

REPORT

Millimeter-scale genetic gradients and community-level molecular convergence in a hypersaline microbial mat

Victor Kunin¹, Jeroen Raes², J Kirk Harris³, John R Spear⁴, Jeffrey J Walker⁵, Natalia Ivanova⁶, Christian von Mering⁷, Brad M Bebout⁸, Norman R Pace⁵, Peer Bork² and Philip Hugenholtz^{1,*}

¹ Microbial Ecology Program, DOE Joint Genome Institute, Walnut Creek, CA, USA, ² Structural and Computational Biology Unit, European Molecular Biology Laboratory, Heidelberg, Germany, ³ Department of Pediatrics, University of Colorado Denver, Aurora, CO, USA, ⁴ Division of Environmental Science and Engineering, Colorado School of Mines, Golden, CO, USA, ⁵ Department of Molecular, Cellular and Developmental Biology, University of Colorado, Boulder, CO, USA, ⁶ Genome Biology Program, DOE Joint Genome Institute, Walnut Creek, CA, USA, ⁷ Institute of Molecular Biology and Swiss Institute of Bioinformatics, University of Zurich, Zurich, Switzerland and ⁸ Microbial Ecology/Biogeochemistry Research Laboratory, Exobiology branch, NASA Ames Research Center, Moffett Field, CA, USA

* Corresponding author. Microbial Ecology Program, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA. Tel.: +1 925 296 5725; Fax: +1 925 296 5720; E-mail: phugenholtz@lbl.gov

Received 3.3.08; accepted 28.4.08

To investigate the extent of genetic stratification in structured microbial communities, we compared the metagenomes of 10 successive layers of a phylogenetically complex hypersaline mat from Guerrero Negro, Mexico. We found pronounced millimeter-scale genetic gradients that were consistent with the physicochemical profile of the mat. Despite these gradients, all layers displayed near-identical and acid-shifted isoelectric point profiles due to a molecular convergence of amino-acid usage, indicating that hypersalinity enforces an overriding selective pressure on the mat community.

Molecular Systems Biology 3 June 2008; doi:10.1038/msb.2008.35

Subject Categories: cellular metabolism; microbiology and pathogens

Keywords: metagenomics; hypersalinity; microbial ecology; fine-scale; salt-in

This is an open-access article distributed under the terms of the Creative Commons Attribution Licence, which permits distribution and reproduction in any medium, provided the original author and source are credited. Creation of derivative works is permitted but the resulting work may be distributed only under the same or similar licence to this one. This licence does not permit commercial exploitation without specific permission.

Introduction

Ecosystems often exhibit distinct gradients. Physicochemical gradients have long been documented, but only recently has environmental shotgun sequencing allowed the associated functional (gene-based) gradients of an ecosystem's biota to be addressed. Macroscale functional gradients have been inferred from oceanic metagenomic data sets, both horizontally (Venter *et al*, 2004; Johnson *et al*, 2006; Rusch *et al*, 2007) and vertically (DeLong *et al*, 2006). Many structured microbial communities have been shown to produce steep physicochemical gradients on the scale of millimeters (Jorgensen *et al*, 1979; Schmitt-Wagner and Brune, 1999; Ludemann *et al*, 2000; Ley *et al*, 2006), but associated community-level functional gradients have not been demonstrated to date.

Here, we investigate a complex, stratified, hypersaline microbial mat from Guerrero Negro, Baja California Sur, Mexico, as a model for fine-scale functional variation (Ley *et al*, 2006). The dense, tofu-like texture of this mat allows intact cross-sections to be obtained down to ~1 mm thickness.

The mat shows pronounced physicochemical variation both in space and time: oxygen is detected routinely in the top 2 mm during the day (up to 700 μ M), and the mat is completely anoxic during the night. The permanently anoxic lower layers are characterized by micromolar levels of sulfide that increase with depth. The mat, dominated by bacteria, was reported to be one of the world's richest and most diverse microbial communities, comprising at least 752 observed species from 42 bacterial phyla, including 15 novel candidate phyla (Ley *et al*, 2006). As the mat grows in hypersaline waters (~3 \times the salinity of sea water), we were also interested to look for evidence of molecular adaptations to hypersalinity in the mat community.

Results and discussion

To investigate millimeter-scale genetic and associated functional stratification, we performed a metagenomic analysis of 10 spatially successive layers of the Guerrero Negro mat. Mat

core samples were collected during the day (Supplementary Table S1) and upper layers were sectioned at a finer scale (1 mm slices) than the lower layers (4–15 mm slices) to capture variation associated with the steep oxygen gradient in the upper millimeters of the mat (Supplementary Table S2). DNA from each layer was cloned and shotgun-sequenced using capillary sequencing with an average of ~13 000 reads per layer. No significant assembly of the reads was possible, even when all data were combined (largest contig was 8.4 kb from a combined assembly). We therefore chose to analyze only the unassembled data (average trimmed (Chou and Holmes, 2001) read length 808 bp) to avoid chimerism that has been reported to be frequent in contigs <10 kb (Mavromatis *et al*, 2007). Genes were predicted on vector and quality-trimmed reads with fgenesb (<http://www.softberry.com/>) using a generic bacterial model, resulting in an average of 13 600 genes per layer (Supplementary Table S2). These data have been deposited in genbank under accession numbers ABPP00000000 through ABPY00000000 and are available through the IMG/M system (Markowitz *et al*, 2006) (see SOM for access information).

Using both bulk similarity matches and phylogenetic mapping of conserved marker genes (von Mering *et al*, 2007a), we found strong phylogenetic variation between layers. Cyanobacteria and Alphaproteobacteria were the most abundant lineages in the top two layers (Supplementary Figure S1). Below the upper 2 mm, Proteobacteria, Bacteroidetes, Chloroflexi and Planctomycetes were the most represented phyla, with a notable peak in Bacteroidetes at 3 mm (Supplementary Figure S1). Numerous traces of other bacterial phyla as well as some archaea and eukaryotes were also identified. A large fraction of predicted proteins in layers below 2 mm did not have significant sequence similarity to any protein in public databases, reflecting the high degree of phylum-level novelty in the mat community (Ley *et al*, 2006). These metagenome-based findings are in broad agreement with single-marker gene surveys of the mat (Spear *et al*, 2003; Ley *et al*, 2006).

A rough measure of functional potential per organism can be made by estimating the average effective genome size (Raes *et al*, 2007). Using this method, we predicted an increased average bacterial genome size at the border of the oxic and anoxic zones (1–2 mm depth): 6 Mb at the border versus 3–3.5 Mb for the rest of the mat (Supplementary Figure S2).

This may reflect an increased functional complexity needed for survival in the constantly fluctuating conditions at this depth as was recently observed in the genome of a marine *Beggiotoa* occupying a similar niche (Mussmann *et al*, 2007).

To investigate genetic gradients through the mat, we determined the relative abundances of individual gene families and metabolic pathways between mat layers, and compared the mat data with external data sets for reference. Many gene families were highly abundant in the mat despite high overall functional diversity (Supplementary Figure S3) and very low sequence coverage of individual species. Indeed, the mat data set roughly doubled existing inventories for some of the gene families described below (Table I). This implies that multiple species and likely higher-level taxa contribute representatives of these families, and suggests that there has been a strong selection for a limited number of common functionalities in the mat.

The key aspect of this study was to use the metagenomic data to determine what, if any, millimeter-scale genetic gradients are detectable in this very complex and structured ecosystem. Several gene families and pathways either directly (Figure 1A) or inversely (Figure 1B) tracked the steep oxygen gradient in the top 2 mm of the mat and sulfide gradient below 2 mm. Genes directly involved in photosynthesis (KEGG map 00195) were statistically over-represented in the top two layers relative to lower layers. In addition, an uncharacterized protein domain (pfam05685) highly paralogous in phototrophic lineages (most cyanobacterial and some Chloroflexi genomes) showed a steep declining gradient in the top 6 mm (Figure 1A) consistent with dominance of phototrophs in the same region. Chaperones similarly tracked the oxygen gradient when all gene families with chaperone activity are combined together. The over-representation of chaperones in the top 2 mm relative to the rest of the mat may not be associated with oxygen concentration, but rather with heat stress caused by direct exposure to sunlight.

Gene families and pathways that tracked inversely with oxygen concentration included ferredoxins, trimethylamine methyltransferase (*Mttb*), sulfatases and sugar degradation pathways (Figure 1B). Ferredoxins and associated proteins show a four-fold increase from the top layer down to a depth of 4 mm and thereafter are uniformly over-represented. Two COG families are chiefly responsible for this trend: COG1148

Table I Most prominent gene families and domains in the Guerrero Negro hypersaline mat core relative to other sequenced microbiome samples^a

Gene family or domain ^b	Annotation	Mat	AMD	Soil	Whalefall	Gutless worm	Sludge	IMG
COG3119	Arylsulfatase A and related enzymes	640 (640)	0	195 (145)	46 (165)	16 (77)	32 (127)	1154 (55)
COG5598	Trimethylamine:corrinoid methyltransferase	112 (112)	0	16 (12)	5 (18)	52 (249)	3 (12)	114 (5)
COG1148	Heterodisulfide reductase, subunit A and related polyferredoxins	172 (172)	0	16 (12)	5 (18)	40 (192)	0	185 (9)
COG2414	Aldehyde:ferredoxin oxidoreductase	110 (110)	0	20 (15)	4 (14)	39 (187)	5 (20)	225 (11)
Pfam05685	DUF820 domain	142(142)	3 (32)	63 (47)	0	8 (38)	10 (40)	825 (40)

Numbers represent raw counts and numbers in parentheses are normalized for mat data set size.

^aMat (combined data from all layers; present study), AMD (acid mine drainage biofilm; Tyson *et al*, 2004), soil (Tringe *et al*, 2005), whalefall (sample 3; Tringe *et al*, 2005), gutless worm (Woyke *et al*, 2006), sludge (US; Garcia Martin *et al*, 2006), IMG (version 2.20, combined data from 728 microbial genomes; Markowitz *et al*, 2006).

^bCOG—cluster of orthologous genes (Tatusov *et al*, 1997), pfam (Bateman *et al*, 2002).

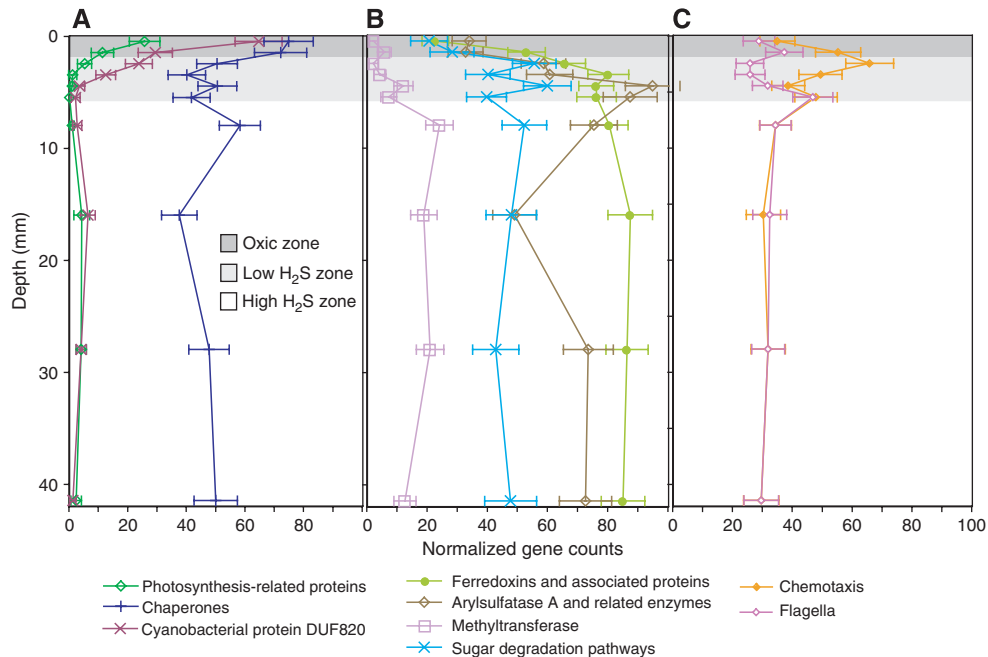


Figure 1 Vertical gradients of gene families or groups of functionally related gene families enriched in the oxic zone (A), anoxic (high H₂S) zone (B) and varying across the oxic–anoxic border (low H₂S) zone (C). Relative abundance is normalized by the average number of genes in a layer. In most cases, these genes and groups of genes were over-represented relative to other metagenomic data sets (Table I). Error bars denote standard deviations calculated from 1000 bootstrap resamplings of predicted proteins, and points with non-overlapping error bars are treated as significantly different. Lists of gene families used in each group (Photosynthesis-related proteins, Chaperones, Ferredoxins and associated proteins, Sugar degradation pathways, Chemotaxis and Flagella) as well as details of the resampling procedure are given in Supplementary information.

(heterodisulfide reductase, subunit A and related polyferredoxins) and COG2414 (aldehyde:ferredoxin oxidoreductase). The expansion of ferredoxins in the anoxic layers likely reflects the diversification of redox reactions required for anaerobic respiration. *Mttb* (pfam06253, COG5598) methyltransferase does not become significantly over-represented until at least 7 mm into the mat (Figure 1B), well below the anoxic boundary. *Mttb* was initially identified as a protein facilitating the first step of methanogenesis from trimethylamine in Methanosarcinaceae (Paul *et al*, 2000). However, this gene family is also found in methylotrophic bacteria (e.g. in Rhodobacteraceae and Rhizobiaceae), suggesting a more generalized role in C1 metabolism.

One of the most pronounced inverse gradients is observed for sulfatases (COG3119) that are involved in the hydrolysis of sulfated organic compounds (Figure 1B). As sulfatases can function in the presence of oxygen, the gradient is presumably a reflection of the availability of sulfated compounds in the mat. Although the concentration gradient of sulfated compounds is not known in the mat, they are produced by phototrophs (Kates, 1986) and are widespread in marine environments (Glockner *et al*, 2003). Sulfatase genes obtained from the mat exhibited extensive sequence divergence, suggesting that a corresponding wide variety of sulfated organic substrates are present in the mat, with the highest concentrations below 2 mm. The over-representation of this gene family may in part be due to an expansion of sulfatase genes in the genomes of Planctomycetes, suggested to be involved particularly in the hydrolysis of sulfated glycopolymers (Glockner *et al*, 2003).

Sugar degradation pathways (glycolysis and pentose and uronic acid degradation) show a two-fold increase with depth through the top 3 mm and maintain high relative representation in the anoxic lower layers (Figure 1B). This suggests that heterotrophic metabolism of sugars, particularly pentoses and uronic acids, is important in the lower layers.

Organisms living at the boundary between the oxic and anoxic zones could potentially accumulate substrates with high reductive potential in the anoxic zone, and then move to the oxic zone to harvest this potential by oxidation (Mussmann *et al*, 2007). This would require boundary zone organisms to be motile and chemotactic. Indeed, we find that chemotaxis signature genes peak sharply at the oxic–anoxic boundary (Figure 1C). Flagella appear not to be the dominant source of motility in these chemotactic organisms as flagellar genes actually dip in this region (Figure 1C). Chemotactic gliding bacteria have been observed in fresh mat cores (Garcia-Pichel *et al*, 1994; Kruschel and Castenholz, 1998) and our molecular data suggest they are most abundant in the boundary zone, bridging the oxic and anoxic layers.

Despite the pronounced phylogenetic and functional gradients in the mat, hypersalinity is a selective pressure common to the whole community. A known adaptation to hypersalinity is enrichment of proteins with acidic amino acids, allowing proteins to function in high cytoplasmic salt concentrations (Soppa, 2006). The resulting acid-shifted protein isoelectric points have been documented in the genomes of only two lineages, the archaeal class Halobacteria (Kennedy *et al*, 2001; Soppa, 2006) and the bacterial species *Salinibacter ruber* (Oren and Mana, 2002; Mongodin *et al*, 2005), so it is

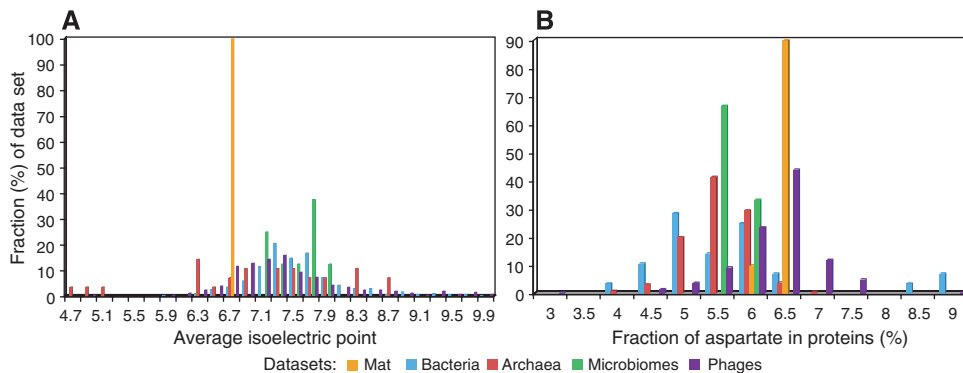


Figure 2 Average isoelectric point (A) and aspartate content (B) of all predicted proteins in the mat layer communities and reference bacteria, archaea, phages and microbiomes available through IMG/M (Markowitz *et al*, 2006). Genomic average was computed for each genome or microbiome, with 10 layers of the mat treated separately. These values were rounded up to the next (larger value) bin in increments of 0.2 and 0.5 in (A) and (B) respectively, and the distribution of the bins was plotted as a fraction of each data set.

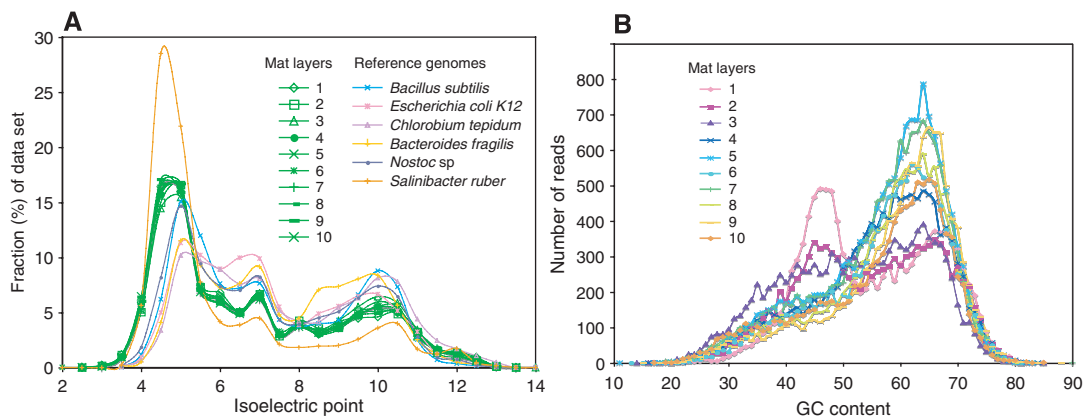


Figure 3 Isoelectric point profiles of predicted proteins (A) and GC content profiles of reads (B) for mat layer communities. In (A), isoelectric point profiles for selected reference genomes are added to highlight the highly overlapping and acid-shifted mat layer profiles.

unclear how widespread this mechanism is in halophilic communities.

The average isoelectric points of the mat layer communities are conspicuously acid-shifted when compared with most bacteria and microbiomes that are non-halophilic (Figure 2A). We determined this to be due primarily to an enrichment in the acidic amino acid aspartate (Figure 2B). Furthermore, the isoelectric profiles of all 10 layers converge on a common acid-shifted profile (Figure 3A) despite a significant variation in GC content between layers (Figure 3B), reflecting differing phylogenetic composition. The latter is consistent with aspartate usage being GC-independent as it can be encoded by both GC-rich and GC-poor codons (GAC and GAT, respectively). As each metagenomic read pair is likely derived from different species and no single species dominates the mat community, we conclude that a significant fraction of the community has converged on the enrichment of low isoelectric point proteins.

In summary, this study demonstrates that millimeter-scale genetic gradients can be readily discerned through a vertical cross-section of a highly structured and complex microbial community using low sequence coverage. Furthermore, we could directly and inversely correlate many of the genetic gradients to the physicochemical profile of the mat. Microbial

biofilms are important in many habitats, including our own bodies (Kroes *et al*, 1999; Eckburg *et al*, 2005), and often display physicochemical gradients at millimeter to centimeter scales. However, few biofilms are as robust as microbial mats and methods may need to be adapted to preserve spatial structure (Webster *et al*, 2006) and allow the relevant fine-scale genetic gradients to be resolved.

Surprisingly, we found that adaptation to hypersalinity by enriching proteins with acidic amino acids is more widespread than previously appreciated. Although this is the first example of species-independent molecular convergence in a microbial community, we predict that similar convergence patterns will be observed in other communities adapted to similar or different environmental conditions, such as temperature (Gianese *et al*, 2001) or pressure (Simonato *et al*, 2006; Lauro and Bartlett, 2008).

Materials and methods

Mat core samples were collected around 1400 hours from pond 4 near pond 5 at the Exportadora de Sal saltworks, Guerrero Negro, Baja California Sur, Mexico. The salinity of the bulk water above the mat was ~9% (~3 × the salinity of sea water). Other metadata for the sample can be found in Supplementary Table S1. Four replicate cores were collected, sectioned into layers with sterile scalpels and

DNA extracted, normalized, pooled and sequenced as described in Supplementary information. Metagenome sequence data are available under the following GenBank accession numbers: ABPP000000000, ABPQ000000000, ABPR000000000, ABPS000000000, ABPT000000000, ABPU000000000, ABPV000000000, ABPW000000000, ABPX000000000, ABPY000000000.

Community composition analysis was performed using the consensus of (i) best BlastP hits (Altschul *et al*, 1997) to the IMG/M database (Markowitz *et al*, 2006) and (ii) phylogenetic mapping of signature genes on a phylogenetic tree (von Mering *et al*, 2007a). See Supplementary information for details.

Gene-based functional gradients were calculated as follows: genes were assigned to their COG families (Tatusov *et al*, 1997) and pfam domains (Bateman *et al*, 2002) based on rpsBLAST (Altschul *et al*, 1997). The gradients were examined for possible over-representation of groups or individual families or domains, and 1000 bootstrap iterations were used to assess the significance of over-representation. The described gradients were independently confirmed using two databases: IMG/M (Markowitz *et al*, 2006) and the STRING database (von Mering *et al*, 2007b). Further details as well as groupings of families/domains are described in Supplementary information.

Isoelectric point distributions, amino-acid composition and GC content were computed using appropriate perl scripts and modules as described in Supplementary information.

Supplementary information

Supplementary information is available at the *Molecular Systems Biology* website (www.nature.com/msb).

Acknowledgements

We thank Amber Hartman for fruitful discussions and the Exportadora de Sal saltworks in Guerrero Negro, Baja California Sur, for access and assistance with the field site. We also thank the NASA funded researchers at NASA Ames who assisted with the field work: David Des Marais, Moira Doty, Tori Hoehler, Mary Hogan and Kendra Turk. Sequencing was provided by the JGI Community Sequencing Program. This work was performed under the auspices of the US Department of Energy's Office of Science, Biological and Environmental Research Program, and the University of California, Lawrence Livermore National Laboratory under contract no. W-7405-Eng-48, Lawrence Berkeley National Laboratory under contract no. DE-AC02-05CH11231 and Los Alamos National Laboratory under contract no. DE-AC02-06NA25396. JR and PB are supported by the European Union 6th Framework Program (contract no. LSHG-CT-2004-503567).

References

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389–3402

Bateman A, Birney E, Cerruti L, Durbin R, Ewinger L, Eddy SR, Griffiths-Jones S, Howe KL, Marshall M, Sonnhammer EL (2002) The Pfam protein families database. *Nucleic Acids Res* **30**: 276–280

Chou HH, Holmes MH (2001) DNA sequence quality trimming and vector removal. *Bioinformatics (Oxford, England)* **17**: 1093–1104

DeLong EF, Preston CM, Mincer T, Rich V, Hallam SJ, Frigaard NU, Martinez A, Sullivan MB, Edwards R, Brito BR, Chisholm SW, Karl DM (2006) Community genomics among stratified microbial assemblages in the ocean's interior. *Science (New York, NY)* **311**: 496–503

Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M, Gill SR, Nelson KE, Relman DA (2005) Diversity of the human intestinal microbial flora. *Science (New York, NY)* **308**: 1635–1638

Garcia Martin H, Ivanova N, Kunin V, Warnecke F, Barry KW, McHardy AC, Yeates C, He S, Salamov AA, Szeto E, Dalin E, Putnam NH,

Shapiro HJ, Pangilinan JL, Rigoutsos I, Kyrpides NC, Blackall LL, McMahon KD, Hugenholtz P (2006) Metagenomic analysis of two enhanced biological phosphorus removal (EBPR) sludge communities. *Nat Biotechnol* **24**: 1263–1269

Garcia-Pichel F, Mechling M, Castenholz RW (1994) Diel migrations of microorganisms within a benthic, hypersaline mat community. *Appl Environ Microbiol* **60**: 1500–1511

Gianese G, Argos P, Pascarella S (2001) Structural adaptation of enzymes to low temperatures. *Protein Eng* **14**: 141–148

Glockner FO, Kube M, Bauer M, Teeling H, Lombardot T, Ludwig W, Gade D, Beck A, Borzym K, Heitmann K, Rabus R, Schlesner H, Amann R, Reinhardt R (2003) Complete genome sequence of the marine planctomycete *Pirellula* sp. strain 1. *Proc Natl Acad Sci USA* **100**: 8298–8303

<http://www.softberry.com/SoftBerry-fgenesb>

Johnson ZI, Zinser ER, Coe A, McNulty NP, Woodward EM, Chisholm SW (2006) Niche partitioning among *Prochlorococcus* ecotypes along ocean-scale environmental gradients. *Science (New York, NY)* **311**: 1737–1740

Jorgensen BB, Revsbech NP, Blackburn TH, Cohen Y (1979) Diurnal cycle of oxygen and sulfide microgradients and microbial photosynthesis in a cyanobacterial mat sediment. *Appl Environ Microbiol* **38**: 46–58

Kates M (1986) *Techniques of Lipidology: Isolation, Analysis, and Identification of Lipids*, 2nd revised edn. Amsterdam, New York: Elsevier

Kennedy SP, Ng WV, Salzberg SL, Hood L, DasSarma S (2001) Understanding the adaptation of *Halobacterium* species NRC-1 to its extreme environment through computational analysis of its genome sequence. *Genome Res* **11**: 1641–1650

Kroes I, Lepp PW, Relman DA (1999) Bacterial diversity within the human subgingival crevice. *Proc Natl Acad Sci USA* **96**: 14547–14552

Kruschel C, Castenholz R (1998) The effect of solar UV and visible irradiance on the vertical movements of cyanobacteria in microbial mats of hypersaline waters. *FEMS Microbiol Ecol* **27**: 53–72

Lauro FM, Bartlett DH (2008) Prokaryotic lifestyles in deep sea habitats. *Extremophiles* **12**: 15–25

Ley RE, Harris JK, Wilcox J, Spear JR, Bebout BM, Maresca JA, Bryant DA, Sogin ML, Pace NR (2006) Unexpected diversity and complexity of the Guerrero Negro hypersaline microbial mat. *Appl Environ Microbiol* **72**: 3685–3695

Ludemann H, Arth I, Liesack W (2000) Spatial changes in the bacterial community structure along a vertical oxygen gradient in flooded paddy soil cores. *Appl Environ Microbiol* **66**: 754–762

Markowitz VM, Ivanova N, Palaniappan K, Szeto E, Korzeniewski F, Lykidis A, Anderson I, Mavrommatis K, Kunin V, Garcia Martin H, Dubchak I, Hugenholtz P, Kyrpides NC (2006) An experimental metagenome data management and analysis system. *Bioinformatics (Oxford, England)* **22**: e359–e367

Mavrommatis K, Ivanova N, Barry K, Shapiro H, Goltsman E, McHardy AC, Rigoutsos I, Salamov A, Korzeniewski F, Land M, Lapidus A, Grigoriev I, Richardson P, Hugenholtz P, Kyrpides NC (2007) Use of simulated data sets to evaluate the fidelity of metagenomic processing methods. *Nat Methods* **4**: 495–500

Mongodin EF, Nelson KE, Daugherty S, Deboy RT, Wister J, Khouri H, Weidman J, Walsh DA, Papke RT, Sanchez Perez G, Sharma AK, Nesbo CL, MacLeod D, Baptiste E, Doolittle WF, Charlebois RL, Legault B, Rodriguez-Valera F (2005) The genome of *Salinibacter ruber*: convergence and gene exchange among hyperhalophilic bacteria and archaea. *Proc Natl Acad Sci USA* **102**: 18147–18152

Mussmann M, Hu FZ, Richter M, de Beer D, Preisler A, Jorgensen BB, Huntemann M, Glockner FO, Amann R, Koopman WJ, Lasken RS, Janto B, Hogg J, Stoodley P, Boissy R, Ehrlich GD (2007) Insights into the genome of large sulfur bacteria revealed by analysis of single filaments. *PLoS Biol* **5**: e230

Oren A, Mana L (2002) Amino acid composition of bulk protein and salt relationships of selected enzymes of *Salinibacter ruber*, an extremely halophilic bacterium. *Extremophiles* **6**: 217–223

- Paul L, Ferguson Jr DJ, Krzycki JA (2000) The trimethylamine methyltransferase gene and multiple dimethylamine methyltransferase genes of *Methanosarcina barkeri* contain in-frame and read-through amber codons. *J Bacteriol* **182**: 2520–2529
- Raes J, Korbel JO, Lercher MJ, von Mering C, Bork P (2007) Prediction of effective genome size in metagenomic samples. *Genome Biol* **8**: R10
- Rusch DB, Halpern AL, Sutton G, Heidelberg KB, Williamson S, Yooseph S, Wu D, Eisen JA, Hoffman JM, Remington K, Beeson K, Tran B, Smith H, Baden-Tillson H, Stewart C, Thorpe J, Freeman J, Andrews-Pfannkoch C, Venter JE, Li K *et al* (2007) The Sorcerer II global ocean sampling expedition: northwest Atlantic through eastern tropical pacific. *PLoS Biol* **5**: e77
- Schmitt-Wagner D, Brune A (1999) Hydrogen profiles and localization of methanogenic activities in the highly compartmentalized hindgut of soil-feeding higher termites (*Cubitermes* spp). *Appl Environ Microbiol* **65**: 4490–4496
- Simonato F, Campanaro S, Lauro FM, Vezzi A, D'Angelo M, Vitulo N, Valle G, Bartlett DH (2006) Piezophilic adaptation: a genomic point of view. *J Biotechnol* **126**: 11–25
- Soppa J (2006) From genomes to function: haloarchaea as model organisms. *Microbiology* **152**: 585–590
- Spear JR, Ley RE, Berger AB, Pace NR (2003) Complexity in natural microbial ecosystems: the Guerrero Negro experience. *Biol Bull* **204**: 168–173
- Tatusov RL, Koonin EV, Lipman DJ (1997) A genomic perspective on protein families. *Science (New York, NY)* **278**: 631–637
- Tringe SG, von Mering C, Kobayashi A, Salamov AA, Chen K, Chang HW, Podar M, Short JM, Mathur EJ, Detter JC, Bork P, Hugenholtz P, Rubin EM (2005) Comparative metagenomics of microbial communities. *Science (New York, NY)* **308**: 554–557
- Tyson GW, Chapman J, Hugenholtz P, Allen EE, Ram RJ, Richardson PM, Solovyev VV, Rubin EM, Rokhsar DS, Banfield JF (2004) Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **428**: 37–43
- Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA, Wu D, Paulsen I, Nelson KE, Nelson W, Fouts DE, Levy S, Knap AH, Lomas MW, Nealson K, White O, Peterson J, Hoffman J, Parsons R, Baden-Tillson H *et al* (2004) Environmental genome shotgun sequencing of the Sargasso Sea. *Science (New York, NY)* **304**: 66–74
- von Mering C, Hugenholtz P, Raes J, Tringe SG, Doerks T, Jensen LJ, Ward N, Bork P (2007a) Quantitative phylogenetic assessment of microbial communities in diverse environments. *Science (New York, NY)* **315**: 1126–1130
- von Mering C, Jensen LJ, Kuhn M, Chaffron S, Doerks T, Kruger B, Snel B, Bork P (2007b) STRING 7—recent developments in the integration and prediction of protein interactions. *Nucleic Acids Res* **35**: D358–D362
- Webster P, Wu S, Gomez G, Apicella M, Plaut AG, St Geme III JW (2006) Distribution of bacterial proteins in biofilms formed by non-typeable *Haemophilus influenzae*. *J Histochem Cytochem* **54**: 829–842
- Woyke T, Teeling H, Ivanova NN, Huntemann M, Richter M, Gloeckner FO, Boffelli D, Anderson IJ, Barry KW, Shapiro HJ, Szeto E, Kyrpides NC, Musmann M, Amann R, Bergin C, Ruehland C, Rubin EM, Dubilier N (2006) Symbiosis insights through metagenomic analysis of a microbial consortium. *Nature* **443**: 950–955



Molecular Systems Biology is an open-access journal published by *European Molecular Biology Organization* and *Nature Publishing Group*.

This article is licensed under a Creative Commons Attribution-NonCommercial-Share Alike 3.0 Licence.