The ISME Journal (2014) 8, 979–990

www.nature.com/ismej

ORIGINAL ARTICLE

Seasonal fluctuations in ionic concentrations drive microbial succession in a hypersaline lake community

Sheila Podell[1], Joanne B Emerson[2,3], Claudia M Jones[2], Juan A Ugalde[1], Sue Welch[4], Karla B Heidelberg[5], Jillian F Banfield[2,6] and Eric E Allen[1,7]

[1]*Marine Biology Research Division, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, CA, USA;* [2]*Department of Earth and Planetary Sciences, University of California, Berkeley, CA, USA;* [3]*Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO, USA;* [4]*School of Earth Sciences, Byrd Polar Research Center, Ohio State University, Columbus, OH, USA;* [5]*Department of Biological Sciences, University of Southern California, Los Angeles, CA, USA;* [6]*Department of Environmental Science, Policy, and Management, University of California, Berkeley, CA, USA and* [7]*Division of Biological Sciences, University of California, San Diego, La Jolla, CA, USA*

Microbial community succession was examined over a two-year period using spatially and temporally coordinated water chemistry measurements, metagenomic sequencing, phylogenetic binning and *de novo* metagenomic assembly in the extreme hypersaline habitat of Lake Tyrrell, Victoria, Australia. Relative abundances of *Haloquadratum*-related sequences were positively correlated with co-varying concentrations of potassium, magnesium and sulfate, but not sodium, chloride or calcium ions, while relative abundances of *Halorubrum, Haloarcula, Halonotius, Halobaculum and Salinibacter*-related sequences correlated negatively with *Haloquadratum* and these same ionic factors. Nanohaloarchaea and *Halorhabdus*-related sequence abundances were inversely correlated with each other, but not other taxonomic groups. These data, along with predicted gene functions from nearly-complete assembled population metagenomes, suggest different ecological phenotypes for Nanohaloarchaea and *Halorhabdus*-related strains versus other community members. Nucleotide percent G + C compositions were consistently lower in community metagenomic reads from summer versus winter samples. The same seasonal G + C trends were observed within taxonomically binned read subsets from each of seven different genus-level archaeal groups. Relative seasonal abundances were also linked to percent G + C for assembled population genomes. Together, these data suggest that extreme ionic conditions may exert selective pressure on archaeal populations at the level of genomic nucleotide composition, thus contributing to seasonal successional processes. Despite the unavailability of cultured representatives for most of the organisms identified in this study, effective coordination of physical and biological measurements has enabled discovery and quantification of unexpected taxon-specific, environmentally mediated factors influencing microbial community structure.

The ISME Journal (2014) **8,** 979–990; doi:10.1038/ismej.2013.221; published online 12 December 2013
Subject Category: Microbial population and community ecology
Keywords: archaea; halophile; hypersaline; metagenome; microbial succession

Introduction

Microbial studies in extreme environments seek to understand how extraordinary physicochemical stresses shape adaptive responses in both individual organisms and entire communities. Achieving this goal requires coordinated, detailed, accurate measurements of both physical and biological factors, but obtaining this information over temporal and spatial scales relevant to natural microbial communities is challenging. Extreme hypersaline aqueous environments harboring limited phylogenetic diversity provide tractable model ecosystems to confront these challenges (Demergasso *et al.*, 2008; Bodaker *et al.*, 2009; Pagaling *et al.*, 2009; Oh *et al.*, 2010; Boujelben *et al.*, 2012; Makhdoumi-Kakhki *et al.*, 2012; Oren, 2013). Although overall salt concentrations in these habitats are, by definition, at or exceeding the limits of ionic solubility, geochemical variation in water sources as well as minerals dissolved from surrounding rocks and

Correspondence: EE Allen, Marine Biology Research Division, University of California, San Diego, Scripps Institution of Oceanography, 9500 Gilman Drive MC0202, La Jolla, CA 92093-0202, USA.
E-mail: eallen@ucsd.edu
Received 20 September 2013; revised 5 November 2013; accepted 10 November 2013; published online 12 December 2013

sediments contribute to variable ratios of different ionic species over space and time (Javor, 1989; Oren, 2013). Evaporative concentration and mineral precipitation, accelerated at higher temperatures, selectively deplete some ionic species while enriching others (Javor, 1989). Intermittent rainfall, agricultural runoff and intrusion from groundwater aquifers provide additional sources of geochemical variability (Macumber, 1992; Ionescu *et al.*, 2012).

The biological effects of dissolved ionic species are mediated not only by absolute concentrations but also by ionic ratios (Park, 2012), which affect the efficiency of cellular ion pumps and antiporters used for balancing intracellular osmolarity and establishing electrochemical gradients for energy production and nutrient transport (Oren, 2013). Ionic concentrations also affect the aqueous solubility of oxygen, which is vanishingly low in extreme hypersaline waters (Sherwood *et al.*, 1991). Paradoxically, the dominant microorganisms in these habitats are aerobic, heterotrophic Archaea, along with bacteria from the genus *Salinibacter*. Although most cultivated microbial strains isolated from extreme hypersaline environments are capable of facultative anaerobic fermentation, they grow optimally in the laboratory under aerobic conditions (Dyall-Smith, 2009; Andrei *et al.*, 2012). The surprising abundance of oxygen-loving organisms in environments with such low oxygen solubility may be explained by the occurrence of non-equilibrium conditions at the air-water interface, providing locally higher oxygen availability. The extent to which temporal differences in mixing efficiency from wind or other factors might affect oxygen distribution through the water column in natural hypersaline environments is largely unknown, but may vary greatly, especially over small distances relevant to microbial localization.

One the most successful microbes in extreme hypersaline waters, the square archaeaon *Haloquadratum walsbyi*, has a specialized cell morphology utilizing large intracellular gas vesicles to facilitate energy-efficient flotation of tightly packed, nearly two-dimensional flat colonies (Walsby, 1980; Bolhuis *et al.*, 2004). This unique morphology may be an adaptation enabling these organisms to take advantage of higher oxygen concentrations at the surface, especially in shallow ponds where currents and wave action are minimized. The importance of near-surface positioning for this species is supported by the observation that laboratory cultures of *Haloquadratum* grow best in unmixed media (Mike Dyall-Smith, personal communication). In addition to greater oxygen availability, surface water localization offers higher levels of incident light, enhancing opportunities for energy production through the action of specialized, proton-pumping rhodopsins (Béjà *et al.*, 2001; Sharma *et al.*, 2006). However, these advantages may be partially offset by higher risks of UV-induced DNA damage, particularly when exposure levels peak during summer seasons (Ruiz-González *et al.*, 2011).

Microorganisms able to survive the most extreme hypersaline environments employ a 'salt-in' strategy to achieve osmotic stability, balancing high extracellular sodium with high intracellular potassium (Oren *et al.*, 2002). Direct measurements in both the archaeaon *Halorubrum sodomense* and the bacterium *Salinibacter ruber* confirm this pattern, demonstrating intracellular potassium levels exceeding 3.0 M (Oren, 1998; Oren *et al.*, 2002). Extreme halophiles also contain much higher intracellular magnesium levels than non-halophiles. Cultured *Halobacterium salinarum* cells, for example, were found to have total intracellular magnesium concentrations exceeding 100 mM, more than three-fold higher than *Escherichia coli* cells (30 mM) measured using the same technique (de Médicis *et al.*, 1986). However, work in this area is not sufficiently complete to determine whether intracellular concentrations of some ions might be better controlled than others under naturally occurring conditions, or whether particular combinations or ratios of ion concentrations might have synergistic effects that are not apparent when varying ionic species one at a time (Park, 2012).

Extensive changes to protein amino acid composition have been observed in microorganisms using a salt-in strategy, presumably to counter potential adverse effects of high salt concentrations on protein structure (Fukuchi *et al.*, 2003; Bolhuis *et al.*, 2008). Specifically, proteins of salt-in halophiles are enriched in amino acids with negatively charged side chains, enhancing solubility and depleted in residues with bulky hydrophobic groups, increasing structural flexibility. Little is known about whether nucleic acid sequences undergoing replication, transcription and/or translation might also need special adaptations to function in organisms using a salt-in strategy for osmotic balance. Nucleic acid structure and stability are known to be exquisitely sensitive to ionic concentrations *in vitro*, especially magnesium, which is orders of magnitude more active than sodium or potassium in increasing double-stranded DNA melting temperature (Hartwig, 2001; Owczarzy *et al.*, 2008) and stabilizing RNA hairpin structures (Bizarro *et al.*, 2012).

Cultured isolates of *Haloquadratum walsbyi*, which thrive in external magnesium concentrations as high as 2 M, have genomic nucleotide compositions with a much lower percent G + C than other halophilic Archaea. It has been proposed that the lower DNA melting temperature associated with this composition is useful in mitigating the effects of high internal magnesium concentrations, which might otherwise prevent strand separation for replication (Bolhuis *et al.*, 2006). Conversely, it has been suggested that high G + C genomes found in extreme hyperthermophiles might help prevent hydrogen bond de-stabilization at higher temperatures (Saunders *et al.*, 2003). Low G + C nucleotide compositions are also characteristic of prokaryotes found in oligotrophic, nitrogen-poor marine habitats (Giovannoni *et al.*, 2005), but the shallow saltern

crystallizer ponds where *Haloquadratum walsbyi* is commonly found generally contain abundant nitrogen and organic nutrients (Javor, 1989).

The microbial community of Lake Tyrrell, an extreme hypersaline habitat in Victoria, Australia, has previously been characterized using a variety of molecular biology techniques, including 16S and 18S rRNA gene amplification, metagenomic sequencing and *de novo* assembly of habitat-specific reference genomes for Archaea, bacteria, and viruses (Emerson *et al.*, 2012; Narasingarao *et al.*, 2012; Emerson *et al.*, 2013; Heidelberg *et al.*, 2013; Podell *et al.*, 2013). The current study of Lake Tyrrell combines detailed ionic composition measurements with long-read metagenomic sequencing from the same geographic location during summer and winter seasons over a two-year period. These studies provide evidence suggesting that seasonal changes in concentrations of specific ionic species drive strain selection both within individual taxa and across the entire microbial community, and that this selection is intimately associated with shifts in genomic nucleotide composition.

Materials and methods

Sample collection, metagenome sequencing and assembly

Surface water samples were collected from at 0.3 m depth from Lake Tyrrell, Victoria, Australia (GPS coordinates −35.32, 142.80). Samples for metagenomic sequencing were passed through filters of decreasing porosity (20 μm > 3 μm > 0.8 μm > 0.1 μm), as previously described (Narasingarao *et al.*, 2012). DNA from 0.8 μm and 0.1 μm filters was sequenced using both paired-end Sanger sequencing and Roche 454 Titanium pyrosequencing at the J Craig Venter Institute, as described in (Goldberg *et al.*, 2006). Sample collection dates and associated sequencing libraries are described in Supplementary Table S1. Each individual lake water sample used for DNA isolation and sequence library construction was also subjected to water chemistry analysis, as described below.

Individual metagenomic reads containing 16S rRNA gene fragments were identified from both Sanger and 454 data sets by low stringency BLASTN searches (e-value of 1e-4 or better) against the GreenGenes database (DeSantis *et al.*, 2006). Reads containing 16S rRNA gene fragments were assembled using Celera Assembler version 5.4 using merSize = 18, utgGenomeSize = 2000 and utgErrorRate = 0.02.

Habitat-specific consensus population genomes were reconstructed using iterative *de novo* assembly techniques, as described previously (Podell *et al.*, 2013). All 232 354 Sanger reads from August 2007 (libraries DAM, EAM and EBM) were combined in a preliminary composite assembly using Celera Assembler version 5.4. with merSize = 15, utgGenomeSize = 500 000 and utgErrorRate = 0.10. Scaffolds from this assembly were binned into provisional groups based on depth of coverage, percent G + C, predicted protein similarity to previously sequenced genomes and number of reads derived from 0.1 micron and 0.8 micron filters, as described in (Podell *et al.*, 2013). August 2008 and January 2009 samples, consisting solely of unpaired 454 Titanium reads, were not assembled.

Preliminary scaffold groups were used to guide selection of August 2007 Sanger and 454 Titanium read subsets for inclusion in iterative, targeted assemblies using more stringent assembly parameters (utgErrorRate = 0.06). Draft population consensus genomes assembled from metagenomic reads were annotated using the IMG-ER pipeline (Markowitz *et al.*, 2009). Genome completeness was estimated based on the recovery of 53 transcription, translation and replication genes universally conserved in Archaea (Ciccarelli *et al.*, 2006; Wu and Eisen, 2008; Puigbò *et al.*, 2009; Narasingarao *et al.*, 2012). 16S rRNA genes have been deposited into NCBI GenBank under accession numbers KF673165-KF673190 and assembled population genomes under AYLL00000000 (A07HR60), AYLK00000000 (A07 HN63), AYLI00000000 (A07HR67) and AYLJ00000000 (A07HB70).

Phylogenetic diversity and abundance

Assembled 16S rRNA genes from January and August 2007 metagenomic data containing 1160 or more nucleotides (>75% complete) were trimmed and aligned using MUSCLE v 3.6 (Edgar, 2004). Maximum likelihood trees were constructed with FastTree version 2.1.1 (Price *et al.*, 2010) using default parameter settings. Relative abundances of microbial taxa represented in unassembled metagenomic reads were estimated using PhymmBL version 3.2 (Brady and Salzberg, 2011). Taxonomic distributions of predicted proteins in metagenomic assemblies were assessed using DarkHorse version 1.4 (Podell and Gaasterland, 2007; Podell *et al.*, 2008). Taxonomic classifications obtained from both PhymmBL and DarkHorse were based on custom reference libraries consisting of bacterial, archaeal, and phage genomes obtained from NCBI GenBank, supplemented with 12 Lake Tyrrell-specific consensus draft genomes from January 2007 (Podell *et al.*, 2013) and 4 from the current study.

Water chemistry

Samples used for water chemistry were filtered through 0.2 μm filters, then diluted from 100- to 100 000-fold with either MilliQ water (for anions) or 1% trace metal grade HNO_3 (for cations). Anion concentrations (F^-, Cl^- and SO_4^{2-}) were determined using a Dionex DX4500i ion chromatograph. Cation concentrations were determined via inductively coupled plasma atomic emission spectroscopy on a Varian Vista Pro axial ICP-AES.

Results

Ionic composition measurements

Seasonal measurements of Lake Tyrrell water chemistry are shown in Table 1. As expected, both temperatures and total ionic strength were higher in samples collected during the Austral summer (January), than those obtained in winter (August). However, concentrations and ratios of individual ionic species varied considerably between samples taken in different years, even for the same season. The extent to which ionic concentrations might have been influenced by diurnal temperature fluctuations and short-term dilutions by small-scale rainfall events was not determined. However, daily air temperature and rainfall records for the 30 days prior to each sampling date from the nearest Australian Government Bureau of Meteorology station provide a contextual background for water measurement snapshots taken at the time of sample collection (Supplementary Table S2).

Sodium, chloride and calcium concentrations were 30–60% lower in summer (January) samples collected in 2007 versus 2009. At the same time, magnesium, potassium and sulfate concentrations increased by 77–168%. Concentrations of individual ions were as much as 40% lower (calcium) or 70% higher (magnesium) in winter (August) samples from 2008 versus 2007. Ionic concentration measurements for water samples taken at two-day intervals varied by less than 5% (Supplementary Table S3), with the exception of magnesium, potassium and sulfate, which were 13–17% lower on the first of two January 2007 sampling dates, consistent with the occurrence of rain two days earlier.

Overall concentration patterns for potassium, magnesium and sulfate correlated strongly with each other across multiple samples, and inversely to sodium, chloride and calcium (Supplementary Figure S1). Although seasonal water temperatures, pH and overall ionic strength were relatively consistent from year to year, variability in concentrations of individual ionic species provided an opportunity to examine the effects of intra- as well as inter-seasonal variation on microbial community composition.

Table 1 Physical properties of Lake Tyrrell sample collection site

Date	*Temp (°C)*	*pH*	*Total ionic strength*[a]	Na^+	K^+	Mg^{2+}	Ca^{2+}	Cl^-	SO_4^{2-}
Jan 2007	24.8	7.2	5835	4251	38	359	10	5318	147
Aug 2007	10.7	7.0	4232	3641	18	122	15	4064	48
Aug 2008	11.5	7.3	4418	3643	23	207	9	4043	66
Jan 2009	23.3	7.0	5965	2886	66	960	5	3975	286

Concentrations are given in units of millimoles/liter. Data for 2007 and 2009 are averages of duplicate water samples collected 2 days apart. Aug 2008 data represent average measurements from samples collected at different times on the same day. Detailed data for individual water samples are presented in Supplementary Table S3.
[a]Total ionic strength I is calculated based on the following equation, where c_i = molar concentration and z_i = charge number for each ionic component of the mixture.

$$I = \frac{1}{2}\sum_{i=1}^{n} c_i z_i^2$$

Microbial community diversity

To supplement Lake Tyrrell-specific genomes previously obtained in summer (January) 2007 (Podell *et al*., 2013), new consensus genomes were assembled for the most dominant microbial species present in winter (August) of the same year (Table 2, Supplementary Table S4). Two of the new genomes from winter 2007 (A07HN63 and A07HR60) included 16S rRNA gene sequences with 97% or greater identity to organisms from the previous summer, suggesting they might belong to the same species. The third genome (A07HR67) contained three 16S rRNA gene copies that were 95% identical to an 820 nt gene fragment from a small, low-coverage, *Halorubrum*-related scaffold from January 2007, suggesting possible membership in the same genus. The fourth new August 2007 genome (A07HB70) contained a 16S rRNA gene that was only 90% identical to previously observed sequences, indicating a more distant relationship.

The four new assembled genomes accounted for approximately 8.4% of the August 2007 metagenomic reads, at coverage depths ranging from 9–14-fold. Assembly of the remaining August 2007 reads under less stringent conditions yielded thousands of additional scaffolds at 3–5-fold coverage depths, with closest BLAST matches in archaeal genera *Haloquadratum*, *Halorhabdus*, *Haloarcula*, phylum Nanohaloarchaea and bacterial genus *Salinibacter*. Most of the lower abundance scaffolds were too short for taxonomic binning methods to distinguish between closely related strains, precluding their assembly into discrete population genomes.

To obtain additional taxonomic diversity information, a stringent gene-specific assembly was performed using only August 2007 reads containing 16S rRNA fragments. This assembly yielded nine new 16S rRNA sequences of 1200 nt or longer, supplementing the 16S rRNA sequences obtained from selective population genome assemblies of A07HB70, A07HR67, A07HR60 and A07HN63. An additional 16S rRNA-specific assembly was performed on selected metagenomic reads from January 2007, yielding more complete versions of 16S rRNA genes than those previously published for halophilic archaeal species J07HX64 and J07HB67. Percent identities of 16S rRNA sequences to previously described environmental samples and cultured isolates are shown in Supplementary Table S5.

A phylogenetic tree of archaeal 16S rRNA genes recovered from both summer (January) and winter (August) 2007 samples is shown in Figure 1. Winter samples included nearly full-length 16S rRNA genes from all major taxonomic groups identified in the previous summer, except '*Candidatus* Nanosalinarum

sp. J07AB56' and *Halonotius* sp. J07HN4. Two new, previously undetected *Halorhabdus*-related 16S rRNA gene sequences (A07Hrhab_scf299 and A07Hrhab_scf310) and one new *Halobaculum*-related 16S rRNA gene sequence (A07HB70) were observed in winter, but not summer samples.

Seasonal variation in genomic nucleotide composition

Average G+C percentages for individual genomes usually fall within a relatively narrow range, and closely related taxa generally have similar nucleotide compositions. These properties make G+C composition of metagenomic reads a useful metric

Table 2 Lake Tyrrell-specific consensus genomes assembled from August 2007 metagenomic reads

Genome name	*Length (nt)*	*G+C pct*	*Coverage depth*	*rRNA operons*	*Closest J07 relative*	*16S pct ident*
Halorubrum sp. A07HR60	2 877 040	59	13.6	3	J07HR59	97%
Halonotius sp. A07HN63	2 394 881	63	10.5	1	J07HN6	98%
Halorubrum sp. A07HR67	2 892 307	67	9.0	3	scf7180000091214	95%
Halophilic archaeon A07HB70	2 389 822	71	9.2	1	J07HB67	90%

Abbreviations: ident, identity; pct, percent. Additional genome statistics are presented in Supplementary Table S3.

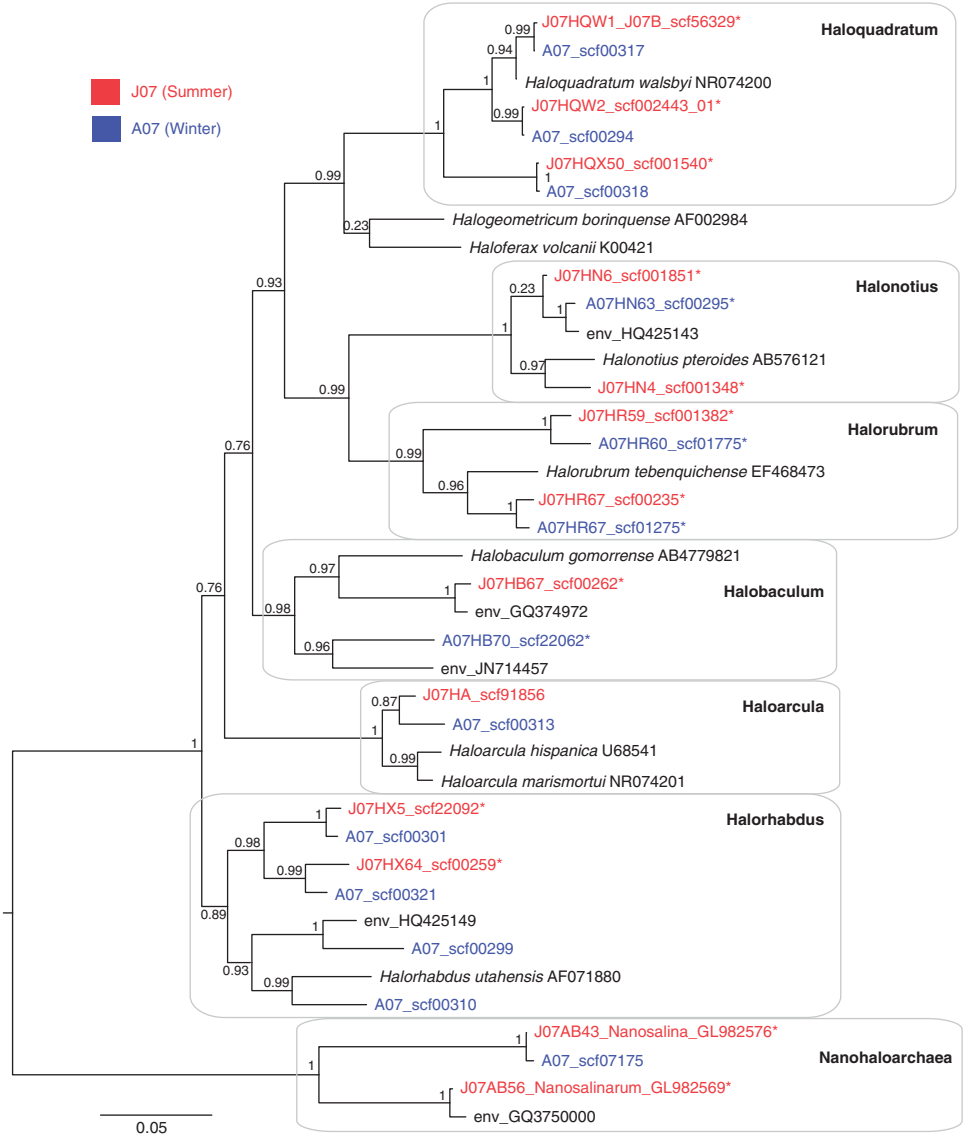

Figure 1 Phylogenetic tree of Archaeal 16S rRNA genes. FastTree confidence values are indicated at nodes. Red sequences were obtained from January 2007 (summer) targeted assemblies and blue from August 2007 (winter) targeted assemblies. Black sequences indicate cultured isolates and environmental sequences from NCBI GenBank. Asterisks indicate sequences included in assembled population genomes (Supplementary Table S4). Additional information on 16S rRNA sequences is provided in Supplementary Table S5.

for observing changes in overall community structure. Figure 2 shows that Lake Tyrrell metagenomic reads from January 2007 and January 2009 had similar, but not identical biphasic distributions, with a smaller peak at 60–65% G+C and a larger peak at approximately 48–50% G+C, respectively. Winter samples contained a greater percentage of sequences in the 60–65% G+C peak. Changes in relative peak heights were much more pronounced in winter 2007 than 2008. These distinctive peak distribution shapes were highly reproducible for samples collected two days apart, and for reads obtained using both Sanger and Roche 454 Titanium technologies (Supplementary Figure S2), confirming that observed seasonal differences were not due to sequencing technology bias.

Seasonal shifts in percent G+C distributions occurred not only in the overall set of metagenomic reads but also within read subsets from multiple pooled libraries that were classified into genus-level taxonomic groups (Figure 3). Sharp peaks in G+C plots coincide with increased relative abundance of individual population genomes reconstructed by *de novo* assembly, for example *Haloquadratum* strain J07HQW2 (47% GC), *Halonotius*-related strain J07HN4 (61% GC), *Halobaculum*-related strain J07HB67 (67% GC) and 'Candidatus *Nanosalina* strain J07AB43' (43% GC) in summer (January) 2007, as well as *Halorubrum*-related strain A07HR60 (59% GC) and *Halobaculum*-related strain A07HB70 (71% GC) in winter (August) 2007. Shifts to lower G+C distributions in summer (January) versus winter (August) 2007 samples were clearly apparent for every taxonomic group except *Salinibacter*, which remained relatively unchanged. Visually observable differences in G+C peak distribution patterns from summers two years apart (January 2007 versus January 2009) were smaller and more subtle than summer versus winter differences within a single year (Supplementary Figure S3). Percent G+C distributions for winter samples taken one year apart appeared to be less consistent than the summer samples, with lower values for *Halonotius, Halorhabdus, Halorubrum, Halobaculum, Haloarcula* and Nanosalina-related groups in 2008 versus 2007 (Supplementary Figure S4).

Evidence linking G+C composition to successional changes based on unassembled reads was corroborated by observations made using metagenomic sequence assembly. The phylogenetic tree of assembled 16S rRNA genes shown in Figure 1 indicated that most genus-level archaeal groups from January (summer) and August (winter) 2007 contained at least two closely related populations whose sequences could be assembled into nearly complete genomes (Supplementary Table S4). In most of these taxonomically related pairs, the population with lower G+C genomic nucleotide composition was more abundant in summer versus winter samples, whereas the converse was true in winter, when the higher genomic G+C composition populations were more abundant. Large summer to winter decreases in relative abundance were observed for *Halonotius* sp. J07HN4 (61% G+C),

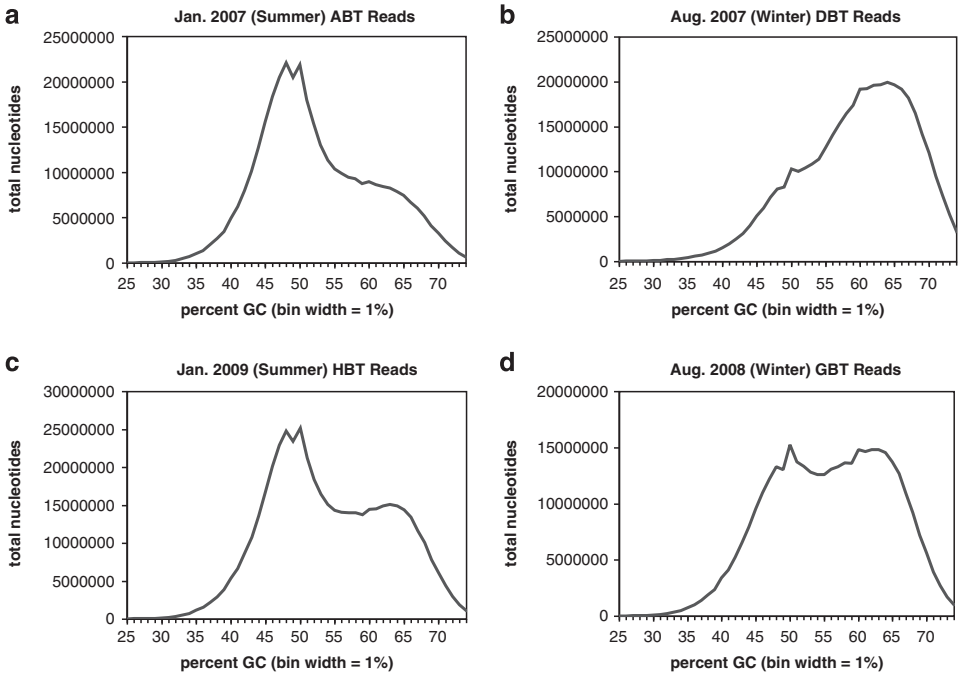

Figure 2 Seasonal variation in percent G+C distribution patterns. Unassembled, unbinned Titanium 454 metagenomic sequencing reads from January 2007 (summer) versus August 2007 (winter) are plotted as smoothed histograms. Panels **a–d** represent sequencing libraries ABT, DBT, HBT and GBT, as described in Supplementary Table S1. Histogram plots for additional libraries are shown in Supplementary Figure S2.

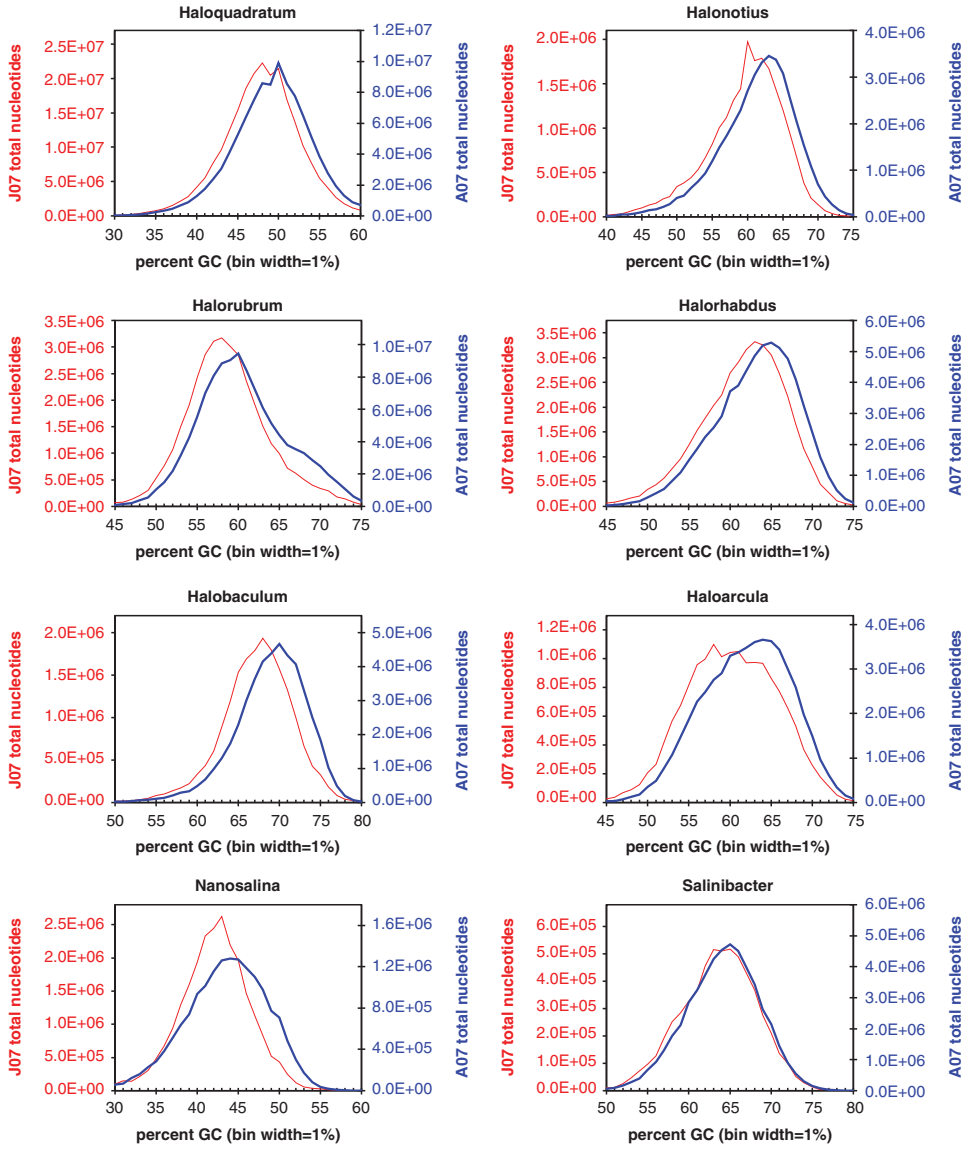

Figure 3 Seasonal shifts in taxon-specific G+C nucleotide distribution patterns. Taxonomic groups are based on PhymmBL classified subsets of unassembled metagenomic reads from January 2007 (summer) and August 2007 (winter). G+C compositions of reads assigned to each taxonomic group have been plotted as smoothed histograms. Separate Y-axis scales have been used on samples from January 2007 (J07) and August 2007 (A07) to facilitate graphical comparisons.

Halorhabus-related species J07HX5 (61% G+C), *Halobaculum*-related species J07HB67 (67% G+C) and '*Candidatus* Nanosalina J07AB43' (43% G+C), which each represented the low end of genomic percent G+C within their taxonomic group. Relative abundances of both *Haloquadratum* sp J07HQW2 (47% G+C) and J07HQW1 (49% G+C) were decreased in winter samples, while *Haloquadratum* sp J07HQX50 (50% G+C) remained relatively constant. Conversely, *Halobaculum*-related species A07HB70 (70% G+C) and *Halorubrum*-related species A07HR67 (67% G+C), which each represented higher G+C composition populations within their taxonomic groups, were much more abundant in winter than summer samples.

Abundance levels of microbial taxa

Relative abundances of different microbial taxa were initially estimated by taxonomically classifying unassembled metagenomic read nucleotide sequences with PhymmBL (Figure 4). Nucleotide-based relative abundances obtained using PhymmBL were confirmed by comparison to amino acid sequence-based taxonomic analysis of predicted proteins from scaffolds obtained in composite assemblies of unbinned reads (Supplementary Figure S5). Both nucleic acid and amino acid-based analysis techniques showed that *Haloquadratum* and Nanohaloarchaea-related sequences were more abundant in summer (January) than winter (August) samples, while *Halorubrum* and *Haloarcula*-related

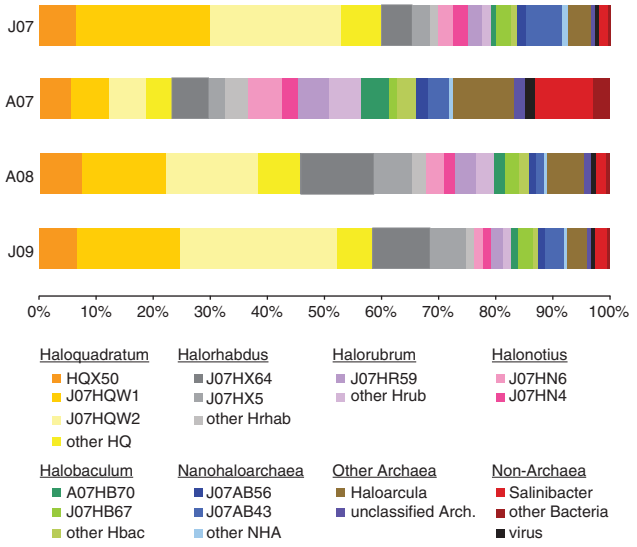

Figure 4 Seasonal taxonomic abundance patterns. Percentage values indicate relative abundance based on PhymmBL classification of unassembled Titanium 454 metagenomic reads. Relative abundance values have not been corrected for different genome sizes.

sequences followed the opposite seasonal pattern. The decrease in *Haloquadratum* abundance was more pronounced in winter 2007 than 2008, consistent with the observed overall G+C peak distributions for unassembled reads (Figure 2). Winter 2007 samples included a larger increase in *Salinibacter*-related sequences than was observed in 2008, and a much more even distribution among different taxa than any of the other sampling dates. *Halorhabdus*-related sequences did not follow any apparent seasonal trend, but were more abundant in both summer and winter samples from 2008 and 2009 than 2007.

Haloquadratum abundance was positively correlated with elevated magnesium concentrations, while *Halorubrum, Haloarcula*, *Halonotius*, *Halobaculum* and *Salinibacter*-related sequences were negatively correlated (Figure 5). Microbial abundance relationships to potassium, sulfate and temperature were nearly identical to those observed for magnesium (Supplementary Figure S6), consistent with the co-variation of these parameters observed in physical water chemistry measurements. Unexpectedly, neither Nanohaloarchaea nor *Halorhabdus*-related sequences showed any correlation with ionic composition, and no correlations were observed for any microbial groups with environmental sodium, chloride or calcium concentrations.

The abundances of *Halorubrum, Haloarcula*, *Halobaculum* and *Salinibacter*-related sequences were all inversely correlated to *Haloquadratum* levels, with R^2 values exceeding 0.985 (Figure 6a). *Salinibacter*-related sequences declined most steeply as *Haloquadratum* increased, following a polynomial rather than a linear curve. *Halonotius* sequences were only moderately correlated with *Haloquadratum*

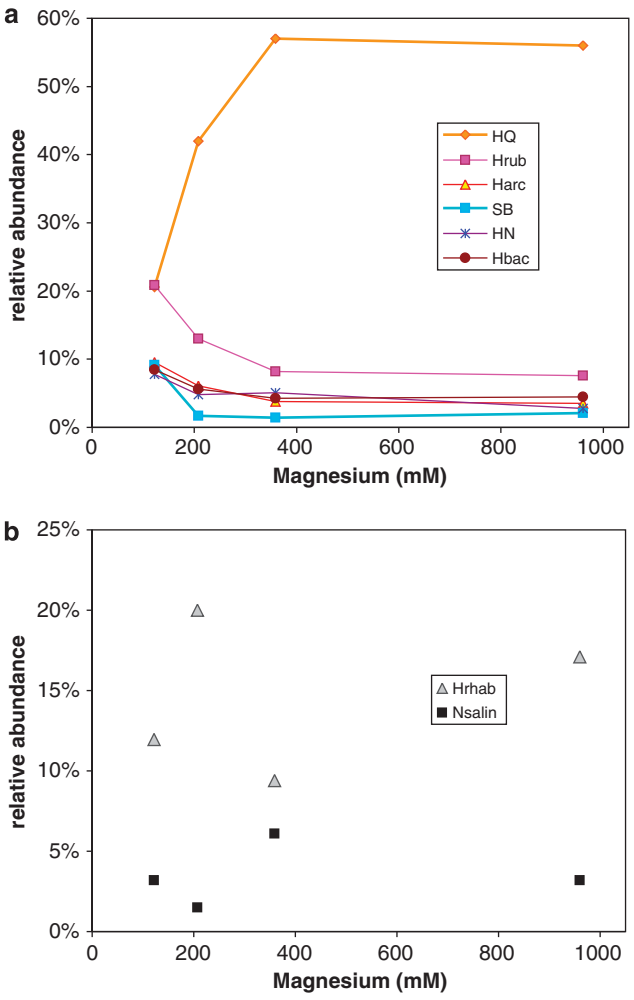

Figure 5 Microbial abundance relationships to environmental magnesium concentrations. Relative abundances of taxonomic groups based on number of nucleotides in PhymmBL classified taxonomic bins for unassembled Titanium 454 metagenomic reads. Abbreviations: Harc, *Haloarcula*; Hbac, *Halobaculum*; HN, *Halonotius*; HQ, *Haloquadratum*; Hrhab, *Halorhabdus*; Hrub, *Halorubrum*; Nsalin, Nanosalina; SB, *Salinibacter*.

abundance ($R^2 = 0.740$), while Nanosalina-related ($R^2 = 0.181$) and *Halorhabdus*-related ($R^2 = 0.003$) sequences appeared to be completely independent (Figure 6b). In contrast, Nanosalina-related sequences correlated negatively with *Halorhabdus*-related sequences ($R^2 = 0.805$), but were the only taxonomic group that showed any such correlation (Figures 6c and d). The disparate correlation levels observed between microbial groups may reflect shared sensitivities to environmental conditions and/or competitive versus non-competitive community relationships.

Discussion

This study has revealed strong correlations between concentrations of specific ions, genomic nucleotide compositions, and the relative abundance of different microbial community members over time in a

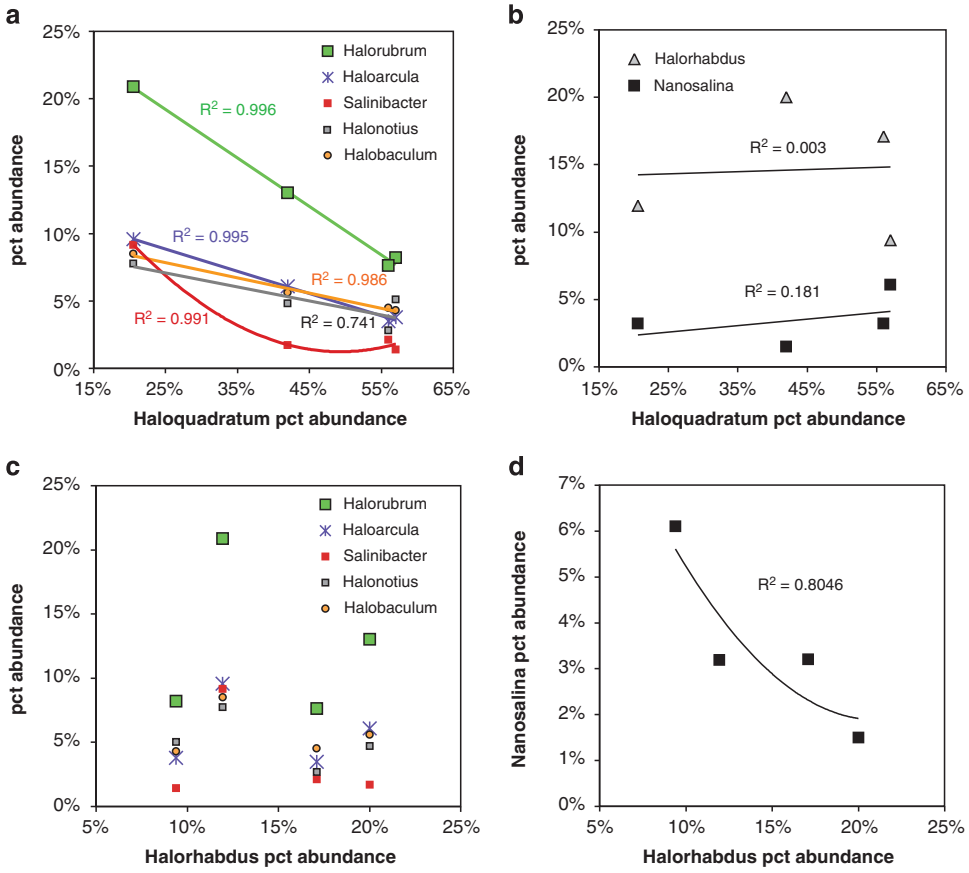

Figure 6 Relative sequence abundance correlations between taxonomic groups. Panels **a** and **b** show abundances of taxonomic groups relative to *Haloquadratum*-like sequences, based on number of nucleotides in PhymmBL classified bins for unassembled Titanium 454 metagenomic reads. Panels **c** and **d** show relative abundances of taxonomic groups compared to *Halorhabdus*-like sequences.

single geographic location. Taxonomic classification of raw metagenomic reads, 16S rRNA gene sequences and predicted proteins from assembled scaffolds demonstrated that similar archaeal populations from class Haloarchaea, phylum Nanohaloarchaeota and the bacterial genus *Salinibacter* were present in both summer (January) and winter (August) samples, but at different relative abundance levels. Although the total number of environmental samples included in this study does not allow rigorous statistical testing of quantitative values, the patterns observed are supported by consistency of water chemistry measurements, nucleotide composition profiles, phylogenetic binning and *de novo* metagenomic assembly of multiple samples obtained at two-day intervals, repeated over two summer and two winter seasons. Further experiments will be required to determine whether the same relationships will be observed over a more extensive collection of samples, including a wider range of ionic concentrations, time intervals and geographical locations.

Several physical variables that appeared to influence taxonomic distribution in the Lake Tyrrell microbial community were linked to each other, complicating interpretation of the results. Relative abundance differences were initially observed to track both temperature and total ionic strength. However, significant differences in community abundance were discovered in August 2007 versus August 2008, when temperatures were similar but ionic conditions were different. Although these results do not explicitly rule out a role for temperature, they suggest that ionic composition plays a particularly important role in shaping microbial community structure.

On closer inspection, ionic strength effects on microbial composition were found to correlate with co-varying concentrations of potassium, magnesium and sulfate, but not calcium, sodium or chloride. Although no independent evidence is available suggesting that potassium or sulfate concentrations might influence microbial selection within the ranges observed, cultured isolates of *Haloquadratum walsbyi* have previously been shown to be much more tolerant of high magnesium than other halophilic Archaea (Bolhuis *et al.*, 2004; Burns *et al.*, 2004). These results support the hypothesis that observed changes in microbial community composition with ionic strength may be driven primarily by elevated magnesium concentrations, rather than other ionic species.

To the best of our knowledge, the current study provides the first reported evidence of selection for microbial strains with lower genomic G + C nucleotide compositions linked to extreme ionic stress in a natural environment. Seasonally linked G + C bias was detected at three different levels of granularity, including total raw reads, taxonomically binned archaeal read subgroups and relative abundance of assembled population genomes. The extent to which the extremely high intracellular magnesium concentrations observed in cultured halophilic Archaea might or might not be controlled by homeostatic mechanisms is unknown. If increased external magnesium concentrations elicit even higher internal magnesium levels, over-stabilization of nucleic acid hydrogen bonds might provide a significant selective disadvantage to microbial strains with higher G + C compositions. The observation that G + C nucleotide compositions were consistently lower in summer and higher in winter for multiple taxa suggests the possibility that seasonally linked selective pressures may contribute to population-level selection and successional changes in microbial community structure. This hypothesis should be amenable to future experimental testing under controlled laboratory conditions.

Results from the current study extend previous reports of an inverse seasonal relationship between the abundance of *Haloquadratum* and *Halorubrum*-related strains in Lake Tyrrell (Emerson *et al.*, 2013) as well as other extreme hypersaline environments (Boujelben *et al.*, 2012) by determining quantitative correlation values that also include the rest of the microbial community. This quantitative analysis led to the unexpected discovery that two relatively abundant Lake Tyrrell groups containing Nanohaloarchaea and *Halorhabdus*-related species varied independently of *Haloquadratum,* while the rest of the microbial community showed an inverse correlation with *Haloquadratum* abundance.

In retrospect, marked ecological phenotype differences between Nanohaloarchaea and other halophilic Archaea should not be surprising, considering their disparate evolutionary histories and cellular morphologies. Nanohaloarchaea are characterized by exceptionally small cells ($\sim$0.6 μm in diameter) lacking gas vesicles, flagella and typical archaeal light-driven proton-pumping rhodopsin family genes (Ugalde *et al.*, 2011; Narasingarao *et al.*, 2012). These attributes are consistent with a reduced dependence on highly aerobic processes and light-driven energy production activities requiring close proximity to the air-water interface. Nearly complete population genomes for '*Candidatus* Nanosalina J07AB43' and '*Candidatus* Nanosalinarum J07AB56' (Supplementary Table S4) include genes suggesting availability of Embden–Meyerhoff glycolysis, oxidative pentose phosphate and glycogen catabolism pathways. However, the highly atypical amino acid compositions of predicted proteins from these genomes (Narasingarao *et al.*, 2012) and their taxonomic distance from previously characterized database sequences (Rinke *et al.*, 2013) have impeded confident determination of metabolic capabilities through bioinformatic inference alone. Since no cultured isolates have yet been obtained for any species in phylum Nanohaloarchaeota and no gene expression data is available, conditions required for optimal growth remain unknown.

Halorhabdus-related strains are more closely related to other Lake Tyrrell halophilic Archaea, and share many common core metabolic pathways. The genome of *Halorhabdus*-related strain J07HX64, estimated to be approximately 92% complete (Podell *et al.*, 2013), encodes gas vesicle and flagellar synthesis genes and a chloride-pumping halorhodopsin, but no bacteriorhodopsin-type light-driven proton pumps like those found in *Haloquadratum*, *Halonotius*, *Halorubrum*, *Haloarcula* and *Halobaculum*-related genomes. However, failure to detect specific genes in unfinished genomes cannot be conclusive, and inferred metabolic functions based on genomic annotation must be verified using experimental methods.

It has not been determined whether the atypical abundance patterns of Nanohaloarchaea and *Halorhabdus*-related strains might be associated with additional environmental parameters not included in this study, for example, concentrations of nitrogen, phosphorus and dissolved organic material, or biological predation stresses imposed by viruses and/or picoeukaryotes. Uncharacterized halocin-like antimicrobial activities recently observed in a broad collection of cultured hypersaline Archaea and bacteria (Atanasova *et al.*, 2013) may further influence community structure in natural environments.

It is tempting to speculate that the lack of apparent correlation between relative abundances of Nanosalina and *Halorhabdus* versus *Haloquadratum*-related populations might be due to greater reliance on fermentation rather than oxygen-requiring metabolic processes, removing the need to compete for near-surface positioning. Adaptive phenotype differences might also include dissimilar levels of daytime versus nighttime activity. *Halorhabdus*-related 16S rRNA sequences have been recovered at high abundance from the deep anoxic Discovery Basin in the Eastern Mediterranean Sea (van der Wielen *et al.*, 2005), and cultured isolates of *Halorhabdus tiamatea* are highly unusual among Halobacteriaceae in preferring anaerobic conditions for optimal growth (Antunes *et al.*, 2008). Although the unavailability of cultured isolates for the Nanohaloarchaea and *Halorhabdus*-related species present in Lake Tyrell currently precludes direct testing of these hypotheses, the linkage of detailed geochemical measurements and abundance data with the habitat-specific consensus population genomic sequences obtained in this study should provide valuable assistance to future cultivation efforts aimed at determining metabolic phenotypes experimentally.

The success of the current study in discovering and quantifying competitive and environmentally mediated effects on specific microbial taxa lacking cultured representatives would not have been possible without the reconstruction of multiple habitat-specific genomes via *de novo* metagenomic assembly. These newly reconstructed genomes from natural populations included many unique sequences absent from previously characterized microbial isolates, helping to ensure that databases used for metagenomic read classification encompassed sufficient community breadth to enable sensitive, accurate phylogenetic binning.

Despite these advances, metagenomic sequence assembly, phylogenetic binning, taxonomic abundance measurements and inferred predictions of gene function cannot, by themselves, completely describe the complexity of natural microbial communities. However, the application of these techniques in the current study has laid the essential groundwork necessary for future experiments using metatranscriptomic and metaproteomic techniques to capture gene expression levels linked to specific community members. These results can then be combined with long-term ecological monitoring and high-resolution sampling to reveal the contributions of both major taxonomic groups and individual microbial strains to essential functional activities and environmental adaptations of the entire community.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgements

We thank Dawn Cardace and Jochen Brocks for sample collection assistance at Lake Tyrrell; Mike Dyall-Smith for generous access to reagents and laboratory equipment; Cheetham Salt Works (Lake Tyrrell, Australia) for permission to collect samples; Matt Lewis and the J Craig Venter Institute for library construction and sequencing; and the US Department of Energy Joint Genomes Institute for genome annotation support via the Integrated Microbial Genome Expert Review (IMG-ER) resource. Funding for this work was provided by NSF award number 0626526 (JFB, KBH and EEA). JAU was supported by a Fulbright–Conicyt fellowship.

References

Andrei AS, Banciu HL, Oren A. (2012). Living with salt: metabolic and phylogenetic diversity of archaea inhabiting saline ecosystems. *FEMS Microbiol Lett* **330**: 1–9.

Antunes A, Taborda M, Huber R, Moissl C, Nobre MF, da Costa MS. (2008). *Halorhabdus tiamatea* sp. nov., a non-pigmented, extremely halophilic archaeon from a deep-sea, hypersaline anoxic basin of the Red Sea, and emended description of the genus *Halorhabdus*. *Int J Syst Evol Microbiol* **58**: 215–220.

Atanasova NS, Pietilä MK, Oksanen HM. (2013). Diverse antimicrobial interactions of halophilic archaea and bacteria extend over geographical distances and cross the domain barrier. *MicrobiologyOpen* **2**: 811–825.

Béjà O, Spudich EN, Spudich JL, Leclerc M, DeLong EF. (2001). Proteorhodopsin phototrophy in the ocean. *Nature* **411**: 786–789.

Bizarro CV, Alemany A, Ritort F. (2012). Non-specific binding of Na^+ and Mg^{2+} to RNA determined by force spectroscopy methods. *Nucleic Acids Res* **40**: 6922–6935.

Bodaker I, Sharon I, Suzuki MT, Feingersch R, Shmoish M, Andreishcheva E *et al.* (2009). Comparative community genomics in the Dead Sea: an increasingly extreme environment. *Isme J* **4**: 399–407.

Bolhuis A, Kwan D, Thomas JR. (2008). Halophilic Adaptations of Proteins. In: Siddiqui KS, Thomas T (eds) *Protein Adaptation in Extremophiles*. Nova Biomedical Books, pp 71–104.

Bolhuis H, Palm P, Wende A, Falb M, Rampp M, Rodriguez-Valera F *et al.* (2006). The genome of the square archaeon *Haloquadratum walsbyi*: life at the limits of water activity. *BMC Genomics* **7**: 169.

Bolhuis H, Poele EM, Rodriguez-Valera F. (2004). Isolation and cultivation of Walsby's square archaeon. *Environ Microbiol* **6**: 1287–1291.

Boujelben I, Gomariz M, Martinez-Garcia M, Santos F, Peña A, López C *et al.* (2012). Spatial and seasonal prokaryotic community dynamics in ponds of increasing salinity of Sfax solar saltern in Tunisia. *Antonie Van Leeuwenhoek* **101**: 845–857.

Brady A, Salzberg S. (2011). PhymmBL expanded: confidence scores, custom databases, parallelization and more. *Nat Methods* **8**: 367.

Burns DG, Camakaris HM, Janssen PH, Dyall-Smith ML. (2004). Cultivation of Walsby's square haloarchaeon. *FEMS Microbiol Lett* **238**: 469–473.

Ciccarelli FD, Doerks T, von Mering C, Creevey CJ, Snel B, Bork P. (2006). Toward automatic reconstruction of a highly resolved tree of life. *Science* **311**: 1283–1287.

de Médicis E, Paquette J, Gauthier JJ, Shapcott D. (1986). Magnesium and manganese content of halophilic bacteria. *Appl Environ Microbiol* **52**: 567–573.

Demergasso C, Escudero L, Casamayor EO, Chong G, Balague V, Pedrós-Alió C. (2008). Novelty and spatiotemporal heterogeneity in the bacterial diversity of hypersaline Lake Tebenquiche (Salar de Atacama). *Extremophiles* **12**: 491–504.

DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K *et al.* (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* **72**: 5069–5072.

Dyall-Smith M. (2009). The Halohandbook. Protocols for haloarchaeal genetics, version 7.2, http://www.haloarchaea.com/resources/halohandbook/.

Edgar RC. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**: 1792–1797.

Emerson JB, Andrade K, Thomas BC, Norman A, Allen EE, Heidelberg KB *et al.* (2013). Virus-host and CRISPR dynamics in archaea-dominated hypersaline Lake Tyrrell, Victoria, Australia. *Archaea* **2013**: 370871.

Emerson JB, Thomas BC, Andrade K, Allen EE, Heidelberg KB, Banfield JF. (2012). Dynamic viral populations in hypersaline systems as revealed by

metagenomic assembly. *Appl Environ Microbiol* **78**: 6309–6320.
Fukuchi S, Yoshimune K, Wakayama M, Moriguchi M, Nishikawa K. (2003). Unique amino acid composition of proteins in halophilic bacteria. *J Mol Biol* **327**: 347–357.
Giovannoni SJ, Tripp HJ, Givan S, Podar M, Vergin KL, Baptista D *et al.* (2005). Genome streamlining in a cosmopolitan oceanic bacterium. *Science* **309**: 1242–1245.
Goldberg SM, Johnson J, Busam D, Feldblyum T, Ferriera S, Friedman R *et al.* (2006). A Sanger/pyrosequencing hybrid approach for the generation of high-quality draft assemblies of marine microbial genomes. *Proc Natl Acad Sci USA* **103**: 11240–11245.
Hartwig A. (2001). Role of magnesium in genomic stability. *Mutat Res* **475**: 113–121.
Heidelberg KB, Nelson WC, Holm JB, Eisenkolb N, Andrade K, Emerson JB. (2013). Characterization of eukaryotic microbial diversity in hypersaline Lake Tyrrell, Australia. *Front Microbiol* **4**: 115.
Ionescu D, Siebert C, Polerecky L, Munwes YY, Lott C, Häusler S *et al.* (2012). Microbial and chemical characterization of underwater fresh water springs in the Dead Sea. *PLoS One* **7**: e38319.
Javor B. (1989). *Hypersaline Environments: Microbiology and Biogeochemistry*. Springer-Verlag: Berlin; New York.
Macumber PG. (1992). Hydrological processes in the Tyrrell Basin, southeastern Australia. *Chemical Geology* **96**: 1–18.
Makhdoumi-Kakhki A, Amoozegar MA, Kazemi B, Pašic L, Ventosa A. (2012). Prokaryotic diversity in Aran-Bidgol salt lake, the largest hypersaline playa in Iran. *Microbes Environ* **27**: 87–93.
Markowitz VM, Mavromatis K, Ivanova NN, Chen IM, Chu K, Kyrpides NC. (2009). IMG ER: a system for microbial genome annotation expert review and curation. *Bioinformatics* **25**: 2271–2278.
Narasingarao P, Podell S, Ugalde JA, Brochier-Armanet C, Emerson JB, Brocks JJ *et al.* (2012). De novo metagenomic assembly reveals abundant novel major lineage of Archaea in hypersaline microbial communities. *Isme J* **6**: 81–93.
Oh D, Porter K, Russ B, Burns D, Dyall-Smith M (2010). Diversity of *Haloquadratum* and other haloarchaea in three, geographically distant, Australian saltern crystallizer ponds. *Extremophiles* **14**: 161–169.
Oren A. (1998). Life and survival in a magnesium chloride brine: the biology of the Dead Sea. *Instrum Meth Missions Astrobiol* **1998**: 44–54.
Oren A. (2013). Life at High Salt Concentrations. In: Rosenberg E (ed) *The Prokaryotes – Prokaryotic Communities and Ecophysiology (4th edition)*. Springer-Verlag: Berlin: Heidelberg, pp 421–440.
Oren A, Heldal M, Norland S, Galinski EA. (2002). Intracellular ion and organic solute concentrations of the extremely halophilic bacterium *Salinibacter ruber*. *Extremophiles* **6**: 491–498.
Owczarzy R, Moreira BG, You Y, Behlke MA, Walder JA. (2008). Predicting stability of DNA duplexes in solutions containing magnesium and monovalent cations. *Biochemistry* **47**: 5336–5353.
Pagaling E, Wang H, Venables M, Wallace A, Grant WD, Cowan DA *et al.* (2009). Microbial biogeography of six salt lakes in Inner Mongolia, China, and a salt lake in Argentina. *Appl Environ Microbiol* **75**: 5750–5760.
Park JS. (2012). Effects of different ion compositions on growth of obligately halophilic protozoan *Halocafeteria seosinensis*. *Extremophiles* **16**: 161–164.
Podell S, Gaasterland T. (2007). DarkHorse: a method for genome-wide prediction of horizontal gene transfer. *Genome Biol* **8**: R16.
Podell S, Gaasterland T, Allen EE. (2008). A database of phylogenetically atypical genes in archaeal and bacterial genomes, identified using the DarkHorse algorithm. *BMC Bioinformatics* **9**: 419.
Podell S, Ugalde JA, Narasingarao P, Banfield JF, Heidelberg KB, Allen EE. (2013). Assembly-driven community genomics of a hypersaline microbial ecosystem. *PLoS One* **8**: e61692.
Price MN, Dehal PS, Arkin AP. (2010). FastTree 2–approximately maximum-likelihood trees for large alignments. *PLoS One* **5**: e9490.
Puigbò P, Wolf YI, Koonin EV. (2009). Search for a 'Tree of Life' in the thicket of the phylogenetic forest. *J Biol* **8**: 59.
Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng JF *et al.* (2013). Insights into the phylogeny and coding potential of microbial dark matter. *Nature* **499**: 431–437.
Ruiz-González C, Lefort T, Gali M, Montserrat Sala M, Sommaruga R, Simó R *et al.* (2011). Seasonal patterns in the sunlight sensitivity of bacterioplankton from Mediterranean surface coastal waters. *FEMS Microbiol Ecol* **79**: 661–674.
Saunders NF, Thomas T, Curmi PM, Mattick JS, Kuczek E, Slade R *et al.* (2003). Mechanisms of thermal adaptation revealed from the genomes of the Antarctic Archaea *Methanogenium frigidum* and *Methanococcoides burtonii*. *Genome Res* **13**: 1580–1588.
Sharma AK, Spudich JL, Doolittle WF. (2006). Microbial rhodopsins: functional versatility and genetic mobility. *Trends Microbiol* **14**: 463–469.
Sherwood JE, Stagnitti F, Kokkinn MJ. (1991). Dissolved oxygen concentrations in hypersaline waters. *Limnoi Oceanogr* **36**: 235–250.
Ugalde JA, Podell S, Narasingarao P, Allen EE. (2011). Xenorhodopsins, an enigmatic new class of microbial rhodopsins horizontally transferred between archaea and bacteria. *Biol Direct* **6**: 52.
van der Wielen PW, Bolhuis H, Borin S, Daffonchio D, Corselli C, Giuliano L *et al.* (2005). The enigma of prokaryotic life in deep hypersaline anoxic basins. *Science* **307**: 121–123.
Walsby AE. (1980). A square bacterium. *Nature* **283**: 69–71.
Wu M, Eisen JA. (2008). A simple, fast, and accurate method of phylogenomic inference. *Genome Biol* **9**: R151.

Supplementary Information accompanies this paper on The ISME Journal website (http://www.nature.com/ismej)