

The Pennsylvania State University

The Graduate School

Department of Geosciences

**THE DIVERSITY AND COPING MECHANISMS OF LIFE INHABITING THE  
HYPERHALINE DEAD SEA**

A Dissertation in

Geosciences

by

Matthew E. Rhodes

© 2011 Matthew E. Rhodes

Submitted in Partial Fulfillment  
of the Requirements  
for the Degree of

Doctor of Philosophy

December 2011

The dissertation of Matthew E. Rhodes was reviewed and approved\* by the following:

Christopher H. House  
Associate Professor of the Geosciences  
Dissertation Advisor  
Co-Chair of Committee

Jennifer L. Macalady  
Associate Professor of the Geosciences  
Co-Chair of Committee

Mark E. Patzkowsky  
Associate Professor of the Geosciences

Ming Tien  
Professor of Biochemistry

Chris Marone  
Professor of Geosciences  
Associate Head for Graduate Programs and Research in Geosciences

\*Signatures are on file in the Graduate School

## ABSTRACT

Salinity has been shown to be a highly important determinant in microbial community composition. Consequently the mechanisms and genes that allow microorganisms to cope with salinity are fundamental for understanding microbial biogeography and evolution. At a salinity approaching 350‰, or approximately ten times that of the Earth's oceans, the Dead Sea represents one of the most saline naturally occurring bodies of water on our planet. When combined with a slightly acidic pH, near toxic magnesium levels, and the dominance of divalent cations, the Dead Sea becomes a truly inhospitable ecosystem. Nevertheless, the Dead Sea is home to a select few halophilic microbes and with a roughly 33% decrease in the salinity of its surface waters, the Dead Sea teems with microbial life. Here we employ "next generation" metagenomic and bioinformatic techniques to explore the limits and evolution of hyperhalophilic life inhabiting the Dead Sea and other hypersaline bodies.

To that end we extracted and/or obtained DNA from the March 2007 Dead Sea, four artificial Dead Sea blooms, and cryopreserved samples from both June and September 1992. We amplified and sequenced portions of the 16S rRNA gene for all samples and sequenced metagenomes for both the March 2007 and September 1992 samples. The amplicons revealed a significant population shift between the 1992 samples. Most striking was the virtual disappearance of the major bacterial lineages that were present in the June sample. The amplicons also revealed major differences between all bloom samples and the residual 2007 population. This indicates that the

archaeal taxa capable of surviving under the most extreme conditions are not the ones that flourish under relatively mild hypersaline conditions.

The shift in populations between the bloom and non-bloom environments was also observable in the translated amino acid profiles of both environments. Many hyperhalophiles balance the osmotic gradient present in their native environments with multi-molar quantities of KCl necessitating radical protein alterations. These alterations were more prevalent in the 2007 metagenome. We analyzed the metagenomes of a number of hypersaline environments, including both Dead Sea metagenomes and found that the degree of protein alteration plots linearly with salinity, suggesting its use as a salinity proxy. We also utilized the protein alterations inherent in hyperhalophilic proteins in combination with the amplicon data sets to identify both lateral gene transfer events involving hyperhalophilic organisms and a number of bacterial lineages with putative hyperhalophilic members.

From there we chose to delve deeper into lateral gene transfer amongst the Halobacteria. Halobacteria are notorious for lateral gene transfer and lateral gene transfer has played a significant role in their evolution. We analyzed over 1,000 genomes for instances of lateral gene transfer and discovered that the Halobacteria in contrast to other microbial lineages often participate in lateral gene transfer with non-halophiles. This paints the picture of the Halobacteria as the consummate opportunists, utilizing DNA from all sources they encounter.

## TABLE OF CONTENTS

LIST OF FIGURES .....	viii
LIST OF TABLES .....	xi
ACKNOWLEDGEMENTS .....	xiii
Chapter 1 Introduction .....	1
The Dead Sea .....	3
The Dead Sea in danger .....	4
Genomic attributes of hyperhalophiles .....	9
Horizontal gene transfer .....	10
Extremophile horizontal gene transfer .....	11
Authorship statement .....	12
References .....	13
Chapter 2 The rise and fall of Dead Sea microbial blooms .....	16
Abstract .....	16
Introduction .....	17
Materials and Methods .....	21
Sample collection and DNA extraction .....	21
Amplicon preparation .....	22
Sequencing and sequence quality control .....	23
Sequence analysis .....	24
Metagenomic analysis .....	25
Results .....	25
Rarefaction curves (sample richness) .....	26
Simpson index (sample diversity) .....	29
Taxonomic assignment .....	30
2007 metagenomic results .....	33
Community sources .....	34
Discussion .....	36
Acknowledgements .....	43
References .....	43
Chapter 3 Amino acid signatures of salinity on an environmental scale with a focus on the Dead Sea .....	47
Summary .....	47
Introduction .....	48
Results .....	51

Cluster analysis .....	52
Redundancy analysis .....	56
Environmental amino acid profiles versus salinity ... ..	58
Archaea in the modern Dead Sea .....	59
Screening of bacterial taxa .....	62
Discussion .....	69
Experimental procedures .....	72
DNA extraction .....	72
Sequencing and sequence analysis .....	73
Bioinformatic analysis .....	73
Acknowledgements .....	74
References .....	75
Chapter 4 Differences in lateral gene transfer in hypersaline versus thermal environments .....	79
Abstract .....	79
Background .....	79
Results .....	80
Conclusions .....	80
Background .....	81
Halophiles and thermophiles .....	82
Identifying lateral gene transfer events .....	86
Results .....	86
Assignment to Clusters of Orthologous Groups of proteins (COGs) .....	90
Homologue taxonomy .....	92
Database bias .....	92
Environmental halophile examples .....	98
Discussion .....	99
Conclusions .....	102
Methods .....	104
DNA extraction and fosmid preparation .....	104
Fosmid analysis .....	104
Genome analysis .....	105
Phylogenetic tree building .....	106
Authors' Contributions .....	107
Acknowledgements .....	107
References .....	107
Appendix A Supplementary information for Chapter 2 .....	116
Supporting information A-1 .....	116
Appendix B Supplementary information for Chapter 3 .....	117

Supporting information B-1 .....	117
Supporting information B-2 .....	119
Appendix C Supplementary information for Chapter 4 .....	125
Supporting information C-1 .....	125
Supporting information C-2 .....	126
Supporting information C-3 .....	127

## LIST OF FIGURES

- Figure 1-1.** (Source: NASA Archives) Satellite image of the Sinai Peninsula. The Dead Sea is located in upper right corner of the image. It is a tectonic basin formed in the basin between the Sinai-Israel Plate and the Arabian-Jordanian Plate. ....2
- Figure 1-2.** Photograph of the North Basin of the Dead Sea (07/26/2007). Each terrace is roughly a meter in height and represents the decline in water level from a single year. ....5
- Figure 1-3.** Photograph depicting the installation of Dead Sea research pools located at the Dead Sea Works facility, Sdom, Israel (07/22/2007). Each pool contains approximately 1 m<sup>3</sup> of water of various mixtures of Dead Sea and Red Sea waters. ..7
- Figure 1-4.** Photograph depicting a close up of one of the Dead Sea research pools located at the Dead Sea Works facility, Sdom, Israel (07/22/2007). This pool, Pool 10, contains approximately 1 m<sup>3</sup> of water of which 80% is from the Dead Sea (salinity ≈350‰) and 20% is from the Red Sea (salinity ≈41‰). The pool was fertilized with 1 μM phosphate to promote microbial growth. The red color of the pool is due to the carotenoids present in the membranes of member of the Haloarchaea. ....8
- Figure 2-1.** Rarefaction curves for all 7 amplicon datasets at (A) the 3% difference level, (B) the 5% difference level, and (C) the 10% difference level. ....28
- Figure 2-2.** Bar charts depicting the percentages of the major genera comprising the microbial populations of (A) the Dead Sea samples and (B) the mesocosm pool samples. Genera are organized according to their prevalence in the March 2007 inter-bloom sample. ....31
- Figure 2-3.** Flow chart of the seeding potential of the natural Dead Sea and mesocosm pool samples through time. Seeding potentials progressing forward in time are shown larger and in bold font. Seeding potentials progressing backward in time are shown smaller and italicized. ....37
- Figure 3-1.** Two-dimensional ‘heat plot’ showing the hierarchical clustering of environments (*y*-axis) and amino acids (*x*-axis). The two clusterings are based on comparisons of the amino acid content of each metagenome. The scales adjacent to the dendrograms give the Euclidean distances. Also shown along the right side, for reference, are the salinities and the lysine and aspartic acid contents of the environments. ....53
- Figure 3-2.** Redundancy analysis of the amino acid content of the environments with salinity and GC content encoded as the dependent variables. The arrows show the

directions of increasing GC content and salinity respectively. The bottom and left-hand axes provide scales for the loading of the amino acids onto the restricted axes and the top and right-hand axes provide scales for the loading of the environmental variables onto the restricted axes. ....57

**Figure 3-3.** Correlation between the ratio of acidic to basic amino acids encoded in an environment and the salinity of that environment. The salinity values range from marine ( $\approx 28$  g/l) to the extreme of hypersalinity ( $\approx 370$  g/l). ....60

**Figure 3-4.** Plot of the observed ratio of 16s rRNA amplicon reads to metagenomic reads versus the calculated salinity adaptation quotient (SAQ) for bacterial taxonomic groups found in the 2007 Dead Sea datasets. Taxonomic groups distributed along the x-axis are those found only in the metagenomic dataset indicating that they are represented by reads of protein-encoding genes that have been transferred into indigenous Dead Sea taxa. Taxonomic groups plotting vertically along an SAQ of about 1, indicating a lack of adaptation to salinity, are interpreted to be reads from either taxa foreign to the Dead Sea or indigenous salt-out taxa. Taxonomic groups represented by 16s rRNA amplicons and showing an elevated SAQ are interpreted to be microorganisms indigenous in the modern Dead Sea. ....65

**Figure 3-5.** Trees depicting the phylogeny of an AAA Family ATPase and an Adenine-Specific DNA Methylase with relevant bootstrap values included. Bacterial species (*Alcanivorax borkumensis*, “*Anaerocellum thermophilum*” (*Caldicellulosiruptor bescii*), *Clostridium botulinum*, “*Desulfococcus oleovorans*”, “*Methylacidiphilum infernorum*”, *Methylobacterium populi*, *Methylococcus capsulatus*, *Moorella thermoacetica*, *Nitrosococcus oceani*, *Pelobacter propionicus*, *Pelodictyon phaeoclathratiforme* (*Chlorobium clathratiforme*), *Pelotomaculum thermopropionicum*, *Photorhabdus luminescens*, *Thermotoga petrophila*, *Thermus thermophilus*, and *Verminephrobacter eiseniae*) are given in blue, haloarchaeal species (*Haloarcula marismortui*, *Halorubrum lacusprofundi*, and *Natrialba magadii*) in orange, other archaeal species (*Archaeoglobus fulgidus*, *Candidatus Korarchaeum cryptofilum*, *Methanocaldococcus vulcanius*, *Metahnopyrus kandleri*, *Methanospirillum hungatei*, *Pyrococcus abyssi*, *Pyrococcus furiosus*, *Pyrococcus horikoshii*, *Staphylothermus marinus*, *Sulfolobus islandicus* L.S.2.15, *Sulfolobus islandicus* Y.G.57.14, and *Thermococcus gammatolerans*) in red, and a representative metagenomic read is provided in green. Along the bottom is the region encoding both genes in the thermophilic bacterium *Thermotoga petrophila* and the halophilic archaeon *Halorubrum lacusprofundi*. Orthologous genes are connected by arrows. ....67-68

**Figure 4-1.** Depiction of 15 randomly selected trees of LGT genes for the Halobacteria. The leftmost column depicts genes from the upper third of BLAST scores, middle column from the middle third, and rightmost column from the bottom third. Target gene is depicted in red and named below the tree, other genes from the same class

are shown in orange, genes from the same phylum but not the same class are shown in yellow, genes from the same domain but not the same phylum are shown in green, and genes from a different domain are shown in blue. Circles next to a node indicate a bootstrap value of greater than 50% and stars indicate bootstrap values of greater than 75%. .....88

**Figure 4-2.** Depiction of 15 randomly selected trees of LGT genes for the Thermoprotei. ....89

**Figure 4-3.** Chart depicting the percentage of genes assigned to various COG categories for both regular genes and genes identified as LGT events into the (a) Halobacteria and (b) Thermoprotei. The COG categories shown demonstrated a statistically significant disparity between LGT genes and non-LGT genes beyond a 95% confidence interval. The categories depicted with an \* fell just below the 95% threshold. ....91

**Figure 4-4.** Pie charts depicting the proportion of inter-class LGT events from halophiles into the Halobacteria and from thermophiles into the Thermoprotei for all genes (a and d), only hits with bit scores > 500 (b and e), and only instances where multiple genes were transferred (c and f). ....94

## LIST OF TABLES

<b>TABLE 2-1.</b> Salinity, sample size, subsampled Inverse Simpson Index, and Inverse Simpson Index for amplicon data sets. Ninety-five percent confidence intervals are provided for both the Subsampled Inverse Simpson Index and the Inverse Simpson Index. ....	27
<b>Table 2-2.</b> Percentages of the microbial populations of the “sink” environments that can be constructed by OTUs occurring in the “source” environments. This analysis was performed at the 0% or “unique” level. ....	35
<b>Table 4-1.</b> Top five BLAST hits to genes identified as LGT events involving the Halobacteria, Thermoprotei, and a collection of Dead Sea fosmids. ....	95
<b>Table 4-2.</b> Percent differences for various indicators of halophilicity between LGT genes and non-LGT genes for both the Halobacteria and Thermoprotei. Values for which the LGT genes are less than the non-LGT genes are shaded in green. Values for which the LGT genes are greater than the non-LGT genes are shaded in red. All values that exhibited greater than 95% confidence are shown in bold font. The indicators are: GC - GC content of the gene pools. CG Di’s, GA/TC Di’s, AC/GT Di’s– Difference in preference for ‘CG’, ‘GA/TC’, and ‘AC/GT’ dinucleotides given nucleotide frequencies. Arginine (CGA and CGG), Cysteine (UGU), Leucine (CUC), Threonine (ACG), and Valine (GUC) – Difference in preference for respective codons given nucleotide abundances. AA Bias – Ratio of aspartic acid and glutamic acid to arginine, lysine, and histidine. ....	97
<b>Table A-1.</b> Table providing the taxonomic assignment of 16S rRNA gene fragments identified in a metagenome from the surface waters of the 2007 Dead Sea. ....	116
<b>Table B-1.</b> Amino acid and GC content of microorganisms and the Dead Sea metagenomes. The taxa name, acidic to basic amino acid ratio, percent encoded glutamic acid, aspartic acid, lysine, histidine, and arginine, and percent GC for the Dead Sea metagenomes (black) and seven sequenced halophiles (maroon; archaeal & orange; bacterial). <i>E. coli</i> and <i>T. volcanium</i> were added as reference non-halophilic microorganims (blue). ....	118
<b>Table B-2.</b> Foreign taxa and salt-out halophilic taxonomic groups. ....	122
<b>Table B-3.</b> Bacterial taxonomic groups represented through lateral gene transfer events into Dead Sea salt-in halophiles. ....	123
<b>Table B-4.</b> Salt-in bacterial taxonomic groups displaying evidence of being indigenous to the Dead Sea. ....	124

**Table C-1.** Table providing the proportion of LGT events originating from a donor inhabiting a similar environment for a spectrum of BLAST cutoff percentages. ....126

**Table C-2.** Table showing the 22 LGT events identified within environmental fosmid clones of the Dead Sea. The DNA sequences are too large to be reproduced here. They can be accessed as part of the supplemental information in association with the published manuscript *BMC Evolutionary Biology* 2011, **11**:199doi:10.1186/1471-2148-11-199. ....127

## ACKNOWLEDGEMENTS

First and foremost I thank my advisor, Christopher House. His contributions to this work are innumerable. The many hours spent in discussion in his office and in the astrobiology lounge (better known as the “Habitable Zone”) have not only contributed greatly to this dissertation but were also invaluable in my development as a scientist. Dr. House’s exceptional accessibility and encouragement of new ideas both fostered creativity and allowed me to forge my own path. In short Dr. House truly defines an advisor who considers his students to be his number one priority.

Christopher House, Jennifer Macalady, Mark Patzkowsky, and Ming Tien served on my doctoral committee and have all aided in the editing of this dissertation. Aharon Oren of the Hebrew University of Jerusalem, John Spear of the Colorado School of Mines, and Sorel Fitz-Gibbon of UCLA have collaborated with me on the work included in this manuscript.

A number of graduate school classes laid the foundation for my doctoral work. Two in particular are Mark Patzkowsky’s class on numerical methods in ecology which provided the methodology for much of Chapter 3 and Jennifer Macalady’s class on geobiology which provided the background knowledge for Chapter 2. Other influential classes were taught by Lee Kump, Katherine Freeman, and Mike Arthur.

Aharon Oren and Oded Béjà aided me in my sample collection and preparation in Israel. Lynn Tomsho, Jennifer Biddle, and Vanya Miteva aided me in sample preparation here at Penn State. Tom Canich, Damian Futrick, and Sorel Fitz-Gibbon troubleshot my

plethora of computer related issues and finally, Kat Dawson and Dan Jones were always available to bounce ideas off of.

This work was partly supported by the National Aeronautics and Space Administration (NASA) Astrobiology Institute (NAI) under NASA-Ames Cooperative Agreement NNA09DA76A and in particular through a NAI DDF award and by the Agriculture and Food Research Initiative Competitive Grants Program Grant no. 2010-65110-20488 from the USDA National Institute of Food and Agriculture. Additional support was provided by the Pennsylvania State University via an Anne c. Wilson Research Award, the Tait Scholarship in Microbial Biogeochemistry, and the Biogeochemical Research Initiative for Education (BRIE) Fellowship.

## Chapter 1

### Introduction

Microbiological research on the athalassic, hypersaline, terminal lake, the Dead Sea (Figure 1), has a rich tradition dating back to the ground breaking work of the late Benjamin Elazari Volcani in the 1930s and continued through the present day by the ever prolific Aharon Oren. In contrast comparative genomics and metagenomics are young and burgeoning scientific fields. This dissertation is an effort to apply the latest genomic and metagenomic technologies to the microbial populations of the Dead Sea. In the course of the metagenomic work on the Dead Sea a number of general parameters pertaining to the evolution of halophilic life were deduced as well.

More specifically, this dissertation utilizes both “next generation” DNA sequencing technology and the databases of fully sequenced microbial genomes to investigate the community structure and evolutionary processes of hyperhalophilic ecosystems. Field studies primarily focused on the aforementioned Dead Sea and a variety of artificial Dead Sea analogue pools. The intentions of this work were [1] to determine the community composition of the Dead Sea under both bloom and non-bloom conditions (Chapter 2), [2] to investigate, on an environmental scale, the degree of amino acid alteration necessary to maintain proper protein function at various salinities (Chapter 3), and [3] to begin to elucidate the extent that and mechanisms by which halophiles participate in lateral gene transfer (Chapter 4).



**Figure 1-1.** (Source: NASA Archives) Satellite image of the Sinai Peninsula. The Dead Sea is located in upper right corner of the image. It is a tectonic basin formed in the basin between the Sinai-Israel Plate and the Arabian-Jordanian Plate.

## The Dead Sea

The Dead Sea is located at the northern edge of the Great Rift Valley bordering Israel, the West Bank, and Jordan. Its current elevation of -343 meters and falling make it the lowest point on the surface of the planet. At a salinity of approximately 350 g/L, the current surface waters of the Dead Sea present a formidable obstacle to any potential colonizers. The dominance of divalent cations and near toxic magnesium levels (currently about 2.0 M  $Mg^{2+}$ ) (Oren, 1999a), test the limits of even the most hearty of halophiles to survive. Current cell counts are at most a paltry  $5 \times 10^5$ /ml (Oren *et al.*, 1990).

Not surprisingly, for the vast majority of recorded history, scientists presumed the Dead Sea to be completely devoid of life. It was the pioneering work of Benjamin Elazari Volcani in the 1930s that first demonstrated the existence of microorganisms in the waters of the Dead Sea (Wilkansky, 1936). Over the past 70 years, our knowledge of the Dead Sea as a habitat for halophilic algae, Archaea, and Bacteria has increased exponentially. The first quantitative studies of the microbial lake communities were performed by Kaplan and Friedmann in the late 1960's (Kaplan & Friedman, 1970). However, it was not until 1980 that systematic monitoring of Dead Sea biota began. Since then work on isolates and lipids have provided a valuable foundation for understanding Dead Sea ecosystem.

It is thought that the algae, *Dunaliella* sp., is the only primary producer in the Dead Sea. However, a dilution of 10-20% in the surface waters of the lake is necessary

to stimulate a bloom of *Dunaliella* (Oren, 2002b). At all other times *Dunaliella* appears absent from the Dead Sea water column (Oren & Shilo, 1982). The most recent *Dunaliella* blooms occurred after heavy winter rains in the years 1980 and 1992. The blooms were limited only by the availability of inorganic phosphate and began nearly simultaneously throughout much of the shallow portions of the Dead Sea. This would indicate that *Dunaliella* sp. cysts preserved in the shallow sediments were the source of the bloom (Oren & Ben-Yosef, 1997).

Coinciding with these algal blooms were massive growths of red halophilic Archaea and other prokaryotic organisms with cell densities upwards of  $3.5 \times 10^7$  cells  $\text{ml}^{-1}$ , sufficiently dense to color the surface waters of the Dead Sea a deep red (Oren, 1983a). The energy sources and inoculums of these Halobacterial blooms are difficult to determine. At least some of these microorganisms likely grew heterotrophically using glycerol, which is produced by *Dunaliella*. With the disappearance of *Dunaliella* as evaporation increased the salinity of the lake, the prokaryotic communities also declined significantly. Although a number of microbial species have been isolated and cultured from the Dead Sea, until recently little was known about the makeup of these microbial communities including the identity of the dominant species both during bloom and non-bloom periods.

### **The Dead Sea in danger**

Due to changing weather patterns and increased diversion of its source waters, the volume of the Dead Sea has been dropping for the past century (Figure 2). The



**Figure 1-2.** Photograph of the North Basin of the Dead Sea (07/26/2007). Each terrace is roughly a meter in height and represents the decline in water level from a single year.

repercussions of this can be observed in a number of ways. When Volcani originally identified life in the Dead Sea, the salinity of the surface waters was roughly  $270 \text{ g l}^{-1}$ . Since then the salinity of the lake has increased by approximately 80%. Secondly, over the past 40 years the surface level of the Dead Sea has dropped over 20 meters. Combined, these two factors have resulted in both the massive precipitation of halite, further increasing the disparity between monovalent and divalent cations, and the complete overturn of the lake in February 1979 (Oren, 1999a). Up until February 1979, the Dead Sea was meromictic with a permanent pycnocline at a depth of 40 meters. Below the pycnocline was an anoxic bottom layer (Neev & Emery, 1967). However, after the 1979 overturn, the pycnocline and anoxic bottom layer ceased to exist (Oren, 1983a).

With the very existence of the Dead Sea in jeopardy, the "Peace Conduit," a water channel linking the Gulf of Aqaba (Red Sea) and the Dead Sea, has been proposed as a potential solution. Water from the Red Sea, with a salinity of 40 g/L, would be channeled toward the Dead Sea and the 416 meter difference in elevation would be utilized to desalinate Red Sea water via reverse osmosis. The Red Sea concentrate would then be transferred to the Dead Sea (Oren *et al.*, 2004). The environmental impact of this impending dilution upon the Dead Sea is unknown. Potentially, the Dead Sea could evolve into a permanent bloom state. To model the potential effects of the Red Sea effluent on the Dead Sea, Israeli scientists have created a series of  $1 \text{ m}^3$  artificial pools filled with various mixtures of Red Sea and Dead Sea waters (Figures 3 and 4).



**Figure 1-3.** Photograph depicting the installation of Dead Sea research pools located at the Dead Sea Works facility, Sdom, Israel (07/22/2007). Each pool contains approximately  $1 \text{ m}^3$  of water of various mixtures of Dead Sea and Red Sea waters.



**Figure 1-4.** Photograph depicting a close up of one of the Dead Sea research pools located at the Dead Sea Works facility, Sdom, Israel (07/22/2007). This pool, Pool 10, contains approximately 1 m<sup>3</sup> of water of which 80% is from the Dead Sea (salinity ≈350‰) and 20% is from the Red Sea (salinity ≈41‰). The pool was fertilized with 1 μM phosphate to promote microbial growth. The red color of the pool is due to the carotenoids present in the membranes of member of the Haloarchaea.

### Genomic attributes of hyperhalophiles

The principle obstacle faced by all organisms that inhabit the Dead Sea and other hypersaline environments, is their high ionic strength. Microorganisms have developed two different strategies to deal with extreme salinity. All eukaryotic species, including *Dunaliella*, most halophilic Bacteria, and the halophilic methanogenic Archaea build up concentrations of organic solutes to balance the osmotic pressure. This “organic-in” method allows the internal mechanisms of the organisms to remain relatively intact but it requires a high energy cost to manufacture the organic molecules. In contrast the halophilic Archaea (*Halobacteriaceae*), and the anaerobic halophilic Bacteria (*Halanaerobiales*), accumulate salts, typically KCl, in their cytoplasm. This “salt-in” method is energetically more efficient but requires the adaptation of all intracellular enzymes to high salt concentrations (Lanyi, 1974; Oren, 2001; Reistad, 1970). The proteins of these microorganisms contain an excess of acidic amino acids and low amounts of hydrophobic amino acids. Generally speaking, the microorganisms that utilize salts to balance the osmotic pressure can survive at a greater range of salinities than those that utilize organic molecules (Oren *et al.*, 1995a). Microorganisms that employ both methods have been isolated from the Dead Sea.

A number of additional genomic characteristics common to halophilicity have recently been revealed by comparative genomic analyses of halophiles, the most ubiquitous being a preference for triply bonded nucleotides (Kennedy *et al.*, 2001). This propensity for GC nucleotides is likely for protection against thymine dimerization and is

present in all known Halobacteria except for *Haloquadratum walsbyi*. Other genomic propensities include a variety of dinucleotide and codon preferences, presumably for the maintenance of protein structure and for stability in protein folding (Paul *et al.*, 2008). At the same time however, Halobacteria genomes have highly varied compositions with GC poor regions interspersed amongst the more common GC rich sequences (Kennedy *et al.*, 2001). This variation along with the regular occurrence of insertion sequences and the diversity of metabolic functions employed by the Halobacteria suggest that the Halobacteria are frequent participants in horizontal gene transfer (Kennedy *et al.*, 2001).

### **Horizontal gene transfer**

As the name implies, horizontal gene transfer refers to the transfer of genetic material from one organism to another organism that is not one of its offspring. While horizontal gene transfer was first described in 1951 (Freeman, 1951), its potential to act as a driver of evolution was not appreciated until relatively recently. It was the sequencing of genomes such as that of *Thermotoga maritima* in the late 1990's that thrust horizontal gene transfer into the limelight (Nelson *et al.*, 1999). The initial estimates revealed that at least one quarter of *Thermotoga maritima*'s genome originated via horizontal gene transfer events.

Horizontal gene transfer itself is driven by four distinct mechanisms: [1] conjugation, or the transfer of genetic material via direct contact (Tatum & Lederberg, 1947), [2] transduction, or the viral mediated transfer of DNA (Zinder & Lederberg,

1952), [3] transformation, or the uptake and incorporation of naked DNA from an environment (Avery *et al.*, 1944), and [4] via virus-like genetic transfer agents (McDaniel *et al.*). However the relative importance of the various mechanisms is not known. While in some environments, such as microbial biofilms, conjugation may dominate, in others, such as the open ocean, transformation or transduction may be the primary force behind horizontal gene transfer.

Furthermore, the identification of horizontal gene transfer events is a difficult and complex task. Horizontal gene transfer events are usually identified either phylogenetically, by analyzing phylogenetic trees for genes that do not match the established phylogeny, or compositionally, by analyzing genomes for regions with anomalous proportions of nucleotides, codons, or other similar genomic properties. The two methods often detect complementary subsets of horizontal gene transfer events. While phylogenetic methods are more adept at detecting ancient horizontal gene transfer events, compositional techniques are better at identifying more recent transfers (Ragan *et al.*, 2006; Tamames & Moya, 2008). Further complicating the issue is the fact that the more closely related two species are the more likely they are to participate in horizontal gene transfer with each other. However, at the same time, the more closely related two participants in horizontal gene transfer are the more difficult for either method to identify the horizontal gene transfer event.

### **Extremophile horizontal gene transfer**

Halophiles and thermophiles represent two classes of organisms that are known to be adept at horizontal gene transfer (Aravind *et al.*, 1998; Mongodin *et al.*, 2005; Nelson *et al.*, 1999). This observation is most likely due to the extreme nature of their environments. Any acquired gene that provides even a slight advantage in coping with the challenges presented by thermal and hypersaline environments is highly prized and ripe for transfer.

However, with regard to the mechanisms of horizontal gene transfer, there are number of differences between thermal and hypersaline environments that must be taken into account. For instance, while DNA rapidly degrades when exposed to extreme temperatures, extreme salinities can protect environmental DNA for extended periods of time (Borin *et al.*, 2008; Marguet & Forterre, 1994; Tehei *et al.*, 2002). Furthermore, hypersaline environments are necessarily topographical minima and consequently natural collectors of cellular debris. Together these facts imply that halophiles should be exposed to greater quantities and a wider variety of environmental DNA than other extremophiles.

### **Authorship statement**

Chapters 2 and 4 have already been published. Chapter 3 is intended for submission as a journal article. While all chapters have multiple contributors, I (Matthew E. Rhodes) am the primary author for all portions of this dissertation. Additionally all analyses conducted and the majority of the respective interpretations are my own.

## References

**Aravind, L., Tatusov, R. L., Wolf, Y. I., Walker, D. R. & Koonin, E. V. (1998).** Evidence for massive gene exchange between archaeal and bacterial hyperthermophiles. *Trends in Genetics* **14**, 442-444.

**Avery, O. T., MacLeod, C. M. & McCarty, M. (1944).** Studies on the chemical nature of the substance inducing transformation of pneumococcal types. *The Journal of Experimental Medicine* **79**, 137-158.

**Borin, S., Crotti, E., Mapelli, F., Tamagnini, I., Corselli, C. & Daffonchio, D. (2008).** DNA is preserved and maintains transforming potential after contact with brines of the deep anoxic hypersaline lakes of the Eastern Mediterranean Sea. *Saline Systems* **4**, 10.

**Freeman, V. J. (1951).** Studies on the virulence of bacteriophage-infected strains of *Corynebacterium diphtheriae*. *Journal of Bacteriology* **61**, 675-688.

**Kaplan, I. R. & Friedman, A. (1970).** Biological productivity in the Dead Sea. Part I: Microorganisms in the water column. *Israel Journal of Chemistry* **8**, 513-528.

**Kennedy, S. P., Ng, W. V., Salzberg, S. L., 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 Research* **11**, 1641-1650.

**Lanyi, J. K. (1974).** Salt-dependent properties of proteins from extremely halophilic bacteria. *Bacteriology Reviews* **28**, 272-290.

**Marguet, E. & Forterre, P. (1994).** DNA stability at temperatures typical for hyperthermophiles. *Nucleic Acids Research* **22**, 1681-1686.

**McDaniel, L. D., Young, E., Delaney, J., Ruhnau, F., Ritchie, K. B. & Paul, J. H. (2010).** High frequency of horizontal gene transfer in the oceans. *Science* **330**, 50.

**Mongodin, E. F., Nelson, K. E., Daugherty, S. & other authors (2005).** The genome of *Salinibacter ruber*: Convergence and gene exchange among hyperhalophilic bacteria and archaea. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 18147-18152.

**Neev, D. & Emery, K. O. (1967).** The Dead Sea depositional processes and environments of evaporates: State of Israel Ministry of Development Geological Survey.

- Nelson, K. E., Clayton, R. A., Gill, S. R. & other authors (1999).** Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of *Thermotoga maritima*. *Nature* **399**, 323-329.
- Oren, A. & Shilo, M. (1982).** Population dynamics of *Dunaliella parva* in the Dead Sea. *Limnology and Oceanography* **27**, 201-211.
- Oren, A. (1983).** Population dynamics of halobacteria in the Dead Sea water column. *Limnology and Oceanography* **28**, 1094-1103.
- Oren, A., Ginzburg, M., Ginzburg, B. Z., Hochstein, L. I. & Volcani, B. E. (1990).** *Haloarcula marismortui* (Volcani) sp. nov., nom. rev., an extremely halophilic bacterium from the Dead Sea. *International Journal of Systematic Bacteriology* **40**, 209-210.
- Oren, A., Gurevich, P., Anati, D. A., Barkan, E. & Luz, B. (1995).** A bloom of *Dunaliella parva* in the Dead Sea in 1992: Biological and biogeochemical aspects. *Hydrobiologia* **197**, 173-185.
- Oren, A. & Ben-Yosef, N. (1997).** Development and spatial distribution of an algal bloom in the Dead Sea: A remote sensing study. *Aquatic Microbial Ecology* **13**, 219-223.
- Oren, A. (1999).** Microbiological studies in the Dead Sea: Future challenges toward the understanding of life at the limit of salt concentrations. *Hydrobiologia* **405**, 1-9.
- Oren, A. (2001).** The bioenergetic basis for the decrease in metabolic diversity at increasing salt concentrations: Implications for the functioning of salt lake ecosystems. *Hydrobiologia* **466**, 61-72.
- Oren, A. (2002).** Molecular ecology of extremely halophilic Archaea and Bacteria. *FEMS Microbiology Ecology* **39**, 1-7.
- Oren, A., Gavrieli, I., Gavrieli, J., Kohen, M., Lati, J. & Aharoni, M. (2004).** Biological effects of dilution of Dead Sea brine with seawater: Implications for the planning of the Red Sea-Dead Sea "Peace Conduit". *Journal of Marine Systems* **46**, 121-131.
- Paul, S., Bag, S., Das, S., Harvill, E. & Dutta, C. (2008).** Molecular signature of hypersaline adaptation: Insights from genome and proteome composition of halophilic prokaryotes. *Genome Biology* **9**, R70.
- Ragan, M. A., Harlow, T. J. & Beiko, R. G. (2006).** Do different surrogate methods detect lateral genetic transfer events of different relative ages? *Trends in Microbiology* **14**, 4-8.
- Reistad, R. (1970).** On the composition and nature of the bulk protein of extremely halophilic bacteria. *Archives of Microbiology* **71**, 353-360.
- Tamames, J. & Moya, A. (2008).** Estimating the extent of horizontal gene transfer in metagenomic sequences. *BMC Genomics* **9**, 136.

**Tatum, E. L. & Lederberg, J. (1947).** Gene recombination in the bacterium *Escherichia coli*. *Journal of Bacteriology* **53**, 673-684.

**Tehei, M., Franzetti, B., Maurel, M., Vergne, J., Hountondji, C. & Zaccai, G. (2002).** The search for traces of life: the protective effect of salt on biological macromolecules. *Extremophiles* **6**, 427-430.

**Wilkansky, B. (1936).** Life in the Dead Sea. *Nature* **138**, 467.

**Zinder, N. D. & Lederberg, J. (1952).** Genetic exchange in *Salmonella*. *Journal of Bacteriology* **64**, 679-699.

## Chapter 2

### The Rise and Fall of Dead Sea Microbial Blooms

Matthew E. Rhodes<sup>1,2\*</sup>, Aharon Oren<sup>3</sup>, and Christopher H. House<sup>1,2\*</sup>

<sup>1</sup>Penn State Astrobiology Research Center, Pennsylvania State University, University Park, PA 16802, USA

<sup>2</sup>Department of Geosciences, Pennsylvania State University, University Park, PA 16802, USA

<sup>3</sup>The Institute of Life Sciences, and the Moshe Shilo Minerva Center for Marine Biogeochemistry, The Hebrew University of Jerusalem, 91904 Jerusalem, Israel

\*To whom correspondence should be addressed:

Under Review by *Applied and Environmental Microbiology*

#### Abstract

Over the past century, a decrease in water input has caused the salinity of the surface waters of the Dead Sea to increase almost 30%, from approximately 270 g/liter to nearly 350 g/liter. This extreme rise in salinity has decimated the microbial population, with cell counts currently under  $10^5$  cells/ml. Only in years with exceptionally heavy rains are the surface waters of the Dead Sea diluted sufficiently to

allow for the establishment of a haloarchaeal bloom. The last such event began in the spring of 1992, peaked in June, and then declined. Since then the Dead Sea has deteriorated into its current near-lifeless state awaiting the opportunity to return to life again. Using “next-generation” sequencing technologies we have obtained 16S rRNA amplicon libraries from both the peak and declining stages of the 1992 bloom and from the modern inter-bloom Dead Sea. We have also sequenced amplicons from a number of simulated Dead Sea blooms produced in experimental mesocosm pools in which Dead Sea water was diluted with freshwater or with Red Sea water. We therefore were able to attain an in depth look at the evolution of a haloarchaeal bloom in the Dead Sea as well as gain insight into the future of the Dead Sea microbial population. Our results reveal significant population shifts between the early and late stages of the 1992 bloom. They also indicate that the residual inter-bloom population though largely distinct from the bloom population, retains a signature from the latter stages of the 1992 bloom.

## **Introduction**

In the 1930s when Benjamin Elazari-Volcani (Wilkansky) first identified life in the Dead Sea (Wilkansky, 1936), the surface waters of the Dead Sea were relatively hospitable environments. However, due to the increased diversion of its source waters and pumping of water into the evaporation ponds of the industries at the southern end of the lake for the production of potash, bromine and other chemicals, the volume of the Dead Sea has been decreasing for the past century (Gavrieli *et al.*, 1999). The repercussions of this can be observed in a number of ways. When Volcani first began his

work, the salinity of the surface water was roughly 270 g/liter. Since then the salinity has increased by approximately one third, to just under 350 g/liter. Secondly, over the past 40 years alone, the water level of the Dead Sea has dropped over 20 meters. Thus the Dead Sea, already the lowest exposed point on the surface of our planet at an elevation of 425 meters below sea level, continues to fall.

The combination of these two factors resulted in the complete overturn of the lake in February 1979 (Beyth, 1980) and in the massive precipitation of halite since the early 1980s. Up until February 1979, the Dead Sea was meromictic with a permanent pycnocline at a depth of 40 meters. Below the pycnocline was an anoxic bottom layer (Neev & Emery, 1967). However, after the 1979 overturn, the pycnocline and anoxic bottom layer ceased to exist (Beyth, 1980). Unlike the Dead Sea of the mid 19<sup>th</sup> century, the modern Dead Sea, at a salinity over 340 g/liter and dominated by divalent cations replete with almost 2 M Mg<sup>2+</sup>, is a truly inhospitable environment with little cellular activity and cell counts below 5x10<sup>5</sup> cells/ml (Oren *et al.*, 1990).

Only twice in the past 30 years have the surface waters of the Dead Sea been diluted to a sufficient extent to allow for the reestablishment of dense microbial communities. In the years 1980 and 1992, runoff from heavy winter rains reformed the pycnocline to depths of between 5 and 15 meters (Gavrieli *et al.*, 1999), and allowed for the establishment of a bloom of the halophilic alga *Dunaliella* sp. Blooms of *Dunaliella* sp. reached cell densities of more than 10<sup>4</sup> cells/ml and were primarily limited by the availability of inorganic phosphate (Oren, 1993b). Coinciding with these algal blooms were massive growths of red halophilic archaea and other prokaryotic organisms, with

cell counts reaching upwards of  $3.5 \times 10^7$  cells/ml (Oren, 1993a). Many of these microorganisms likely grew heterotrophically using glycerol, which is produced by *Dunaliella* sp. for use as an osmolyte.

During the 1980 bloom, the *Dunaliella* population persisted in the surface waters from approximately June until the end of the year. By the end of 1981 *Dunaliella* sp. individuals were virtually non-existent (Oren, 1983a). The halobacterial population as well declined considerably toward the end of 1980. However, the population eventually stabilized at approximately  $5 \times 10^6$  cells/ml for roughly another two years (Oren, 1983a). The 1992 bloom began earlier in the year. The surface *Dunaliella* bloom began in April, peaked in May, and declined rapidly from late May through early June (Oren, 1993a). The halobacterial bloom peaked in early June and persisted at more than  $10^7$  cells/ml well through September. In September of 1992 a bloom of *Dunaliella* sp. reappeared, this time at a depth of between six and ten meters (Oren, 1993a). The cause and dynamics of this secondary bloom are not well understood.

With the Dead Sea water level continuing to decline at a rate of approximately one meter per year and the Dead Sea ecology significantly altered from that of the early 1900s, plans exist for the construction of a canal connecting the Gulf of Aqaba to the Dead Sea (Gavrieli & Oren, 2004; Oren *et al.*, 2004). The canal would introduce copious quantities of marine (40 g/liter) water to the surface of the hypersaline Dead Sea. The effects of this potential dilution are to some degree uncertain. Consequently, for the past few decades, researches have been filling artificial one cubic meter pools with various mixtures of Gulf of Aqaba and Dead Sea waters. These pools inevitably form

*Dunaliella* and halobacterial blooms and allow scientists to investigate both the potential implications of the Dead Sea – Gulf of Aqaba canal as well as the nature of blooms inoculated with modern Dead Sea water.

As *Dunaliella* sp. appears to be wholly absent from the inter-bloom Dead Sea environments, one question that naturally arose is where the *Dunaliella* sp. inocula originated. LANDSAT remote sensing studies revealed that the *Dunaliella* blooms began simultaneously in much of the shallow portions of the Dead Sea, thereby indicating that the source of the bloom were *Dunaliella* sp. cysts preserved in the sediments (Oren & Ben-Yosef, 1997). The sources of the inocula for the haloarchaeal bloom are far less clear. At no recent point in time has the Dead Sea been completely devoid of haloarchaeal life. Thus it difficult to ascertain how much of the haloarchaeal blooms originate from residual inter-bloom populations and how much is introduced via the surrounding environs and sediments.

Similarly it is unclear to what extent the residual inter-bloom populations reflect the early stages and end stages of the blooms. Recent studies using modern metagenomic and amplicon based techniques have begun to take in depth looks at the make ups of both the bloom and residual populations (Bodaker *et al.*, 2010; Rhodes *et al.*, 2010). Both studies concluded that the modern Dead Sea is primarily inhabited by heretofore uncultured and unisolated representatives. Additionally, Bodaker *et al.* (Bodaker *et al.*, 2010) utilized a collection of short (~60 bp) amplicons from the V6 hypervariable region of the 16S rRNA gene to conclude that a number of the taxa represented in the early stages of the 1992 bloom were not present in the residual

population. They also suggested that the modern Dead Sea demonstrated both more species evenness and diversity than the 1992 bloom.

Here, in an effort to understand the evolution of a halobacterial bloom we have deep sequenced a suite of 16S rRNA amplicons from a variety of Dead Sea and simulated bloom environments. Our samples consist of material collected from the June 1992 Dead Sea, at the peak of the last haloarchaeal bloom, the September 1992 Dead Sea, a few months into the decline of the last haloarchaeal bloom, the March 2007 inter-bloom Dead Sea, and four distinct Dead Sea mesocosm pools. These amplicons span both the V6 and V7 hypervariable regions and are a minimum of 275 base pairs in length, allowing for more accurate phylogenetic assignment. We also analyzed a metagenomic data set from the March 2007 sample focusing solely on 16S rRNA sequences for perhaps the least biased view of the Dead Sea microbial community.

## **Materials and Methods**

### *Sample collection and DNA extraction*

The June 1992 and March 2007 samples were collected and processed according to the protocols described by Bodaker *et al.* (Bodaker *et al.*, 2010). The September 1992 sample was collected and processed according to the protocol described by Rhodes *et al.* (Rhodes *et al.*, 2010). The Dead Sea mesocosm pool samples were collected in July of 2007 from pools 1, 2, 9, and 10 at the research station at the Dead Sea Works facility, Sedom, Israel (17). Pools no. 1 and 2 were filled in April 2007 with 80% (v/v) Dead Sea water and 20% deionized water and 20% Red Sea water, respectively. Pool 9 was filled

in March 2003 with 70% Dead Sea water and 30% Red Sea water and was enriched with 5.5  $\mu\text{M}$   $\text{KH}_2\text{PO}_4$ , and pool 10 was set up in March 2004 with 80% Dead Sea water, 20% Red Sea water and 1  $\mu\text{M}$   $\text{KH}_2\text{PO}_4$ . The water volume in each pool was 900 liter. To compensate for evaporation, deionized water was added every few days to keep the water level constant. Microscopic enumeration of prokaryotes in these pools in July 2007 yielded estimates of  $17 \times 10^6$ ,  $7 \times 10^6$ ,  $34 \times 10^6$  and  $44 \times 10^6$  cells/ml respectively. All pools also had *Dunaliella* cells ( $13.3 \times 10^3$ ,  $2.1 \times 10^3$ ,  $0.6 \times 10^3$  and  $4.6 \times 10^3$  cells/ml). Pool 9 also contained large numbers of *Aphanothece*-like cyanobacteria. For each sample approximately 250 ml of water was pumped through a 0.7 micron glass fiber filter using a 60 ml hand syringe. The filters were immediately placed on ice. They were then stored at the Hebrew University of Jerusalem at  $-80^\circ\text{C}$  until shipment to Penn State on dry ice. The filters remained frozen at  $-80^\circ\text{C}$  at Penn State for approximately three years. In the fall of 2010 DNA was extracted from half of each filter using the protocol described in Macalady *et al.* (Macalady *et al.*, 2008).

#### *Amplicon preparation*

The V6-V9 hypervariable regions of the 16S rRNA gene were amplified using the primer pair 926F 5'-AACTYAAAKGAATTGRCGG-3' which we obtained by reverse complementing 907R (Muyzer *et al.*, 1996) and an adapted version of 1392R 5'-ACGGGCGGTGTGTRC-3' (Field *et al.*, 1988) with a degeneracy added at the 5' end. Over the past two years our group has initiated a diverse set of projects using these primers. Our unpublished results have shown good amplification of a broad number of prokaryotic groups from both the Bacteria and the Archaea. These primers were

independently developed for amplification of DNA from oil sands with promising results (Park *et al.*, 2011). For each sample, the forward primer was augmented with both the 454 Titanium 'A' adapter 5'-CCATCTCATCCCTGCGTGTCTCCGAC-3' and a unique ten base pair barcode. The reverse primers were augmented only with the 454 Titanium 'B' adapter 5'-CCTATCCCCTGTGTGCCTTGGCAGTC-3'. The approximately 500 base pair sequence was amplified and purified according to the procedure found in Rhodes *et al.* 2010 (Rhodes *et al.*, 2010) with the following modification: the samples from the mesocosm pools were cycled 29 times.

#### *Sequencing and sequence quality control*

The samples were then sequenced, along with an additional six unrelated samples, on a quarter plate of a 454 FLX Titanium sequencer (454 Life Sciences). Sequence quality control was performed using primarily the mothur platform (Schloss *et al.*, 2009). All sequences with greater than one error in the primer or barcode were removed with the following exception: a large number of sequences contained only two rather than three 'A' nucleotides at the start of the primer. This anomaly was ignored. Additionally all regions with ambiguous bases, extended homopolymer repeats (>8 base pairs), and low scoring regions (regions of 50 nucleotides where the average quality score was <35) were also removed. The remaining sequences and sequence fragments over 305 nucleotides were aligned using the Silva bacterial and archaeal alignments and preclustered using the mothur preclustering program (Pruesse *et al.*, 2007). The sequences were then chimera checked using the UCHIME algorithm and with the

sequences themselves as the database (Edgar *et al.*). All chimeric sequences were removed.

### *Sequence analysis*

Sequence analysis was performed using a combination of home written scripts and the mothur platform (Schloss *et al.*, 2009). The pairwise distances between all quality controlled sequences were calculated and the sequences were clustered using the average neighbor method. The clusterings were then used to create rarefaction curves and to calculate the Inverse Simpson Index. The Inverse Simpson calculations were repeated 1,000 times. To ensure that the Simpson diversity indices were not biased by sample size, 1,400 sequences were randomly subsampled from each sample and the Simpson Index was calculated for each subsample. This procedure was also repeated 1,000 times. The operational taxonomic units (OTUs) were then assigned to taxa at the 3% distance level using a combined Silva and RDP taxonomic database (Cole *et al.*, 2009; Pruesse *et al.*, 2007). All OTUs with greater than 10 sequences were checked manually using BLASTN (Altschul *et al.*, 1997) searches of representative sequences for each OTU to both the nr database and the combined Silva and RDP database. Taxonomic assignments to known genera for which sequence identity was less than 95% were disregarded and the sequences were classified as uncultured. The proportions of the major taxa for each sample were then tabulated. Finally the seeding potentials of each sample were calculated.

### *Metagenomic analysis*

A BLASTN search was performed on a metagenome sequenced from the March 2007 Dead Sea sample (Rhodes *et al.*, 2010) against a collection of full length representative 16S rRNA genes. All reads with hits with an e-value of  $10^{-10}$  or greater were extracted. The resulting 143 sequences were assigned to taxa using the RDP classifier (Cole *et al.*, 2009) and the assignments were checked as above.

## Results

DNA was extracted from surface water samples collected from the Dead Sea in June of 1992, September of 1992, and March of 2007. DNA was also extracted from samples collected from four mesocosm pools containing a variety of mixtures of water from the modern Dead Sea and the Gulf of Aqaba. The two 1992 samples and all four mesocosm pools samples were collected at various stages of haloarchaeal blooms, and all of these samples demonstrated the characteristic reddish coloring associated with the carotenoids of the Haloarchaea. The 2007 sample, in contrast, was collected during an inter-bloom period with cell counts roughly three to four orders of magnitude lower than the other samples.

For all seven samples an approximately 500 base pair portion of the 16S rRNA gene incorporating the V6-V9 hypervariable regions was amplified using distinct barcoded primers. The resulting amplification products were sequenced unidirectionally from the V6 end on one quarter plate of a 454 FLX Titanium sequencer. The sequencing run produced 281,931 raw sequences with an average read length of 416.5 base pairs of which the first 30 base pairs consisted of the barcode and forward primer. Therefore as

expected the vast majority of sequences did not successfully span the entire amplified region.

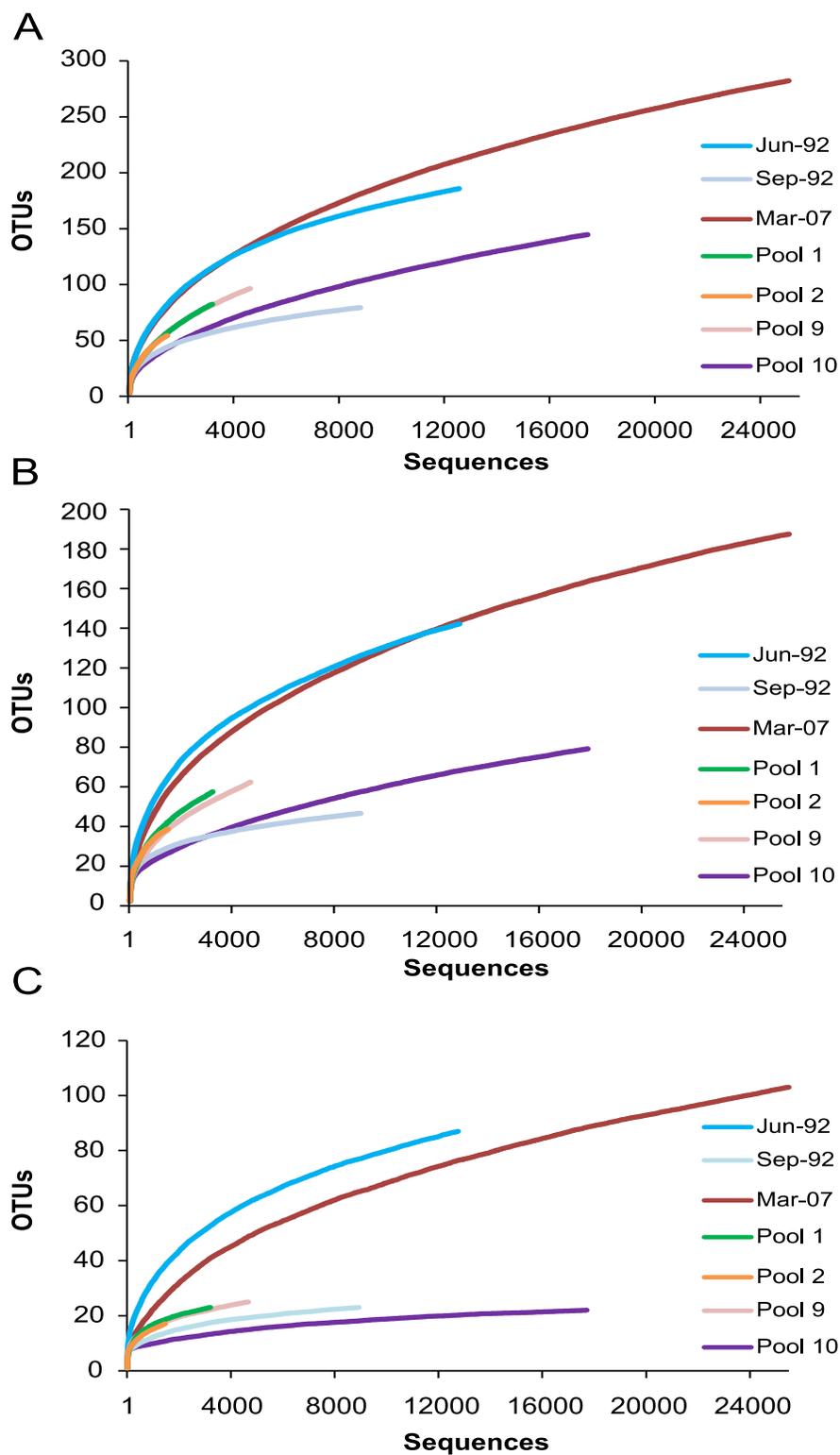
In an effort to balance read length and sample size, we restricted our analyses to the first 305 base pairs of the amplicon products. This ensured that the amplicons spanned the entirety of the V7 hypervariable region. Thus the amplicons included both the V6 and V7 hypervariable regions, the conserved region between them, and the majority of the conserved region between the V5 and V6 hypervariable regions. Quality checking, chimera checking, and preclustering left 1,765 unique sequences representing 74,096 total sequences (Table 1). Sample sizes spanned roughly an order of magnitude with the March 2007 sample containing the most sequences (25,586) and the Pool 2 sample containing the least (1,452).

#### *Rarefaction curves (sample richness)*

The remaining sequences were then clustered and a rarefaction analysis was performed at what we infer to be the strain (3% difference), species (5% difference), and genus (10% difference) levels (Figure 1). The 2007 Dead Sea displayed far greater richness at the strain level than either the September 1992 Dead Sea bloom sample or the mesocosm pool samples (Figure 1a). It displayed slightly greater richness than the June 1992 sample. The June 1992 bloom in turn displayed significantly greater strain level richness than the September 1992 sample or any of the mesocosm pools. The September 1992 sample displayed the least strain level richness of any sample. None of

**TABLE 2-1.** Salinity, sample size, subsampled Inverse Simpson Index, and Inverse Simpson Index for amplicon data sets. Ninety-five percent confidence intervals are provided for both the Subsampled Inverse Simpson Index and the Inverse Simpson Index.

Sample	Salinity	Number of Sequences	Subsampled Inverse Simpson			Inverse Simpson		
			0.03	0.05	0.1	0.03	0.05	0.1
Jun-92	249‰	12733	7.45 ±0.56	7.31 ±0.56	6.36 ±0.38	7.46 ±.20	7.33 ±.19	6.37 ±.13
Sep-92	288‰	8904	5.58 ±0.30	5.43 ±0.30	4.22 ±0.18	5.59 ±.13	5.45 ±.13	4.22 ±.08
Mar-07	347‰	25471	10.90 ±0.70	9.20 ±0.54	4.09 ±0.22	10.98 ±.17	9.24 ±.13	4.10 ±.05
Pool 1	287‰	3188	2.97 ±0.18	2.92 ±0.18	2.67 ±0.12	2.97 ±.15	2.92 ±.15	2.67 ±.11
Pool 2	279‰	1452	4.82 ±0.08	4.78 ±0.08	4.06 ±0.04	4.83 ±.41	4.79 ±.40	4.07 ±.25
Pool 9	225‰	4655	3.90 ±0.18	3.82 ±0.18	3.42 ±0.14	3.91 ±.12	3.82 ±.12	3.42 ±.09
Pool 10	261‰	17693	2.63 ±0.20	2.61 ±0.18	2.03 ±0.12	2.63 ±.06	2.62 ±.06	2.02 ±.04



**Figure 2-1.** Rarefaction curves for all 7 amplicon datasets at (A) the 3% difference level, (B) the 5% difference level, and (C) the 10% difference level.

the samples with the possible exception of the September 1992 bloom sample showed any indication of saturation.

The richness appears much the same at the species level (Figure 1b). The only difference being that the March 2007 Dead Sea sample and the June 1992 Dead Sea sample appeared to have comparable levels of richness. Finally, at the genus level, the June 1992 bloom sample actually slightly surpassed the 2007 sample (Figure 1c). Despite the over 12,000 sequences for the June 1992 samples and the over 25,000 sequences for the March 2007 sample, neither appeared to be sampled to saturation. In contrast, the September 1992 sample and all mesocosm bloom samples exhibited similar levels of richness and all appeared to max out at between 20 and 30 total genera.

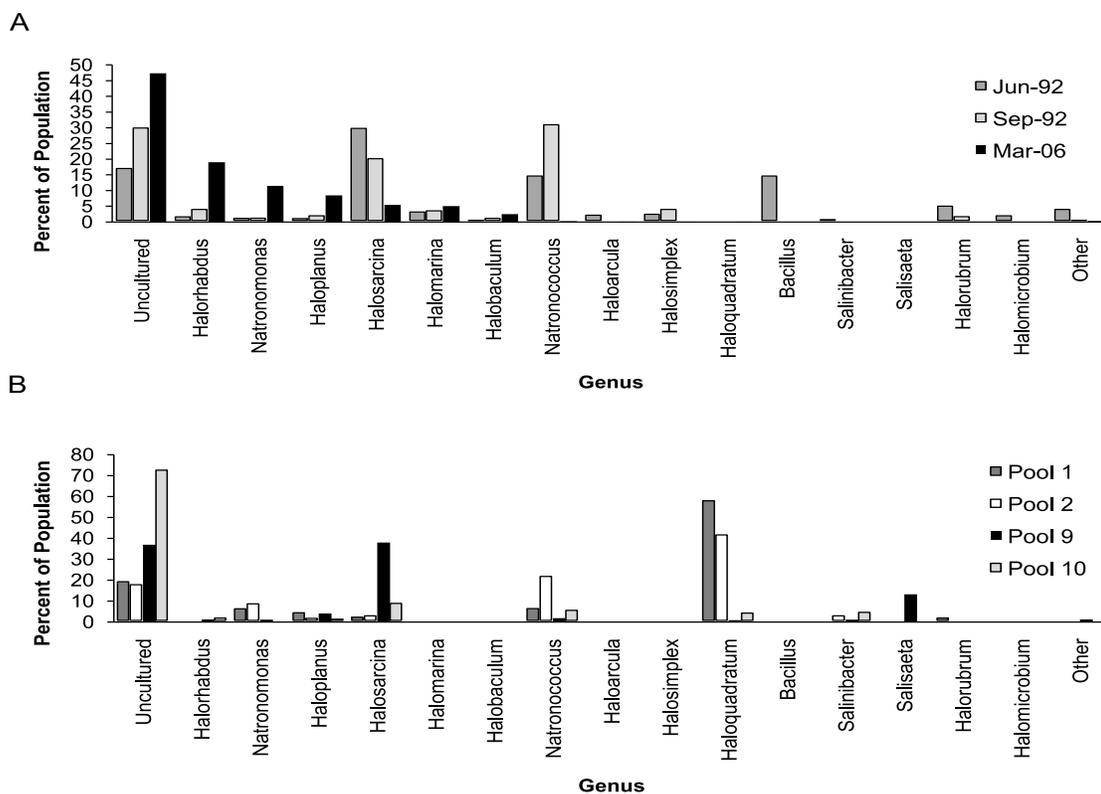
#### *Simpson index (sample diversity)*

While rarefaction curves are a good representation of sample richness comparatively little information regarding species evenness and thus overall species diversity can be inferred from the shape of the curve. Therefore, in order to analyze the overall species diversity we calculated the inverse of the Simpson Index for all samples (Table 1). The calculations were repeated 1,000 times and the mean values and 95% confidence intervals are provided in Figure 2. Once again, at the strain level, the March 2007 sample displayed far greater diversity than any other sample. The June 1992 sample displayed greater diversity than the September 1992 sample, and all Dead Sea samples displayed greater diversity than the mesocosm pools. These trends maintained themselves at the species level.

However, by the genus level the diversity of the March 2007 sample had dropped precipitously, to the point that the June 1992 sample actually displayed the most diversity, and the March 2007 sample and the September 1992 sample displayed comparable levels of diversity with the September 1992 sample displaying slightly greater diversity. The mesocosm pools all displayed less diversity than the Dead Sea samples. However, Pool 2 displayed nearly as much diversity as both the March 2007 and September 1992 samples. The Simpson Index can be skewed greatly by sample size, which as noted above spans over an order of magnitude. Therefore to verify our results we randomly subsampled 1,400 sequences from each sample and calculated the inverse Simpson Index for those 1,400 sequences alone. This exercise was repeated 1,000 times and the mean subsampled Simpson Indices and the respective 95% confidence intervals are also provided in Table 1. The results from the subsampling coincide well with the Simpson Index values calculated from the entire sample pools.

#### *Taxonomic assignment*

The OTUs were then assigned to taxa at the strain (3% difference) level and the proportions of sequences belonging to the major lineages were tabulated for each sample (Figure 2). All samples contained considerable portions of heretofore uncultured microorganisms, with the Pool 10 and March 2007 samples containing the highest percentages of uncultured haloarchaea. In particular a single sequence displaying 99% homology to an uncultured haloarchaeon clone from an evaporitic crust in Guerrero Negro, Baja California (Sahl *et al.*, 2008) comprised over 25% of the total data set and represented over 68% of the Pool 10 community.



**Figure 2-2.** Bar charts depicting the percentages of the major genera comprising the microbial populations of (A) the Dead Sea samples and (B) the mesocosm pool samples. Genera are organized according to their prevalence in the March 2007 inter-bloom sample.

Taxonomic assignment also revealed a number of population shifts between the samples. The identifiable portions of the haloarchaeal populations in both Dead Sea bloom samples were dominated by members of the genera *Natronococcus* and *Halosarcina* (Figure 2a). The identifiable portion of the March 2007 residual Dead Sea haloarchaeal population, however, was dominated by microorganisms belonging to the genera *Halorhabdus*, and *Natronomonas*. For the March 2007 Dead Sea, our taxonomic assignment largely confirms the results obtained by Bodaker *et al.* (Bodaker *et al.*, 2010) from fosmid and environmental clones. However, with respect to the early stages of the 1992 bloom, our amplicon data sets revealed different and far greater taxonomic diversity. Furthermore, it is our belief that the amplicons sequenced by us provide a far more accurate representation of the Dead Sea microbial communities than the three to four times shorter amplicons used in other Dead Sea investigations (Bodaker *et al.*, 2010; Rhodes *et al.*, 2010).

The bacterial Dead Sea populations displayed a more pronounced shift. Only in the early bloom, June 1992 sample, did the Bacteria compose a significant proportion of the overall microbial community. Microbes apparently related to the genus *Bacillus* comprised over 13% of the June 1992 sample and sequences affiliated with the genera *Burkholderia*, *Chryseobacterium*, and *Salinibacter* were identified as well. In total more than 18% of the June 1992 microbial community was bacterial in origin. In stark contrast, the latter Dead Sea bloom sample, September 1992, and the residual Dead Sea sample, March 2007, each had less than 0.5% of identifiable sequences originating from the Bacteria.

There were also significant population shifts amongst the mesocosm Dead Sea pools (Figure 2b). While the more saline and younger pools, Pools 1 and 2, were primarily dominated by *Haloquadratum*, the less saline and older pools, Pools 9 and 10, were primarily dominated by uncultured haloarchaea. Pool 9, the oldest and least saline of the pools, also contained significant quantities of *Halosarcina* and exhibited significant proportions of the bacterium *Salisaeta*. All other pools contained only trace amounts of identifiable Bacteria.

#### *2007 metagenomic results*

We had previously sequenced on a 454 FLX sequencer a metagenome from the March 2007 Dead Sea (Rhodes *et al.*, 2010). A BLAST comparison of this dataset to a library of full length 16S rRNA genes revealed a total of 142 distinct sequence fragments of the 16S rRNA gene. As these sequences were sequenced directly from extracted DNA, they potentially offer the least biased representation of the Dead Sea microbial community. Though there inevitably remain a number of potential biases, these sequences avoid the possibilities of both PCR bias and cloning bias, biases common to other 16S rRNA datasets.

Of the 142 16S rRNA sequences, 133 originated in the haloarchaea, 7 in the Bacteria, and 2 were undeterminable. Unfortunately the random locations of the metagenomic 16S fragments along the 16S rRNA gene made it comparatively difficult to accurately assign genera to the majority of the sequences (Table A-1). Nevertheless, to an extent the metagenomic sequences confirm the results of the March 2007 amplicon dataset. Both the genera *Halorhabdus* and *Natronomonas* are well represented amongst

the assignable sequences. Also present with multiple sequences are *Haloplanus*, *Halosarcina*, and *Halomarina*.

#### *Community sources*

The various natural and artificial Dead Sea environments are intimately linked with one another. The June 1992 Dead Sea evolved into the September 1992 sample over the course of three months. The September 1992 Dead Sea in turn deteriorated into the March 2007 sample over the course of fifteen years. Finally the mesocosm pools were constructed with inter-bloom Dead Sea water much akin to the March 2007 Dead Sea. At the same time, however, the various environments were exposed to microbial input from the surrounding environs. Thus our samples represent some combination of inherited and introduced biota. Our final analysis focused on the ability of one Dead Sea microbial community to seed another. We therefore determined the percentage of the “sink” microbial community that could be constructed from microbes occurring in the “source” microbial community (Table 2). Essentially this amounts to comparing the presence or absence of a group in the “source” community with the percent community composition in the “sink” community. The logic behind this analysis being that as long as an organism is present in an inoculum, it can proceed to establish a population.

As expected the best source for the September 1992 sample was the June 1992 sample. Similarly the best source for the March 2007 was the September 1992 sample. This suggests that even after 15 years the residual population in the Dead Sea is still partly a reflection of the latter stages of the 1992 bloom. However, despite

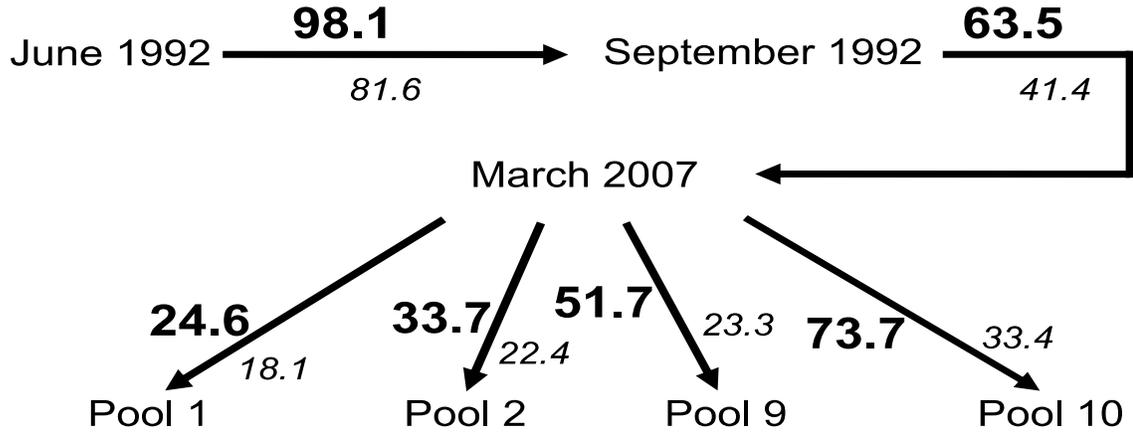
**Table 2-2.** Percentages of the microbial populations of the “sink” environments that can be constructed by OTUs occurring in the “source” environments. This analysis was performed at the 0% or “unique” level.

Sink	Source						
	Jun-92	Sep-92	Mar-07	Pool 10	Pool 9	Pool 2	Pool 1
Jun-92	100.0	81.6	26.5	68.4	65.9	66.0	61.7
Sep-92	98.1	100.0	41.4	91.6	86.7	87.3	80.1
Mar-07	58.6	63.5	100.0	33.4	23.3	22.4	18.1
Pool 10	82.5	86.9	73.7	100.0	84.7	94.4	92.1
Pool 9	77.4	90.4	51.7	94.2	100.0	75.6	74.4
Pool 2	46.3	70.4	33.7	93.0	44.2	100.0	89.2
Pool 1	50.0	45.0	24.6	89.5	51.0	92.7	100.0

the degree of sequence diversity observed in the March 2007 Dead Sea, the March 2007 sample is a poor source for the mesocosm pools. Nevertheless, even for the poorly seeded samples, the source-sink values do appear to adhere to a temporal directionality (Figure 3). The ability of a given sample to seed its succeeding environment was always better than the ability of the succeeding environment to seed its predecessor. For example, the March 2007 Dead Sea was a better source for the mesocosm pools than the pools were for the 2007 Dead Sea. Likewise the June 1992 sample was a better source for the September 1992 sample than the September 1992 sample was a source for the June 1992 sample and the September 1992 sample was a better source for the 2007 sample than the 2007 sample was a source for the September 1992 sample. To avoid inconsistencies due to the random nature of average neighbor clustering, this analysis was performed on groupings of unique sequences. However, similar results were obtained for clusters of OTUs at the 3% and 5% difference levels.

## **Discussion**

Our results indicate that at the strain level the residual Dead Sea population from March 2007 displayed greater richness than the microbial bloom of 1992. This coincides well with the results of Bodaker *et al.* (Bodaker *et al.*, 2009b) and is likely due to the gradual accumulation of strains in the fifteen years that elapsed since the 1992 bloom. Likewise the 2007 Dead Sea displayed greater diversity at the strain level than the 1992 Dead Sea. This result is not particularly surprising as during a bloom a select



**Figure 2-3.** Flow chart of the seeding potential of the natural Dead Sea and mesocosm pool samples through time. Seeding potentials progressing forward in time are shown larger and in bold font. Seeding potentials progressing backward in time are shown smaller and italicized.

group of species often dominate a population thereby reducing overall diversity.

However at the genus level the early stages of the 1992 halobacterial bloom contained comparable if not greater richness and diversity than the residual 2007 population. Thus while the residual population was relatively rich in strains it was relatively poor with respect to broader phylogenetic categories. The overwhelming challenges presented by the modern Dead Sea appear to have restricted the residual microbial community to a select few branches. Nevertheless, within each branch there was considerable diversity.

At the same time, the latter stages of the bloom exhibited far less species richness and diversity than the early bloom at the strain, species, and genus levels. Whether this was due to some sort of microbial succession or environmental factors is difficult to say. The early stages of the bloom may have been dominated by species and strains that are more adept at rapidly exploiting the suddenly hospitable environment. Later more slow-growing species may have out-competed the initial bloom population. Alternatively increasing salinity and decreasing environmental resources, such as depletion of glycerol and other organic nutrients, may have contributed to a population shift.

As for the composition of the respective microbial communities, the residual Dead Sea microbial community appeared to be significantly different than any of the other Dead Sea microbial communities. The identifiable portion of the 1992 bloom samples were dominated by organisms related to the alkaliphilic halophilic genus *Natronococcus* as well as the haloarchaeal genus, *Halosarcina*. A limited number of sequences were recovered of *Halobaculum gomorreense*, an organism isolated from the

1992 bloom (Oren & Gurevich, 1995). Our results lie in stark contrast to those obtained by Bodaker *et al.* (Bodaker *et al.*, 2010) from fosmid and environmental sequences cloned from the June 1992 Dead Sea bloom sample. The fosmid and environmental clones were restricted to a single genus, *Halobacterium*. *Halobacterium*, while present in our amplicon data set, comprised only a very small portion of the Dead Sea bloom population. Our results remain distinct yet are more congruous with the amplicon library sequenced for the same study (Bodaker *et al.*, 2010). This suggests the occurrence of a significant cloning bias in the preparation of the 1992 bloom clones. Meanwhile, the 2007 sample contained a higher proportion of uncultured haloarchaea and the identifiable portion was dominated by the genus *Halorhabdus* and the alkaliphilic halophilic genus *Natronomonas*. Only the June 1992 sample contained significant quantities of bacteria, primarily from the genus *Bacillus*, and showed any indication of the presence of archaea not belonging to the haloarchaea (family *Halobacteriaceae*).

The *Bacillus* sequences were 100% identical to the species *Bacillus selenatarseatis* and *Bacillus jeotgali*. Nevertheless, the finding of *Bacillus* as a possibly dominant organism in the June 1992 sample was highly unexpected. We know of no isolated *Bacillus* species that can grow at Dead Sea water salinity. *Virgibacillus marismortui*, isolated in the 1990s from an enrichment culture set up in the 1930s with Dead Sea water, grows optimally at 10% salt and cannot grow above 25% total salinity (Arahal *et al.*, 1999), and the polar lipid chromatographic patterns obtained in 1992 (Oren & Gurevich, 1993) did not indicate a massive presence of bacteria in the community.

The identifiable portions of Pools 1 and 2, the more saline of the Dead Sea experimental pools, were dominated by the genus *Haloquadratum*. The genus *Haloquadratum* is thus far represented by a single species, *H. walsbyi*, with a world-wide distribution (Dyall-Smith *et al.*, 2011) and described as very tolerant to high magnesium concentrations and life at low water activities (Bolhuis *et al.*, 2006), properties essential for growth and survival in the Dead Sea. Cells of *H. walsbyi* are flat square or rectangular in shape, and such cells can be recognized in micrographs and electron micrographs of the microbial community of the 1992 bloom Dead Sea (Oren & Gurevich, 1993; Oren *et al.*, 1997). Therefore while it is not surprising to find relatives of *H. walsbyi* in the mesocosm pools, it is somewhat surprising that relatively few representatives of *Haloquadratum* were identified in the 1992 bloom Dead Sea samples. Additionally, Pools 1 and 2 exhibited considerable amounts of the alkaliphilic genera *Natronomonas* and *Natronococcus* despite the fact that the pH of the waters was near-neutral (6.5-7.2). Pool 10, a pool with middling salinity, also contained significant quantities of both *Natronomonas* and *Natronococcus*. However the microbial population of pool 10 was largely dominated by a single uncultured strain which comprised almost 70% of the entire population.

As with the bacterium *Bacillus*, it is interesting to compare the outcome of the phylogenetic 16S rRNA sequence based analysis of the microbial communities in the Dead Sea and in the Sedom experimental ponds with the analysis of the polar lipid composition of the communities. Polar lipids are excellent chemotaxonomic markers for different genera of Archaea, which differ in the type(s) of glycolipids present (non-

sulfated, mono-sulfated or bi-sulfated diglycosyl, triglycosyl, and/or tetraglycosyl glycolipids, or no glycolipids at all – as characteristic for the haloalkaliphilic Archaea) and in the presence or absence of the phytanyl diether derivative of phosphatidylglycerol sulfate. Four major polar lipid components were detected in extracts of the biomass collected during the 1992 bloom: the diphytanyl diether derivatives of phosphatidylglycerol, of the methyl ester of phosphatidylglycerol phosphate, of phosphatidylglycerol sulfate, and a single glycolipid, identified as a monosulfated diglycosyl diether lipid on the basis of its chromatographic properties (Oren & Gurevich, 1993). The same components also dominated the communities that developed in the Sedom mesocosm pools under a variety of conditions (Oren *et al.*, 2004). This is exactly the pattern characteristic of *Haloquadratum* (Lobasso *et al.*, 2008; Oren *et al.*, 1996), one of the genera identified most frequently in the Pool 1 and 2 samples on the basis of 16S rRNA gene sequences.

Finally, only pool 9, the oldest and least saline of the pools, exhibited significant quantities of both *Halosarcina* and also of the bacterial genus *Salisaeta* (*Bacteroidetes*). The latter was not unexpected, as *Salisaeta longa*, the only described species within the genus, was isolated from one of the experimental pools at Sedom (Vaisman & Oren, 2009). It therefore looks as if only pool 9 accurately reflected the early to mid stages of the 1992 bloom and that a considerable dilution of the Dead Sea surface waters would be required to promote the growth of bacterial halophiles. It likewise appears that the residual Dead Sea population was not merely a reflection of the bloom population and

that the introduction of Red Sea brine to the Dead Sea would cause a significant population shift.

However, our results also indicate that it would be incorrect to assume that the 1992 bloom population had no impact on the residual Dead Sea population. Despite the significant decrease in diversity from the June 1992 sample to the September 1992 sample, microbes present in the September 1992 sample do a better job of seeding the residual population. As microbes present in the June 1992 sample can account for more than 98% of the September 1992 community, the microbes present in both the residual population and the September 1992 population but not in March 1992 are relatively sparse. Nevertheless, there is in fact some indication that the 1992 bloom began to evolve to a state more representative of the residual population. However, it is not these remnants of the latter stages of the 1992 bloom in the residual 2007 population that dominate the bloom in the mesocosm pools. Despite its considerable species richness, the March 2007 Dead Sea sample does a very poor job of seeding the blooms in all of the mesocosm blooms. The inocula for these blooms are either microbes present at astonishingly low concentrations in the modern Dead Sea or are introduced from elsewhere, most likely via aeolian deposition. We therefore conclude that while some of the bloom taxa are preserved in the residual Dead Sea population and are available to bloom, a significant proportion of bloom taxa are unable to survive the harsh inter-bloom conditions and must be reintroduced from the surrounding environs.

## Acknowledgements

We thank Oded Béjà, Idan Bodaker, and the late Moti Gonen for their assistance in collecting and processing the March 2007 and June 1992 samples and Lynn Tomsho and Stephan Schuster for sequencing. We further thank the Geological Survey of Israel and the Dead Sea Works, Ltd., for setting up and maintaining the experimental mesocosms at Sedom. This work was supported in part by the Agriculture and Food Research Initiative Competitive Grants Program Grant no. 2010-65110-20488 from the USDA National Institute of Food and Agriculture and by the National Aeronautics and Space Administration (NASA) Astrobiology Institute (NAI) under NASA-Ames Cooperative Agreement NNA09DA76A (C.H.H.). The 454 facility at the Pennsylvania State University Center for Genome Analysis is funded, in part, using Tobacco Settlement Funds appropriated by the legislature provided by a grant from the Pennsylvania Department of Health.

## References

- Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997).** Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**, 3389-3402.
- Arahal, D. R., Márquez, M. C., Volcani, B. E., Schleifer, K. H. & Ventosa, A. (1999).** *Bacillus marismortui* sp. nov., a new moderately halophilic species from the Dead Sea. *International Journal of Systematic Bacteriology* **49**, 521-530.
- Beyth, M. (1980).** Recent evolution and present stage of Dead Sea brines. In *Hypersaline Brines and Evaporitic Environments*, pp. 155-165. Edited by A. Nissenbaum: Elsevier.
- Bodaker, I., Béjà, O., Feingersh, R., Rosenberg, M., Oren, A., Hindiyyeh, M. Y. & Malkawi, H. I. (2009).** Archaeal diversity in the Dead Sea: Microbial survival under increasingly harsh conditions. In *Saline lakes around the world: Unique systems with unique values*, pp. 137-143.

Edited by A. Oren, D. Naftz, P. Palacios & W. A. Wurtsbaugh. Logan, Utah: S. J. and Jessie E. Quinney Natural Resources Research Library.

**Bodaker, I., Sharon, I., Suzuki, M. T. & other authors (2010).** Comparative community genomics in the Dead Sea: An increasingly extreme environment. *The ISME Journal* **4**, 399-407.

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

**Cole, J. R., Wang, Q., Cardenas, E. & other authors (2009).** The Ribosomal Database Project: Improved alignments and new tools for rRNA analysis. *Nucleic Acids Research* **37**, D141-D145.

**Dyall-Smith, M. L., Pfeiffer, F., Klee, K., Palm, P., Gross, K., Schuster, S. C., Rampp, M. & Oesterhelt, D. (2011).** *Haloquadratum walsbyi*: Limited diversity in a global pond. *PLoS ONE* **6**, e20968.

**Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C. & Knight, R.** 23 June 2011, posting date. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*. doi:10.1093/bioinformatics/btr381.

**Field, K., Olsen, G., Lane, D., Giovannoni, S., Ghiselin, M., Raff, E., Pace, N. & Raff, R. (1988).** Molecular phylogeny of the animal kingdom. *Science* **239**, 748-753.

**Gavrieli, I., Beyth, M. & Yechieli, Y. (1999).** The Dead Sea - A terminal lake in the Dead Sea rift: A short overview. In *Microbiology and biogeochemistry of hypersaline environments*, pp. 121-127. Edited by A. Oren. Boca Raton, Florida: CRC Press.

**Gavrieli, I. & Oren, A. (2004).** The Dead Sea as a dying lake. In *Dying and dead seas Climactic versus anthropic causes*, pp. 287-305. Edited by J. C. J. Nihoul, P. O. Zavialov & P. P. Micklin. Dordrecht: Kluwer Academic Publishers.

**Lobasso, S., Lopalco, P., Mascolo, G. & Corcelli, A. (2008).** Lipids of the ultra-thin square halophilic archaeon *Haloquadratum walsbyi*. *Archaea* **2**, 177-183.

**Macalady, J. L., Dattagupta, S., Schaperdoth, I., Jones, D. S., Druschel, G. K. & Eastman, D. (2008).** Niche differentiation among sulfur-oxidizing bacterial populations in cave waters. *The ISME Journal* **2**, 590-601.

**Muyzer, G., Hottentrager, S., Teske, A. & Wawer, C. (1996).** Denaturing gradient gel electrophoresis of PCR-amplified 16S rDNA. A new molecular approach to analyze the genetic diversity of mixed microbial communities. In *Molecular Microbial Ecology Manual*, pp. 3.4.4.1-3.4.4.22. Edited by A. D. L. Akkermans, J. D. van Elsas & F. J. de Bruijn. Dordrecht: Kluwer Academic Publishers.

**Neev, D. & Emery, K. O. (1967).** The Dead Sea depositional processes and environments of evaporates: State of Israel Ministry of Development Geological Survey.

- Oren, A. (1983).** Population dynamics of halobacteria in the Dead Sea water column. *Limnology and Oceanography* **28**, 1094-1103.
- Oren, A., Ginzburg, M., Ginzburg, B. Z., Hochstein, L. I. & Volcani, B. E. (1990).** *Haloarcula marismortui* (Volcani) sp. nov., nom. rev., an extremely halophilic bacterium from the Dead Sea. *International Journal of Systematic Bacteriology* **40**, 209-210.
- Oren, A. (1993a).** Ecology of extremely halophilic microorganisms. In *The biology of halophilic bacteria*, pp. 25-53. Edited by R. H. Vreeland & L. I. Hochstein. Boca Raton: CRC Press.
- Oren, A. (1993b).** The Dead Sea - alive again. *Experientia* **49**, 518-522.
- Oren, A. & Gurevich, P. (1993).** Characterization of the dominant halophilic archaea in a bacterial bloom in the Dead Sea. *FEMS Microbiology Ecology* **12**, 249-256.
- Oren, A. & Gurevich, P. (1995).** Dynamics of a bloom of halophilic archaea in the Dead Sea. *Hydrobiologia* **315**, 149-158.
- Oren, A., Duker, S. & Ritter, S. (1996).** The polar lipid composition of Walsby's square bacterium. *FEMS Microbiology Letters* **138**, 135-140.
- Oren, A. & Ben-Yosef, N. (1997).** Development and spatial distribution of an algal bloom in the Dead Sea: A remote sensing study. *Aquatic Microbial Ecology* **13**, 219-223.
- Oren, A., Bratbak, G. & Haldal, M. (1997).** Occurrence of virus-like particles in the Dead Sea. *Extremophiles* **1**, 143-149.
- Oren, A., Gavrieli, I., Gavrieli, J., Kohen, M., Lati, J. & Aharoni, M. (2004).** Biological effects of dilution of Dead Sea brine with seawater: Implications for the planning of the Red Sea-Dead Sea "Peace Conduit". *Journal of Marine Systems* **46**, 121-131.
- Park, H. S., Chatterjee, I., Johnston, S., Wong, M., Jack, T. R. & Voordouw, G. (2011).** Microbially influenced corrosion in oil sands water handling facilities. In *Corrosion 2011*. Houston, Texas: NACE International
- Pruesse, E., Quast, C., Knittel, K., Fuchs, B. M., Ludwig, W., Peplies, J. & Glöckner, F. O. (2007).** SILVA: A comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Research* **35**, 7188-7196.
- Rhodes, M. E., Fitz-Gibbon, S. T., Oren, A. & House, C. H. (2010).** Amino acid signatures of salinity on an environmental scale with a focus on the Dead Sea. *Environmental Microbiology* **12**, 2613-2623.
- Sahl, J. W., Pace, N. R. & Spear, J. R. (2008).** Comparative molecular analysis of endoevaporitic microbial communities. *Applied and Environmental Microbiology* **74**, 6444-6446.

**Schloss, P. D., Westcott, S. L., Ryabin, T. & other authors (2009).** Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology* **75**, 7537-7541.

**Vaisman, N. & Oren, A. (2009).** *Salisaeta longa* gen. nov., sp. nov., a red, halophilic member of the Bacteroidetes. *International Journal of Systematic and Evolutionary Microbiology* **59**, 2571-2574.

**Wilkansky, B. (1936).** Life in the Dead Sea. *Nature* **138**, 467.

## Chapter 3

### Amino acid signatures of salinity on an environmental scale

#### with a focus on the Dead Sea

Matthew E. Rhodes<sup>ab1</sup>, Sorel T. Fitz-Gibbon<sup>ac</sup>, Aharon Oren<sup>d</sup>, Christopher H. House<sup>ab</sup>

<sup>a</sup>Penn State Astrobiology Research Center, Pennsylvania State University, University Park, PA 16802

<sup>b</sup>Department of Geosciences, Pennsylvania State University, University Park, PA 16802

<sup>c</sup>Center for Astrobiology, Institute of Geophysics and Planetary Physics, University of California, Los Angeles, CA 90095

<sup>d</sup>The Institute of Life Sciences, and the Moshe Shilo Minerva Center for Marine Biogeochemistry, The Hebrew University of Jerusalem, 91904 Jerusalem, Israel

<sup>1</sup>To whom correspondence should be addressed

Published in *Environmental Microbiology* 7 May, 2010

#### Summary

The increase of the acidic nature of proteins as an adaptation to hypersalinity has been well documented within halophile isolates. Here we explore the effect of salinity on amino acid preference on an environmental scale. Via pyrosequencing, we have obtained two distinct metagenomic data sets from the Dead Sea, one from a 1992

archaeal bloom and one from the modern Dead Sea. Our data, along with metagenomes from environments representing a range of salinities, show a strong linear correlation ( $R^2=0.97$ ) between the salinity of an environment and the ratio of acidic to basic amino acids encoded by its inhabitants. Using the amino acid composition of putative protein-encoding reads and the results of 16S rRNA amplicon sequencing, we differentiate recovered sequences representing microorganisms indigenous to the Dead Sea from lateral gene transfer events and foreign DNA. Our methods demonstrate lateral gene transfer events between a halophilic archaeon and relatives of the thermophilic bacterial genus *Thermotoga* and suggest the presence of indigenous Dead Sea representatives from 10 traditionally non-hyperhalophilic bacterial lineages. The work suggests the possibility that amino acid bias of hypersaline environments might be preservable in fossil DNA or fossil amino acids, serving as a proxy for the salinity of an ancient environment. Finally, both the amino acid profile of the 2007 Dead Sea metagenome and the V9 amplicon library support the conclusion that the dominant microorganism inhabiting the Dead Sea is most closely related to a thus far uncultured relative of an alkaliphilic haloarchaeon.

## **Introduction**

Advances in DNA sequencing have made it possible to study the genetic makeup of entire environments with theoretically little bias. Thus, the GC content or gene content of an environment can be ascertained yielding valuable information as to the makeup and metabolic capabilities of microorganisms present (Tyson *et al.*, 2004;

Venter *et al.*, 2004). Another parameter of interest is the distribution of amino acids coded within the genomes of an environment. It has been documented for individual species that various environmental stresses, such as extreme acidity or extreme salinity (Goodarzi *et al.*, 2008; Haney *et al.*, 1999; Paul *et al.*, 2008), can bias their amino acid composition due to their desired chemical characteristics. Here we investigate whether the pattern of encoded amino acids can be indicative of environments as well.

Halophilic microorganisms have developed two strategies to deal with the multimolar salinities of their environments. All eukaryotic species, most halophilic Bacteria, and the halophilic methanogenic Archaea build up concentrations of organic solutes (osmolytes), to balance the osmotic pressure. This “salt-out” method allows the internal mechanisms of the cell to remain in their native states but requires a high energy cost to manufacture the organic molecules. In contrast, Archaea of the order Halobacteriales, as well as a limited number of halophilic Bacteria, accumulate high concentrations of salts, typically KCl, within their cytoplasm. This “salt-in” method is energetically more efficient but requires the adaptation of intracellular proteins to high salt concentrations (Lanyi, 1974; Oren, 1986; Oren, 1999b; Oren, 2002a; Reistad, 1970). The presence of high quantities of  $K^+$  alters the intracellular environment, thereby interfering with protein interactions (Lanyi, 1974; Madern *et al.*, 2000). This necessitates a number of changes in protein structure to maintain proper protein function. Lanyi in 1974 summarized the adaptations of proteins to extreme salinity (Lanyi, 1974). Included in Lanyi’s summary is an overall increase in acidic amino acids which is offset by an overall decrease in basic amino acids. This trend has been

demonstrated in the genomes of hyperhalophilic salt-in Archaea such as *Halobacterium* NRC -1 (Ng *et al.*, 2000). In salt-out halophiles, however, only proteins exposed to the hypersaline medium exhibit an excess of acidic amino acids (Oren *et al.*, 2005).

At a salinity of 347 g/l, the modern Dead Sea represents an especially inhospitable environment at the extreme of hypersalinity. Due to changing weather patterns and increased water usage, the salinity of the Dead Sea surface water has risen steadily from 269 g/l in the 1930's (Volcani, 1944) to its current value (Volcani, 1944), with two notable exceptions. In 1980 and 1992 heavy winter rains created a net positive water budget and diluted the surface waters of the Dead Sea to 200 g/l and 170 g/l respectively (Gavrieli *et al.*, 1999). The dilutions allowed the establishment of a bloom of the alga *Dunaliella*. The algal blooms released organic material, most likely including the organic osmolyte glycerol. These compounds in turn provided the energy for subsequent blooms of halophilic Archaea with cell counts exceeding  $3 \times 10^7$  per ml in the late spring of 1992 (Oren & Gurevich, 1995).

We have used metagenomic methods to analyze the encoded amino acid distribution of two disparate Dead Sea ecosystems, that of the modern Dead Sea of March 2007 and that of a properly prepared and frozen sample from the archaeal bloom of September 1992. We have compared these ecosystems to a number of environments for which multiple similar data sets are available: a Spanish saltern (Legault *et al.*, 2006), the deep sea subsurface (Biddle *et al.*, 2008), mammalian guts (Gill *et al.*, 2006; Turnbaugh *et al.*, 2006), whale falls (Tringe *et al.*, 2005), and the moderately saline Guerrero Negro, Baja California microbial mats (Kunin *et al.*, 2008; Ley *et al.*,

2006; Robertson *et al.*, 2009; Spear *et al.*, 2003). We employed these results to demonstrate that a metagenomic amino acid profile is characteristic of an environment and we suggest the use of the ratio of acidic amino acids to basic amino acids encoded within a metagenome or preserved within fossilized peptides as a proxy for the salinity of highly hypersaline paleoenvironments. We also used the acidic nature of proteins in salt-in halophiles to differentiate with sequence data alone, organisms that naturally inhabit the Dead Sea from lateral gene transfer (LGT) events, and/or the presence of foreign species.

## Results

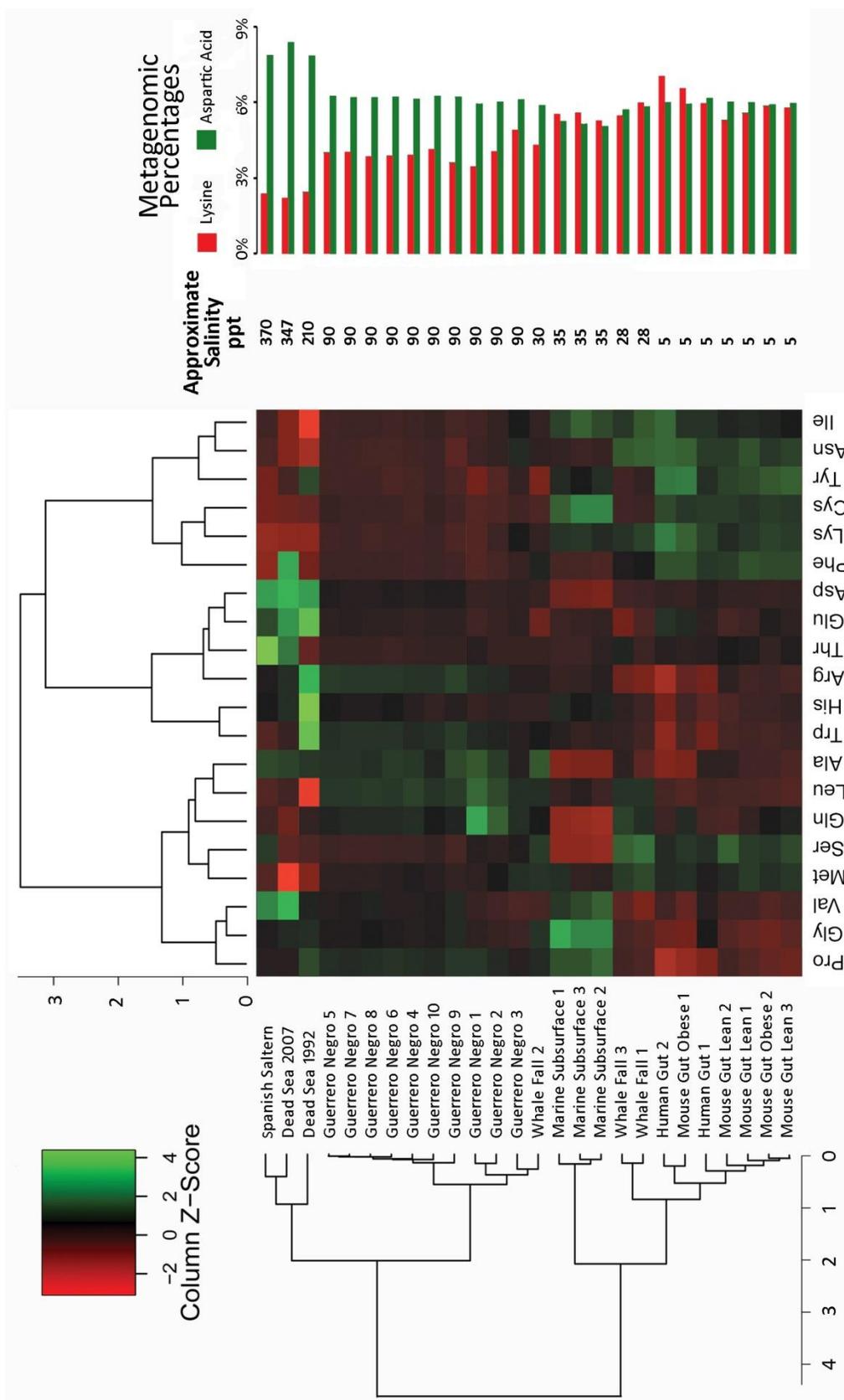
DNA was extracted from two water samples collected from the Dead Sea in September of 1992 and March of 2007 respectively. A portion of the 1992 sample was subjected to whole genome amplification (Dean *et al.*, 2001). Subsequently, a half plate of both the amplified 1992 sample and the unamplified 2007 sample were sequenced on a 454 Life Sciences / Roche FLX sequencer. The 2007 half plate yielded a total of 243,816 unique reads with an average length of 250.8 base pairs. The 1992 half plate yielded a total of 137,137 unique reads with an average read length of 240.1 base pairs.

These datasets were compared to the non-redundant protein database using BLASTX (Altschul *et al.*, 1997). With a cutoff e-value of  $10^{-5}$ , the 2007 dataset returned 139,345 or 57% of the reads as having homology and the 1992 dataset returned 15,301 or 11% of the reads as having homology. The discrepancy in the proportion of hits is likely caused by the amplification process producing chimeric DNA in the 1992 sample.

For each top hit the portion of the read that matched to a homologous protein by BLASTX was extracted, and its amino acid composition was tallied. A similar analysis was performed on BLASTX output from three datasets from the Peru Margin subseafloor (Biddle *et al.*, 2008) as well as a number of publically available protein metagenomes for which there were multiple data sets. These consisted of metagenomes from a Spanish saltern (Legault *et al.*, 2006), ten layers of a moderately saline microbial mat from Guerrero Negro (Kunin *et al.*, 2008), three metagenomes from whale falls (Tringe *et al.*, 2005), and seven metagenomes from mammalian guts (Gill *et al.*, 2006; Turnbaugh *et al.*, 2006). Additionally the approximately 70 base pair V9 hypervariable region of the 16S rRNA gene was amplified from the 2007 sample and the amplification product was sequenced on the 454 FLX sequencer. The resulting 29,673 quality controlled 16S amplicons were assigned to taxa by a BLASTN comparison as described in the methods section.

### *Cluster analysis*

We standardized the raw counts data by first dividing through by the site (row) totals and then by the amino acid (column) maximums. Subsequently we performed a hierarchical cluster analysis on both the site data and the amino acid data using Ward's method (Ward, 1963) and the Euclidean distance (Figure 1). Both analogous environments and different samplings from the related environments cluster together. The two most hypersaline environments, the 2007 Dead Sea and the Spanish saltern, group closely together. These two in turn cluster with the 1992 Dead Sea, albeit rather distantly, and the three hypersaline environments are most closely related to all ten of



**Figure 3-1.** Two-dimensional 'heat plot' showing the hierarchical clustering of environments (y-axis) and amino acids (x-axis). The two clusterings are based on comparisons of the amino acid content of each metagenome. The scales adjacent to the dendrograms give the Euclidean distances. Also shown along the right side, for reference, are the salinities and the lysine and aspartic acid contents of the environments.

the moderately saline Guerrero Negro mat metagenomes. In the bottom half of the dendrogram we see the deep sea subsurface clustering together, the mammalian guts clustering together, and two of the three whale falls clustering together. Presumably selective pressure on amino acid preference imparted by salinity does not become a significant factor until salinities greater than those of the marine subsurface (35‰). Thus, when taken as a whole, environments have characteristic signatures of amino acid usage throughout their entire metagenomes.

The right most portion of the upper dendrogram (Figure 1) includes all the amino acids encoded primarily by A's and T's except for methionine. This clustering is likely created by the extremely high GC content of most hyperhalophilic genomes, often upwards of 60%. The 2007 Dead Sea metagenome and the 1992 Dead Sea metagenome encode 67% and 62% GC's respectively (Table B-1). Due to the dominance of the particularly GC poor hyperhalophile, *Haloquadratum walsbyi*, the Spanish Saltern has an overall GC content of only 54%.

As expected, the aspartic acid values are especially high in the hypersaline environments and decrease with decreasing salinity. Also as expected, the glutamic acid levels are extremely high in both the Spanish saltern and the Dead Sea, yet they do not display an increased prevalence in the moderately saline Guerrero Negro mats. Conversely, the lysine values are especially low in the Dead Sea and increase with decreasing salinity. Arginine, however, displays a counterintuitive trend. As a basic amino acid, arginine would be expected to have a lower proportion in saline environments. Nevertheless, the proportion of encoded arginine increases with salinity.

A potential driver for these opposing tendencies is the GC content of the codons associated with each amino acid. The amino acid lysine is encoded by the two codons AAA and AAG, both of which have a bias towards AT nucleotides. In contrast, arginine is encoded by the six codons, CGT, CGC, CGA, CGG, AGA, and AGG for which five of the six codons have a bias toward GC nucleotides. A number of studies have suggested that high genomic GC-content is an almost universal adaptation to hypersaline environments (Kennedy *et al.*, 2001; Soppa, 2006) and, as mentioned above, the two Dead Sea metagenomes have GC contents of over 60%. In contrast, the Guerrero Negro mat metagenomes have GC contents in the mid fifties, and the non-saline metagenomes have GC contents of about 50%. Thus, based upon mutational bias alone we would expect to see a discrepancy between lysine and arginine in the GC rich Dead Sea.

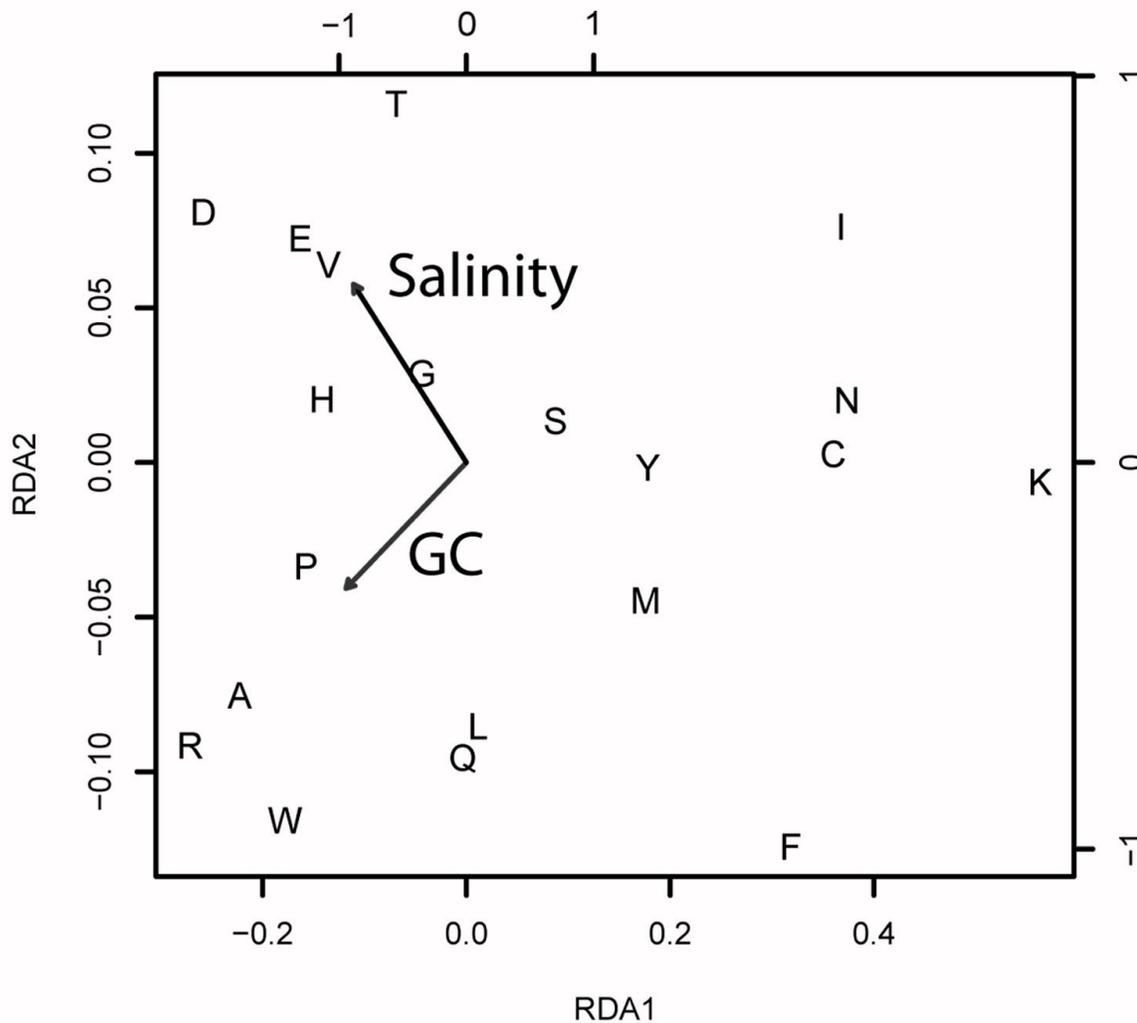
This might suggest that mutational bias alone is the driving force behind the decrease in lysine and the increase in arginine. However there is one known exception to the GC rich salt-in hyperhalophile rule. The hyperhalophilic archaeon *H. walsbyi*, has a genome with a GC content of only 47.9% (Bolhuis *et al.*, 2006). As with other hyperhalophiles, *H. walsbyi* also encodes a low proportion of lysine, 2.33% (Table B-1), suggesting that a decrease in lysine is universal and not merely a product of GC bias. On the other hand, the arginine content encoded by *H. walsbyi* is at relatively normal levels, 5.80%, indicating that the increase in arginine in the Dead Sea is potentially caused by the GC bias of the environment. Other notable trends are the high alanine values in the saline environments, the high cystein values in the deep sea subsurface and the mammalian guts, and the high tyrosine values in the Dead Sea. Tyrosine like lysine is

encoded solely by AT biased codons and would be expected to be relatively uncommon in the GC rich Dead Sea.

#### *Redundancy analysis*

A redundancy analysis was performed with salinity and GC content encoded as the dependent variables (Figure 2). Fifty-two percent of the variance is explained by the first constrained axis (x-axis) and another four percent by the second constrained axis (y-axis). The two environmental variables both plot strongly negative along the first axis, while they differ along the second axis. GC content plots negatively along the second axis and salinity plots positively.

As predicted, aspartic acid and glutamic acid plot tightly along the salinity trend. Meanwhile arginine, alanine, and tryptophan, all amino acids encoded exclusively or primarily by GC biased codons, plot tightly with GC content. In the opposite direction, lysine, aspartic acid, and cysteine appear to be affected by both salinity and GC content, and Ile, an amino acid encoded by three AT biased codons, plots negatively associated with GC content. Finally the amino acids histidine and phenylalanine do not conform to expectation. Although phenylalanine is encoded by two AT biased codons, it plots negatively with salinity, not with GC content. This can potentially be explained by the low proportion of phenylalanine in the GC poor marine subsurface. Histidine, a positively charged amino acid, plots positively with salinity. Unlike arginine this cannot be explained by a GC bias. Histidine is encoded by equally by both AT and GC biased



**Figure 3-2.** Redundancy analysis of the amino acid content of the environments with salinity and GC content encoded as the dependent variables. The arrows show the directions of increasing GC content and salinity respectively. The bottom and left-hand axes provide scales for the loading of the amino acids onto the restricted axes and the top and right-hand axes provide scales for the loading of the environmental variables onto the restricted axes.

codons. Nevertheless halophilic Archaea do appear to encode greater proportions of histidine than non-halophilic Archaea (Table B-1).

#### *Environmental amino acid profiles versus salinity*

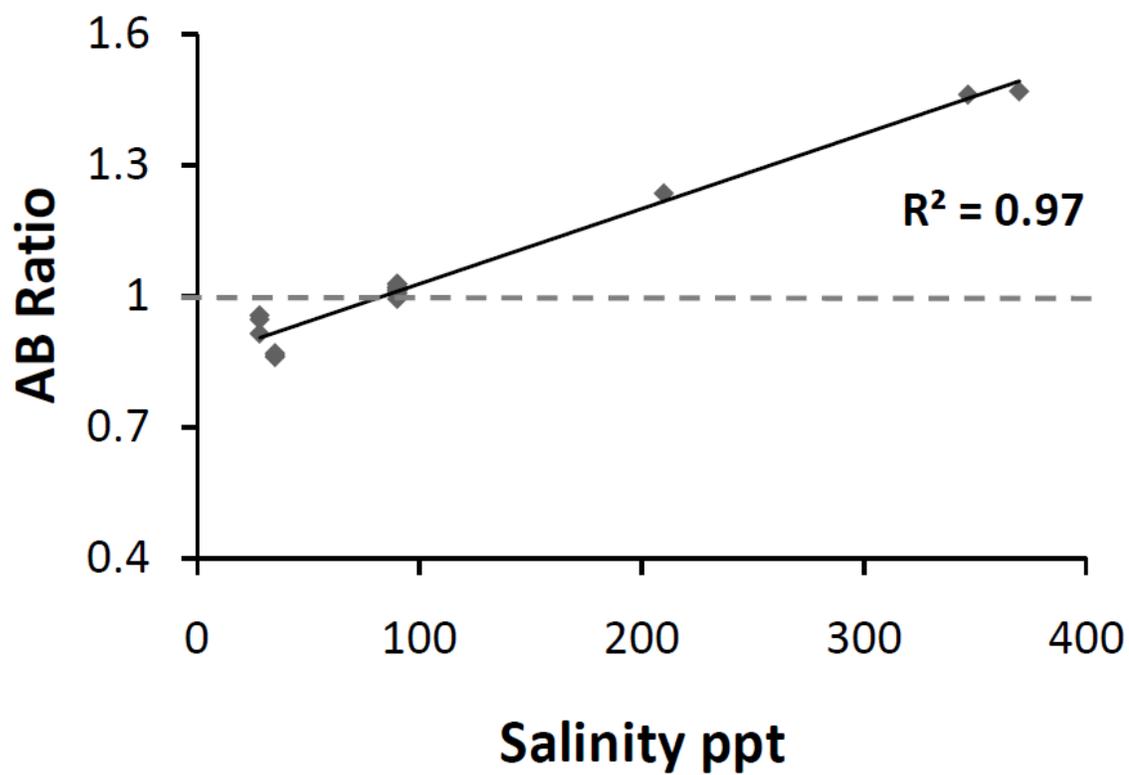
We observe significant deviation in the percentage of encoded glutamic acid, aspartic acid, and lysine between the hypersaline metagenomes and all other metagenomes (Figure 1). Furthermore, the moderately saline, Guerrero Negro microbial mat metagenomes, display a slight excess of aspartic acid and a more moderate deficit of lysine relative to non-saline environments. The excess of glutamic acid and aspartic acid and the deficit of lysine in the saline metagenomes are products of the survival mechanisms used by salt-in halophiles to cope with the stresses of hypersaline environments. The adaptations generally include an increase of the acidic nature of intracellular proteins. The radical nature of the salt-in method makes it likely that the three distantly related halophilic lineages adopting this strategy derived it independently (Santos & Da Costa, 2002). Thus, the salt-in method potentially marks an example of convergent evolution. However, Kunin *et al.* observed the acidic nature of proteins in a moderately saline environment to be far more widespread than previously thought, indicating that the salt-in method of halophilicity may in fact be relatively common in halophiles (Kunin *et al.*, 2008).

Consequently, the degree of excess acidic amino acids and dearth of basic amino acids reflects the prevalence of the “salt-in” strategy in an environment and the amount of adaptation necessary to cope with the environmental stress. This can be quantified by looking at the ratio of the acidic amino acids glutamic acid and aspartic acid to the

basic amino acids lysine, histidine, and arginine which we term the AB ratio. Here the difference between the 1992 Dead Sea and the 2007 Dead Sea becomes apparent. The 2007 Dead Sea and the Spanish saltern both have AB ratios of approximately 1.46. Meanwhile the 1992 dilution of the surface waters of the Dead Sea leading to the haloarchaeal bloom lowered the AB ratio of 1.24 (Table B-1). Since both the 1992 Dead Sea and 2007 Dead Sea are archaeal dominated environments (MEGAN assigned >90% of reads to Archaea for both Dead Sea metagenomes), this shift is presumably a reflection of the change in the haloarchaeal community concurrent with salinity. The Guerrero Negro mats fall on a range between 1.00 and 1.03 with an average of 1.01, and the marine environments range between .86 and .96 with an average of .90. We therefore see a distinct correlation as we increase from marine to hypersaline salinity levels between the amino acid proportions and the salinity of an environment (Figure 3). At lower salinities, however, it is unclear whether this linear relationship holds. The mammalian gut metagenomes observe a dichotomy and are unexpectedly high. Both human guts have an AB ratio of 1.02 while the mouse guts range from .95 to .97. The cause for this discrepancy is unclear and more work is required to ascertain the total spread of AB ratios for non-saline and slightly saline environments.

#### *Archaea in the modern Dead Sea*

The MEGAN program (Huson *et al.*, 2007) assigned 77% of the 2007 Dead Sea reads to the order Halobacteriales and a number of halobacterial species have been



**Figure 3-3.** Correlation between the ratio of acidic to basic amino acids encoded in an environment and the salinity of that environment. The salinity values range from marine ( $\approx 28$  g/l) to the extreme of hypersalinity ( $\approx 370$  g/l).

isolated from its waters. We compared the 2007 Dead Sea metagenomic amino acid values to those of the fully sequenced Dead Sea isolates and other sequenced halophiles. We chose to focus on the five charged amino acids, aspartic acid, glutamic acid, lysine, histidine, and arginine found in the 2007 Dead Sea metagenome as it is the more robust dataset. The metagenomic aspartic acid value falls comfortably within the range of values for the fully sequenced hyperhalophiles (Table B-1). In contrast the glutamic acid value exceeds those of all known hyperhalophiles except the alkaliphile, *Natronomonas pharaonis*. The Dead Sea itself is slightly acidic, with a pH of approximately 6. For the basic amino acids, the arginine value exceeds those of all but the hyperhalophilic bacterium *Salinibacter ruber*, which has not been observed in the Dead Sea. The lysine value exceeds those of all but *H. walsbyi*, and the high GC content of the Dead Sea metagenome precludes *H. walsbyi* from composing of the majority of the ecosystem. Finally, the histidine value exceeds those of all hyperhalophiles.

Therefore, based upon the amino acid profile data, we can conclude that the hyperhalophilic communities within the 2007 Dead Sea are predominantly composed of as of yet genomically unsequenced organisms. The 16S rRNA amplicon library of the 2007 Dead Sea supports this assertion. The majority of the classifiable 16S tags, 66%, are most closely related to uncultured haloarchaea in the RDP database (Cole *et al.*, 2007). This includes the most frequently observed amplicon sequence which alone composes greater than one fourth of the archaeal community and matches identically to uncultured sequences extracted from the archaeal community of the alkaline-saline soil of the former lake, Texcoco (Valenzuela-Encinas *et al.*, 2008). When taken together

the V9 amplicon library and the amino acid profile of the 2007 Dead Sea suggest that the dominant organism inhabiting the modern Dead Sea is an uncultured and unisolated relative of an alkaliphilic haloarchaeon.

### *Screening of bacterial taxa*

A total of 6.7% of the 2007 metagenomic reads were assigned by MEGAN to bacterial taxa. These bacterial reads can be broken down into three categories:

- I: Reads from bacteria indigenous to the ecosystem.
- II: Reads of DNA foreign to the ecosystem.
- III: Reads showing LGT events involving indigenous members of the ecosystem.

Here we have used the metagenomic data, a collection of 16S rRNA tags from the V9 hypervariable region, and the amino acid bias employed by salt-in halophiles in a novel method to differentiate between the three categories of bacterial reads.

Representatives of salt-in taxa from category I and LGT events involving salt-in taxa from category III should demonstrate an overall signal of protein adaptation to salinity. All representatives of category II taxa and salt-out representatives of the other two categories should not. We quantified the salinity adaptation quotient (SAQ) by first pooling all hits to a given taxonomic group, in essence creating an artificial genome. We then compared the amino acid ratios of the metagenomic reads with those of the homologous portions of the respective best BLAST hits:

$$\text{SAQ} = \frac{\text{AB Reads}}{\text{AB BLAST}} = \frac{(\text{Asp} + \text{Glu (Reads)} / \text{His} + \text{Arg} + \text{Lys (Reads)})}{(\text{Asp} + \text{Glu (BLAST)} / \text{His} + \text{Arg} + \text{Lys (BLAST)})}$$

Essentially, an SAQ of greater than one indicates that the taxonomic group as a whole has undergone some adaptation to salinity. We chose an SAQ of greater than 1.2 to be indicative of significant adaptation to salinity (See Appendix A, Supplemental Information 2). Furthermore, it is accepted that informational genes including 16S rRNA genes exhibit a lower propensity for LGT than operational genes (Rivera *et al.*, 1998). Therefore, the presence of an organism's 16S rRNA gene in an environment indicates the presence of genomic DNA from that organism.

Thus, while all three categories of bacterial taxa should be present in the metagenome, only categories I and II should be represented by 16S rRNA amplicons. However, while the presence of particular 16S rRNA amplicons presents convincing evidence of the presence of specific genomic DNA in a sample, the absence of 16S rRNA amplicons does not necessarily eliminate the possibility of a taxon from occurring in a metagenome. Extraction biases and/or PCR biases could potentially prevent a taxon from showing up in a 16S rRNA library.

After scaling the number of reads for the 16S amplicons to the number of reads in the metagenome, we defined a ratio of greater than 0.15 of 16S amplicons to metagenomic reads to indicate the presence of the taxa in the environment. This value allows for the potential occurrence of sequencing errors and or misidentifications.

Therefore:

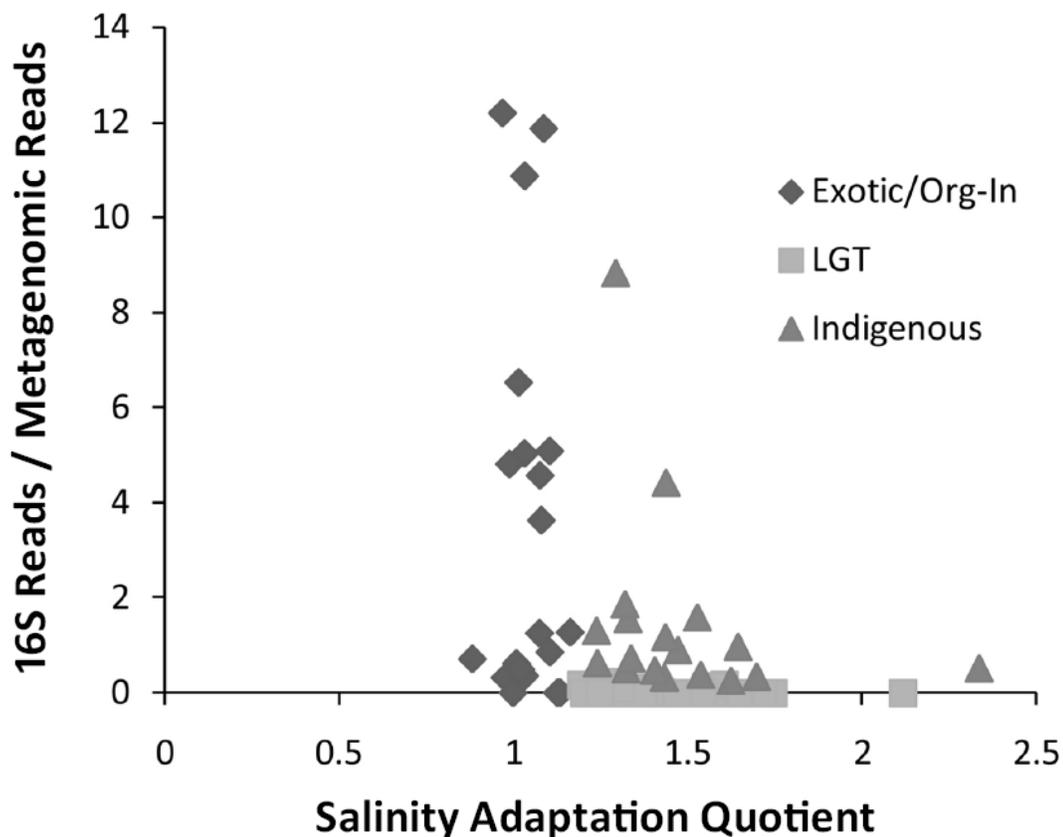
All salt-in category I taxa should be represented in both the metagenome and the 16S tags and should demonstrate adaptation to hypersalinity.

All category II taxa and all salt-out associated taxa should be represented in both the metagenome and the 16S tags but should not demonstrate adaptation to hypersalinity.

All category III taxa participating in LGT with salt-in halophiles should only be present in the metagenome, not in the 16S tags and should demonstrate evidence of adaptation to hypersalinity.

Figure 4 depicts a plot of the ratio of bacterial 16S reads versus bacterial metagenomic reads against the SAQ values. The analysis separates the chart into three regions. The putative category II taxa and salt-out taxa plot vertically along an SAQ of approximately one. The putative category III salt-in taxa plot horizontally along the x-axis and the putative category I salt-in taxa plot in a cloud between them. Amongst the category III taxa we observe significant LGT between the haloarchaea and both the halophilic Bacteria of the genus *Salinibacter* and the thermophilic Bacteria of the genus *Thermotoga*. Meanwhile included within the category I taxa are 10 bacterial lineages that are either not commonly associated with extreme hypersalinity and/or found in the Dead Sea (Table B-4).

As a case sample we chose to investigate the purported occurrences of lateral gene transfer involving *Thermotoga* related species. In total there were 105 reads that were assigned by MEGAN to *Thermotoga* related species. Of these 105 reads, 26 were assigned to an AAA family ATPase (YP\_001245352.1, Tpet\_1776) and 57 were assigned

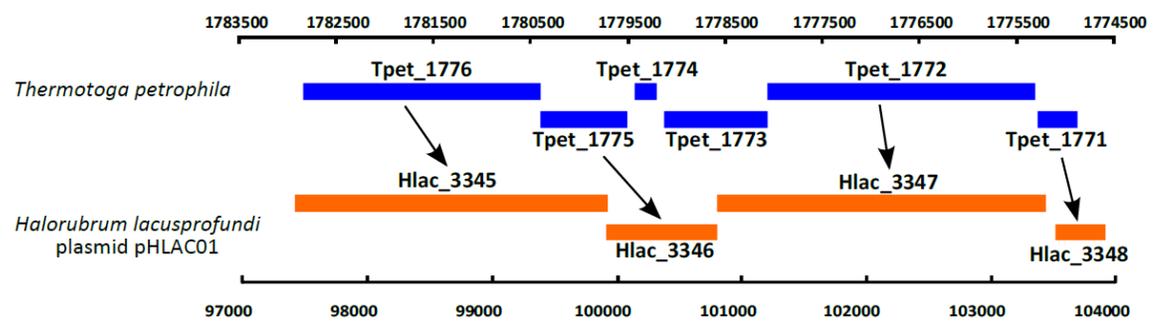
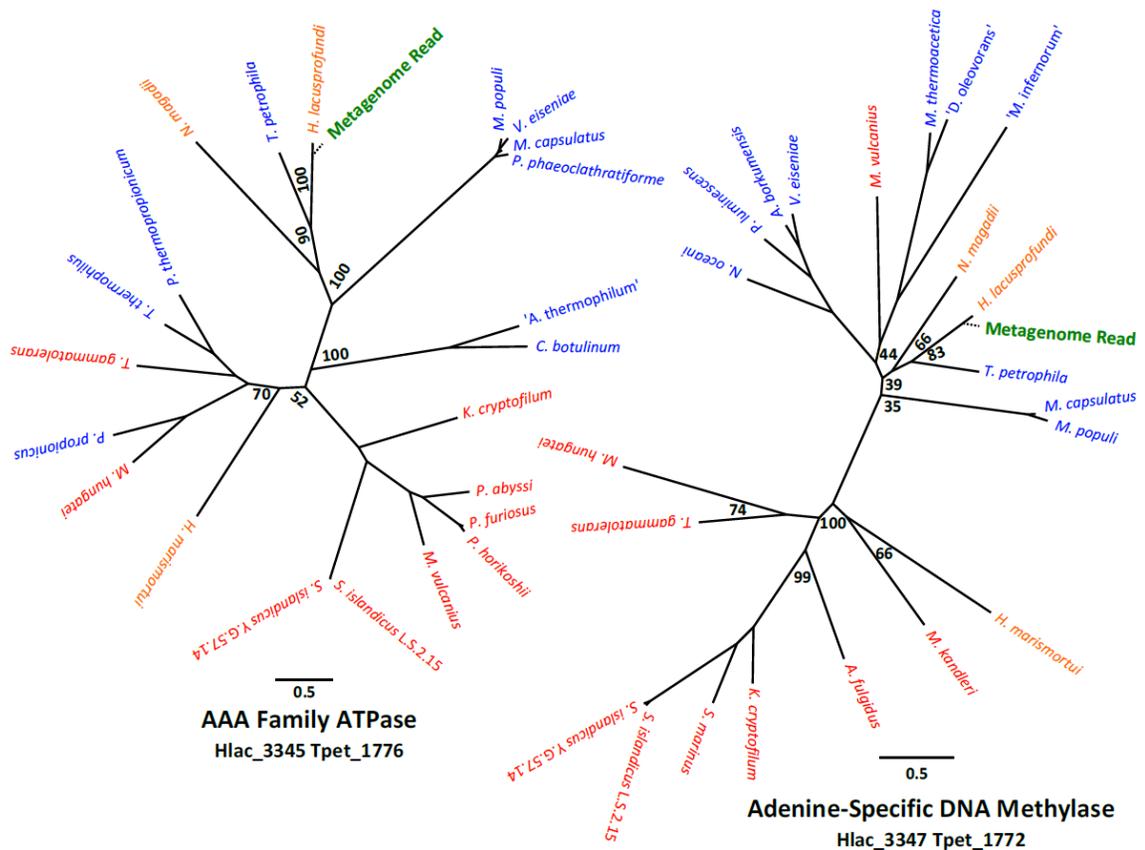


**Figure 3-4.** Plot of the observed ratio of 16s rRNA amplicon reads to metagenomic reads versus the calculated salinity adaptation quotient (SAQ) for bacterial taxonomic groups found in the 2007 Dead Sea datasets. Taxonomic groups distributed along the x-axis are those found only in the metagenomic dataset indicating that they are represented by reads of protein-encoding genes that have been transferred into indigenous Dead Sea taxa. Taxonomic groups plotting vertically along an SAQ of about 1, indicating a lack of adaptation to salinity, are interpreted to be reads from either taxa foreign to the Dead Sea or indigenous salt-out taxa. Taxonomic groups represented by 16s rRNA amplicons and showing an elevated SAQ are interpreted to be microorganisms indigenous in the modern Dead Sea.

to an adenine-specific DNA methylase (YP\_001245348.1, Tpet\_1772), both from the species *Thermotoga petrophila*. The very fact that these two genes are so overrepresented in the metagenomic dataset is another indication that these reads did not originate from *T. petrophila*. It is more likely that these reads belong to a species composing a significant proportion of the Dead Sea microbiome that received or donated these genes through lateral gene transfer events.

Consequently it is not surprising that in the year since the initial metagenomic analysis, additional halophiles have been added to the databases and the majority of the reads originally assigned to *T. petrophila* are now assigned to a haloarchaeal species, *Halorubrum lacusprofundi*. *T. petrophila* now generates the second strongest BLAST hits. It thus appears that a lateral gene transfer event has taken place between the thermophilic bacterium *T. petrophila* and the halophilic archaeon *H. lacusprofundi*. This also indicates that as more genomes are added to the databases, including the recipients of LGT events, phylogenetic based lateral gene finding methods, such as the one described above, will need to be adapted accordingly.

To further test our hypothesis that a LGT event has occurred between *H. lacusprofundi* and *T. petrophila* we generated phylogenetic trees for both the AAA family ATPase and the adenine-specific DNA methylase from *H. lacusprofundi*. We included the ten most homologous orthologs to these genes found in the KEGG database, from both the archaeal and bacterial domains. We also included the sequences from the haloarchaea *Natrialba magadii*, as they appeared highly



**Figure 3-5.** Trees depicting the phylogeny of an AAA Family ATPase and an Adenine-Specific DNA Methylase with relevant bootstrap values included. Bacterial species (*Alcanivorax borkumensis*, “*Anaerocellum thermophilum*” (*Caldicellulosiruptor bescii*), *Clostridium botulinum*, “*Desulfococcus oleovorans*”, “*Methylacidiphilum infernorum*”, *Methylobacterium populi*, *Methylococcus capsulatus*, *Moorella thermoacetica*, *Nitrosococcus oceani*, *Pelobacter propionicus*, *Pelodictyon phaeoclathratiforme* (*Chlorobium clathratiforme*), *Pelotomaculum thermopropionicum*, *Photorhabdus luminescens*, *Thermotoga petrophila*, *Thermus thermophilus*, and *Verminephrobacter eiseniae*) are given in blue, haloarchaeal species (*Haloarcula marismortui*, *Halorubrum lacusprofundi*, and *Natrialba magadii*) in orange, other archaeal species (*Archaeoglobus fulgidus*, *Candidatus Korarchaeum cryptofilum*, *Methanocaldococcus vulcanius*, *Methanopyrus kandleri*, *Methanospirillum hungatei*, *Pyrococcus abyssi*, *Pyrococcus furiosus*, *Pyrococcus horikoshii*, *Staphylothermus marinus*, *Sulfolobus islandicus* L.S.2.15, *Sulfolobus islandicus* Y.G.57.14, and *Thermococcus gammatolerans*) in red, and a representative metagenomic read is provided in green. Along the bottom is the region encoding both genes in the thermophilic bacterium *Thermotoga petrophila* and the halophilic archaeon *Halorubrum lacusprofundi*. Orthologous genes are connected by arrows.

homologous to *H. lacusprofundi*, but were not included in the KEGG database (Figure 5). In both trees *T. petrophila* and *H. lacusprofundi* are tightly linked. Additionally in both instances the genes appear to fall in clades included in the bacterial domain, indicating that genetic material was transferred from *T. petrophila* to *H. lacusprofundi*.

Finally, we investigated the location of both genes in the genomes of *H. lacusprofundi* and *T. petrophila*. In both species the genes are located almost adjacent to each other (Figure 5). In *H. lacusprofundi*, there is one small intermediate gene (YP\_002567463, Hlac\_3346) encoding a PglZ domain protein. In *T. petrophila*, there are three small intermediate genes, one of which (YP\_001245351.1, Tpet\_1775) is most homologous to Hlac\_3346. Downstream of Hlac\_3347 there is another small hypothetical gene (YP\_002567465, Hlac\_3348) which is also most homologous to the corresponding gene in *T. petrophila*. Thus, in total there exists a cassette of four genes in *H. lacusprofundi* that are both adjacent to one another and are most homologous to four genes in *T. petrophila* confirming the occurrence of a inter-domain LGT event.

## Discussion

While the acidic enrichment of proteins has been well documented within individual halophilic species, this phenomenon was first demonstrated on an environmental scale by Kunin *et al.* In their work, the environmental pressures imparted by a salinity of approximately 90 g/l appeared to have caused widespread interspecific convergent evolution toward proteins with increased proportions of aspartic acid. Here we have investigated the amino acid coding bias within a number of

environments. Included within our sample sites are environments displaying a wide range of salinities, from non-saline to over 300 g/l. Our analysis demonstrates that the amino acid coding pattern within an environment is sufficient to distinguish between saline and non-saline environments and can potentially offer finer levels of differentiation. We, therefore, propose the use of the amino acid coding profile as a summary statistic of an environment.

The distinction between saline and non-saline environments is primarily caused by aspartic acid and lysine with other amino acids such as glutamic acid and cysteine contributing as well. At the same time, there is a strong association of increased GC content and hypersalinity. It is therefore important to differentiate between amino acids actually coevolving with salinity and amino acids such as arginine and tryptophan which appear to be influenced largely by nucleotide mutational bias. Furthermore, we identified a strong linear relationship between the salinity of an environment and the ratio of acidic to basic amino acids encoded within its metagenome. This relationship suggests the use of the acidic to basic amino acid coding ratio as a potential salinity proxy for environments with moderate to high salinities (>100‰). Assuming the acid residue bias of genes is preserved in expression, the salinity of a paleoenvironment could potentially be preserved in fossil DNA or fossil amino acids.

The amino acid profiles of the metagenome should mirror the amino acid profiles of the dominant organisms. The encoded amino acid profile of the modern Dead Sea, however, does not match the profiles of the hyperhalophilic organisms sequenced thus far. The implication is that the halophilic isolates currently available are

not the dominant microorganisms inhabiting the Dead Sea. This assertion is corroborated by our 16S Dead Sea amplicons.

Finally, with only two distinct types of sequence data, we can differentiate members of the indigenous population from both probable foreign organisms and LGT events. While both the indigenous salt-in population and LGT events to salt-in halophiles should display a molecular adaptation to salinity, only the indigenous population and the foreign organisms should be represented in 16S clone libraries. We utilized this technique to identify the exchange of a roughly 7 kilobase region of DNA between the bacterial species *T. petrophila* and the haloarchaeal species *H. lacusprofundi*. The occurrence of such an inter-domain transfer of genetic material involving *Thermotoga* related species is not without precedent. Nelson *et al.* (Nelson *et al.*, 1999) demonstrated numerous such events between *Thermotoga maritima* and thermophilic Archaea, including the transfer of 15 regions greater than 4 kilobases in size. The example of LGT presented here is different in that aside from being an inter-domain transfer of genetic material, it also represents the transfer of genetic material between distinct extreme environments.

We can also use the method presented above to make educated guesses concerning the identities of rare but functioning organisms within a hypersaline ecosystem. We can then use the information provided as leads for molecular approaches that may confirm the existence of these species. Molecular convergence on an environmental scale has, to date, only been documented in hypersaline environments. However, with the indications of similar amino acid biases in acidophiles

and thermophiles (Goodarzi *et al.*, 2008; Haney *et al.*, 1999), these approaches may be applicable to other extreme environments as well.

## **Experimental procedures**

### *DNA extraction*

The 1992 sample consisted of roughly 10 ml of a bright red solution obtained via centrifugation (15 min, 8,000 x g) of approximately 5 liters of brine. It remained frozen at -20°C for five years, and then further at -80° until being shipped on dry ice in the spring of 2007. The sample was then stored at -80°C until analysis. DNA was extracted from 3ml of sample using the MoBio UltraClean Microbial DNA kit (MoBio Laboratories). The 2007 sample was collected and processed by the lab of Oded Béjà (the Technion, Haifa, Israel) according to the protocol described in Bodaker *et al.* (Bodaker *et al.*, 2009a). A sufficient quantity of DNA for pyrosequencing was shipped in an agarose plug which was subsequently digested using the Gelase enzyme (Epicentre).

### *Whole genome amplification and preparation for pyrosequencing*

Approximately 40 ng of the 1992 DNA was subjected to whole-genome amplification using the REPLI-g Mini kit (Qiagen). The V9 hypervariable region of the 16S rRNA gene was amplified using the primer set 5'-gcctccctcgcgcatcag-TGYACACACCGCCCGTC-3' and 5'-gccttgccagcccgctcag-ACGGNWACCTTGTTACGACTT-3' adapted from primers 1407f and 1492r (Lane, 1991) and augmented with 454 Life Science's A or B sequencing adapters. The amplification mix consisted of 1.25 units of PfuUltra Hotstart DNA polymerase (Stratagene), 1X PfuUltra HF reaction buffer, a 200

$\mu\text{M}$  concentration of dNTP's, a 0.2  $\mu\text{M}$  of each primer, and approximately 5 ng of genomic DNA. The PCR reaction was run according to the conditions of Sogin *et al.* with 27 cycles (Sogin *et al.*, 2006). The resulting product was gel purified using a 1% low melting point agarose gel. Samples were then sequenced on a GS FLX sequencer (454 Life Sciences).

#### *Sequencing and sequence analysis*

Both genomic samples were sequenced on a half pico-titer plate. The V9 amplicons were sequenced with a number of additional amplicon sets on a separate half pico-titer plate. All amplicons with errors in the primer sequence, bad calls within the sequence, or of highly unusual length were discarded. BLASTX analyses against the nr database were performed on both genomic data sets. For each top hit with an e-value of  $10^{-5}$  or less the portion of the read matching a homologous protein in the database was extracted and subjected to the analysis described in this paper. Additionally, approximately 6.5% of the hits showed evidence of a frameshift within the read. For these sequences as well, the portions of the reads matching a homologous protein in the database was extracted independently and the intervening amino acids were discarded. A BLASTN analysis was performed on the V9 amplicons against both the RDP 16S database and a collection of the genomes of all fully sequenced microbes augmented with the 16S genes of members of taxa hit in the genomic analysis.

#### *Bioinformatic analysis*

The cluster analysis and redundancy analysis were performed using the statistical package R (Ihaka & Gentleman, 1996). The genomic BLAST hits were assigned

to taxa using the MEGAN program, and a top percentage of 10%. The amplicons were assigned to taxa according to their respective best hit as matched by BLAST. Sequences were aligned using the MUSCLE program (Edgar, 2004) with default parameters. The alignments were then filtered to only include locations for which the majority of species contained codons and for which at least three species encoded the same amino acid. Trees were then constructed using RAxML (Stamatakis *et al.*, 2005), and 100 bootstrap iterations were performed. Finally the metagenomic reads were inserted using the maximum parsimony feature of the ARB program (Ludwig *et al.*, 2004). All other analyses of the 2007 Dead Sea data set were performed using home-written scripts in Perl and/or Python. The scripts are available upon request. The 1992 Dead Sea data set was not similarly analyzed because the reduced quantity of hits did not enable individual non-halophilic taxa to be assessed.

### **Acknowledgments**

We thank Lynn Tomsho, Stephen Schuster, Tom Canich, Mark Patzkowsky, Oded Béjà, Idan Bodaker and the late Moti Gonen for assistance. This work was supported by the National Aeronautics and Space Administration (NASA) Astrobiology Institute (NAI) under NASA–Ames Cooperative Agreement NNA04CC06A (C.H.H. and S.T.F.) and in particular through a NAI DDF award. The GS20 facility at the Pennsylvania State University Center for Genome Analysis is funded, in part, by a grant from the Pennsylvania Department of Health using Tobacco Settlement Funds appropriated by the legislature.

## References

- Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997).** Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**, 3389-3402.
- Biddle, J. F., Fitz-Gibbon, S., Schuster, S. C., Brenchley, J. E. & House, C. H. (2008).** Metagenomic signatures of the Peru Margin subseafloor biosphere show a genetically distinct environment. *Proceedings of the National Academy of Sciences of the United States of America*.
- Bodaker, I., Béjà, O., Feingersch, R., Rosenber, M., Oren, A., Hindiyeh, M. Y. & Malkawi, H. I. (2009).** Archaeal diversity in the Dead Sea: Microbial survival under increasingly harsh conditions. In *Saline lakes around the world: Unique systems with unique values*, pp. 137-143. Edited by A. Oren, D. Naftz, P. Palacios & W. A. Wurtsbaugh. Logan, Utah: S. J. and Jessie E. Quinney Natural Resources Research Library.
- Bolhuis, H., Palm, P., Wende, A., Falb, M., Rampp, M., Rodriguez-Valera, F., Pfeiffer, F. & Oesterhelt, D. (2006).** The genome of the square archaeon *Haloquadratum walsbyi*: Life at the limits of water activity. *BMC Genomics* **7**, 169.
- Cole, J. R., Chai, B., Farris, R. J. & other authors (2007).** The ribosomal database project (RDP-II): introducing myRDP space and quality controlled public data. *Nucleic Acids Research* **35**, D169-D172.
- Dean, F. B., Nelson, J. R., Giesler, T. L. & Lasken, R. S. (2001).** Rapid amplification of plasmid and phage DNA using Phi29 DNA polymerase and multiply-primed rolling circle amplification. *Genome Research* **11**, 1095-1099.
- Edgar, R. C. (2004).** MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**, 1792-1797.
- Gavrieli, I., Beyth, M. & Yechieli, Y. (1999).** The Dead Sea - A terminal lake in the Dead Sea rift: A short overview. In *Microbiology and biogeochemistry of hypersaline environments*, pp. 121-127. Edited by A. Oren. Boca Raton, Florida: CRC Press.
- Gill, S. R., Pop, M., DeBoy, R. T. & other authors (2006).** Metagenomic analysis of the human distal gut microbiome. *Science* **312**, 1355-1359.
- Goodarzi, H., Torabi, N., Najafabadi, H. S. & Archetti, M. (2008).** Amino acid and codon usage profiles: Adaptive changes in the frequency of amino acids and codons. *Gene* **407**, 30-41.
- Haney, P. J., Badger, J. H., Buldak, G. L., Reich, C. I., Woese, C. R. & Olsen, G. J. (1999).** Thermal adaptation analyzed by comparison of protein sequences from mesophilic and extremely thermophilic *Methanococcus* species. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 3578-3583.
- Huson, D. H., Auch, A. F., Qi, J. & Schuster, S. C. (2007).** MEGAN analysis of metagenomic data. *Genome Research* **17**, 377-386.

- Ihaka, R. & Gentleman, R. (1996).** R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics* **5**, 299-314.
- Kennedy, S. P., Ng, W. V., Salzberg, S. L., 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 Research* **11**, 1641-1650.
- Kunin, V., Raes, J., Harris, J. K. & other authors (2008).** Millimeter-scale genetic gradients and community-level molecular convergence in a hypersaline microbial mat. *Molecular Systems Biology* **4**.
- Lane, D. J. (1991).** 16S/23S rRNA sequencing. In *Nucleic acid techniques in bacterial systematics*, pp. 115-175. Edited by E. Stackebrandt & M. Goodfellow. New York, New York: John Wiley and Sons.
- Lanyi, J. K. (1974).** Salt-dependent properties of proteins from extremely halophilic bacteria. *Bacteriology Reviews* **28**, 272-290.
- Legault, B., Lopez-Lopez, A., Alba-Casado, J., Doolittle, W. F., Bolhuis, H., Rodriguez-Valera, F. & Papke, R. T. (2006).** Environmental genomics of "*Haloquadratum walsbyi*" in a saltern crystallizer indicates a large pool of accessory genes in an otherwise coherent species. *BMC Genomics* **7**, 171.
- Ley, R. E., Harris, J. K., Wilcox, J. & other authors (2006).** Unexpected diversity and complexity of the Guerrero Negro hypersaline microbial mat. *Applied and Environmental Microbiology* **72**, 3685-3695.
- Ludwig, W., Strunk, O., Westram, R. & other authors (2004).** ARB:A software environment for sequence data. *Nucleic Acids Research* **32**, 1363-1371.
- Madern, D., Ebel, C. & Zaccai, G. (2000).** Halophilic adaptation of enzymes. *Extremophiles* **4**, 91-98.
- Nelson, K. E., Clayton, R. A., Gill, S. R. & other authors (1999).** Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of *Thermotoga maritima*. *Nature* **399**, 323-329.
- Ng, W. V., Kennedy, S. P., Mahairas, G. G. & other authors (2000).** Genome sequence of *Halobacterium* species NRC-1. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 12176-12181.
- Oren, A. (1986).** Intracellular salt concentrations of the anaerobic halophilic eubacteria *Haloanaerobium praevalens* and *Halobacteroides halobius*. *Canadian Journal of Microbiology* **32**, 4-9.
- Oren, A. & Gurevich, P. (1995).** Dynamics of a bloom of halophilic archaea in the Dead Sea. *Hydrobiologia* **315**, 149-158.

- Oren, A. (1999).** Bioenergetic aspects of halophilism. *Microbiology and Molecular Biology Reviews* **63**, 334-348.
- Oren, A. (2002).** Properties of halophilic proteins. In *Halophilic microorganisms and their environments*, pp. 233-278. Edited by A. Oren. Dodrecht, The Netherlands: Kluwer Academic.
- Oren, A., Larimer, F., Richardson, P., Lapidus, A. & Csonka, L. N. (2005).** How to be moderately halophilic with broad salt tolerance: clues from the genome of *Chromohalobacter salexigens*. *Extremophiles* **9**, 275-279.
- Paul, S., Bag, S., Das, S., Harvill, E. & Dutta, C. (2008).** Molecular signature of hypersaline adaptation: Insights from genome and proteome composition of halophilic prokaryotes. *Genome Biology* **9**, R70.
- Reistad, R. (1970).** On the composition and nature of the bulk protein of extremely halophilic bacteria. *Archives of Microbiology* **71**, 353-360.
- Rivera, M. C., Jain, R., Moore, J. E. & Lake, J. A. (1998).** Genomic evidence for two functionally distinct gene classes. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 6239-6244.
- Robertson, C. E., Spear, J. R., Harris, J. K. & Pace, N. R. (2009).** Diversity and stratification of archaea in a hypersaline microbial mat. *Applied and Environmental Microbiology* **75**, 1801-1810.
- Santos, H. & Da Costa, M. S. (2002).** Compatible solutes of organisms that live in hot saline environments. *Environmental Microbiology* **4**, 501-509.
- Sogin, M. L., Morrison, H. G., Huber, J. A., Welch, D. M., Huse, S. M., Neal, P. R., Arrieta, J. M. & Herndl, G. J. (2006).** Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proceedings of the National Academy of Sciences of the United States of America* **103**, 12115-12120.
- Soppa, J. (2006).** From genomes to function: Haloarchaea as model organisms. *Microbiology* **152**, 585-590.
- Spear, J. R., Ley, R. E., Berger, A. B. & Pace, N. R. (2003).** Complexity in natural microbial ecosystems: The Guerrero Negro experience. *The Biological Bulletin* **204**, 168-173.
- Stamatakis, A., Ludqig, T. & Meler, H. (2005).** Raxml-III: A fast program for maximum likelihood-based inference of large phylogenetic trees. *Bioinformatics* **21**, 456-463.
- Tringe, S. G., von Mering, C., Kobayashi, A. & other authors (2005).** Comparative metagenomics of microbial communities. *Science* **308**, 554-557.
- Turnbaugh, P. J., Ley, R. E., Mahowald, M. A., Magrini, V., Mardis, E. R. & Gordon, J. I. (2006).** An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**, 1027-1131.

**Tyson, G. W., Chapman, J., Hugenholtz, P. & other authors (2004).** Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **428**, 37-43.

**Valenzuela-Encinas, C., Neria-González, I., Alcántara-Hernández, R., Enríquez-Aragón, J., Estrada-Alvarado, I., Hernández-Rodríguez, C., Dendooven, L. & Marsch, R. (2008).** Phylogenetic analysis of the archaeal community in an alkaline-saline soil of the former lake Texcoco (Mexico). *Extremophiles* **12**, 247-254.

**Venter, J. C., Remington, K., Heidelberg, J. F. & other authors (2004).** Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**, 66-74.

**Volcani, B. (1944).** The microorganisms of the Dead Sea. In *Papers collected to commemorate the 70th anniversary of Dr Chaim Weizmann*, pp. 71-85. Rehovoth: Daniel Sieff Research Institute.

**Ward, J. H. (1963).** Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association* **58**, 236-244.

## Chapter 4

### Differences in lateral gene transfer in hypersaline versus thermal environments

Matthew E. Rhodes<sup>1\*</sup>, John R. Spear<sup>2</sup>, Aharon Oren<sup>3</sup>, Chris H. House<sup>1</sup>

<sup>1</sup>Penn State Astrobiology Research Center and Department of Geosciences, The  
Pennsylvania State University, University Park, PA 16802, USA

<sup>2</sup>Division of Environmental Science and Engineering, Colorado School of Mines, Golden,  
CO 80401, USA

<sup>3</sup>The Institute of Life Sciences and the Moshe Shilo Minerva Center for Marine  
Biogeochemistry, The Hebrew Institute of Jerusalem, Jerusalem 91904, Israel

\*To whom correspondence should be addressed

Published in *BMC Evolutionary Biology* 8 July, 2011

#### **Abstract**

##### *Background*

The role of lateral gene transfer (LGT) in the evolution of microorganisms is only beginning to be understood. While most LGT events occur between closely related

individuals, inter-phylum and inter-domain LGT events are not uncommon. These distant transfer events offer potentially greater fitness advantages and it is for this reason that these “long distance” LGT events may have significantly impacted the evolution of microbes. One mechanism driving distant LGT events is microbial transformation. Theoretically, transformative events can occur between any two species provided that the DNA of one enters the habitat of the other. Two categories of microorganisms that are well-known for LGT are the thermophiles and halophiles.

### *Results*

We identified potential inter-class LGT events into both a thermophilic class of Archaea (Thermoprotei) and a halophilic class of Archaea (Halobacteria). We then categorized these LGT genes as originating in thermophiles and halophiles respectively. While more than 68% of transfer events into Thermoprotei taxa originated in other thermophiles, less than 11% of transfer events into Halobacteria taxa originated in other halophiles.

### *Conclusions*

Our results suggest that there is a fundamental difference between LGT in thermophiles and halophiles. We theorize that the difference lies in the different natures of the environments. While DNA degrades rapidly in thermal environments due to temperature-driven denaturation, hypersaline environments are adept at preserving DNA. Furthermore, most hypersaline environments, as topographical minima, are natural collectors of cellular debris. Thus halophiles would in theory be exposed to a greater diversity and quantity of extracellular DNA than thermophiles.

## Background

The extent and role of lateral gene transfer (LGT) as a force of evolution has only recently become appreciated. Only in the past couple decades has the sequencing of genomes such as that of *Thermotoga maritima* thrust LGT into the limelight (Nelson *et al.*, 1999). The original estimates suggested that over 20% of *Thermotoga maritima*'s genome was the result of long distance LGT events. This and numerous other results have led to a potential reevaluation of the tree of life and the notion of a Last Universal Common Ancestor (Doolittle, 1999; Doolittle & Bapteste, 2007).

LGT itself is driven by a variety of mechanisms including conjugation, or the transfer of genetic material via direct contact (Tatum & Lederberg, 1947), transduction, or the viral mediated transfer of DNA (Zinder & Lederberg, 1952), and transformation, or the uptake and incorporation of naked DNA from an environment (Avery *et al.*, 1944). Conjugative transfers necessitate the cohabitation of the participants and are generally thought to require the participants to be closely related, although inter-class conjugative events have been shown to occur between members of the Proteobacteria (Bathe *et al.*, 2004). Similarly, while most transductive phages and phage like objects are restricted to infecting members of the same species, phages that infect across classes are known to exist (Jensen *et al.*, 1998). Finally, transformative events present no definitive phylogenetic barrier. Presumably a microorganism can take up virtually any DNA present in its immediate environment. However the probability of a harvested piece of assembled DNA being incorporated into a genome is partially dependent on

sequence similarity between the donor and host DNA and is therefore much greater for closely related individuals (Thomas & Nielsen, 2005). Consequently the vast majority of LGT events are thought to occur between closely related species. Nevertheless inter-phylum and inter-domain transfer events can and do occur (Deppenmeier *et al.*, 2002; Nelson *et al.*, 1999). These “long range” transfer events are partially the result of transformation events and, while relatively rare, offer a potentially significant evolutionary mechanism.

Species within the domain Archaea and a variety of bacterial phyla are known to be capable of transformation (Thomas & Nielsen, 2005). Preliminary estimates indicate that approximately 1% of bacterial species are naturally able to take up DNA (Jonas *et al.*, 2001). The frequency of a transformation event is dependent on a number of factors, including but not limited to, the quantity of DNA in an environment, the rate of DNA degradation in an environment, the frequency of DNA uptake by the recipients, the likelihood of incorporation into a genome, and natural selection on the incorporated DNA (Thomas & Nielsen, 2005). These factors in turn are highly specific to individual species and environments. Here, we have used genomic and metagenomic techniques to test mechanisms of LGT into two phylogenetically coherent clades from different extreme environments.

#### *Halophiles and thermophiles*

Extremophiles, and in particular thermophiles and halophiles, are well-known for participating in rampant LGT (Aravind *et al.*, 1998; Mongodin *et al.*, 2005; Nelson *et al.*, 1999). It is theorized that the very nature of their extreme environments encourages

the exchange of genetic material. Essentially any advantages gained in overcoming the environmental challenges are highly sought after, rapidly exchanged, and potentially accelerate the rate of evolution. In this regard, thermal and saline environments are quite similar. Both offer considerable environmental obstacles to be overcome before life can persist.

The crenarchaeal class, Thermoprotei, consists solely of obligate thermophiles. Similarly, the euryarchaeal class, Halobacteria, consists solely of obligate halophiles. Any LGT event into a member of the Thermoprotei or the Halobacteria necessarily occurred in either a thermal or hypersaline environment respectively. Thus, together these two distinct archaeal lineages offer a naturally occurring evolutionary experiment by which we can study “long range,” inter-class and more distant, LGT events in these specific environments.

However, with regard to transformation, there are some significant differences between these two types of extreme environments. For example, high temperatures rapidly degrade unprotected DNA, both intracellularly and extracellularly, thereby preferentially preserving more thermally protected DNA. Fittingly, certain proteins, enzymes, and specifically salts, such as  $MgCl_2$  and  $KCl$ , can help protect DNA from thermal degradation (Marguet & Forterre, 1994; Tehei *et al.*, 2002). In contrast, high salinities can preserve even naked DNA for exceptionally long periods of time. Borin *et al.* demonstrated that the preservation of naked DNA in deep-sea anoxic hypersaline brines did not depend on the species of origin and that DNA was often capable of participating in natural transformation after weeks of exposure (Borin *et al.*, 2008).

Another fundamental difference between thermal and saline environments is that saline environments almost as a rule are topographical minima. Saline environments such as the Dead Sea are therefore natural collectors of cellular debris and may therefore contain the DNA of a diversity of contaminant species (Rhodes *et al.*). Thermal environments, however, may or may not be topographical minima and therefore may or may not be natural collectors of cellular debris. Environmental factors would therefore serve to increase the diversity of extracellular DNA in a typical saline environment relative to the typical thermal environment.

Analyses of halophiles have revealed a number of genomic characteristics common to halophilicity. Foremost amongst these characteristics is a propensity for GC richness possibly to protect against thymine dimerization due to the intense UV radiation often associated with hypersaline environments (Kennedy *et al.*, 2001). The preference for GC nucleotides is present in virtually all known lineages of halophiles and is nearly ubiquitous amongst the Halobacteria, *Haloquadratum walsbyi* being the sole known exception. However interspersed among the GC rich genomes of the Halobacteria are many GC poor regions (Kennedy *et al.*, 2001). The varied composition of the Halobacteria genomes combined with the diversity of metabolic functions and the frequent occurrence of insertion sequence elements suggested to Kennedy *et al.* that the Halobacteria are particularly adept at procuring novel genes and metabolic pathways (Kennedy *et al.*, 2001).

Other DNA level propensities include an increased abundance of the dinucleotides 'CG', 'GA/TC', and 'AC/GT' and preferences for specific codons for the

amino acids arginine, cysteine, leucine, threonine, and valine, presumably for secondary and tertiary stability in protein folding (Paul *et al.*, 2008). Furthermore, halophiles have developed two distinct strategies to overcome the extreme salinities of their native environments. While the “salt-out” halophiles balance the osmotic pressure of their environments with intra-cellular organic solutes such as betaine, the “salt-in” halophiles use KCl. The presence of often multimolar concentrations of K<sup>+</sup> ions requires radical alterations of protein chemistry. These alterations in protein chemistry include an overall preference for amino acids with acidic residues relative to amino acids with basic residues (Lanyi, 1974). The halophiles of the class Halobacteria are all “salt-in” and they all demonstrate a bias toward amino acids with acidic residues, regardless of their nucleotide composition. Recent metagenomic studies have confirmed this trend on an environmental scale in a number of hypersaline environments (Rhodes *et al.*).

At a salinity of over 340 g/l, the modern surface waters of the Dead Sea represent one of the most saline naturally occurring bodies of water known to harbor life. When combined with a slightly acidic pH ( $\approx 6$ ), near toxic magnesium levels, (currently about 2.0 M Mg<sup>2+</sup>), and dominance of divalent cations over monovalent cations (Oren, 1999a), it becomes a truly unique and inhospitable ecosystem. Current cell counts are well below  $5 \times 10^5 \text{ mL}^{-1}$  (Oren, 2000). A number of species of the Halobacteria have been isolated from the Dead Sea, including *Haloarcula marismortui* (Oren *et al.*, 1990), *Haloferax volcanii* (Mullakhanbhai & Larsen, 1975), *Halorubrum sodomense* (Oren, 1983b), and *Halobaculum gomorrense* (Oren *et al.*, 1995b). However recent metagenomic studies have suggested that the dominant microorganism in the

modern Dead Sea is most closely related to a member of the neutrophilic, halophilic, euryarchaeal genus *Halobacterium* or the alkaliphilic, halophilic, euryarchaeal genus *Natronomonas* (Bodaker *et al.*, 2010; Rhodes *et al.*, 2010).

### *Identifying lateral gene transfer events*

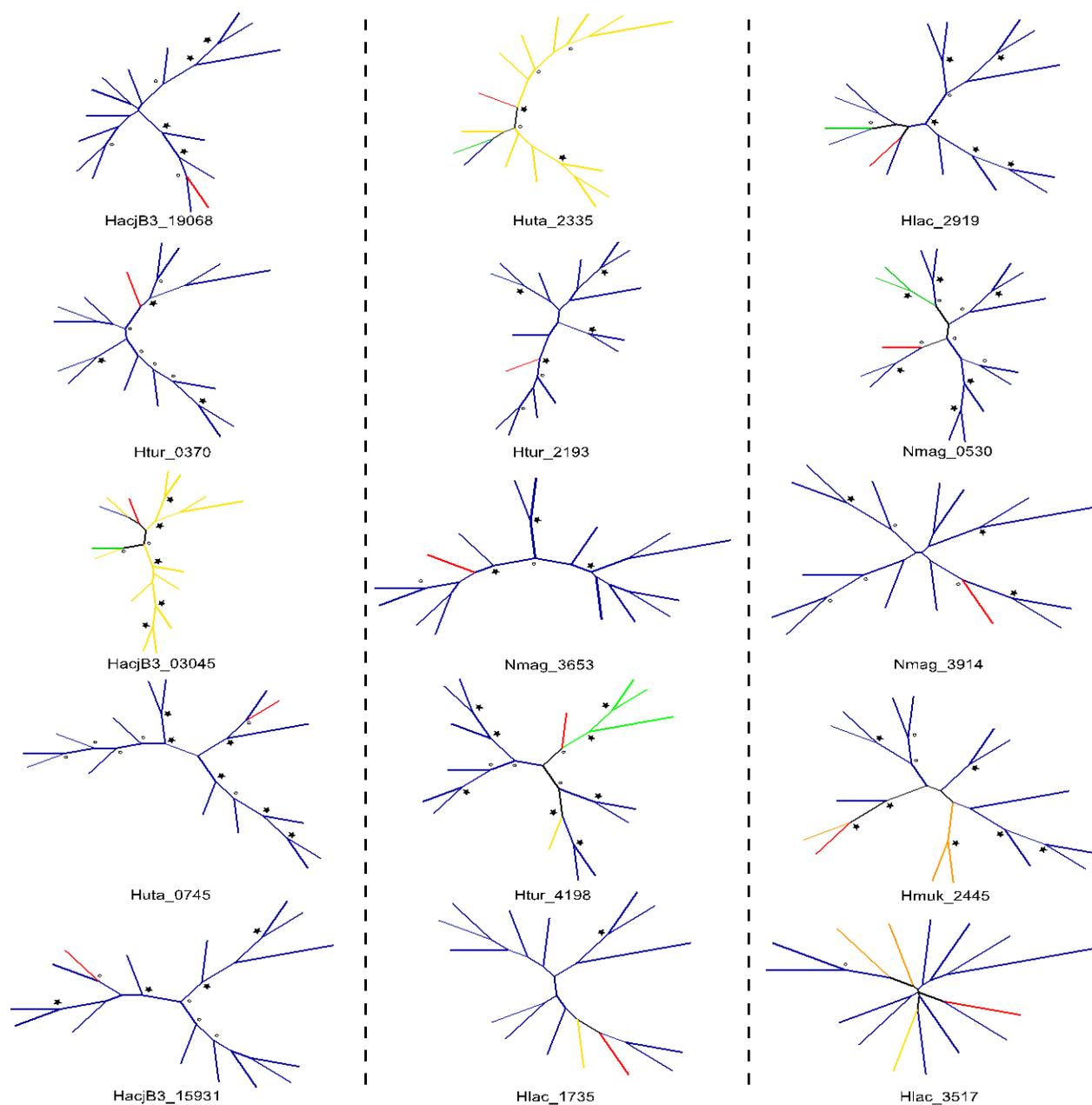
Putative LGT events are generally identified using two distinct methods: phylogenetic methods attempt to identify genes associated with LGT events by constructing and analyzing phylogenies in an effort to find genes that do not conform to the group's established taxonomy. Compositional methods, on the other hand, identify LGT events by searching for genes whose DNA or amino acid signatures do not match those of their host organism. The methods are essentially complementary, in that they use unrelated data to obtain similar conclusions. For this reason the two methods often identify entirely different classes of LGT events (Ragan *et al.*, 2006; Tamames & Moya, 2008). Here, in an attempt to investigate the drivers of LGT in thermal and hypersaline environments, we have employed a predominantly phylogenetic approach to identify putative LGT events involving a thermophilic class of Archaea, the Thermoprotei, and a halophilic class of Archaea, the Halobacteria. We then seek to confirm our results in a collection of environmental fosmids from the Dead Sea.

## **Results**

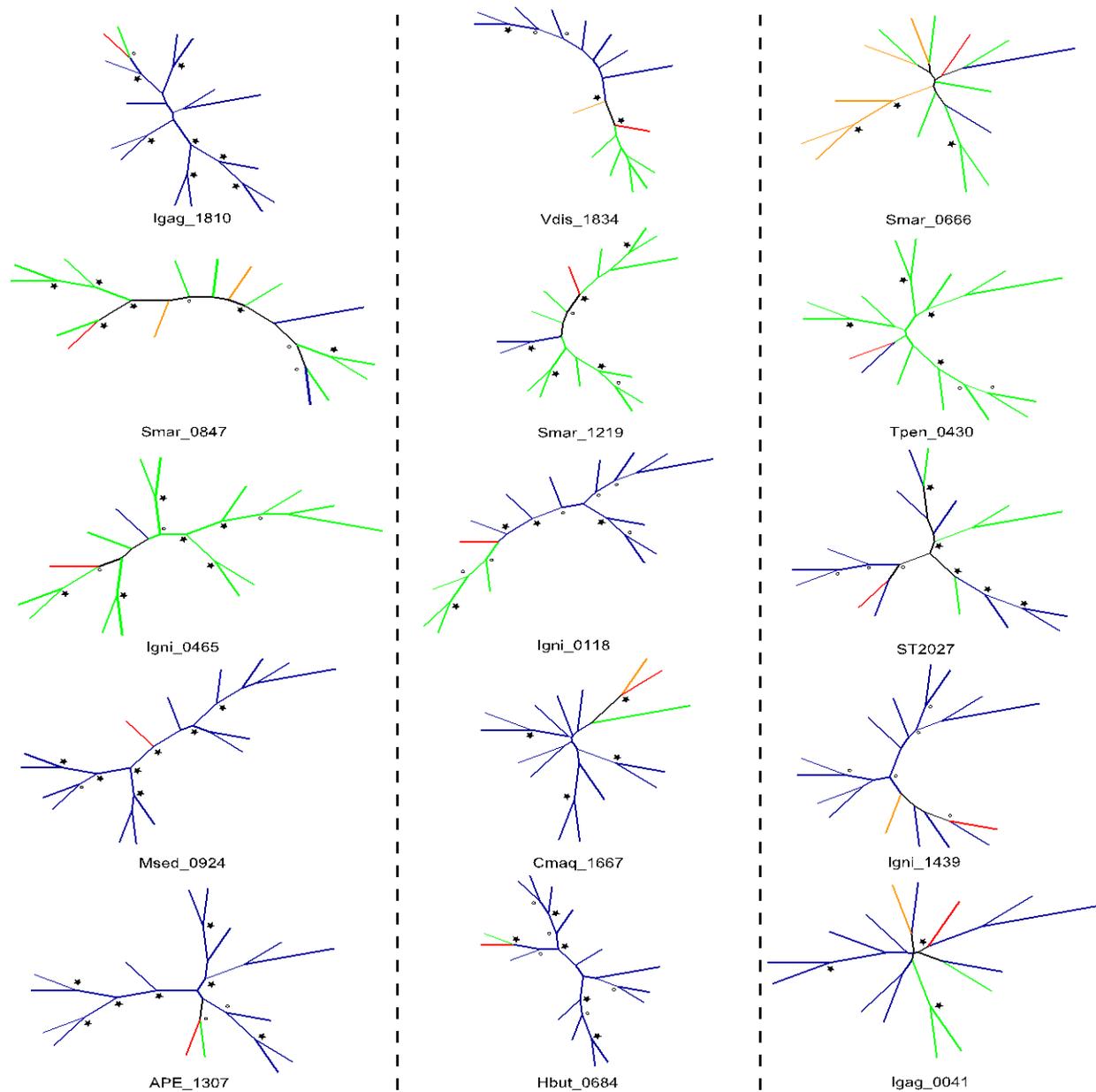
Genomes from all fully sequenced genera of the archaeal classes Thermoprotei (*Acidilobus*, *Aeropyrum*, *Caldivirga*, *Desulfurococcus*, *Hyperthermus*, *Ignicoccus*, *Ignisphaera*, *Metallosphaera*, *Pyrobaculum*, *Staphylothermus*, *Sulfolobus*, *Thermofilum*,

*Thermoproteus*, *Thermosphaera*, and *Vulcanisaeta*) and Halobacteria (*Halalkalicoccus*, *Haloarcula*, *Halobacterium*, *Haloferax*, *Halomicrobium*, *Haloquadratum*, *Halorhabdus*, *Halorubrum*, *Haloterrigena*, *Natrialba*, and *Natronomonas*), were obtained from the NCBI database in November of 2010. These genomes were then compared to the entire collection of fully sequenced microbes using the BLASTP program and default parameters (Altschul *et al.*, 1997). In cases where the normalized best BLAST score to members of its own class but not within its genus was less than 75% of the normalized best BLAST score to non-members of the Thermoprotei or Halobacteria respectively, the gene was flagged as a probable inter-class LGT event. Overall this method identified 1226 genes from Halobacteria and 1279 genes from Thermoprotei as “long distance” LGT events. To test for the possibility of bias in our downstream analyses associated with the 75% BLAST score cutoff, the procedure was repeated with cutoffs ranging from 90% to 50%.

We believe that our algorithm should preferentially identify LGT events into the Halobacteria and Thermoprotei. For the vast majority of LGT events identified the closest homologues were exclusively or almost exclusively from outside of the Halobacteria or Thermoprotei respectively, thereby suggesting that the transfer was into a member of the Halobacteria or Thermoprotei. Nevertheless, to check this we constructed phylogenetic trees for a representative sample of LGT genes from each class using the top homologues in the KEGG database (Figure 1, Figure 2, and Appendix C Supplemental Information 1) (Kanehisa & Goto, 2000). As expected, upon inspection, the vast majority of genes showed evidence of having been transferred into the



**Figure 4-1.** Depiction of 15 randomly selected trees of LGT genes for the Halobacteria. The leftmost column depicts genes from the upper third of BLAST scores, middle column from the middle third, and rightmost column from the bottom third. Target gene is depicted in red and named below the tree, other genes from the same class are shown in orange, genes from the same phylum but not the same class are shown in yellow, genes from the same domain but not the same phylum are shown in green, and genes from a different domain are shown in blue. Circles next to a node indicate a bootstrap value of greater than 50% and stars indicate bootstrap values of greater than 75%.

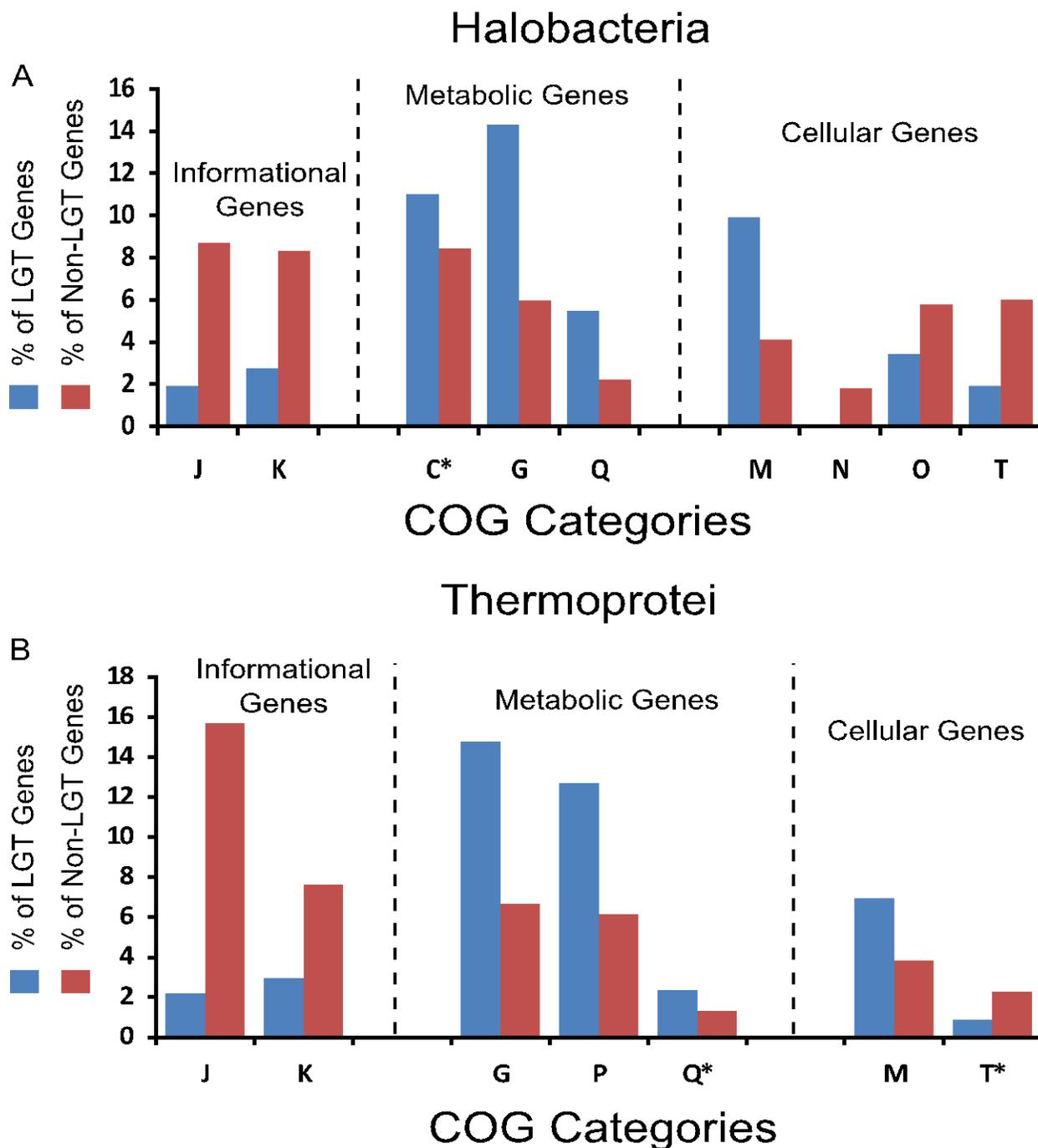


**Figure 4-2.** Depiction of 15 randomly selected trees of LGT genes for the Thermoprotei.

Halobacteria and Thermoprotei. For the remainder, the majority showed phylogenies too disordered to make an accurate statement.

*Assignment to Clusters of Orthologous Groups of proteins (COGs)*

Both the genes associated with long distance LGT events and the genes not associated with long distance LGT events were assigned to functional categories according to the classification of the Clusters of Orthologous Groups of proteins (COG) database (Figure 3) (Tatusov *et al.*, 2003). For both the Thermoprotei genomes and the Halobacteria genomes, “Information storage and processing” genes belonging to categories, J (Translation, ribosomal structure and biogenesis) and K (Transcription) are considerably underrepresented amongst genes presumed to have undergone LGT. In a similar vein, “Metabolic” genes belonging to categories C (Energy production and conversion), G (Carbohydrate transport and metabolism), P (Inorganic ion transport and metabolism), and Q (Secondary metabolites biosynthesis, transport and catabolism) are overrepresented amongst genes that have undergone LGT in either the Thermoprotei, the Halobacteria, or both. Genes involved in “Cellular processes and signaling” are inconclusive. While category M (Cell wall/membrane/envelope biogenesis) is overrepresented in genes associate with LGT, categories N (Cell motility), O (Posttranslational modification, protein turnover, chaperones), and T (Signal transduction mechanisms) are underrepresented among genes associated with LGT. As a whole, these results coincide well with the conclusions of Rivera *et al.* (Rivera *et al.*, 1998) that LGT should favor operational genes over informational genes. This is especially true for distantly related LGT events and may offer limited non-phylogenetic



**Figure 4-3.** Chart depicting the percentage of genes assigned to various COG categories for both regular genes and genes identified as LGT events into the (a) Halobacteria and (b) Thermoprotei. The COG categories shown demonstrated a statistically significant disparity between LGT genes and non-LGT genes beyond a 95% confidence interval. The categories depicted with an \* fell just below the 95% threshold.

evidence that we have in fact identified long distance LGT events (Zhaxybayeva *et al.*, 2006).

#### *Homologue taxonomy*

For the genes associated with LGT events, we then identified the species representing the closest homologue to the original archaeal gene as matched by BLAST. We categorized these donor species according to their halophilicity and thermophilicity (Figure 4a-f). As expected, the majority, 68%, of the donor species to the Thermoprotei demonstrate thermophilic character themselves (Figure 4a). Genes originating in thermophiles should be pre-adapted to thermal conditions and therefore should present fewer obstacles to incorporation into a Thermoprotei genome. In stark contrast however, the vast majority, >89%, of donor species to the Halobacteria, were not species with known halophilic character (Figure 4d). This appears to suggest that something other than pre-adaptation to a high salt environment is the determining factor in successful LGT events into Halobacteria and presumably other halophiles. It is worth noting that for both the Thermoprotei and the Halobacteria the proportion of genes identified as intra-environmental LGT events remained consistent regardless of the BLAST cutoff used in the LGT identification step (Table C-1).

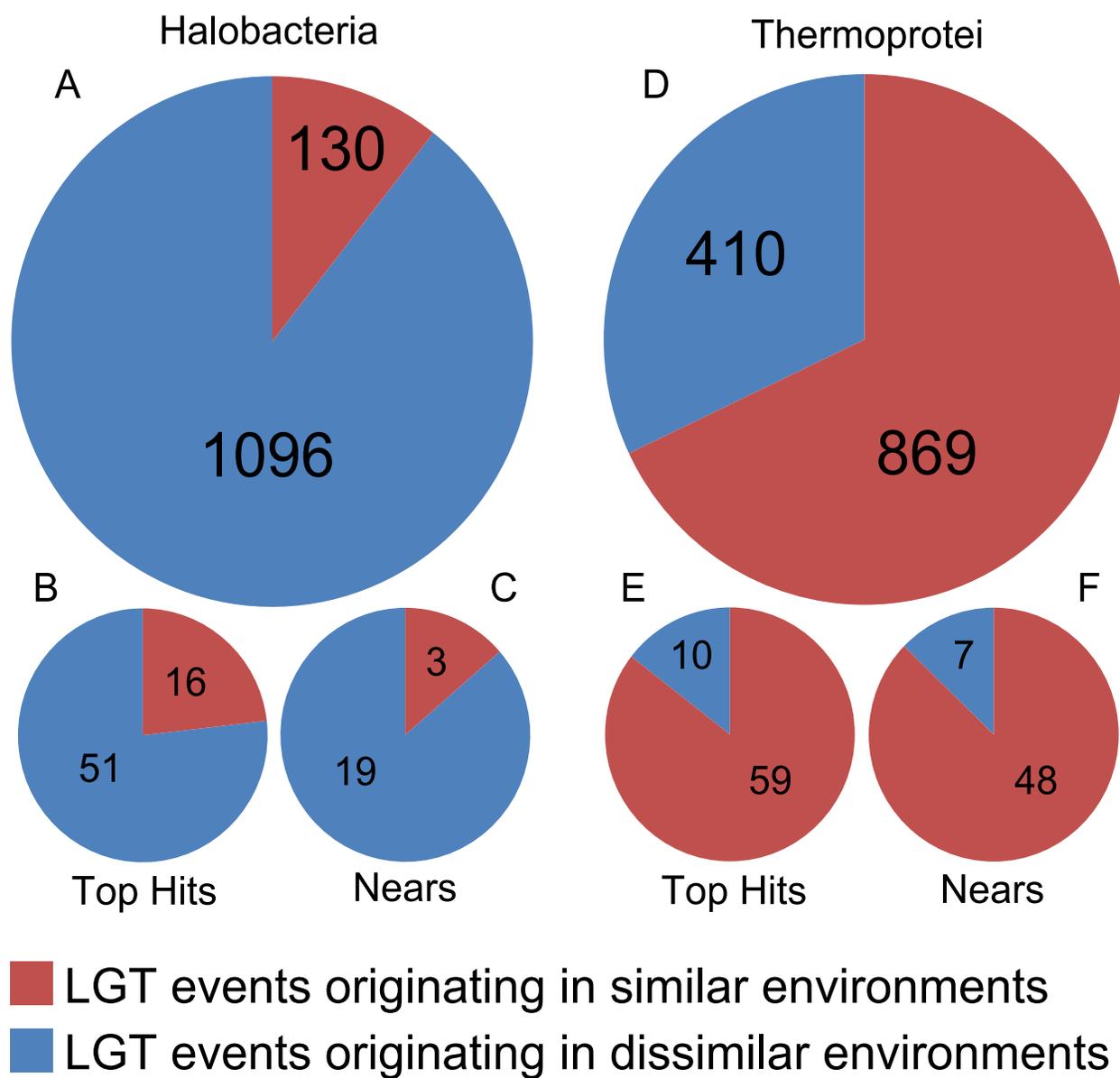
#### *Database bias*

There exists a potential bias in our analysis however, in that there are many more fully sequenced thermophiles in the databases than there are halophiles. While it seems unlikely, most if not all of the LGT events from non-halophiles into the Halobacteria could actually originate in heretofore unidentified and/or unsequenced

halophiles. The complexity and diversity of hypersaline environments, and for that matter the majority of the microbial world, is poorly constrained (Ley *et al.*, 2006). Thus, barring a direct observation of a LGT event from a non-halophile to a member of the Halobacteria, it appears impossible to rule out the possibility that we have not identified the correct donor species. Nevertheless, there are a number of tests that can lend support to our assertion that that the database bias does not account for the vast majority of the discrepancy between LGT into Thermoprotei and Halobacteria. These include:

1) Restricting the analysis to only particularly strong matches -

We restricted the analysis to LGT events whose top homologue had a BLAST bit score of greater than 500. Of the 1226 identified LGT events into the Halobacteria, 67 met this criterion. Of these 67, 51 or approximately 76% were to species which do not demonstrate halophilic tendencies (Figure 4b). Thus, as expected we do observe an increase in the proportion of LGT events originating in halophiles relative to non-halophiles. However, the increase only accounts for a small portion of the putative LGT events originating in non-halophiles. Of the top five examples of putative LGT events, only one is to a known halophile, *Salinibacter ruber* (Bacteroidetes). The other four, all with BLAST scores of 963 or better, are to non-halophiles (Table 1). The same analysis was performed on the Thermoprotei LGT genes. Amongst the 1279 identified LGT events into the Thermoprotei, 69 had BLAST scores of greater than 500. Of these 69, only 10, or approximately 14% were to non-thermophilic species (Figure 4e), yielding an



**Figure 4-4.** Pie charts depicting the proportion of inter-class LGT events from halophiles into the Halobacteria and from thermophiles into the Thermoprotei for all genes (a and d), only hits with bit scores > 500 (b and e), and only instances where multiple genes were transferred (c and f).

**Table 4-1.** Top five BLAST hits to genes identified as LGT events involving the Halobacteria, Thermoprotei, and a collection of Dead Sea fosmids.

Protein Function	Protein ID	Host	Donor	Percent Identical	Score
<b>Thermoprotei</b>					
Hypothetical protein	296242776	<i>Thermosphaera aggregans</i>	<i>Thermococcus sibiricus</i>	75	1667
Cobaltochelataase	146304837	<i>Metallosphaera sedula</i>	<i>Picrophilus torridus</i>	53	1256
Cellobiose phosphorylase	305664199	<i>Ignisphaera aggregans</i>	<i>Thermotoga maritima</i>	61	1055
Carbon-monoxide dehydrogenase	229582559	<i>Sulfolobus islandicus</i>	<i>Rhodothermus marinus</i>	69	1095
Aldehyde dehydrogenase	118431783	<i>Aeropyrum pernix</i>	<i>Thermomicrobium roseum</i>	62	991
<b>Halobacteria</b>					
Homocysteine methyltransferase	110668174	<i>Haloquadratum walsbyi</i>	<i>Salinibacter ruber</i>	76	1196
Molybdopterin oxidoreductase	284164928	<i>Haloterrigena turkmenica</i>	<i>Hydrogenobacter thermophilus</i>	46	1054
Hypothetical protein	222481310	<i>Halorubrum lacusprofundi</i>	<i>Hyphomicrobium denitrificans</i>	44	1029
Alpha amylase catalytic region	257386281	<i>Halomicrobium mukohataei</i>	<i>Thermoanaerobacter mathranii</i>	64	963
Glycosyltransferase	257052439	<i>Halorhabdus utahensis</i>	<i>Ignisphaera aggregans</i>	58	963
<b>Dead Sea Fosmids</b>					
Acetophenone carboxylase	56476760		<i>Azoarcus</i> sp.	45	637
Cytosine Deaminase	221633818		<i>Thermomicrobium roseum</i>	47	395
Isopropylaminohydrolyase	108803077		<i>Rubrobacter xylanophilus</i>	51	375
Alkaline phosphatase	262198659		<i>Haliangium ochraceum</i>	40	365
AMP-dependent synthetase	241661738		<i>Ralstonia pickettii</i>	39	335

even greater increase than for the Halobacteria. Also, all five of the strongest examples of LGT into Thermoprotei originate in thermophilic lineages (Table 1).

### 2) Restricting the analysis to only LGTs of neighboring gene pairs -

We identified 22 instances within the Halobacteria where adjacent genes or genes separated by a single gene were apparently transferred together and showed conservation not only of gene content, but also of gene order. The inter-class conservation of gene order offers concrete proof that these genes have undergone a LGT event. Of the 22 gene pairings, 19 originated in species with no known halophilic tendencies (Figure 4c), again suggesting that a significant portion of long range LGT events into the Halobacteria did not originate in halophiles. For the Thermoprotei we identified 55 multiple gene transfers, of these, only 7 or 13% originated in non-thermophiles (Figure 4f).

### 3) Looking for the presence of distinctly halophilic traits within the transferred genes-

As mentioned above, there have been a number of reported genomic indicators of halophilicity. These indicators include genome wide GC content, a preference for the dinucleotides 'CG', 'GA/TC', and 'AC/GT', a number of codon preferences, and an overall bias toward amino acids with acidic residues (Lanyi, 1974; Paul *et al.*, 2008). If a significant portion of the LGT genes did originate in non-halophiles, then values for these indicators would be expected to be higher in the non-LGT halobacterial genes than in the LGT halobacterial genes. These trends should not necessarily be observed for the Thermoprotei. Figure 5 shows

**Table 4-2.** Percent differences for various indicators of halophilicity between LGT genes and non-LGT genes for both the Halobacteria and Thermoprotei. Values for which the LGT genes are less than the non-LGT genes are shaded in green. Values for which the LGT genes are greater than the non-LGT genes are shaded in red. All values that exhibited greater than 95% confidence are shown in bold font. The indicators are: GC - GC content of the gene pools. CG Di's, GA/TC Di's, AC/GT Di's- Difference in preference for 'CG', 'GA/TC', and 'AC/GT' dinucleotides given nucleotide frequencies. Arginine (CGA and CGG), Cysteine (UGU), Leucine (CUC), Threonine (ACG), and Valine (GUC) - Difference in preference for respective codons given nucleotide abundances. AA Bias - Ratio of aspartic acid and glutamic acid to arginine, lysine, and histidine.

Halobacteria										
Genera	GC	CG Di's	GA/TC Di's	AC/GT Di's	Cysteine	Leucine	Threonine	Valine	Arginine	AB Ratio
<i>Halalkalicoccus</i>	3.8	0.7	0.6	-0.11	2.16	2.03	-2.04	3.91	-5.24	-0.1
<i>Haloarcula</i>	4.06	0.69	0.22	0.31	-0.99	-1.09	0.35	3.68	-2.05	0.11
<i>Halobacterium</i>	3.51	0.7	-0.38	0.07	-16.45	-3.59	0.64	0.18	-5.84	0.02
<i>Haloferax</i>	4.04	1.03	0.69	0.49	6.03	5.87	1.67	3.61	-3.53	-0.11
<i>Halomicrobium</i>	3.96	0.62	0.73	-0.05	2.8	0.42	3.17	3.05	-4.88	-0.06
<i>Haloquadratum</i>	1	0.21	-0.25	0.39	4.12	-1.67	-0.87	-0.78	0.7	0.01
<i>Halorhabdus</i>	0.86	0.43	0.49	0.03	1.17	1.54	2.56	-0.64	2.17	-0.19
<i>Halorubrum</i>	7.5	1.85	0.97	0.59	7.17	8.51	3.05	7.74	-4.85	-0.04
<i>Haloterrigena</i>	2.08	0.33	0.74	-0.44	3.61	1	-1.08	6.46	-4.31	0.1
<i>Natrialba</i>	1.82	0.51	0.37	0.16	2.8	5	0.87	2.72	-5.72	0.09
<i>Natronomonas</i>	4.05	0.99	0.16	0.54	4.88	2.3	4.14	7.84	-3.96	0.03
Thermoprotei										
Genera	GC	CG Di's	GA/TC Di's	AC/GT Di's	Cysteine	Leucine	Threonine	Valine	Arginine	AB Ratio
<i>Acidilobus</i>	-1.3	-0.11	0.26	-0.58	0.64	-2.21	-3.61	0.37	1.89	0.02
<i>Aeropyrum</i>	3.67	-0.41	0.66	0.03	-6.46	3.65	-0.52	1.91	-2.98	-0.01
<i>Caldivirga</i>	0.1	-0.01	-0.31	-0.02	-5.14	0.45	0.22	-0.04	-0.04	-0.05
<i>Desulfurococcus</i>	-2.92	-0.26	0.17	-0.53	9.87	-1.27	-1.54	-2.69	2	0.05
<i>Hyperthermus</i>	-1.69	0.11	0.01	-0.21	7.53	-1.9	-3.22	0.07	0.53	0.01
<i>Ignicoccus</i>	-0.24	-0.14	0.34	-0.17	-0.16	1.16	-0.4	0.76	-2.7	0.04
<i>Ignisphaera</i>	-1.54	0.06	0.13	-0.33	6.12	-0.96	-0.68	-1.03	0.77	-0.07
<i>Metallosphaera</i>	-1.79	0.18	0.05	-0.03	-0.3	-2.78	-0.85	1.34	0.45	-0.02
<i>Pyrobaculum</i>	2.02	-0.05	0.5	0.04	-0.14	2.08	2.5	0.46	-0.09	0.04
<i>Staphylothermus</i>	-0.15	0.12	0.3	-0.14	-2.67	-0.36	0.12	0.41	0.63	0.03
<i>Sulfolobus</i>	-1.17	0.04	0.11	-0.7	1.58	-1.09	-0.85	-1.24	0.2	0.02
<i>Thermofilum</i>	0.17	-0.03	0.24	-0.27	-0.26	1.34	-0.68	0.85	1.26	0.05
<i>Thermoproteus</i>	1.59	-0.44	0.65	0.55	8.37	-0.85	2.13	3.09	-3.53	0.01

a genus by genus breakdown of all values in LGT genes and non-LGT genes for both the Halobacteria and the Thermoprotei. The most robust trends are observed for the GC content and 'CG' dinucleotide preference. All 11 Halobacteria genera demonstrate a statistically significant increase in both GC content and 'CG' dinucleotide preference for the non-LGT genes relative to the LGT genes. This includes the genus *Haloquadratum* which has an anomalously low genomic GC content of 48%. However, as the Thermoprotei do not demonstrate a similar trend, the lowered GC content does not appear to be inherent to LGT events, nor can it be attributed to an artifact of our algorithm.

Both 'GA/TC' and 'AC/GT', also demonstrate a trend toward an increase among the non-LGT genes for the Halobacteria. The trends while not quite as strong, are still apparent.

The Thermoprotei show a similar trend for 'GA/TC' and a reverse trend for 'AC/GT'.

Amongst the codon biases, cysteine, leucine, threonine, and valine all show a trend toward an increased preference in the halobacterial non-LGT genes. Only arginine demonstrates a reverse trend. Meanwhile, the Thermoprotei do not appear to exhibit any particular trends. Finally, neither the Halobacteria nor the Thermoprotei show a particularly strong trend between LGT genes and non-LGT genes for the amino acid bias.

#### *Environmental halophiles examples*

Twenty-five 40 kb fosmids from the surface water of the 2006 Dead Sea were sequenced on a quarter plate of a 454FLX sequencer. The sequencing run produced a total of 90,479 reads with an average read length of 237 base pairs, for a total of approximately 21 million base pairs of sequence. The sequences were then assembled

and a total of 95 contigs with greater than 2,000 base pairs were produced. These contigs were compared to the collection of fully sequenced genomes using a BLASTX search. The contigs were then scanned for the presence of Halobacteria genes, and all contigs without a majority of Halobacteria genes were discarded. The remaining contigs were searched for the presence of genes whose top normalized BLAST score to a member of the Halobacteria was less than 75% of the top normalized BLAST score to any non-member of the Halobacteria. Twenty-two putative “long distance” LGT genes were identified in this manner of which only two were from known halophiles (Table C-2). The top five instances are provided in Table 1.

## **Discussion**

Using a homology-based approach we identified 1,226 putative inter-class LGT events involving members of the obligatory halophilic archaeal class Halobacteria and 1,269 putative inter-class LGT events involving members of the obligatory thermophilic archaeal class Thermoprotei. The vast majority of these LGT events consisted of gene transfers into the Halobacteria and Thermoprotei. Furthermore, the phylogenetic distance between the donor species and the recipient species suggests that the majority of these LGT events were the result of natural transformation. As the Halobacteria are all obligate halophiles and the Thermoprotei are all obligate thermophiles the transformative events must have occurred in saline and thermal environments respectively.

Conventional thinking would suggest that the Halobacteria would be exposed to naked DNA from predominately other halophiles and that the Thermoprotei would be exposed to naked DNA from predominately other thermophiles. Additionally, genes originating in other halophiles and thermophiles would be preadapted to the particular environmental conditions and would therefore be more likely to be successfully transferred. Thus we would expect the majority of LGT events into the Thermoprotei to originate in other thermophiles and the majority of LGT events into the Halobacteria to originate in other halophiles. However, we found that while the majority of these transformational events into the Thermoprotei did in fact originate in other thermophiles, the majority of these transformational events into the Halobacteria did not originate in other known halophiles. This suggests that there is something fundamentally different between LGT in thermophiles and LGT in halophiles.

Unfortunately, as with all studies relying on genomic databases, there is the potential for distortion from database bias. In our study we face three disparate database issues. First, there is always the possibility that we may have misidentified LGT events. Second, while the fully sequenced Thermoprotei originate from a number of distinct orders, the fully sequenced Halobacteria all belong to a single family, the Halobacteriaceae. It is unclear how the reduced phylogenetic diversity of the Halobacteria would affect our analysis. Finally, the relative paucity of fully sequenced halophiles as compared to thermophiles may explain our observation that more Halobacteria donors are non-halophiles than Thermoprotei donors are non-thermophiles. Nevertheless, a number of halophiles have been sequenced in non-

Halobacterial lineages including *Bacillus halodurans* of the bacterial class Bacilli, *Methanohalophilus mahii* and *Methanohalobium evestigatum* of the euryarchaeal class Methanomicrobia, and *Chromohalobacter salexigens* of the bacterial class Gammaproteobacteria. In all four of these cases there remain numerous LGT events apparently originating in close non-halophilic relatives. At the same time we also went to great lengths to seek out additional lines of evidence that would help confirm our findings.

We restricted the analysis to LGT genes with especially strong matches and to instances where multiple genes were transferred together and gene order was conserved. Restricting the analysis to especially strong matches increases the likelihood of having identified both an LGT event and the correct donor species. Restricting the analysis to multiple gene transfers virtually guarantees that a LGT event took place. In both cases we achieved similar results to our initial analysis.

We then investigated independent genomic indicators of halophilicity for each genus of the Halobacteria and Thermoprotei. These indicators consisted of GC content, 'CG', 'GA/TC', and 'AC/GT' dinucleotide content, codon preferences, and amino acid preferences. If the LGT genes into the Halobacteria did in fact originate in non-halophiles, some residual signature of non-halophilicity could remain. Essentially this amounts to an independent assessment of LGT events specifically targeting halophiles. Of the ten indicators investigated for the Halobacteria, eight supported our assertion, one did not indicate a clear trend, and only the codon preference for arginine refuted our conclusion.

The strongest support came from genomic GC content and 'CG' dinucleotide content. Both indicators showed a statistically significant increase from non-LGT genes to LGT genes for every genus and represent strong support for the correct identification of LGT genes and for the non-halophile origin of the majority of them. This includes the genus *Haloquadratum* which is unique among the Halobacteria for having a relatively low genomic GC content of approximately 48%. It therefore might appear that there is something inherent to genes associated with LGT that accounts for the differences in GC content. However, for the Thermoprotei there was no clear trend for the indicators as a whole and for genomic GC content there appeared to be a decrease in genomic GC content from LGT genes to non-LGT genes.

Finally we sought confirmation of our genome based results from within metagenomic samples. We used fully and partially assembled fosmid inserts from the surface waters of the Dead Sea, a highly saline environment, to identify inter-class LGT events in environmental halophiles. Once again the vast majority of the donor species were non-halophiles.

## **Conclusions**

In this study we provide a number of lines of evidence that suggest that the mechanisms and origins of "long distance" LGT events into the Thermoprotei and Halobacteria are different. We theorize that the difference in the origin of LGT genes lies in the differing natures of hypersaline and thermal environments with respect to naked DNA. Hypersaline environments are often adept at preserving both naked DNA

and intact microorganisms. There have even been claims of intact DNA and viable bacteria preserved in 200 million year old salt crystals (Fish *et al.*, 2002; Mormile *et al.*, 2003; Vreeland *et al.*, 2007). In contrast, thermal environments rapidly degrade DNA. Thermophilic organisms, therefore, must go to great lengths to protect and stabilize their DNA from the environment. Thus intracellular degradation of DNA would be expected to be greater for non-thermophiles, and thermally stable DNA would be in better condition upon release to the environment. Furthermore various mechanisms of DNA protection, such as association with DNA binding proteins, may provide transient protection extracellularly. The net effect of these and other protective methods would lead to an increase in intact thermophilic DNA in a thermal environment relative to non-thermophilic DNA.

In a related vein, hypersaline environments generally occupy topographic minima. This makes hypersaline environments such as the deep Mediterranean basins and the Dead Sea natural collectors of debris, cellular and otherwise (Borin *et al.*, 2008; Rhodes *et al.*, 2010). Thus the average halophilic microorganism should be exposed to a much greater diversity of DNA than an average thermophilic microbe. Together these facts suggest that halophilic microorganisms are exposed to a greater proportion of intact non-halophilic DNA than thermophiles are exposed to intact non-thermophilic DNA. This suggestion combined with the relatively large genome size, diverse genomic composition, and broad range of metabolic capabilities of the Halobacteria paint the picture of the Halobacteria potentially acting as the consummate opportunists, incorporating and utilizing genes from a great variety of organisms. However, in order

to better identify and understand LGT events amongst halophiles many more halophiles must be sequenced and the dynamics of naked DNA in a variety of naturally occurring settings must be studied.

## **Methods**

### *DNA extraction and fosmid preparation*

The Dead Sea environmental sample was collected and processed in 2007 by the Béjà lab group (Technion, Haifa, Israel) according to the protocol of Bodaker *et al.* (Bodaker *et al.*, 2010). The fosmid inserts were then shipped frozen to Penn State. The inserts were run on a 1% low melting point agarose gel to remove residual contamination and the 40 kb band was extracted and digested with the Gelase enzyme (Epicentre). The fosmids were then sequenced on a GS FLX sequencer (454 Life Sciences) on one quarter of a pico-titre plate.

### *Fosmid analysis*

The fosmid sequences were assembled using the 454 assembler program. All contigs of greater than 2,000 base pairs were compared to the collection of fully sequenced Bacteria and Archaea using the BLASTX program, an e-value of  $10^{-5}$ , and default parameters. The contigs were then spliced according to gene location and another identical BLASTX comparison was conducted on each gene. Each gene whose top hit was not to a member of the Halobacteria, had a normalized bit score (BLAST bit score to homologue divided by BLAST bit score to self) more than 25% greater than the best hit to a Halobacteria gene, and had a bit score greater than 67 was flagged as a

putative inter-class LGT event. Then all contigs were scanned for genes belonging to the Halobacteria, and contigs without a majority of genes assigned to Halobacteria species were discarded. Finally, a number of genes demonstrated near perfect, upwards of 95%, identity to likely laboratory contaminants such as *Escherichia coli*. These genes were also removed from the analysis. The remaining 22 genes were considered LGT events and the donor species were assigned according to the best hit as matched by BLAST. A web search was then conducted to identify whether the donor species was a known halophile.

#### *Genome analysis*

All fully sequenced Thermoprotei and Halobacteria genomes were compared to the collection of all fully sequenced Bacteria and Archaea using BLASTP, an e-value of  $10^{-5}$ , and default parameters. Each gene whose top non-identical hit was not to a member of the Halobacteria or Thermoprotei respectively, had a normalized bit score more than 25% greater than the best non-identical hit to a member of the Halobacteria or Thermoprotei, and had a bit score greater than 67 was flagged as an inter-class LGT event. In cases such as *Pyrobaculum* where more than one species has been sequenced, the analysis was conducted on the species with the most genes and all hits to members of its genus were masked out. The remaining species were subjected to the usual analysis and any additional LGT genes were included in the analysis of the genus. In cases such as *Sulfolobus islandicus* where multiple strains have been sequenced a similar masking was performed. The donor species were assigned according to the best hit as matched by BLAST and a web search was then conducted to identify whether the donor

species was a known halophile or thermophile respectively. Genes were assigned to COGs based upon the NCBI annotation. All additional analysis was performed using home-written scripts in Perl and/or Python. The scripts are available upon request. For the purposes of our statistical analysis the values of the non-LGT gene pool were taken as representative of the taxon as a whole. We then used a chi square test to assess the likelihood of the LGT genes originating in the same population.

#### *Phylogenetic tree building*

The phylogenetic trees included in the supplemental material were constructed using the online tools available in association with the KEGG database. For each gene with a BLAST score of over 500, its twenty closest homologues were selected. The CLUSTALW tool was then used to create an alignment, and an unrooted neighbor joining tree was constructed. The phylogenetic trees were inspected manually for indicators of LGT directionality. For the phylogenetic trees depicted in Figures 1 and 2, LGT genes from both the Thermoprotei and Halobacteria were pooled into three pools depending on BLAST bit score. Five genes at random were chosen from each pool and the top 14 homologues from distinct genera were selected from the KEGG database. The CLUSTALW tool was again used to create an alignment. The alignments were then loaded into PHYLIP and trees were constructed with 100 bootstraps and the mean-least-squared method (Felsenstein, 1989).

### **Authors' Contributions**

MR performed the bioinformatic analysis and laboratory work and drafted the manuscript. JS and CH jointly oversaw this study and aided in the revising process. AO provided expert guidance and also assisted in the editing of this manuscript. All authors have read and approved of the manuscript.

### **Acknowledgements**

We thank I. Bodaker and O. Béjà for their efforts in collecting and processing the fosmid samples, L. Tomsho for sequencing, and S. Fitz-Gibbon for bioinformatic support. This work was supported in part by the National Aeronautics and Space Administration (NASA) Astrobiology Institute (NAI) under NASA-Ames Cooperative Agreement NNA09DA76A (C.H.H.) and by the Agriculture and Food Research Initiative Competitive Grants Program Grant no. 2010-65110-20488 from the USDA National Institute of Food and Agriculture. The 454 facility at the Pennsylvania State University Center for Genome Analysis is funded, in part, by a grant from the Pennsylvania Department of Health using Tobacco Settlement Funds appropriated by the legislature.

### **References**

- Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997).** Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**, 3389-3402.
- Arahal, D. R., Márquez, M. C., Volcani, B. E., Schleifer, K. H. & Ventosa, A. (1999).** *Bacillus marismortui* sp. nov., a new moderately halophilic species from the Dead Sea. *International Journal of Systematic Bacteriology* **49**, 521-530.

**Aravind, L., Tatusov, R. L., Wolf, Y. I., Walker, D. R. & Koonin, E. V. (1998).** Evidence for massive gene exchange between archaeal and bacterial hyperthermophiles. *Trends in Genetics* **14**, 442-444.

**Avery, O. T., MacLeod, C. M. & McCarty, M. (1944).** Studies on the chemical nature of the substance inducing transformation of pneumococcal types. *The Journal of Experimental Medicine* **79**, 137-158.

**Bathe, S., Lebuhn, M., Ellwart, J. W., Wuertz, S. & Hausner, M. (2004).** High phylogenetic diversity of transconjugants carrying plasmid pJP4 in an activated sludge-derived microbial community. *FEMS Microbiology Letters* **235**, 215-219.

**Beyth, M. (1980).** Recent evolution and present stage of Dead Sea brines. In *Hypersaline Brines and Evaporitic Environments*, pp. 155-165. Edited by A. Nissenbaum: Elsevier.

**Biddle, J. F., Fitz-Gibbon, S., Schuster, S. C., Brenchley, J. E. & House, C. H. (2008).** Metagenomic signatures of the Peru Margin seafloor biosphere show a genetically distinct environment. *Proceedings of the National Academy of Sciences of the United States of America*.

**Bodaker, I., Béjà, O., Feingersch, R., Rosenber, M., Oren, A., Hindiyeh, M. Y. & Malkawi, H. I. (2009a).** Archaeal diversity in the Dead Sea: Microbial survival under increasingly harsh conditions. In *Saline lakes around the world: Unique systems with unique values*, pp. 137-143. Edited by A. Oren, D. Naftz, P. Palacios & W. A. Wurtsbaugh. Logan, Utah: S. J. and Jessie E. Quinney Natural Resources Research Library.

**Bodaker, I., Béjà, O., Feingersch, R., Rosenberg, M., Oren, A., Hindiyeh, M. Y. & Malkawi, H. I. (2009b).** Archaeal diversity in the Dead Sea: Microbial survival under increasingly harsh conditions. In *Saline lakes around the world: Unique systems with unique values*, pp. 137-143. Edited by A. Oren, D. Naftz, P. Palacios & W. A. Wurtsbaugh. Logan, Utah: S. J. and Jessie E. Quinney Natural Resources Research Library.

**Bodaker, I., Sharon, I., Suzuki, M. T. & other authors (2010).** Comparative community genomics in the Dead Sea: An increasingly extreme environment. *The ISME Journal* **4**, 399-407.

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

**Borin, S., Crotti, E., Mapelli, F., Tamagnini, I., Corselli, C. & Daffonchio, D. (2008).** DNA is preserved and maintains transforming potential after contact with brines of the deep anoxic hypersaline lakes of the Eastern Mediterranean Sea. *Saline Systems* **4**, 10.

**Cole, J. R., Chai, B., Farris, R. J. & other authors (2007).** The ribosomal database project (RDP-II): introducing myRDP space and quality controlled public data. *Nucleic Acids Research* **35**, D169-D172.

**Cole, J. R., Wang, Q., Cardenas, E. & other authors (2009).** The Ribosomal Database Project: Improved alignments and new tools for rRNA analysis. *Nucleic Acids Research* **37**, D141-D145.

**Dean, F. B., Nelson, J. R., Giesler, T. L. & Lasken, R. S. (2001).** Rapid amplification of plasmid and phage DNA using Phi29 DNA polymerase and multiply-primed rolling circle amplification. *Genome Research* **11**, 1095-1099.

**Deppenmeier, U., Johann, A., Hartsch, T. & other authors (2002).** The genome of *Methanosarcina mazei*: Evidence for lateral gene transfer between bacteria and archaea. *Journal of Molecular Microbiology and Biotechnology* **4**, 453-461.

**Doolittle, W. F. (1999).** Phylogenetic classification and the universal tree. *Science* **284**, 2124-2129.

**Doolittle, W. F. & Baptiste, E. (2007).** Pattern pluralism and the Tree of Life hypothesis. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 2043-2049.

**Dyall-Smith, M. L., Pfeiffer, F., Klee, K., Palm, P., Gross, K., Schuster, S. C., Rampp, M. & Oesterhelt, D. (2011).** *Haloquadratum walsbyi*: Limited diversity in a global pond. *PLoS ONE* **6**, e20968.

**Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C. & Knight, R. (2011).** UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*.

**Edgar, R. C. (2004).** MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**, 1792-1797.

**Felsenstein, J. (1989).** PHYLIP - Phylogeny Inference Package (Version 3.2). *Cladistics* **5**, 164-166.

**Field, K., Olsen, G., Lane, D., Giovannoni, S., Ghiselin, M., Raff, E., Pace, N. & Raff, R. (1988).** Molecular phylogeny of the animal kingdom. *Science* **239**, 748-753.

**Fish, S. A., Shepherd, T. J., McGenity, T. J. & Grant, W. D. (2002).** Recovery of 16S ribosomal RNA gene fragments from ancient halite. *Nature* **417**, 432-436.

**Freeman, V. J. (1951).** Studies on the virulence of bacteriophage-infected strains of *Corynebacterium diphtheriae*. *Journal of Bacteriology* **61**, 675-688.

**Gavrieli, I., Beyth, M. & Yechieli, Y. (1999).** The Dead Sea - A terminal lake in the Dead Sea rift: A short overview. In *Microbiology and biogeochemistry of hypersaline environments*, pp. 121-127. Edited by A. Oren. Boca Raton, Florida: CRC Press.

**Gavrieli, I. & Oren, A. (2004).** The Dead Sea as a dying lake. In *Dying and dead seas Climactic versus anthropic causes*, pp. 287-305. Edited by J. C. J. Nihoul, P. O. Zavialov & P. P. Micklin. Dordrecht: Kluwer Academic Publishers.

**Gill, S. R., Pop, M., DeBoy, R. T. & other authors (2006).** Metagenomic analysis of the human distal gut microbiome. *Science* **312**, 1355-1359.

- Goodarzi, H., Torabi, N., Najafabadi, H. S. & Archetti, M. (2008).** Amino acid and codon usage profiles: Adaptive changes in the frequency of amino acids and codons. *Gene* **407**, 30-41.
- Haney, P. J., Badger, J. H., Buldak, G. L., Reich, C. I., Woese, C. R. & Olsen, G. J. (1999).** Thermal adaptation analyzed by comparison of protein sequences from mesophilic and extremely thermophilic *Methanococcus* species. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 3578-3583.
- Huson, D. H., Auch, A. F., Qi, J. & Schuster, S. C. (2007).** MEGAN analysis of metagenomic data. *Genome Research* **17**, 377-386.
- Ihaka, R. & Gentleman, R. (1996).** R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics* **5**, 299-314.
- Jensen, E. C., Schrader, H. S., Rieland, B., Thompson, T. L., Lee, K. W., Nickerson, K. W. & Kokjohn, T. A. (1998).** Prevalence of broad-host-range lytic bacteriophages of *Sphaerotilus natans*, *Escherichia coli*, and *Pseudomonas aeruginosa*. *Applied and Environmental Microbiology* **64**, 575-580.
- Jonas, D. A., Elmadfa, I., Engel, K. & other authors (2001).** Safety considerations of DNA in food. *Annals of Nutrition and Metabolism* **45**, 235-254.
- Kanehisa, M. & Goto, S. (2000).** Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research* **28**.
- Kaplan, I. R. & Friedman, A. (1970).** Biological productivity in the Dead Sea. Part I: Microorganisms in the water column. *Israel Journal of Chemistry* **8**, 513-528.
- Kennedy, S. P., Ng, W. V., Salzberg, S. L., 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 Research* **11**, 1641-1650.
- Kunin, V., Raes, J., Harris, J. K. & other authors (2008).** Millimeter-scale genetic gradients and community-level molecular convergence in a hypersaline microbial mat. *Molecular Systems Biology* **4**.
- Lane, D. J. (1991).** 16S/23S rRNA sequencing. In *Nucleic acid techniques in bacterial systematics*, pp. 115-175. Edited by E. Stackebrandt & M. Goodfellow. New York, New York: John Wiley and Sons.
- Lanyi, J. K. (1974).** Salt-dependent properties of proteins from extremely halophilic bacteria. *Bacteriology Reviews* **28**, 272-290.
- Legault, B., Lopez-Lopez, A., Alba-Casado, J., Doolittle, W. F., Bolhuis, H., Rodriguez-Valera, F. & Papke, R. T. (2006).** Environmental genomics of "*Haloquadratum walsbyi*" in a saltern crystallizer indicates a large pool of accessory genes in an otherwise coherent species. *BMC Genomics* **7**, 171.

- Ley, R. E., Harris, J. K., Wilcox, J. & other authors (2006).** Unexpected diversity and complexity of the Guerrero Negro hypersaline microbial mat. *Applied and Environmental Microbiology* **72**, 3685-3695.
- Lobasso, S., Lopalco, P., Mascolo, G. & Corcelli, A. (2008).** Lipids of the ultra-thin square halophilic archaeon *Haloquadratum walsbyi*. *Archaea* **2**, 177-183.
- Ludwig, W., Strunk, O., Westram, R. & other authors (2004).** ARB:A software environment for sequence data. *Nucleic Acids Research* **32**, 1363-1371.
- Macalady, J. L., Dattagupta, S., Schaperdoth, I., Jones, D. S., Druschel, G. K. & Eastman, D. (2008).** Niche differentiation among sulfur-oxidizing bacterial populations in cave waters. *The ISME Journal* **2**, 590-601.
- Madern, D., Ebel, C. & Zaccai, G. (2000).** Halophilic adaptation of enzymes. *Extremophiles* **4**, 91-98.
- Marguet, E. & Forterre, P. (1994).** DNA stability at temperatures typical for hyperthermophiles. *Nucleic Acids Research* **22**, 1681-1686.
- McDaniel, L. D., Young, E., Delaney, J., Ruhnau, F., Ritchie, K. B. & Paul, J. H. (2010).** High frequency of horizontal gene transfer in the oceans. *Science* **330**, 50.
- Mongodin, E. F., Nelson, K. E., Daugherty, S. & other authors (2005).** The genome of *Salinibacter ruber*: Convergence and gene exchange among hyperhalophilic bacteria and archaea. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 18147-18152.
- Mormile, M. R., Biesen, M. A., Gutierrez, M. C., Ventosa, A., Pavlovich, J. B., Onstott, T. C. & Fredrickson, J. K. (2003).** Isolation of *Halobacterium salinarum* retrieved directly from halite brine inclusions. *Environmental Microbiology* **5**, 1094-1102.
- Mullakhanbhai, M. F. & Larsen, H. (1975).** *Halobacterium volcani* spec. nov., a Dead sea halobacterium with a moderate salt requirement. *Archives of Microbiology* **104**, 207-214.
- Muyzer, G., Hottentrager, S., Teske, A. & Wawer, C. (1996).** Denaturing gradient gel electrophoresis of PCR-amplified 16S rDNA. A new molecular approach to analyze the genetic diversity of mixed microbial communities. In *Molecular Microbial Ecology Manual*, pp. 3.4.4.1-3.4.4.22. Edited by A. D. L. Akkermans, J. D. van Elsas & F. J. de Bruijn. Dordrecht: Kluwer Academic Publishers.
- Neev, D. & Emery, K. O. (1967).** The Dead Sea depositional processes and environments of evaporates: State of Israel Ministry of Development Geological Survey.
- Nelson, K. E., Clayton, R. A., Gill, S. R. & other authors (1999).** Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of *Thermotoga maritima*. *Nature* **399**, 323-329.

- Ng, W. V., Kennedy, S. P., Mahairas, G. G. & other authors (2000).** Genome sequence of *Halobacterium* species NRC-1. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 12176-12181.
- Oren, A. & Shilo, M. (1982).** Population dynamics of *Dunaliella parva* in the Dead Sea. *Limnology and Oceanography* **27**, 201-211.
- Oren, A. (1983a).** Population dynamics of halobacteria in the Dead Sea water column. *Limnology and Oceanography* **28**, 1094-1103.
- Oren, A. (1983b).** *Halobacterium sodomense* sp. nov., a Dead Sea halobacterium with an extremely high magnesium requirement. *International Journal of Systematic Bacteriology* **33**, 381-386.
- Oren, A. (1986).** Intracellular salt concentrations of the anaerobic halophilic eubacteria *Haloanaerobium praevalens* and *Halobacteroides halobius*. *Canadian Journal of Microbiology* **32**, 4-9.
- Oren, A., Ginzburg, M., Ginzburg, B. Z., Hochstein, L. I. & Volcani, B. E. (1990).** *Haloarcula marismortui* (Volcani) sp. nov., nom. rev., an extremely halophilic bacterium from the Dead Sea. *International Journal of Systematic Bacteriology* **40**, 209-210.
- Oren, A. (1993a).** The Dead Sea - alive again. *Experientia* **49**, 518-522.
- Oren, A. (1993b).** Ecology of extremely halophilic microorganisms. In *The biology of halophilic bacteria*, pp. 25-53. Edited by R. H. Vreeland & L. I. Hochstein. Boca Raton: CRC Press.
- Oren, A. & Gurevich, P. (1993).** Characterization of the dominant halophilic archaea in a bacterial bloom in the Dead Sea. *FEMS Microbiology Ecology* **12**, 249-256.
- Oren, A. & Gurevich, P. (1995).** Dynamics of a bloom of halophilic archaea in the Dead Sea. *Hydrobiologia* **315**, 149-158.
- Oren, A., Gurevich, P., Anati, D. A., Barkan, E. & Luz, B. (1995a).** A bloom of *Dunaliella parva* in the Dead Sea in 1992: Biological and biogeochemical aspects. *Hydrobiologia* **197**, 173-185.
- Oren, A., Gurevich, P., Gemmell, R. T. & Teske, A. (1995b).** *Halobaculum gomorrhense* gen. nov., sp. nov., a novel extremely halophilic archaeon from the Dead Sea. *International Journal of Systematic Bacteriology* **45**, 747-754.
- Oren, A., Duker, S. & Ritter, S. (1996).** The polar lipid composition of Walsby's square bacterium. *FEMS Microbiology Letters* **138**, 135-140.
- Oren, A. & Ben-Yosef, N. (1997).** Development and spatial distribution of an algal bloom in the Dead Sea: A remote sensing study. *Aquatic Microbial Ecology* **13**, 219-223.
- Oren, A., Bratbak, G. & Heldal, M. (1997).** Occurrence of virus-like particles in the Dead Sea. *Extremophiles* **1**, 143-149.

- Oren, A. (1999a).** Microbiological studies in the Dead Sea: Future challenges toward the understanding of life at the limit of salt concentrations. *Hydrobiologia* **405**, 1-9.
- Oren, A. (1999b).** Bioenergetic aspects of halophilism. *Microbiology and Molecular Biology Reviews* **63**, 334-348.
- Oren, A. (2000).** Biological processes in the Dead Sea as influenced by short-term and long-term salinity changes. *Archives of Hydrobiology Special Issues on Advanced Limnology* **55**, 531-542.
- Oren, A. (2001).** The bioenergetic basis for the decrease in metabolic diversity at increasing salt concentrations: Implications for the functioning of salt lake ecosystems. *Hydrobiologia* **466**, 61-72.
- Oren, A. (2002a).** Properties of halophilic proteins. In *Halophilic microorganisms and their environments*, pp. 233-278. Edited by A. Oren. Dodrecht, The Netherlands: Kluwer Academic.
- Oren, A. (2002b).** Molecular ecology of extremely halophilic Archaea and Bacteria. *FEMS Microbiology Ecology* **39**, 1-7.
- Oren, A., Gavrieli, I., Gavrieli, J., Kohen, M., Lati, J. & Aharoni, M. (2004).** Biological effects of dilution of Dead Sea brine with seawater: Implications for the planning of the Red Sea-Dead Sea "Peace Conduit". *Journal of Marine Systems* **46**, 121-131.
- Oren, A., Larimer, F., Richardson, P., Lapidus, A. & Csonka, L. N. (2005).** How to be moderately halophilic with broad salt tolerance: clues from the genome of *Chromohalobacter salexigens*. *Extremophiles* **9**, 275-279.
- Park, H. S., Chatterjee, I., Johnston, S., Wong, M., Jack, T. R. & Voordouw, G. (2011).** Microbially influenced corrosion in oil sands water handling facilities. In *Corrosion 2011*. Houston, Texas: NACE International
- Paul, S., Bag, S., Das, S., Harvill, E. & Dutta, C. (2008).** Molecular signature of hypersaline adaptation: Insights from genome and proteome composition of halophilic prokaryotes. *Genome Biology* **9**, R70.
- Pruesse, E., Quast, C., Knittel, K., Fuchs, B. M., Ludwig, W., Peplies, J. & Glöckner, F. O. (2007).** SILVA: A comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Research* **35**, 7188-7196.
- Ragan, M. A., Harlow, T. J. & Beiko, R. G. (2006).** Do different surrogate methods detect lateral genetic transfer events of different relative ages? *Trends in Microbiology* **14**, 4-8.
- Reistad, R. (1970).** On the composition and nature of the bulk protein of extremely halophilic bacteria. *Archives of Microbiology* **71**, 353-360.

**Rhodes, M. E., Fitz-Gibbon, S. T., Oren, A. & House, C. H. (2010).** Amino acid signatures of salinity on an environmental scale with a focus on the Dead Sea. *Environmental Microbiology* **12**, 2613-2623.

**Rivera, M. C., Jain, R., Moore, J. E. & Lake, J. A. (1998).** Genomic evidence for two functionally distinct gene classes. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 6239-6244.

**Robertson, C. E., Spear, J. R., Harris, J. K. & Pace, N. R. (2009).** Diversity and stratification of archaea in a hypersaline microbial mat. *Applied and Environmental Microbiology* **75**, 1801-1810.

**Sahl, J. W., Pace, N. R. & Spear, J. R. (2008).** Comparative molecular analysis of endoevaporitic microbial communities. *Applied and Environmental Microbiology* **74**, 6444-6446.

**Santos, H. & Da Costa, M. S. (2002).** Compatible solutes of organisms that live in hot saline environments. *Environmental Microbiology* **4**, 501-509.

**Schloss, P. D., Westcott, S. L., Ryabin, T. & other authors (2009).** Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology* **75**, 7537-7541.

**Sogin, M. L., Morrison, H. G., Huber, J. A., Welch, D. M., Huse, S. M., Neal, P. R., Arrieta, J. M. & Herndl, G. J. (2006).** Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proceedings of the National Academy of Sciences of the United States of America* **103**, 12115-12120.

**Soppa, J. (2006).** From genomes to function: Haloarchaea as model organisms. *Microbiology* **152**, 585-590.

**Spear, J. R., Ley, R. E., Berger, A. B. & Pace, N. R. (2003).** Complexity in natural microbial ecosystems: The Guerrero Negro experience. *The Biological Bulletin* **204**, 168-173.

**Stamatakis, A., Ludqig, T. & Meler, H. (2005).** Raxml-III: A fast program for maximum likelihood-based inference of large phylogenetic trees. *Bioinformatics* **21**, 456-463.

**Tamames, J. & Moya, A. (2008).** Estimating the extent of horizontal gene transfer in metagenomic sequences. *BMC Genomics* **9**, 136.

**Tatum, E. L. & Lederberg, J. (1947).** Gene recombination in the bacterium *Escherichia coli*. *Journal of Bacteriology* **53**, 673-684.

**Tatusov, R., Fedorova, N., Jackson, J. & other authors (2003).** The COG database: An updated version includes eukaryotes. *BMC Bioinformatics* **4**, 41.

**Tehei, M., Franzetti, B., Maurel, M., Vergne, J., Hountondji, C. & Zaccai, G. (2002).** The search for traces of life: the protective effect of salt on biological macromolecules. *Extremophiles* **6**, 427-430.

- Thomas, C. M. & Nielsen, K. M. (2005).** Mechanisms of, and Barriers to, Horizontal Gene Transfer between Bacteria. *Nat Rev Micro* **3**, 711-721.
- Tringe, S. G., von Mering, C., Kobayashi, A. & other authors (2005).** Comparative metagenomics of microbial communities. *Science* **308**, 554-557.
- Turnbaugh, P. J., Ley, R. E., Mahowald, M. A., Magrini, V., Mardis, E. R. & Gordon, J. I. (2006).** An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**, 1027-1131.
- Tyson, G. W., Chapman, J., Hugenholtz, P. & other authors (2004).** Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **428**, 37-43.
- Vaisman, N. & Oren, A. (2009).** *Salisaeta longa* gen. nov., sp. nov., a red, halophilic member of the Bacteroidetes. *International Journal of Systematic and Evolutionary Microbiology* **59**, 2571-2574.
- Valenzuela-Encinas, C., Neria-González, I., Alcántara-Hernández, R., Enríquez-Aragón, J., Estrada-Alvarado, I., Hernández-Rodríguez, C., Dendooven, L. & Marsch, R. (2008).** Phylogenetic analysis of the archaeal community in an alkaline-saline soil of the former lake Texcoco (Mexico). *Extremophiles* **12**, 247-254.
- Venter, J. C., Remington, K., Heidelberg, J. F. & other authors (2004).** Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**, 66-74.
- Volcani, B. (1944).** The microorganisms of the Dead Sea. In *Papers collected to commemorate the 70th anniversary of Dr Chaim Weizmann*, pp. 71-85. Rehovoth: Daniel Sieff Research Institute.
- Vreeland, R. H., Jones, J., Monson, A. & other authors (2007).** Isolation of live Cretaceous (121-112 million years old) halophilic archaea from primary salt crystals. *Geomicrobiology Journal* **24**, 275-282.
- Ward, J. H. (1963).** Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association* **58**, 236-244.
- Wilkansky, B. (1936).** Life in the Dead Sea. *Nature* **138**, 467.
- Zhaxybayeva, O., Gogarten, J. P., Charlebois, R. L., Doolittle, W. F. & Papke, R. T. (2006).** Phylogenetic analyses of cyanobacterial genomes: Quantification of horizontal gene transfer events. *Genome Research* **16**, 1099-1108.
- Zinder, N. D. & Lederberg, J. (1952).** Genetic exchange in *Salmonella*. *Journal of Bacteriology* **64**, 679-699.

## Appendix A

## Supplemental Information for Chapter 2

## Supporting Information A-1

**Table A-1.** Table providing the taxonomic assignment of 16S rRNA gene fragments identified in a metagenome from the surface waters of the 2007 Dead Sea.

Genus	Sequences
Uncultured Archaea	63
<i>Natronomonas</i>	20
<i>Halorhabdus</i>	18
<i>Haloplanus</i>	9
<i>Haloarcula</i>	6
<i>Halomarina</i>	4
<i>Natrinema</i>	3
<i>Halorubrum</i>	3
Uncultured Bacteria	3
<i>Halobacterium</i>	2
<i>Azoarcus</i>	2
<i>Haloquadratum</i>	2
<i>Acinetobacter</i>	1
<i>Halalkalicoccus</i>	1
<i>Halobiforma</i>	1
<i>Natronobacterium</i>	1
<i>Halostagnicola</i>	1
<i>Halosimplex</i>	1
<i>Chlamydomonad</i>	1
Total	142

## Appendix B

### Supplemental Information for Chapter 3

#### Supporting Information B-1

Table B-1 shows the acid and basic amino acids ratio and compositions, and the GC content for our Dead Sea metagenomic datasets, for a suite of fully sequenced halophiles, and for two reference non-halophilic microorganisms (*E. coli* and *T. volcanium*). The hyperhalophilic bacterium *S. ruber* has not been identified in the Dead Sea nor is it represented in our 16S amplicon library. The acidic to basic amino acid (AB) ratio for *S. ruber*, though higher than both that of the non-halophiles and moderate halophiles, is still significantly lower than the AB ratio for the haloarchaea and the Dead Sea metagenomes. Thus bacterial genes transferred from *Salinibacter* to haloarchaea are identifiable by a salinity adaptation quotient elevated beyond  $\approx 1.2$ .

We also observe that the amino acid proportions of the two Dead Sea metagenomes do not correspond well with those of the archaeal halophiles. The 1992 Dead Sea metagenome encodes more arginine, lysine, histidine, and glutamic acid than any of the sequenced archaeal hyperhalophiles and the 2007 Dead Sea metagenome encodes more histidine and arginine than all six, more lysine than all but *H. walsbyi*, and more glutamic acid than all but *N. pharaonis*.

**Table B-1.** Amino acid and GC content of microorganisms and the Dead Sea metagenomes. The taxa name, acidic to basic amino acid ratio, percent encoded glutamic acid, aspartic acid, lysine, histidine, and arginine, and percent GC for the Dead Sea metagenomes (black) and seven sequenced halophiles (maroon; archaeal & orange; bacterial). *E. coli* and *T. volcanium* were added as reference non-halophilic microorganisms (blue).

	Acidic/Basic	Glutamic Acid	Aspartic Acid	Lysine	Histidine	Arginine	GC
Dead Sea 2007	1.46	8.63	8.41	2.23	2.35	7.09	67
Dead Sea 1992	1.24	9.57	7.87	2.44	3.32	8.35	62
<i>Natronomonas pharaonis</i>	1.73	8.82	8.74	1.86	1.96	6.35	63
<i>Haloarcula marismortui</i> (Dead Sea isolate)	1.62	8.07	8.30	2.00	2.00	6.10	62
<i>Halorubrum lacusprofundi</i>	1.60	8.28	8.80	1.74	1.92	7.00	67
<i>Halobacterium salinarum</i>	1.53	7.05	8.88	1.77	2.20	6.47	68
<i>Halobacterium</i> sp. NRC-1	1.50	6.97	8.86	1.78	2.23	6.53	67
<i>Haloquadratum walsbyi</i>	1.47	7.42	7.77	2.38	2.13	5.82	47
<i>Salinibacter ruber</i>	1.16	7.01	6.97	2.12	2.22	7.73	66
<i>Chromohalobacter salexigens</i> (moderate halophile)	0.98	6.19	6.09	2.50	2.54	7.45	64
<i>Thermoplasma volcanium</i> (Archaea)	0.91	6.36	5.48	6.86	1.50	4.67	39
<i>Escherichia coli</i> (Bacteria)	0.90	5.77	5.15	4.38	2.24	5.52	50

**Supporting Information B-2:**

We divided the amino acid profiles into three categories, acidic, basic, and other. Then, treating the amino acid profiles of the BLAST best hits as the expected values, we used a chi-squared test with two degrees of freedom to calculate the probability of observing the profiles of the reads by chance. The tables below display the taxon name, the taxon category, the chi-squared value, the SA quotient, the standardized ratio of 16S hits to metagenomic hits (16S/Meta), and the number of metagenomic hits. For the most part an SA value of 1.2 corresponded with at worst a 5% chance of observing salinity adaptation of the metagenomic amino acid profile by chance. However, large discrepancies between the total number of amino acids for the metagenomic reads and their homologues can lead to false positives. When possible, taxa were analyzed at the genera level. In cases where there was not sufficient reads tagged to a specific genera or when higher level taxa displayed an increased signal of halophilicity, the higher level taxa were provided. The results are shown in the following tables.

Taxonomic groups represented by metagenomic reads that do not show a significant increase in the acidic nature of their proteins are interpreted to represent either foreign taxa, halophilic taxa that use the salt-out strategy of salt tolerance, or lateral gene transfer events to salt-out halophilic taxa (Table B-2). Interestingly, of all taxa studied, only *Alkaliphilus* displays a statistically significant shift toward basic amino acids in the Dead Sea metagenome.

Taxonomic groups represented exclusively or nearly so by metagenomic reads (i.e., taxonomic groups with few amplicon 16s rRNA reads) and whose metagenomic

reads display a bias toward acidic amino acids are interpreted to represent possible cases where genes have been transferred into indigenous salt-in Dead Sea halophiles (Table B-3). Due to both extraction biases and PCR biases, the absence of a taxon in the amplicon library, however, does not necessarily imply its absence in the environment. For that reason alternative methods should be employed to confirm the classifications presented below. The results show a number of cases of bacterial genes transferred into the indigenous haloarchaea. However, the genera *Salinibacter* and *Thermotoga* stand out as having donated many genes through rampant lateral gene transfer with haloarchaea. They are represented by 191 and 105 metagenomic reads respectively, but neither is represented by a single 16S amplicon. Furthermore both genera show significant adaptation to salinity, have never been observed in the Dead Sea, and are likely participants in LGT with halophiles.

Taxonomic groups represented by both 16s rRNA amplicon reads from the Dead Sea and by metagenomic reads that show a bias toward acidic amino acids are interpreted as potentially representing lineages indigenous to the Dead Sea (Table B-4). A total of 14 traditionally non-hyperhalophilic taxa are identified in this manner. However included in this list are a number of taxa which for varying reasons are highly unlikely to inhabit the upper portion of the Dead Sea water column (red). These taxa include the anaerobes, *Clostridium* and Ectothiorhodospiraceae, the plant root symbionts, Rhizobiales, and the ammonia-oxidizing, *Nitrosococcus*, as nitrification does not proceed at high salt concentrations. Alternative explanations for the strong salinity adaptation signal for these reads include the possibility that they inhabit alternative

niches in the Dead Sea, are actually represented by both lateral gene transfer events into salt-in halophiles and by the presence of foreign DNA, or they represent misidentifications due to database bias. The remaining taxa (green) represent good candidates for novel salt-in bacterial halophiles, with it likely that many of them are indigenous to the Dead Sea ecosystem.

**Table B-2.** Foreign taxa and salt-out halophilic taxonomic groups

Taxa	Chi-Square	SA Quotient	16S/Meta	Metagenomic Reads
<i>Acidovorax</i>	0.53	1.08	0.74	23
<i>Acinetobacter</i>	4.99	0.98	0.19	378
<i>Aeromonas</i>	1.00	1.03	6.41	46
<i>Alkaliphilus</i>	17.74	0.88	0.42	50
<i>Arcobacter</i>	0.92	0.97	7.19	83
<i>Bacteroides</i>	2.14	1.03	2.97	62
<i>Blastopirellula</i>	0.03	1.00	0.00	19
Comamonadaceae	12.38	1.11	3.00	135
Enterobacteriaceae	8.69	1.08	2.70	56
Flavobacteriales	0.07	0.99	2.84	135
Neisseriaceae	1.11	1.09	7.00	26
Oceanospirillales	17.92	1.02	0.29	55
<i>Plesiocystis</i>	2.26	1.14	0.00	31
Propionibacterineae	5.67	1.03	0.21	19
<i>Pseudomonas</i>	0.16	1.02	3.85	60
<i>Psychrobacter</i>	0.63	1.01	0.36	22
Rhodobacteraceae	36.88	1.11	0.50	115
Rhodocyclaceae	2.39	1.08	2.14	57
<i>Roseobacter</i>	4.17	1.13	0.00	43
<i>Saccharopolyspora</i>	1.75	1.13	0.00	28
Sphingomonadaceae	6.25	1.17	0.75	36

**Table B-3.** Bacterial taxonomic groups represented through lateral gene transfer events into Dead Sea salt-in halophiles

Taxa	Chi-Square	SA Quotient	16S/Meta	Metagenomic Reads
<i>Chloroflexus</i>	89.64	1.56	0.00	42
<i>Frankia</i>	30.49	1.25	0.00	41
<i>Geobacter</i>	30.31	1.31	0.00	35
<i>Halothermothrix</i>	28.69	1.37	0.00	26
<i>Heliobacterium</i>	57.20	1.27	0.08	26
<i>Herpetosiphon</i>	88.54	1.61	0.10	40
<i>Marinobacter</i>	12.03	1.27	0.13	39
<i>Moorella</i>	40.88	1.54	0.00	46
<i>Mycobacterium</i>	23.55	1.34	0.04	48
<i>Roseiflexus</i>	251.69	1.59	0.00	98
<i>Rubrobacter</i>	218.13	1.47	0.00	257
<i>Salinibacter</i>	96.61	1.25	0.00	191
<i>Streptomyces</i>	42.24	1.53	0.00	30
<i>Thermotoga</i>	145.28	1.65	0.00	105

**Table B-4.** Salt-in bacterial taxonomic groups displaying evidence of being indigenous to the Dead Sea

Taxa	Chi-Square	SA Quotient	16S/Meta	Metagenomic Reads
<i>Azoarcus</i>	6.15	1.29	5.21	19
Bacillaceae	262.49	1.54	0.22	233
Burkholderiaceae	34.00	1.24	0.77	78
Lactobacillales	14.29	1.33	0.92	24
<i>Methylobacterium</i>	41.16	1.32	0.29	41
Micrococcineae	54.82	1.32	1.10	40
Phyllobacteriaceae	39.56	1.44	0.69	29
Rhodospirillales	32.64	1.70	0.20	20
<i>Synechococcus</i>	50.49	1.64	0.57	21
<i>Thiomicrospira</i>	36.42	2.34	0.30	23
Verrucomicrobia	12.01	1.47	0.53	17
<i>Clostridium</i>	112.42	1.53	0.94	94
Ectothiorhodospiraceae	35.18	1.43	0.19	37
<i>Nitrosococcus</i>	44.31	1.44	2.61	28
Rhizobiales	116.78	1.34	0.43	209

## Appendix C

### Supplemental Information for Chapter 4

#### Supporting Information C-1

Additional file 1 depicts a collection of 160 trees representing all LGT genes with BLAST scores greater than 500 in both the Thermoprotei and Halobacteria. The KEGG database three letter genome code is given before the colon and can be found here [http://www.genome.jp/kegg/catalog/org\\_list.html](http://www.genome.jp/kegg/catalog/org_list.html). The corresponding gene locus tags are provided after the colon. The collection is too large for reproduction in this manuscript. It can be accessed as part of the supplemental information in association with the published manuscript *BMC Evolutionary Biology* 2011, **11**:199doi:10.1186/1471-2148-11-199.

**Supporting Information C-2****Table C-1.** Table providing the proportion of LGT events originating from a donor inhabiting a similar environment for a spectrum of BLAST cutoff percentages

	90%	80%	75%	70%	60%	50%
Halobacteria	10.5	10.4	10.6	10.7	10.7	10.9
Thermoprotei	69.6	67.8	67.9	68.3	68.1	68.3

### Supporting Information C-3

**Table C-2.** Table showing the 22 LGT events identified within environmental fosmid clones of the Dead Sea. The DNA sequences are too large to be reproduced here. They can be accessed as part of the supplemental information in association with the published manuscript *BMC Evolutionary Biology* 2011, **11**:199doi:10.1186/1471-2148-11-199.

Closest Species	Halophilicity	BLAST Score of Best Hit	GI of Best Hit
Azoarcus sp.	no	637	56476760
Thermomicrobium roseum	no	395	221633818
Rubrobacter xylanophilus DSM	no	375	108803077
Haliangium ochraceum	yes	365	262198659
Ralstonia pickettii 12D	no	335	241661738
Desulfitobacterium hafniense Y51	no	319	89892989
Burkholderia multivorans ATCC 17616	no	301	189353841
Pseudomonas fluorescens	no	267	77460000
Thermobaculum terrenum ATCC BAA-798	no	258	269839071
Methanosarcina mazei	no	239	21226751
Geobacter sp. M21	no	238	253700569
Rhodococcus erythropolis	no	231	226308790
Methanothermobacter thermautotrophicus str. Delta H	no	206	15679216
Cyanothece sp. PCC 7822	no	178	307155289
Chromohalobacter salexigens DSM 3043	yes	174	92115006
Cronobacter turicensis z3032	no	153	260596449
Geobacter uraniireducens Rf4	no	145	148262178
Chloroflexus sp. Y-400-fl	no	107	222523386
Pyrococcus horikoshii OT3	no	99.8	33359265
Pyrococcus horikoshii OT3	no	78.6	33359265
Heliobacterium modesticaldum Ice1	no	69.7	167630419
Methanoculleus marisnigri JR1	no	65.1	126178014

## VITA

### Matthew Elliot Rhodes

#### EDUCATION

**The Pennsylvania State University**, University Park, PA 2006-present  
Candidate for a Doctoral degree in the geosciences, Dual title astrobiology

**Cornell University**, Ithaca, NY 2000-2004  
B.A in Mathematics and Computational Biology

#### AWARDS

- USDA-AFRI Training Grant Fellowship 2010-2011
- Tait Scholarship in Microbial Biogeochemistry 2008-2009
- Biogeochemical *Research* Initiative for Education (*BRIE*) Fellowship 2006-2007
- Anne C. Wilson Graduate Research Award 2006-2007

#### PUBLICATIONS

- **Rhodes, M. E.**, Oren, A., and House, C. H., "The Rise and Fall of Dead Sea Microbial Blooms." Under Review by *Applied and Environmental Microbiology*.
- Frank, L., and others including **Rhodes, M. E.**, "Scientific Background." In, *Human Microbiome Research: Ethical, Legal and Social Implications*, Edited by R. Rhodes, N. Gligorov, and A. Schwab. Commissioned by Oxford University Press, New York, NY.
- Martino, A. J., **Rhodes, M. E.**, Biddle, J. F., Brandt, L. D., Tomsho, L. and House, C. H., "Development of a Degenerate PCR Method for Whole Genome Amplification of DNA from Very Low Biomass Marine Subsurface Sediments" Under Review by *Frontiers in Extreme Microbiology*.
- Rhodes, R., and twenty-three others including **Rhodes, M. E.**, "De Minimis Risk: A New Category of Research Risk." Under Review by the *Journal of the American Medical Association*.
- **Rhodes, M. E.**, Spear, J. R., Oren, A., and House, C. H., 2011, "Differences in Lateral Gene Transfer in Hypersaline versus Thermal Environments," *BMC Evolutionary Biology*, 11:199
- **Rhodes, M. E.**, Fitz-Gibbon, S., Oren, A., and House, C. H., 2010, "Amino Acid Signatures of Salinity on an Environmental Scale with a Focus on the Dead Sea," *Environmental Microbiology*, 12:2613-2623
- Bersaglieri, T., Sabeti, P. C., Patterson, N., Vanderploeg, T., Schaffner, S. F., Drake, J. A., **Rhodes, M. E.**, Reich, D. E., and Hirschhorn, J. N., 2004, "Genetic Signatures of Strong Recent Positive Selection at the Lactase Gene," *American Journal of Human Genetics*, 74:1111-1120.