

# Metagenomics of extreme environments

Cowan, D.A\*, Ramond, J-B, Makhalanyane, T.P, De Maayer, P.

Centre of Microbial Ecology and Genetics, Department of Genetics, University of Pretoria, Pretoria,  
South Africa

\*Corresponding author: [don.cowan@up.ac.za](mailto:don.cowan@up.ac.za)

## Abstract

Whether they are exposed to extremes of heat, cold, or buried deep beneath the Earth's surface, microorganisms have an uncanny ability to survive under these conditions. This ability to survive has fascinated scientists for nearly a century, but the recent development of metagenomics and 'omics tools has allowed us to make huge leaps in understanding the remarkable complexity and versatility of extremophile communities. Here, in the context of the recently developed metagenomic tools, we discuss recent research on the community composition, adaptive strategies and biological functions of extremophiles.

## Introduction

The advent of genome sequencing has opened new windows into the global microbiome. The genomes of > 30,000 Bacteria and >700 Archaea (Genomes OnLine Database; <https://gold.jgi-psf.org>; December 2014) have provided key insights into, for example, their evolution and role in biogeochemical cycles, their modes of pathogenesis and antimicrobial resistance and the genetic determinants underlying their biotechnological potential. However, only a small fraction of the microorganisms on Earth are culturable [less than 1%: Amann R [1]]. Until very recently culturability was a pre-requisite for genome sequencing and for full access to the genetic complement of individual organisms. To some extent this limitation has been resolved by the development of the field of metagenomics [2] and single cell genomics. Metagenome sequencing, coupled with other 'omic' technologies, such as transcriptomics (measuring of mRNA transcript levels), proteomics (study

of the protein complement) and metabolomics (study of cellular metabolites), has led to the development of sophisticated systems biology approaches [3] which facilitate the combinatorial study of microbial community functions and their interactions within, and with, the environment. To date, more than 5,000 full metagenomes and over 16,000 amplicon libraries (Metagenomics RAST server; <http://metagenomics.anl.gov>; publically available metagenomes as of December 2014) have been sequenced and analysed, covering a broad cross-section of the Earth's biomes. Here we discuss current aspects of metagenomic analyses and the impacts of these state-of-the-art technologies on our understanding of the unique elements of some of the most well-known extremophilic microbial communities.

### **Thermophiles and thermophilic environments**

The field of thermophilic microbiology was effectively initiated in the late 1970's with the pioneering work of Thomas Brock [4] and dramatically expanded through the 1980's with the isolation of an extensive array of novel hyperthermophiles by Karl Stetter and his team at the University of Regensburg [5]. The development of ssu rRNA phylogenetics effectively revealed the true complexity and diversity of prokaryotic phylotypes across the Earth's widely differing thermophilic biotopes (spanning a pH gradient from less than 0 to more than 10, and a temperature gradient from 60°C to over 120°C). The discovery of deep sea hydrothermal vents (DSHV) added a unique ultra-high temperature dimension to 'thermophile' studies, and led to a range of isolation of hyperthermophilic archaeal species which steadily increased the known upper temperature limit of life [6].

There is a natural progression in environmental microbiology research, following a path of increasing technical and system complexity; from community diversity → environmental drivers of community structure → community function. As with all studies of environmental microbiology, our understanding of the function of thermophilic microbial consortia and their individual members, and of the interactions between organisms, has lagged substantially behind our understanding of basic species diversity. However, recent advances in 'omics' technologies, particularly when applied in a *systems biology* context, have made significant inroads into the prediction (if not the actual measurement) of *in situ* functionality.

The issue of which factors actually determine the composition (and function) of thermophilic microbial communities, traditionally restricted to temperature, pH and to a lesser extent  $dO_2$ , can be addressed at much a deeper and more sophisticated level by inclusion of extensive geochemical datasets and full community metagenome data [7,8], rather than just phylotypic data. Cluster analysis of orthologous genes and gene products, correlated with environmental variables, provides an elegant (and potentially testable) approach to understanding how metabolic functional capacity and functional guilds vary across gradients.

Community metagenome analysis can provide key insights into metabolic capacity in systems which are not readily accessed or subject to *in situ* studies. For example, sulphur-cycling genes dominated in the most abundant genera in sulphidic deep sea hydrothermal vent high temperature communities [9], while genes for  $H_2$  oxidation were prevalent in organisms in  $H_2$ -rich serpentinite hydrothermal vent communities [10]. Metagenome sequence data can confirm metabolic functionalities predicted from physicochemical analyses, such as the prevalence of lipid oxidation genes in DSHV communities in oil-rich substrates [11]. The coordinated use of phylogenetic data and transcript analysis confirmed that dominant epsilon proteobacteria in submarine hydrothermal field seeps were also the key functional taxa, using the abundant reduced sulphur compounds as the primary source of chemoautotrophic energy [12]. Genes for ammonia-oxidation (*amoA*) were prevalent in the most abundant genomes in ammonia-rich end member fluids in the Guaymas Basin [13,14], but novel mechanisms of N-assimilation, such as urea utilisation, were also implicated [14].

Metagenomic datasets can provide novel, and even unexpected insights into community structure and function. While it is widely acknowledged that the most understudied (and least understood) element of high temperature microbial communities is the phage/virus component, community metagenomes provide catalogues of integrated virus genomic elements, along with some host taxonomic identity data [15]. Metaviromic sequence datasets, admittedly most usually restricted to the dsDNA virus/phage fraction, nevertheless both complement and greatly extend 'classical' virology studies [15]. The plethora of integrative and Lateral Gene Transfer elements identified in a DSHV metavirome [15] is indicative of the adaptive benefits of gene acquisition.

New genomic approaches, such as analysis of CRISPR sequences, can provide valuable insights into host-virus pairs and virus host specificity [16]. Comparing the database of CRISPR spacers from all published archaeal and bacterial genomes (identified using a CRISPR RECOGNITION TOOL (CRT) [17] with reads from a viral metagenome demonstrated, surprisingly, that thermophiles showed a statistically significant higher number of CRISPR sequences in their genomes than mesophiles [18], suggesting that viruses/phage may play a more important role in thermophile community structures and functions than previously suspected.

### **Psychrophiles and Cold environments**

On the opposite end of the thermal spectrum, cold environments (polar, montane, deep marine and man-made) are populated by members of all three Kingdoms, and have been the focus of intensive research for decades. The biodiversity, physiology and adaptive strategies of psychrophiles have been extensively reviewed (see, for example: Margesin R, Schinner F, Marx J-C, Gerday C [19], Bej AK, Aislabie J, Atlas RM [20], Seckbach J, Oren A, Stan-Lotter H [21], Cowan DA [22],[23,24]. Here we highlight a few specific recent studies which have stimulated the reconsideration of established paradigms.

The cold oligotrophic desert soils of the Antarctic continent are widely considered to be one of the extreme environments on earth and, until as little as 15 years ago, largely unstudied by molecular phylogeneticists. Recent studies have yielded some surprising revelations: (i) standing microbial biomass is orders-of-magnitude higher than originally thought [25]; (ii) microbial community complexity and diversity is much higher than predicted and soils have a high level of spatial heterogeneity [26-28] despite high levels of Aeolian mixing and the absence of higher eukaryotes, and (iii) soil microbial communities exhibit surprisingly rapid structural changes in response to changing environmental conditions [29,30].

The trophic complexity and dynamics of aquatic psychrophilic communities, the extent of niche partitioning and the relationship between these factors and the specific physiological capacity of individual species has recently been demonstrated in elegant studies relying essentially on comparative genomics or metaproteogenomics [31,32].

The Antarctic continent harbours other unique niches, most notably the sub-glacial ice habitats, of which the Blood Falls discharge flows and Lake Vostok represent the most notable. Analyses of the former have revealed a community dominated by chemoautotrophs capable of acquiring energy from reduced iron and sulphur compounds (in an interesting parallel to DSHV hyperthermophile communities), while the latter awaits the acquisition of uncontaminated lake water or accretion ice samples [33]. Samples recovered from the much shallower Lake Whillans in West Antarctica [34] show a diverse but recognisable community of bacteria and archaea probably supported by chemolithoautotrophy, with important numbers of OTUs closely related to microorganisms able to use reduced nitrogen, iron and sulphur compounds as energy sources.

Deep cold marine habitats poly-extreme environments, where microbial populations exist under stable conditions of low temperature, high pressure, oligotrophy and super-saturated oxygen [35,36], nevertheless support substantial populations of bacteria and archaea, including members of the *Thaumarchaeota* and phylotypes related to the deep subterranean South African Gold Mine Euryarchaeotal Group [37]. It is generally believed that the availability of organic carbon is the dominant driver of microbial activity in deep marine sediment communities.

### **Extremely halophilic habitats**

The microbial communities associated with extreme halophilic environments have long been the subject of classical culturing and phylotypic analyses. Full random metagenome sequencing of multiple hypersaline systems (Santa Polasaltern in Spain - 13 to 37% salinity), the hypersaline lake Tyrell in Australia (29% salinity) and crystallizer ponds in the USA (18 to 38% salinity) show equally high phylotypic diversity and the general dominance of Archaea, in particular the square-shaped halophilic archaeon *Haloquadratum walsbyi* [38-41]. Remarkably, *de novo* sequence assembly of metagenomic samples from hypersaline environments has led to the discovery of a novel uncultivated class, the 'Nanohaloarchaea', previously detected only once by "classic" 16S rRNA amplicon sequencing [42] but now known to be dominant in these systems across the world [39,40]. *De novo* genome annotation has revealed the projected metabolic

capacities of the uncultured Nanohaloarchaea with (i) a unique combination of protein amino acids to increase both osmo-resistance and protein flexibility, (ii) the first complete archaeal pentose-phosphate-pathway and (iii) the absence of Gvp gas vesicle protein orthologs which are otherwise highly conserved in halophilic archaeal genomes [40]. Single-cell genomic analyses have further characterized a member of this euryarchaeal class as a low-GC photoheterotrophic, *salt-in* strategist with the ability to degrade polysaccharides [39].

The haloviral populations from saturated NaCl brines have only recently been characterized, and bacteriophage diversity closely reflects the dominant microbial phylotypes associated with hypersaline environments [43].

Detailed functional annotation of a hypersaline metagenome has shown a generally simplistic C and N biogeochemical cycling capacity, but with the capacity to use light as an energy source via various bacteriorhodopsins [41]. Haloresistance mechanisms are dominated by the capacity to synthesise compatible solutes (such as glycine, betaine, ectoine and trehalose) [41].

### **Acid environments and acidophiles**

The first full metagenome to be sequenced was that of a natural acidophilic biofilm [44]. As the major focus of JF Banfield's group [45-49], acid mine drainage (AMD) metagenomes revealed low-diversity microbial communities assembled into self-sufficient chemoautotrophic consortia [44,48]. From these data, the complete genomes of the bacterial AMD-specialists *Leptospirillum* group II and III, the archaeon *Ferroplasma* type II and the unculturable A- and G-Plasma archaeal clades were assembled [44]. These genomes provided invaluable insights into the molecular mechanisms for acidophilic survival, specific metabolic pathways (e.g. C fixation by the Calvin-Benson-Bassham cycle and/or the acetyl-CoA pathway, N fixation via a single *nif* operon or energy generation by ferrous iron oxidation) and the functioning of acidophilic communities as whole [44,49]. Moreover, the metagenomic data could be employed to develop novel culturing strategies for the uncultured diazotrophic, free-living, iron-oxidizing strain, *Leptospirillum ferrodiazotrophum* [47].

Highly acidic (pH 0–1) biofilms (‘snottites’) were recently investigated using metagenomic sequencing [50]. In contrast to the AMD biofilms, these are largely dominated by *Acidithiobacillus thiooxidans*, *Acidimicrobium* and *Ferrimicrobium* spp. and the archaeal ‘G-plasma clade’ of the Thermoplasmatales. Functional annotation revealed that the autotrophic *A. thiooxidans* can be considered as the snottite community ‘architect’ with the capacity to oxidize reduced S and to fix CO<sub>2</sub>. The absence of N-fixation gene cassettes reflects the ready availability of nitrogen in this system.

### **Deep subterranean habitats**

Deep subterranean habitats are arguably the least understood of the ‘extreme’ environments. Extremely high pressures, nutrient limitation, and osmotic stressors combine to produce an extreme environment dominated by microbial communities. It has been estimated that up to  $25 \times 10^{29}$  bacterial cells may be present in the terrestrial subsurface [51] and it is therefore likely that these populations play an important role in biogeochemical cycling [52,53]. However, despite this numerical abundance of cells, crucial questions regarding deep subterranean habitats remain largely unaddressed. For instance, we know surprisingly little of the functional diversity of organisms in deep subterranean habitats, of the drivers of community structure and function (e.g. environmental parameters), or how these organisms survive (and function) under apparent conditions of extreme oligotrophy [54]. Unsurprisingly, given the technical difficulties associated with sample acquisition, very few explorations of deep subterranean habitats have been undertaken [55,56].

Recently reports of the first subsurface metatranscriptomes, representing over 1 billion cDNA sequence reads from anaerobic sediments 159 metres beneath the sea floor, demonstrated that anaerobic metabolism of amino acids, carbohydrates and lipids were the main drivers of metabolic processes [53]. The study also confirmed that dissimilatory sulphite reductase (*dsr*) catalysed sulphate reduction by subsurface microbial guilds may be a key source of energy [57,58]. Interestingly, the majority of the cDNA transcripts were of bacterial origin, with fungal and archaeal transcripts formally minor components. What this study, and others, does confirm is that deep subsurface microbial communities are indeed active [55,57,59,60] and that metabolic activity and growth is apparently limited by energy and not the availability of carbon

and nitrogen compounds [59]. It remains unclear, however, which archaea or bacteria mediate key trophic functions.

Some key insights into nitrogen cycling processes in subsurface habitats have been gained via the application of metagenome sequencing [61]. Analysis of water samples retrieved from ultra-deep South African mines showed the ubiquitous presence of known variants of nitrogen cycling genes such as *NarV*, *NPD*, PAPS reductase, *NifH*, *NifD*, *NifK*, *NifE*, and *NifN*. Interestingly, the authors found no significant correlation in these samples to geographical or subsurface environmental parameters [61]. While the results of this study and ocean subsurface metagenomic surveys (reviewed by Lever MA [62]) have important implications for *in situ* biogeochemical cycling, it is still not clear to what extent such phylogenetic sequences merely represent historical signatures of extant species. It is therefore critical to link metagenomic surveys with metatranscriptomic and functional assay data in order to accurately quantify the functional capacity of microbial communities in subterranean habitats.

## **Conclusions**

The advent of modern metagenomic approaches has undoubtedly led to new insights into how microbial communities are shaped, the physiological roles of community members, and how they adapt to environmental extremes. Such sequence data, particularly when coupled with transcript sequences and with functional assay data have the capacity to generate sophisticated community network structures, which include elements of trophic interactions, community dynamics and metabolic capacities. Nevertheless, such studies even at the current level of sophistication, barely ‘scratch the surface’ of the true complexity of communities occupying hundreds or even thousands of different microbial species from two of the three kingdoms. The democratisation of ‘omics’ technologies, in terms of pricing, analytical tools, as well as the integration of ‘omic’ technologies into all-encompassing systems biology approaches, will almost certainly continue to challenge scientific paradigms on the ecology of extremophiles



## References and recommended reading

Papers of particular interest, published within the last two years, have been highlighted as:

\* of special interest

\*\* of outstanding interest

## Acknowledgements

We wish to thank the following organisations for financial support: South African National Research Foundation (DA, J-BR, PDM), the University of Pretoria (Research Development funding for JBR and TPM) and the Genomics Research Institute. We appreciate comments from the editor, which helped improve the article. The Centre for Microbial Ecology and Genomics supports a large research program on extreme environments with a focus on hot and cold deserts.

## References

1. Amann R: **Who is out there? Microbial aspects of biodiversity.** *Syst Appl Microbiol* (2000) **23**(1):1-8.
2. De Maayer P, Valverde A, Cowan DA: **The current state of metagenomic analysis.** *Genome Analysis: Current Procedures and Applications* (2014) 183.
3. Zhang W, Li F, Nie L: **Integrating multiple 'omics' analysis for microbial biology: Application and methodologies.** *Microbiology* (2010) **156**(2):287-301.
4. Brock TD: **The origins of research on thermophiles.** In: *Thermophiles biodiversity, ecology, and evolution.* Springer, (2001):1-9.
5. Stetter KO: **Hyperthermophilic procaryotes.** *FEMS Microbiol Rev* (1996) **18**(2-3):149-158.
6. Cowan D: **The upper limit of life: How far can we go.** *Trends Microbiol* (2004) **12**(58-60).

7. Xie W, Wang F, Guo L, Chen Z, Sievert SM, Meng J, Huang G, Li Y, Yan Q, Wu S: **Comparative metagenomics of microbial communities inhabiting deep-sea hydrothermal vent chimneys with contrasting chemistries.** *Isme J* (2010) 5(3):414-426.
  
8. Alsop EB, Boyd ES, Raymond J: **Merging metagenomics and geochemistry reveals environmental controls on biological diversity and evolution.** *BMC Ecol* (2014) 14(1):16.
  
9. Cao H, Wang Y, Lee OO, Zeng X, Shao Z, Qian P-Y: **Microbial sulfur cycle in two hydrothermal chimneys on the southwest indian ridge.** *MBio* (2014) 5(1):e00980-00913.  
 \* This paper shows a significant role of microbial sulphur cycle in low-temperature hydrothermal systems
  
10. Brazelton WJ, Nelson B, Schrenk MO: **Metagenomic evidence for h<sub>2</sub> oxidation and h<sub>2</sub> production by serpentinite-hosted subsurface microbial communities.** *Front Microbiol* (2011) 2
  
11. He Y, Xiao X, Wang F: **Metagenome reveals potential microbial degradation of hydrocarbon coupled with sulfate reduction in an oil-immersed chimney from guaymas basin.** *Front Microbiol* (2013) 4
  
12. Urich T, Lanzén A, Stokke R, Pedersen RB, Bayer C, Thorseth IH, Schleper C, Steen IH, Øvreas L: **Microbial community structure and functioning in marine sediments associated with diffuse hydrothermal venting assessed by integrated meta-omics.** *Environ Microbiol* (2013).  
 \*\* This is an exceptional study and details functioning of microbial communities in diffuse hydrothermal systems
  
13. Lesniewski RA, Jain S, Anantharaman K, Schloss PD, Dick GJ: **The metatranscriptome of a deep-sea hydrothermal plume is dominated by water column methanotrophs and lithotrophs.** *Isme J* (2012) 6(12):2257-2268.
  
14. Baker BJ, Lesniewski RA, Dick GJ: **Genome-enabled transcriptomics reveals archaeal populations that drive nitrification in a deep-sea hydrothermal plume.** *Isme J* (2012) 6(12):2269-2279.  
 \* A very interesting study which highlights the use of metatranscriptomic assembly as an approach for understanding microbial communities
  
15. Zillig W, Prangishvili D, Schleper C, Elferink M, Holz I, Albers S, Janekovic D, Götz D: **Viruses, plasmids and other genetic elements of thermophilic and hyperthermophilic archaea.** *FEMS Microbiol Rev* (1996) 18(2-3):225-236.

16. Anderson RE, Sogin ML, Baross JA: **Evolutionary strategies of viruses, bacteria and archaea in hydrothermal vent ecosystems revealed through metagenomics.** *PLoS One* (2014) **9**(10):e109696.
17. Bland C, Ramsey TL, Sabree F, Lowe M, Brown K, Kyrpides NC, Hugenholtz P: **Crispr recognition tool (crt): A tool for automatic detection of clustered regularly interspaced palindromic repeats.** *BMC Bioinformatics* (2007) **8**(1):209.
18. Anderson RE, Brazelton WJ, Baross JA: **Using crisprs as a metagenomic tool to identify microbial hosts of a diffuse flow hydrothermal vent viral assemblage.** *FEMS Microbiol Ecol* (2011) **77**(1):120-133.
19. Margesin R, Schinner F, Marx J-C, Gerday C: *Psychrophiles: From biodiversity to biotechnology.* Springer, (2008).
20. Bej AK, Aislabie J, Atlas RM: *Polar microbiology: The ecology, biodiversity and bioremediation potential of microorganisms in extremely cold environments.* CRC Press, (2009).
21. Seckbach J, Oren A, Stan-Lotter H: *Polyextremophiles: Life under multiple forms of stress.* 27. Springer, (2013).
22. Cowan DA: *Antarctic terrestrial microbiology: Physical and biological properties of antarctic soils* Springer, (2014).
23. Cowan DA, Makhalanyane TP, Dennis PG, Hopkins DW: **Microbial ecology and biogeochemistry of continental antarctic soils.** *Front Microbiol* (2014) **5**(154).
24. Makhalanyane TP, Valverde A, Velázquez D, Gunnigle E, Van Goethem MW, Quesada A, Cowan DA: **Ecology and biogeochemistry of cyanobacteria in soils, permafrost, aquatic and cryptic polar habitats.** *Biodiversity and Conservation* (2015) **24**(4):819-840.
25. Cowan DA, Russell NJ, Mamais A, Sheppard DM: **Antarctic dry valley mineral soils contain unexpectedly high levels of microbial biomass.** *Extremophiles* (2002) **6**(5):431-436.
26. Chong CW, Pearce DA, Convey P, Tan GYA, Wong RCS, Tan IKP: **High levels of spatial heterogeneity in the biodiversity of soil prokaryotes on signy island, antarctica.** *Soil Biology and Biochemistry* (2010) **42**(4):601-610.
27. Lee CK, Barbier BA, Bottos EM, McDonald IR, Cary SC: **The inter-valley soil comparative survey: The ecology of dry valley edaphic microbial communities.** *Isme J* (2012) **6**(5):1046-1057.

28. Makhalanya TP, Valverde A, Birkeland N-K, Cary SC, Marla Tuffin I, Cowan DA: **Evidence for successional development in antarctic hypolithic bacterial communities.** *Isme J* (2013) **7**(11):2080-2090.
29. Yergeau E, Bokhorst S, Kang S, Zhou J, Greer CW, Aerts R, Kowalchuk GA: **Shifts in soil microorganisms in response to warming are consistent across a range of antarctic environments.** *Isme J* (2012).
30. Tiao G, Lee CK, McDonald IR, Cowan DA, Cary SC: **Rapid microbial response to the presence of an ancient relic in the antarctic dry valleys.** *Nat Commun* (2012) **3**
- \* This study showed for the first time, that edaphic communities rapidly respond to environmental changes in cold deserts
31. Lauro FM, DeMaere MZ, Yau S, Brown MV, Ng C, Wilkins D, Raftery MJ, Gibson JA, Andrews-Pfannkoch C, Lewis M: **An integrative study of a meromictic lake ecosystem in antarctica.** *Isme J* (2010) **5**(5):879-895.
32. Williams TJ, Allen MA, DeMaere MZ, Kyrpides NC, Tringe SG, Woyke T, Cavicchioli R: **Microbial ecology of an antarctic hypersaline lake: Genomic assessment of ecophysiology among dominant haloarchaea.** *Isme J* (2014).
- \*\* This is a facinating study which uses comparative genomics together with growth assays to analyse the distribution of four Antarctic haloarchaea
33. Lukin VV, Vasiliev NI: **Technological aspects of the final phase of drilling borehole 5g and unsealing vostok subglacial lake, east antarctica.** *Ann Glaciol* (2014) **55**(65):83-89.
34. Christner BC, Priscu JC, Achberger AM, Barbante C, Carter SP, Christianson K, Michaud AB, Mikucki JA, Mitchell AC, Skidmore ML: **A microbial ecosystem beneath the west antarctic ice sheet.** *Nature* (2014) **512**(7514):310-313.
35. Lauro FM, Bartlett DH: **Prokaryotic lifestyles in deep sea habitats.** *Extremophiles* (2008) **12**(1):15-25.
36. Nogi Y: **Psychrophilic organisms in marine environments.** In: *Cold-adapted microorganisms.* Yunoto I (Ed) Caister Academic Press, Norfolk, UK, (2013):33-50.
37. Wang P, Wei Y, Li T, Li F, Meng J, Zhang CL: **Archaeal diversity and spatial distribution in the surface sediment of the south china sea.** *Geomicrobiol J* (2014) **31**(1):1-11.
- \* This paper provides an interesting profile of archaea from shallow and deep water

38. Pašić L, Rodriguez-Mueller B, Martin-Cuadrado A-B, Mira A, Rohwer F, Rodriguez-Valera F: **Metagenomic islands of hyperhalophiles: The case of *salinibacter ruber***. *BMC Genomics* (2009) **10**(1):570.
39. Ghai R, Pašić L, Fernández AB, Martin-Cuadrado A-B, Mizuno CM, McMahon KD, Papke RT, Stepanauskas R, Rodriguez-Brito B, Rohwer F: **New abundant microbial groups in aquatic hypersaline environments**. *Sci Rep* (2011) **1**(1):1-6.
40. Narasingarao P, Podell S, Ugalde JA, Brochier-Armanet C, Emerson JB, Brocks JJ, Heidelberg KB, Banfield JF, Allen EE: **De novo metagenomic assembly reveals abundant novel major lineage of archaea in hypersaline microbial communities**. *Isme J* (2011) **6**(1):81-93.
41. Fernández AB, Ghai R, Martin-Cuadrado AB, Sánchez-Porro C, Rodriguez-Valera F, Ventosa A: **Prokaryotic taxonomic and metabolic diversity of an intermediate salinity hypersaline habitat assessed by metagenomics**. *FEMS Microbiol Ecol* (2014).
- \* A well executed study providing metagenomic analysis of a hypersaline habitat
42. Casanueva A, Galada N, Baker GC, Grant WD, Heaphy S, Jones B, Yanhe M, Ventosa A, Blamey J, Cowan DA: **Nanoarchaeal 16s rRNA gene sequences are widely dispersed in hyperthermophilic and mesophilic halophilic environments**. *Extremophiles* (2008) **12**(5):651-656.
43. Garcia-Heredia I, Martin-Cuadrado A-B, Mojica FJ, Santos F, Mira A, Antón J, Rodriguez-Valera F: **Reconstructing viral genomes from the environment using fosmid clones: The case of haloviruses**. *PLoS One* (2012) **7**(3):e33802.
- \*\* This is a very interesting read and is one of the early metaviromic studies
44. Tyson GW, Chapman J, Hugenholtz P, Allen EE, Ram RJ, Richardson PM, Solovyev VV, Rubin EM, Rokhsar DS, Banfield JF: **Community structure and metabolism through reconstruction of microbial genomes from the environment**. *Nature* (2004) **428**(6978):37-43.
45. Ruepp A, Graml W, Santos-Martinez M-L, Koretke KK, Volker C, Mewes HW, Frishman D, Stocker S, Lupas AN, Baumeister W: **The genome sequence of the thermoacidophilic scavenger *Thermoplasma acidophilum***. *Nature* (2000) **407**(6803):508-513.
46. Baker BJ, Comolli LR, Dick GJ, Hauser LJ, Hyatt D, Dill BD, Land ML, VerBerkmoes NC, Hettich RL, Banfield JF: **Enigmatic, ultrasmall, uncultivated archaea**. *Proceedings of the National Academy of Sciences* (2010) **107**(19):8806-8811.
47. Tyson GW, Lo I, Baker BJ, Allen EE, Hugenholtz P, Banfield JF: **Genome-directed isolation of the key nitrogen fixer *Leptospirillum ferrodiazotrophum* sp. nov.**

- From an acidophilic microbial community.** *Appl Environ Microbiol* (2005) **71**(10):6319-6324.
48. Baker BJ, Banfield JF: **Microbial communities in acid mine drainage.** *FEMS Microbiol Ecol* (2003) **44**(2):139-152.
49. Yelton AP, Comolli LR, Justice NB, Castelle C, Denev VJ, Thomas BC, Banfield JF: **Comparative genomics in acid mine drainage biofilm communities reveals metabolic and structural differentiation of co-occurring archaea.** *BMC Genomics* (2013) **14**(1):485.
50. Jones DS, Albrecht HL, Dawson KS, Schaperdoth I, Freeman KH, Pi Y, Pearson A, Macalady JL: **Community genomic analysis of an extremely acidophilic sulfur-oxidizing biofilm.** *Isme J* (2012) **6**(1):158-170.
51. Whitman WB, Coleman DC, Wiebe WJ: **Prokaryotes: The unseen majority.** *Proceedings of the National Academy of Sciences* (1998) **95**(12):6578-6583.
52. Hinrichs K-U, Inagaki F: **Downsizing the deep biosphere.** *Science* (2012) **338**(6104):204-205.
53. Orsi WD, Edgcomb VP, Christman GD, Biddle JF: **Gene expression in the deep biosphere.** *Nature* (2013) **499**(7457):205-208.
- \*\* A well executed study looking at gene expression in the deep biosphere
54. Edwards KJ, Becker K, Colwell F: **The deep, dark energy biosphere: Intraterrestrial life on earth.** *Annual Review of Earth and Planetary Sciences* (2012) **40**(551-568).
55. Orcutt BN, LaRowe DE, Biddle JF, Colwell FS, Glazer BT, Reese BK, Kirkpatrick JB, Lapham LL, Mills HJ, Sylvan JB, Wankel SD *et al*: **Microbial activity in the marine deep biosphere: Progress and prospects.** *Front Microbiol* (2013) **4**(189).
56. McMahon S, Parnell J: **Weighing the deep continental biosphere.** *FEMS Microbiol Ecol* (2014) **87**(1):113-120.
- \* This work shows evidence of microbial activity in the deep biosphere
57. D'Hondt S, Jørgensen BB, Miller DJ, Batzke A, Blake R, Cragg BA, Cypionka H, Dickens GR, Ferdelman T, Hinrichs K-U: **Distributions of microbial activities in deep seafloor sediments.** *Science* (2004) **306**(5705):2216-2221.
58. Schrenk MO, Huber JA, Edwards KJ: **Microbial provinces in the seafloor.** *Annual Review of Marine Science* (2010) **2**(279-304).

59. Morono Y, Terada T, Nishizawa M, Ito M, Hillion F, Takahata N, Sano Y, Inagaki F: **Carbon and nitrogen assimilation in deep seafloor microbial cells.** *Proceedings of the National Academy of Sciences* (2011) **108**(45):18295-18300.
60. Orsi W, Biddle JF, Edgcomb V: **Deep sequencing of seafloor eukaryotic rRNA reveals active fungi across marine subsurface provinces.** *PLoS One* (2013) **8**(2):e56335.
61. Lau MCY, Cameron C, Magnabosco C, Brown CT, Schilkey F, Grim S, Hendrickson S, Pullin M, Sherwood Lollar B, van Heerden E, Kieft TL *et al*: **Phylogeny and phylogeography of functional genes shared among seven terrestrial subsurface metagenomes reveal n-cycling and microbial evolutionary relationships.** *Front Microbiol* (2014) **5**(531).
- \*\* A very comprehensive study assessing functional gene diversity in subterranean habitats
62. Lever MA: **Functional gene surveys from ocean drilling expeditions – a review and perspective.** *FEMS Microbiol Ecol* (2013) **84**(1):1-23.