-related polymeric materials and (ii) a two-dimensional description of the organic matter continuum (Chin et al. 1998; Azam 1998). This is also consistent with the similar patch sizes observed for chlorophyll concentration and excess viscosity before and after the bloom, and the larger excess-viscosity patch sizes observed after the foam formation, when phytoplankton biomass and seawater viscosity appear to be uncoupled.

More generally, microscale hot spots of phytoplankton abundance and/or organic matter concentration may well represent the basis of microhabitats for the development of microbial foodwebs (Thingstad and Billen 1994). Microscale patchiness of organisms attached to aggregates and in the surrounding water column will also fundamentally alter predator-prey and virus-host dynamics, and may subsequently modify the flow of matter through the microbial loop (Azam et al. 1983). The intensity and frequency of these hotspots will also influence the sedimentation processes characterising most *Phaeocystis* sp. blooms (e.g. Andreassen and Wassman 1998) and ultimately determine the proportion of carbon that is sequestered to the sediments or retained in the pelagic food web.

Small-scale versus microscale variability

The temporal patterns observed in the present work for chlorophyll *a* concentration and seawater excess viscosity (Fig. 5a, b) are very similar to those obtained in the coastal waters of the Eastern English Channel, roughly 2 nautical miles offshore, from a standard sampling strategy based on 5 replicate samples taken at three different depth (subsurface, intermediate and bottom waters) using Niskin bottles (Fig. 5c, d; Seuront et al. 2006). This ensures the relevance of the previously proposed mechanisms and generalises them from the scale of the water column down to the microscale. The main difference is the significantly higher chlorophyll concentrations and excess viscosity observed at each sampling date from our intertidal station than from the coastal station (WMW test, $P < 0.05$), which is consistent with the previously observed differences in hydrological properties between intertidal and coastal waters of the Eastern English Channel (Seuront and Spilmont 2002).

The variability observed in chlorophyll *a* concentration and seawater excess viscosity from our microscale (5 cm resolution) two-dimensional sampling is significantly higher than the variability previously estimated at the scale of the water column with a resolution of a few meters (Seuront et al. 2006; see Fig. 5). The coefficient of variation and the ratios between maximum and minimum values are, thus significantly higher at the microscale than at the small-scale (WMW test, $P < 0.01$). More specifically, the ratios between the coefficients of variation (CV) obtained at the microscale and the small-scale range from 2.6 to 42.3 for chlorophyll concentration and from 4.5 to 243.2 for seawater excess viscosity. The ratios between maximum and minimum values are 2.1 to 7.4-fold larger at the microscale than at the small-scale for chlorophyll concentration and 3.1- to 14.3-fold larger at the microscale than at the small scale for excess viscosity. These results are in accordance with previous work showing increased variability and structural complexity with decreasing scale (Seuront and Spilmont 2002; Waters et al. 2003), and indicate that extent of microscale variability amongst marine phytoplankton communities can be greater than the variability observed at larger scales.

These findings are particularly pertinent within the Eastern English Channel as the intense tidal mixing, characterised by turbulent kinetic energy dissipation rates ranging from 10^{-16} to 10^{-4} m^2 s^{-3} (Seuront 2005), is expected to homogenise phytoplankton distributions for scales typically smaller than the depth of the vertically well-mixed water column. Both the extent and the non-random spatial structure of the microscale variability observed here in terms of chlorophyll biomass and seawater excess viscosity imply that phytoplankton organisms, and more specifically *Phaeocystis globosa*, can be independent of the surrounding turbulent flows and induce high levels of spatial heterogeneity, even below the 5 cm scale resolution achieved here.

On the potential role of biologically increased seawater viscosity in *P. globosa* ecology

The impact of turbulence on microorganisms is dependent on its relative size to the smallest

Kolmogorov turbulent eddies defined as $l_k = (v^3/\varepsilon)^{0.25}$, where v is the kinematic viscosity ($m^2\ s^{-1}$) related to the measured viscosity η_m (see Eq. 1) by $v = \eta_m/\rho$ (where η_m and ρ are the fluid viscosity and density, respectively) and ε is the turbulent energy dissipation rate ($m^2\ s^{-3}$). Above this scale the flow is turbulent, while below it, viscosity dominates resulting in a laminar shear. The coastal waters of the Eastern English Channel and the Southern North Sea where *P. globosa* typically flourishes are characterised by elevated turbulent dissipation rates (typically bounded between 10^{-7} and 10^{-4} $m^2\ s^{-3}$; Seuront 2005). These high turbulence intensities and the related sub-millimetre Kolmogorov eddies, ranging from 3.2×10^{-1} to 1.8 mm, are not compatible with the size nor the breakability of *P. globosa* colonies (Peperzak 2002). Mucous secretion, and any subsequent increase in seawater viscosity and Kolmogorov eddy size, may be an environmental engineering strategy that *P. globosa* uses to dampen turbulence, create a favourable, turbulent-free physical habitat to protect colony integrity.

At then end of the bloom, the release of nanoplanktonic solitary cells under nutrient limitation would lead to dramatic losses to small protozoan grazers, viruses and zooplankton that would easily offset population growth (Turner et al. 2002). However, excess viscosity remains high, or even increases (Seuront et al. 2006; present work), after colony destruction. This is consistent with the excretion of mucus and dissolved organic carbon observed under nutrient limitation observed in a single cell culture of *P. globosa* (Janse et al. 1999), in colonial cultures of *P. globosa* (van Rijssel et al. 2000) and at the end of *P. pouchetii* blooms in mesocosms (Alderkamp et al. 2005) as well as in many phytoplankton species (Myklestad 1974, 1988). The released flagellated cells may still avoid grazing and viral infection via an alteration of motility and diffusion processes, and a decrease in encounter probabilities and/or in grazing rates attributed, although never demonstrated, to a mechanical hindrance due to increased viscosity (Schoemann et al. 2005). The high-shear environment related to suspensions of aggregates may also be used by *P. globosa* flagellates released from colonies to minimise predation as high shear will decrease their conspicuousness to predators, and not all the perceived preys will be accessible (Seuront et al. 2006). In addition to colony formation, the exudates released by *P. globosa* and the subsequent increase in viscosity might then also be considered as a potential antipredator adaptive strategy that ultimately ensures the completion of its life cycle in highly turbulent environments.

In the sea, the viscous properties of most polymers and suspensions of aggregates are generally dependent on deformation rate. The biologically induced excess viscosity η_{Bio} (see Eq. 3) is then related to the shear γ (s^{-1}) as (Jenkinson 1986) $\eta_{Bio} = k\gamma^{-P}$, where k is a constant, $\gamma = (\varepsilon/v)^{0.5}$ with $v = 10^{-6}$ m s^{-1}, and P has so far been found to lie between 0 and 1.6 (e.g. Jenkinson and Biddanda 1995), or even as low as −0.2 (Jenkinson et al. 1998). A single value of P ($P = 1.11$) has been derived from 15 samples taken in the German Bight when *Phaeocystis* sp. was blooming (Jenkinson 1993). Considering the lack of information related to the value of P for *Phaeocystis globosa*, the intrinsic plurispecific and dynamic nature of the phytoplankton assemblages, the range of P values proposed in the literature and the range of turbulence intensities found in *P. globosa* natural environment (i.e. $\varepsilon = 10^{-7}$to 10^{-4} $m^2\ s^{-3}$, Seuront 2005), any attempt to quantify the nontrivial effect of excess viscosity on the Kolmogorov scale and related microscale processes is still unreasonable at this time. Future investigations should, however, focus on the potential differential contribution of the different carbohydrates that are differentially produced during a *Phaeocystis globosa* bloom (see Alderkamp et al., 2007) to the observed increase in seawater viscosity.

Conclusions

The present work suggests that *Phaeocystis globosa* has developed specific adaptive strategies to favourably modify its immediate microenvironment. The consequences of this strategy are far-reaching as they have been shown to modify the microscale variability of chlorophyll concentration and seawater excess viscosity for scales ranging from 5 to 45 cm. In particular, the modification of the extent and the non-random spatial structure of the microscale variability observed here in terms of chlorophyll biomass and seawater excess viscosity over the course of a *Phaeocystis globosa* spring bloom imply that phytoplankton in general, and more specifically *P. globosa*,

can induce high levels of spatial heterogeneity. The microscale adaptive processes discussed here are thus likely to cascade from the cell and/or colony scales up to larger scale and to influence critical processes such as biogeochemical fluxes through their impact on, e.g. sedimentation and remineralisation processes, and predator-prey and virus-host dynamics. It is finally stressed that our journey to elucidate the relationship between seawater viscous properties and the plankton components is only beginning, and that the future of plankton rheology should rely on interdisciplinary efforts focusing on seawater rheological properties (including viscosity and elasticity) as well as on phytoplankton taxonomy and standing stock, *P. globosa* colony size, turbulence intensity and the quality and quantity of exopolymeric materials.

Acknowledgements We thank D. Menu, who built the two-dimensional sampler, and D. Menu, D. Hilde, T. Caron, M. Priem, B. Thullier and D. Devreker for their assistance during the survey. Ian R. Jenkinson is acknowledged for his enlightening comments on topics related to the present work. This work has been financially and infrastructurally supported by a grant (Action Concertée Incitative ''Jeunes Chercheurs'' #3058) from the French Ministry of Research to L. Seuront, the CPER '*Phaeocystis*' (France), PNEC '*Chantier Manche Orientale-Sud Mer du Nord*' (France), Université des Sciences et Technologies de Lille (France), Australian Research Council (Australia) and Flinders University (Australia).

References

Al-Hasan RH, Ali AM, Radwan SS (1990) Lipids, and their constituent fatty acids, of *Phaeocystis* sp. From the Arabian Gulf. Mar Biol 105:9–14

Alderkamp AC, Buma AGJ, van Rijssel M (2007) The carbohydrates of Phaeocystis and their degradation in the microbial food web. Biogeochemistry, doi:10.1007/s10533-007-9078-2

Alderkamp AC, Nejstgaard JC, Verity PG, Zirbel MJ, Sazhin AF, van Rijssel M (2005) Dynamics in carbohydrate composition of Phaeocystis pouchetii colonies during spring blooms in mesocosm. J Sea Res 55:169–181

Andreassen IJ, Wassman P (1998) Vertical flux of phytoplankton and particulate biogenic matter in the marginal zone of the Barents Sea in May 1993. Mar Ecol Prog Ser 170:1–14

Azam F (1998) Microbial control of oceanic carbon flux: the plot thickens. Science 280:694–696

Azam F, Fenchel T, Field JG, Gray JS, Meyer-Reil LA, Thingstad F (1983) The ecological role of water-column microbes in the sea. Mar Ecol Prog Ser 10:257–263

Brylinski JM, Lagadeuc Y, Gentilhomme V, Dupont JP, Lafite R, Dupeuple PA, Huault MF, Auger Y, Puskaric E, Wartel M, Cabioch L (1991) Le 'fleuve côtier': un phénomène hydrologique important en Manche orientale (exemple du Pas de Calais). Oceanol Acta 11:197–203

Buskey EJ 1998. Components of mating behavior in planktonic copepods. J Mar Syst 15:13–21

Cartamil DP, Lowe CG (2004) Diel movement patterns of ocean sunfish *Mola mola* off southern California. Mar Ecol Prog Ser 266:245–253

Chin WC, Orellana MV, Verdugo P (1998) Spontaneous assembly of marine dissolved organic matter into polymer gels. Nature 391:568–572

Decho A, Fleeger J (1988) Microscale dispersion of meiobenthic copepods in response to food-resource patchiness. J Exp Mar Biol Ecol 118:229–243

Dekshenieks MM, Donaghay PL, Sullivan JM, Rines JEB, Osborn TR, Twardowski MS (2001) Temporal and spatial occurrence of thin phytoplankton layers in relation to physical processes. Mar Ecol Prog Ser 223:61–71

Dreyfuss R (1962) Note biologique à propos des eaux rouges. Cah Cent Rech Biol Océanogr Méd Nice 1:14–15

Geary RC (1954) The contiguity ratio and statistical mapping. Incop Statist 5:115–145

Janse I, van Rijssel M, Ottema A, Gottschal JC (1999) Microbial breakdown of *Phaeocystis* mucopolysaccharides. Limnol Oceanogr 44:1447–1457

Jenkinson IR (1986) Oceanographic implications of non-Newtonian properties found in phytoplankton cultures. Nature 323:435–437

Jenkinson IR (1993) Bulk-phase viscoelastic properties of seawater. Oceanol Acta, 16:317–334

Jenkinson IR, Biddanda BA (1995) Bulk-phase viscoelastic properties of seawater: relationship with plankton components. J Plankton Res 17:2251–2274

Jenkinson IR, Wyatt T, Malej T (1998) How viscoelastic effects of colloidal biopolymers modify rheological properties of seawater. In: Emri I, Cvelbar R (eds) Proceedings of the 5th European Rheology Conference, Portoroz, Slovenia, September 6–11, 1998, Progress and Trends in Rheology 5:57–58

Koike I, Shigemitsu H, Kazuki T, Kogure K (1990) Role of sub-micrometer particles in the ocean. Nature 345:242–244

Lancelot C, Billen G, Sournia A, Weisse T, Colijn F, Veldhuis MJW, Davies A, Wassman P (1987) *Phaeocystis* blooms and nutrient enrichment in the continental zones of the North Sea. Ambio 16:38–46

Legendre P, Legendre L (1998) Numerical ecology. Elsevier, New York

Levin SA (1992) The problem of patterns and scale in ecology. Ecology 73:1943–1967

Logan BE, Passow U, Alldredge AL, Grossard HP, Simon M (1995) Rapid formation and sedimentation of large aggregates is predictable from coagulation rates (half-lives) of transparent exopolymer particles (TEP). Deep-Sea Res 42:203–214

Mari X, Rassoulzadegan F, Brussaard CPD, Wassmann P (2005) Dynamics of transparent exopolymeric particles (TEP) production by *Phaeocystis globosa* under N- or P-limitation: a controlling factor of the retention/export balance. Harmful Algae 4:895–914

Martin AP (2003) Phytoplankton patchiness: the role of lateral stirring and mixing. Prog Oceanogr 57:125–174

Miyake Y, Koizumi M (1948). The measurements of the viscosity coefficient of sea water. J Mar Res 7:63–66

Myklestad S (1974) Production of carbohydrates by marine planktonic diatoms. I. Comparison of nine different species in culture. J Exp Mar Biol Ecol 15:261–274

Myklestad SM (1988) Production, chemical structure, metabolism, and biological function of the (1,3)-linked, beta-D-glucans in diatoms. BiolOceanogr 6:313–326

Moran PAP (1950) Notes on continuous stochastic phenomena. Biometrika 37:17–23

Peperzak L (2002) The wax and wane of *Phaeocystis globosa* blooms. Ph.D. Thesis, University of Groningen, The Netherland

Schoemann V, Becquevort S, Stefels J, Rousseau V, Lancelot C (2005) *Phaeocystis* blooms in the global ocean and their controlling mechanisms: a review. J Sea Res 53:43–66

Seuront L (2005) Hydrodynamic and tidal controls of small-scale phytoplankton patchiness. Mar Ecol Prog Ser 302:93–101

Seuront L, Spilmont N (2002) Self-organized criticality in intertidal microphytobenthos patch patterns. Physica A 313:513–539

Seuront L, Schmitt F, Lagadeuc Y (2001) Turbulence intermittency, small-scale phytoplankton patchiness and encounter rates in plankton: where do we go from here? Deep-Sea Res I 43:1199–1215

Seuront L, Vincent D, Mitchell JG (2006) Biologically-induced modification of seawater viscosity in the Eastern English Channel during a *Phaeocystis globosa* spring bloom. J Mar Syst 61:118–133

Seymour JR, Mitchell JG, Pearson L, Waters RL (2000) Heterogeneity in bacterioplankton abundance from 4.5 millimetre resolution sampling. Aquat Microb Ecol 22:143–153

Sokal R, Oden N (1978a) Spatial autocorrelation in biology. I Methodology Biol J Linn Soc London 10:199–228

Sokal R, Oden N (1978b) Spatial autocorrelation in biology. II. Some biological implications and four applications of evolutionary and ecological interest. Biol J Linn Soc London 10:229–249

Sokal RR, Rohlf FJ (1995) Biometry. The principles and practice of statistics in biological research. Freeman, San Francisco

Strickland JDH, Parsons TR (1972) A practical handbook of seawater analysis. Bull Fish Res Bd Canada 167:1–311

Suzuki R, Ishimaru T (1990) An improved method for the determination of phytoplankton chlorophyll using N, N-Dimethylformamide. J Oceanogr Soc Japan 46:190–194

Tiselius P (1992) Behavior of *Acartia tonsa* in patchy food environments. Limnol Oceanogr 8:1640–1651

Thingstad TF, Billen G (1994) Microbial degradation of *Phaeocystis* material in the water column. J Mar Syst 5:55–65

Turner JT, Ianora A, Esposito F, Carotenuto Y, Miralto A (2002) Zooplankton feeding ecology: does a diet of *Phaeocystis* support good copepod grazing, survival, egg production and egg hatching success? J. Plankton Res 24:1185–1195

van Rijssel M, Janse I, Noordkamp DJB, Gieskes WWC (2000) An inventory of factors that affect polysaccharide production by *Phaeocystis globosa*. J Sea Res 43:297–306

Verity PG, Medlin LK (2003) Observations on colony formation by the cosmopolitan phytoplankton genus *Phaeocystis*. J Mar Syst 43:153–164

Waters RL, Mitchell JG, Seymour JR (2003) Geostatistical characterisation of centimetre-scale spatial stucture of in vivo fluorescence. Mar Ecol Prog Ser 251:49–58

Zar JH (1996) Biostatistical analysis. Prentice Hall, NJ