Phenotypic and genomic stability in a halophilic model organism

Karlyn D. Beer

A dissertation

submitted in partial fulfillment of the

requirements of the degree of

Doctor of Philosophy

University of Washington

2013

Reading Committee:

Benjamin Kerr, Chair

Nitin Baliga

Jody Deming

Program Authorized to Offer Degree:

Molecular and Cellular Biology

University of Washington

Abstract

Phenotypic and genomic stability in a halophilic model organism

Karlyn Doraldyne Beer

Chair of the Supervisory Committee:

Assistant Professor Christine Queitsch

Department of Genome Sciences

Although microbial adaptation to new environments has been studied extensively in laboratory evolution experiments, less is understood about how natural strains adapt to new, constant laboratory conditions after transitioning to become model organisms. In this dissertation, I took a cross-sectional approach to query the osmotolerance repertoire of a decades-old model halophilic archaeon compared with recently isolated conspecific natural isolates. Using growth rate reaction norms and whole genome sequencing, I found that fitness is highly variable across gradients in total salinity and ion composition for both laboratory and natural strains. Although we expected divergent reaction norms and genome sequences to support evidence of specialization in the model organism, both the model and natural isolates shared similar growth patterns across conditions and shared >95% sequence similarity. In the following chapters, I characterize the hypersaline environment in the Great Salt Lake, quantify osmotolerance phenotypes across three gradients in salinity, and I explore gene expression

signatures across an ion composition gradient in *H. salinarum* NRC-1. I close by comparing fitness profiles and genome sequences of the model and natural strains, and I address the unexpected similarity among strains, even after eliminating contamination as an explanation. Ultimately, despite at least 50 years of evolutionary opportunity between the lab and natural isolates, we find that osmotolerance phenotypes and nucleotide sequence is largely maintained in the laboratory, and that a decades-old model organism remains a relevant study organism for microbial-environmental response.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I thank all members of the Baliga Lab at the Institute for Systems Biology for their friendship, support and advice. I have been intensely challenged intellectually, and also unwaveringly encouraged along the way. In particular, I thank Elisabeth Wurtmann, Nicolás Pinel, Amardeep Kaur and Min Pan for their hard work and thoughtful collaboration on this project. I thank Fang Yin Lo, Aaron Brooks, Serdar Turkarslan, Mónica Orellana, Dave Reiss, Chris Plaisier, Claudia Ludwig, Chris Bare, Lee Pang, and Danielle Durudas for insight, friendship and conversation over the years. I am especially grateful to Nitin Baliga for his patience, scientific insight and training in the elusive art and skill of Thinking Deeply. I am grateful for his willingness to take on a student who had little computational experience and whose scientific interests seemed far afield of our lab's work, and for the flexibility I had to follow my interests in shaping my project.

ISB was a place I enjoyed being every day, because of Allison Lee, Cora Chadick, Michelle Hays, Zhihao Tan, Greg Cary, Debbie Chang, Stephanie Skelton and all of the graduate students and postdocs who have come together to make this a fun and vibrant place to learn.

I am immensely grateful for my advisory committee, Christine Queitsch, Maitreya Dunham, Benjamin Kerr, Jody Deming and Herbert Sauro. I could not have asked for a more critical yet supportive group of faculty to keep and eye on my progress during graduate school. They kept me challenged and excited for the work ahead, but always helped me to see reasons to be proud of my accomplishments even when I could not.

I thank our collaborators in Salt Lake City, Utah, without whom I would not have been able to incorporate environmental samples and context into this project: David Naftz and Mike Freeman (USGS), Bonnie Baxter and Jaimi Butler (Westminster College). I thank Goodwin Gibbins for her work on the Growth Curve Analysis Function R package, and Brooks Miner and Mónica Orellana for lending me field equipment for the Utah field trip. Special thanks again to Cora Chadick and Noel Blake for enabling the Hamilton Robot to be compatible with Bioscreen plates. The staff of the MCB program, including MaryEllin Robinson, Terry Duffey, Diane Darling, Milli Morris and Maria Sanders were wonderfully supportive of me during my time in graduate school, and I thank them all of their work and help.

Most importantly, I thank my friends and family. I thank Simina, Sarah, Leslie, Kyle, Nisha, Emily P., and Emily K. for their friendship and empathy as we all navigated this process together. I thank Diane and Charles for their mentorship and friendship, bikerides and Halloween parties. My family, Jennifer, Mike and Kendal have been supportive the whole way, and have always reminded me that I can do whatever I set my mind and heart to. I have them to thank for the tenacity/stubbornness that got me here. Finally, Ed Gano deserves a world of gratitude for his unwavering support, love and friendship, for keeping me so well fed, and for reminding me that I can have a valuable impact on the world.

Chapter 1: Introduction

Abstract

The evolutionary success of an organism is a testament to its inherent capacity to keep pace with environmental conditions that change over short and long periods. Mechanisms underlying adaptive processes are being investigated with renewed interest and excitement. This revival is partly fueled by powerful technologies that can probe molecular phenomena at a systems scale. Such studies with microbes provide spectacular insight into the mechanisms of adaptation, including rewiring of regulatory networks via natural selection of horizontal gene transfers, gene duplication, deletion, readjustment of kinetic parameters, and myriad other genetic reorganizational events. A fundamental understanding of adaptation is needed to guide our activities on a planet with limited resources, on which we are but one among millions of species biologically wired to compete and persist. Here, I review our understanding of microbial adaptation, mechanisms of adaptation, and the challenges associated with understanding adaptation in the natural environment and in the laboratory. I will end with a review of prokaryotic osmotolerance strategies and what is known about salinity adaptation in extreme halophilic archaea.

Introduction to microbial adaptation

The following text and figures contain sections slightly adapted from a review article published in Wiley Interdisciplinary Reviews of Systems Biology and Medicine: Brooks, A. N., Turkarslan, S., **Beer, K. D**., Yin Lo, F. & Baliga, N. S. Adaptation of cells to new environments. *Wiley Interdiscip Rev Syst Biol Med* (2010). doi:10.1002/wsbm.136.

Microorganisms experience myriad environmental factors over their evolutionary history, including those that remain essentially constant over long periods (e.g. geological epochs), change slowly (e.g. general increase in annual temperatures), fluctuate periodically (e.g. day-night cycles and seasonal variations), or change frequently and somewhat randomly (e.g. unpredictable nutrient loading). These changes occur over diverse timescales, ranging from the lifetime of an individual cell to multiple generations. Accordingly, microbes have evolved unique strategies to deal with the peculiarities of their environment[1]. Characteristic examples include adaptation to extreme environmental niches[2,3], entrainment of phototrophs to day-night

cycles[4,5], and the physiological adjustment of *E. coli* as it passes through the human intestine[6]. Organisms respond to short term environmental changes by reversibly adjusting their physiology to maximize resource utilization while maintaining structural and genetic integrity by repairing and minimizing damage to cellular infrastructure[7,8] thereby balancing innovation with robustness. Naturally, physiological response networks emerge as products of evolution by natural selection where they can lend reproductive or fitness advantage, particularly when the environmental change is recurrent[9].

Environmental adaptation of biological systems can be considered from three evolutionary perspectives: (i) acclimation of existing cellular machinery to operate optimally in a new environmental niche; (ii) acquisition of entirely new capabilities through horizontal gene transfer or neofunctionalization of gene duplications and (iii) reorganization of network dynamics to appropriately adjust existing physiological processes to match dynamic environmental changes. The first type of adaptation can arise through two types of events that differ dramatically in duration. Simple mutations can greatly increase fitness over very short time frames (within one or few generations). Prominent examples of short-term adaptive events include resistance to drugs[10,11] and altered nutrient conditions[12]. Alternatively, complex mutations in multiple loci may accumulate over very long time frames, such as the evolution of acidic protein surfaces in halophilic archaea[2,13,14]. While the initial transfer of adaptive genes by HGT occurs rapidly[15], full integration of laterally transferred component(s) typically occurs over longer time frames (10s of millions of years), where HGT events often require regulatory rewiring to function optimally in the context of existing cellular networks[16]. Finally, physiological readjustment occurs both because of genetic and physiological robustness to withstand stress that accumulates over many generations and latent genetic variance that is revealed after environmental perturbation[17].

This introduction focuses on the evolution of adaptive mechanisms for acclimation to recurrent yet transient environmental changes. When transient changes are recurrent they select for genetic traits that confer fitness by improving the ability of an organism to rapidly and reversibly adjust physiology to match current environmental conditions. These traits manifest at varying hierarchies of genetic information processing, from receptors for sensing environmental factors, to signal relay, transcriptional, post-transcriptional, translational and post-translational control mechanisms, and also at the metabolic level through modulation of enzyme function

(affinities, kinetics etc.). Such adaptive changes occur over intermediate time frames (upwards of 100s of generations in *E. coli* K-12 MG1655; [18]) and, surprisingly, they arise repeatedly and in some cases with some regularity in distinct lineages[19]. Fitness, or the number of surviving offspring after one generation[20], is a complex property that emerges from the integration of changes at all these levels. A holistic systems approach, therefore, is necessary to fully appreciate how these varied mechanisms work together when an organism adapts to a new environment.

For the purpose of our discussion, we define environmental conditions to include both abiotic physical variables (such as light and temperature) and biotic components (such as other co-inhabiting organisms). Additionally, we restrict our analysis to asexual prokaryotic systems (including both bacteria and archaea) in which the dynamics and mechanisms of genetic evolution lack the pervasive variation that recombination by sexual reproduction promotes. Sexual populations also may not experience signatures of selection prevalent in asexual populations, such as 'classic' sweeps associated with unconditionally advantageous mutations[21]. The reader should note, however, that the physiology of archaea and bacteria diverge substantially, with archaea sharing startling similarity to eukaryotes. Infrequently, we may explicitly refer to findings in eukaryotes that reflect mechanisms that likely also occur in prokaryotic systems.

This introduction will bridge the conceptual gap between *adaptation,* which by definition requires heritable genetic change[20], and *physiological readjustment*, which is a product of adaptation that equips organisms to attune their physiology to dynamic changes in their environments. We will suggest how systems-level methodologies and insights can be applied to better understand the strategies living systems employ to withstand and in some cases take advantage of change in their environments.

The fields of microbiology, molecular evolution, and systems biology are expansive – no single review can possibly cover all adaptive mechanisms and scenarios that may influence the evolution of natural microbial populations. Instead, we will highlight major themes and new insights in microbial evolution while demonstrating how the principles of systems biology can be leveraged to develop a more comprehensive, integrative understanding of cellular adaptation to

new environments. Throughout our analysis we will point the reader to other outstanding reviews that complement our discussion.

TYPES OF ADAPTATIONS AND ASSOCIATED MECHANISMS

What is adaptation?

When we say that an organism has *adapted* to its habitat, we imply that it has evolved molecular mechanisms that allow it to grow optimally under the spatiotemporally varying physicochemical conditions of its environment. Evolution, however, is an unfinished process. Organisms chase fitness optima in constantly changing environments. Subtle fitness differences between individuals (due to genomic plasticity or metabolic flexibility) and phylogenetic complexity (i.e. numbers and diversity of species within a community) can lead to the diversification of species or the extinction of less fit genotypes over time[22]. Although many adaptive mutations are lost by random chance (as a result of genetic drift), mutations that confer significant selective advantage have a greater propensity to become fixed within the population, especially in large populations[23]. If selection imposed by the environment is particularly strong, fitness-enhancing genotypes will rapidly rise in frequency in the population, often carrying associated, possibly detrimental, genes along with them (i.e. selective sweep)[24] and interfering with one another via clonal interference[25].

Adaptation to linked environmental changes: general stress response or anticipatory behavior?

Conditions within natural environments can change continually over long periods, periodically, or transiently and unpredictably. In the context of evolution, these environmental changes occur simultaneously albeit on different timescales and exert selective pressure to enrich genotypes well matched to particular ecosystems. Not surprisingly, the repertoire of analogous solutions that characterizes success in a given environment can be similar across diverse organisms occupying similar habitats, suggesting convergent evolution of adaptive solutions that were discovered independently in divergent lineages[26,27]. While a significant fraction of these responses are condition-specific, most organisms have also evolved robust generalized

mechanisms to deal with shared aspects of stress resulting from diverse kinds of environmental changes. In yeast, for example, a set of ~900 genes responds similarly to a diverse array of environmental stresses, sharing common regulatory themes mediated by Yap1p, Msn2p, and Msn4p[28]. This generalized stress response typically includes activation of heat shock proteins, phage shock proteins, and oxidative stress response proteins[29,30], although there are notable examples in *Candida albicans* and *Halobacterium salinarum sp. NRC-1* where the central role of general stress response has been called into question[31,32]. The alternate perspective offered by these studies is that changes in environmental variables are physically linked and do not occur in isolation. Elevated temperatures, for example, induce a number of associated changes in environmental conditions, including decrease in oxygen solubility[33]. Theoretical and experimental work in diverse species suggests that organisms can learn to take advantage of this natural co-variation between environmental parameters (e.g. temperature and oxygen), thereby displaying "anticipatory behavior"[18,33,34]. *B. subtilis*, for example, retains short-term and long-term memories to inform sporulation dynamics and degradative enzyme synthesis[34].

Coupled environmental variables have important evolutionary implications that can be assessed by assaying the fitness consequences upon artificially decoupling such associations in the lab. Relative to laboratory populations evolved in an inverted environment, wild-type *E. coli* has a fitness disadvantage when naturally linked parameters, such as temperature and oxygen, are decoupled artificially. Likewise, cyanobacteria mutants with non-functional circadian clocks are less fit and out-competed by their wild-type counterparts[35]. The enhanced fitness conferred by diurnal entrainment can be attributed to anticipation of predictable, associated stresses such as damaging UV radiation[4]. Adaptation to stress may also prepare cells to better respond to future stresses. *Vibrio parahaemolyticus*, for example, has a greater tolerance to acidity and temperature stresses following growth in media containing 3% NaCl versus 1% NaCl concentrations[36,37]. On the other hand, adaptation to a subset of environmental factors could come at a cost of sacrificing tolerance to others. Propionate adaptation of *Salmonella enteritidis*, for instance, leads to enhanced resistance to stresses experienced inside the host but overall decrease in infectivity[38]. For all these reasons, we need to appreciate the high degree of connectedness within environmental networks to interpret causes and functional consequences of complex biological responses.

LONG-TERM ADAPTATION

Some environmental conditions seldom change. We define long-term adaptation as the response to environmental conditions that remain relatively constant throughout the lifetime of a species. For organisms, this represents the most fundamental level of adaptation. Over many thousands to millions of years or longer, the internal architecture of the cell changes to accurately match the general features of the habitat it occupies. Typically, the features that reflect long-term environmental adaptation are deeply ingrained in the structure of the genome, regulatory schemas, or even the molecules of the cell themselves. These molecular artifacts reflect the organism's uphill struggle towards increased fitness in an environment that changes (being directly modified by biotic factors and even the organism itself) yet remains mostly constant over long timescales. Complete reversion of these features would be difficult, if not impossible, for an organism to achieve over short time scales, suggesting that cells are poorly adapted to large variations in these environmental parameters. *Halobacterium salinarum* NRC-1, for example, thrives in high salt conditions. While *H. salinarum* can withstand lower salt concentrations (as low as 2.5M), anything below this concentration is lethal to cells. Adaptation to high salt conditions requires a number of significant structural changes in the cell, including global alterations at the level of DNA, RNA and protein composition[39]. In high salt, water activity is low; this has profound consequences for enzymatic activity and the structural integrity of the cell membrane and genome. Adaptation to high-salt conditions has required the evolution of a highly-acidic proteome, high genomic GC content, and increased intracellular concentration of potassium cation[40]. Other factors including gene redundancy have been reported to enhance survival in fluctuating salt conditions. *Salinibacter ruber*, for instance, possesses two or more copies of several essential genes. It has been proposed that slight differences in the amino acid composition of two versions of the same protein (ecoparalogs) might allow *Salinibacter* to survive broader fluctuations in salt concentration[41]. These strategies for thriving in hypersaline environments are generally utilized by diverse halophillic archaea. Halophilic bacteria and eukaryotes, on the other hand, have independently evolved alternate mechanisms such as overproduction of organic osmolytes (sugar and amino acid-derivatives) to live in high salt environments[40]. Adaptation to temperature is another well-studied example of molecular response to long-term environmental pressures. While all microbial species are adapted to some

range of permissive temperatures, interesting mechanistic examples of thermal adaption have been reported in both extreme warm and cold environments. Psychrophillic organisms occupy extremely cold ecosystems (permanent temperatures below 5C) located deep in the ocean, and in polar and alpine regions. The habitats colonized by psychrophiles are of particular interest because they constitute more than three-quarters of the Earth's surface[42]. Similarly, hyperthermophiles (optimal growth temperature >80C) occupy high temperature habitats. These organisms are so well adapted to their environments that, under some conditions, their doubling times approach that of *E. coli* grown at 37C[42]. In both cases, the microorganisms that occupy these habitats have evolved enzymes and metabolic strategies that allow them to survive and proliferate at temperatures that would be restrictive or lethal to mesophilic organisms. To withstand these extreme temperatures, microbial species have evolved molecular mechanisms that regulate membrane fluidity and conformational flexibility of proteins, and produce thermo-stable protein variants, all of which are significantly challenged at extreme temperatures[43,44]. While the specific mechanisms employed to combat extreme temperatures vary across species, common mechanisms, such as increased protein stability through ion pairs, hydrogen bonding, hydrophobic interactions, disulfide bridges, packing, and intersubunit interactions at high temperatures and increased protein flexibility and membrane fluidity and exopolysaccharide production at low temperatures, are common aspects of cellular adaption to extreme environments[45]. Like the halophilic organisms described earlier, thermophiles and psychrophiles occupy environments that impose a number of additional harsh constraints on living systems (including metal ion concentrations, nutrient limitation, and sometimes increased pressure), making them useful models to understand adaptation to stressful conditions.

SHORT-TERM ADAPTATION

Environmental variables can fluctuate in unanticipated ways. Such variations are often stressful and, depending on severity, might drive selection of genotypes that are better suited to readjust physiology to manage and mitigate the consequences of stress. Termed “Phenotypic plasticity,” the ability of a single genotype to express multiple phenotypes in response to stress is a selectable trait itself. (Phenotypic plasticity as a measure of adaptation will be reviewed in more detail below.) Genotypes can be selected on the basis of either genetic or non-genetic components. Asymmetric cell division, in which mother and daughter cells receive

disproportionate numbers of molecules, alters the dynamic behavior of cells in response to environmental change and may lead to the selection of genotypes that do not necessarily contain heritable allelic alterations, but rather exhibit a temporary, dynamic state compatible with the altered environment. Sporulation in *B. subtilis*, for instance, couples asymmetric morphological changes with differential gene expression between the forespore and mother cell (Reviewed in [46]), leading to divergent cell fate. Due to asymmetric cell division, daughter cells may receive more or fewer ribosomes, transcription factors, or other cellular components, each of which might contribute to their success (or failure) in new environments.

Phase variation (or phenotype switching) is another mechanism (non-genetic, though heritable alteration) by which microbial populations leverage stochastic variations in cell components to respond to uncertainty in environmental fluctuations[47]. Genetic alterations, on the other hand, can occur anywhere within the cellular hierarchy; advantageous mutations may be located within proteins, affecting the stability or kinetic parameters of the protein, or within regulatory sequences, affecting when, where, and how much of a biological product is made. A common feature of these alterations, however, is that they tend to be simple, i.e. they are the product of one or a few adaptive mutations that spontaneously occur, or preexisting neutral or buffered mutations whose consequence is revealed by perturbation. In this sense, these adaptations are flexible, being easily gained and subsequently lost by genetic drift. The inherent plasticity of these mutations makes them especially important for physiological adaptation, as slight changes can drastically alter the dynamic response of an organism to stress. Short-term adaptation is tightly coupled to the longer-term adaptive mechanisms previously described. Temporary changes in long-term environmental trends may elicit genetic alterations that are short-lived, only becoming fixed within the population and canalized into regulatory programs if the stress that has elicited the advantageous mutation surpasses a temporal threshold, becoming a regular feature of the environment.

Numerous studies have investigated mechanisms of adaptation to altered growth temperature[48,49], nutrient composition[50-52] and population structure[53]. Universally, these studies find that fitness to a new environment increases rapidly over the first several thousand generations[54,55]. Surprisingly, adaptive mutations discovered in the lab typically occur in one or a few genes, reflecting the earliest events in the adaptive process. Only four mutations in *E. coli*,

for example, are responsible for gaining growth advantage in stationary phase: one in the stationary-phase sigma factor , one in the leucine-responsive regulatory protein, two genomic rearrangements of IS5 transposon insertion sequences, and a mutation in the sgaC gene[53].

Short-term, stressful environmental perturbations may also specifically induce increased rates of mutation, potentially facilitating adaptive evolution by rapid exploration of a broader genotypic space. In response to nutrient starvation, for example, the DNA damage and cell-cycle checkpoint control response (SOS response pathway) is induced in *E. coli*. Following induction of this pathway, cells experience higher rates of mutation due to inhibition of mismatch repair and recombinational break repair, and induction of a mutator DNA polymerase[56]. In addition, some genetic loci exhibit a disproportionatly high frequency of mutation during hypermutation. These mutational 'hot' spots may reflect regions of the genome that are more readily modified and thereby *adaptive.* Taken together, these observations suggest that organisms might possess the capacity to introduce a bias in the distribution of mutational variation along their chromosome(s).

ADAPTATION THROUGH REWIRING OF REGULATORY NETWORKS

Systems-level coordination of cellular functions is accomplished by gene regulatory networks (GRNs)[57]. The general form and features of biological networks are illustrated in Figure 1-1. Central to all GRNs is the interaction of TFs and their cognate DNA binding sites. Remarkably, the origin of DNA-binding domains (DBDs) in all present day TFs can be traced to a few ancestral classes, such as winged-helix and zinc ribbon domains[36]. This raises important questions regarding the evolutionary process(es) that led to the diversification of these few DBDs to create a vast array of distinct DNA binding specificities. It could be argued that the common origins of DBDs within TFs and their division into related classes (protein families) should help characterize GRNs in one organism and suggest projections of that information onto orthologous systems in phylogenetically related species. Functions of even structurally similar transcription factors (TFs), however, can diverge substantially through alterations in regulatory domains of either the transcription factor itself or the cis-regulatory sequences of downstream target genes. The divergence of DNA binding specificities and allosteric domains of two FNR (fumarate and nitrate reduction) family TFs in *E. coli* and *B. subtilis* are case in point. As a

result of subtle DNA binding differences, FNR in *E. coli* controls 135 genes whereas its counterpart in *B. subtilis* regulates only 8 genes[58]. Importantly, this result demonstrates that species-specific coevolution of interacting partners (in this case FNR and the target promoters) prohibits simple projection of TF-binding orthologies across species, even when the genes involved share recent ancestry. On the other hand, this finding underscores the flexibility of GRNs – malleability of regulatory network topology can promote variation in gene expression that acclimates a species to the nuances of its environment[59].

Phenotypic plasticity as a mode of adaptation

PHENOTYPIC PLASTICITY DEFINED

Phenotypic plasticity describes the environment-dependent physiological adjustments reviewed above, and is technically defined as the ability of a single genotype (as individual or isogenic population) to produce multiple phenotypes in response to environmental change[60]. For example, sessile organisms like plants depend on their ability to change the orientation and shape of their leaves when key resources like sunlight and water vary in their availability. The ideal organism would be infinitely plastic as to match its phenotype with its environment under every condition, to maximize fitness. This ideal of course does not exist, and so the tradeoff between plasticity and robustness becomes apparent: organisms must balance the ability to adjust physiology when the environment changes with the risk of adjusting too slowly, and suffering the mismatch of environment with non-optimal physiology. Robustness, or resistance to change, is important for surviving environmental change that occurs more quickly than physiology can accommodate.

The field of phenotypic plasticity arose from the need to partition total phenotypic variation into that attributable to genotype and other external factors. Of the external factors, contributions from the environment, stochastic noise and error had not been previously distinguished from environmental effects, nor had the effect of genotype × environment interactions. The importance of a systematic environmental effect and an interaction – i.e. the idea that the environment's effects are different for certain genotypes over others – opened up the gates for studies of gene regulation (described above) as the mechanism of gene × environmental interactions, and of plasticity itself as a selectable trait. For a thorough review of the genetics underlying selection for plasticity, see Garland *et al* 2006. As a powerful means of

adaptation, the 'how' and 'why' questions surrounding plastic phenotypes are increasingly interesting and important for biology.

Plasticity is measurable as the continuous change in a phenotypic trait value vs. change in environment. Such a plot, called a reaction norm (Figure 1-2), allows plasticity and niche breadth to be quantified, parameterized and compared[60]. The Y-axis on reaction norm plots can measure any physical, behavioral or genetic trait, or it can measure a component of fitness (growth rate, number of offspring). Collectively, these are called "performance." The X-axis is an environmental variable, and may be discrete or continuous. Therefore, "plasticity" broadly describes any change in phenotype with environment, and manifests as a non-horizontal line on the plot.

However, it is important to mention that the meaning of plasticity differs slightly, depending on the phenotype being measured. It is useful to distinguish between specific traits (morphology, behavior, gene expression) vs. phenotypic components of fitness (growth rate, number of offspring)[61,62]. This is because fitness defines the adaptiveness of other traits, and a low value of fitness on a reaction norm plot indicates the presence of traits maladapted for that environment. By contrast, a low value of a particular trait (beak length, leaf shape) on a reaction norm plot may not give any information about the fitness of that phenotype. This also leads to internal semantic inconsistencies when describing "high" and "low" plasticity. On fitness-based plots, a genotype whose fitness changes little across environments (*i.e*. a flat line) is called highly plastic, whereas trait-based plots with large changes in phenotype (*i.e.* non-flat line) are also called highly plastic. Put another way, an organism with a highly plastic morphological trait might be similarly fit across many environments due to the plasticity of that trait.

The generic use of "plasticity" in the literature for both trait- and fitness-based reaction norms has created some confusion, despite the unifying concept of environmental adaptation through physiological flexibility. For microbes especially, where non-fitness based traits are more difficult to measure, plasticity is often considered in terms of growth rate and other survival-based measurements[50,63-67]. These microbiology studies discuss the role of reaction norms and plasticity in environmental adaptation just as botanists and vertebrate biologists do, but they may have inadvertently circumvented the confusion simply by choosing not to classify plasticity as high vs. low.

Another way authors have reconciled the discrepancy between trait- and fitness-based plasticity studies is to consider fitness-based reaction norms separately, as "tolerance curves," where changes in fitness across environments relate to the degree of environmental tolerance or *niche breadth*[62]. The shape of such a curve represents the degree to which an organism is a specialist, well-adapted to a small set of conditions (narrow niche), or a generalist with comparable growth in a wide range of environments (wide niche) (Figure 2-1F). Although asking how niche breadth is adaptive in different environments is a fundamentally different question than asking how phenotypic variation is adaptive across environments, the two concepts are intimately related. Relating tolerance curves and niche breadth back to phenotypic plasticity, a genotype with more highly plastic traits is more likely to be a generalist, with similar fitness in a wide range of environments[61].

The remainder of this dissertation will focus specifically on plasticity in fitness-based reaction norms (tolerance curves), and questions surrounding the evolution of generalists and specialists in response to environments with varying complexity. I will use the term "plasticity" to describe the degree to which any trait, including components of fitness, vary across environments, and I will define it as required in terms of generalist and specialist strategies to avoid confusion over high vs. low plasticity values.

IS PLASTICITY ADAPTIVE IN VARIABLE ENVIRONMENTS?

All organisms are under strong selection to evolve strategies for dealing with environmental change. These strategies differ, depending on how frequently and strongly a given environment varies, but no organism can perfectly match its physiology to the environment for optimal fitness at all times. These inevitable mismatches between physiology and environment produce fitness tradeoffs[68-70]. Selection for high fitness in one prevailing condition comes at the cost of reduced fitness in other environments encountered more rarely. Conversely, if an organism experiences different fluctuating environments, theory and data predict that generalists with average fitness among all environments will prevail[63,67,71,72]. Given that costs are associated with plasticity and fitness across a range of environments, a trade-off between maximum fitness and niche breadth is also expected. Nevertheless, the data supporting these tradeoffs are not unequivocal.

In studies with microbes, experimental evolution has been applied to test the hypothesis that a change in plasticity in the form of generalist-specialist transitions occurs during evolution in constant vs. variable environments. Using Chlamydomonas, cultures of light- and dark-adapted cells were passaged in each of constant light, constant dark, or fluctuating light and dark for 200 generations. As expected, the constant-adapted cultures had specialized to their growth conditions, growing much more densely in one vs. the other. The cultures passaged in fluctuating light and dark grew equally well in both conditions compared with the ancestor, indicating the emergence of generalists[67].

In addition to light, thermal and nutrient fluctuation have been used to explore the evolution of specialists and generalists in constant and variable environments, respectively. In *E. coli*[64,65], bacteriophage[50] and *P. fluorescens*[71], specialists emerged after evolution in a single condition and generalists emerged when conditions fluctuated over time. While generalists and specialists did emerge as expected, the degree to which niche breadth comes at the expense of maximum fitness was not consistent across studies.

The Chlamydomonas studies produced generalists, but these evolved cultures grew just as well in the light and dark as the constant light- and dark-adapted strains[66,67]. The observation that generalism does not come at the cost of optimal fitness has also been observed in multicellular organisms[62,73], but will not be reviewed here. Nonetheless, In Lenski's *E. coli* thermal adaptation experiments, specialization to a single temperature did come at the cost of niche breadth[65,74]. The broad variation across studies of the cost of becoming a generalist in fluctuating suggests that while environmental heterogeneity is necessary to produce generalists, it is not sufficient to induce tradeoffs between niche breadth and maximum fitness. Thus, the production of generalists within fluctuating environments must proceed through several different mechanisms.

The mechanisms underlying the evolution of generalists must support this same outcome both through fitness costs and without them. Briefly, mechanisms that incur costs of adaptation may arise from two sources. The first is antagonistic pleiotropy (AP), in which genes favorable in one environment are deleterious in other environments. The second is the mutation accumulation (MA), or mutations that are neutral in the environment of selection and deleterious elsewhere[40,69,72]. How might generalists arise without incurring fitness costs? First, it is possible that trade-offs (evidence of cost) exist, but were not detected experimentally due to populations

out of equilibrium. The experiments described here were performed by passaging batch cultures and the sampling scheme may have masked any deleterious effects of AP or MA. Secondly, fitness costs for generalists vs. specialists may not be observed if MA is the dominant mechanism and the deleterious effects are not manifest in the experimental conditions. In this case, specialists would not enjoy an advantage in either constant environment, leaving them at the same fitness level as the generalists[72].

In summary, the change in plasticity that enables variable environments to produce generalists is adaptive, as organisms evolve to withstand dynamic exogenous stresses. Generalists are expected to emerge when environments fluctuate, whether or not they suffer a reduced fitness maximum. These insights are derived from powerful experimental evolution studies in the laboratory, but are limited in their generalizability to natural systems. The following section will discuss the importance of understanding adaptation to changing environments in natural systems, and in transitions from natural to laboratory environments. I will also discuss the inherent challenges and possible experimental approaches to studying adaptation in natural systems.

Challenges and approaches for studying adaptation and plasticity in natural systems

Evolution underlies all of biology, but is inherently difficult to investigate because of the time required to measure its effects. This problem has been circumvented by experimenting with microbes, due to their conveniently short generation times, high mutation rates, and simplicity of experimental reproducibility. Experimental evolution with microbes has been extensively reviewed, especially completely, by Elena and Lenski in 2003[69,75]. These approaches have revealed substantial insight into the genetic underpinnings of selection, and the linkages between mutation and fitness. A major limitation of laboratory evolution, however, is that they cannot tell us about the prevalence or importance of these mechanisms in natural systems[72]. A natural system is any non-laboratory microbial ecosystem of interest, including host colonization by pathogens, human microbial commensals, marine environments, and microbes found in soil, water and air. Microbial processes in these natural systems are often of great interest due to their potential to influence the biosphere in ways that are important for humans and animals. Therefore, it is important to pursue an understanding of adaptation in these systems.

Although natural systems are often practically important and of great interest, they are not ideal experimentally because they cannot be easily controlled, growth rates are slower, and

do not have the breadth of tools and experience that laboratory systems afford. Studying phenotypic plasticity and the effects of environmental complexity in model organisms such as *E. coli, S. cereviseae,* and *H. salinarum* is especially fraught, because variability is intentionally removed when model organisms are established. Selection against genes that that would confer plasticity has likely occurred during the culture history of these models[60].

Nonetheless, there are several approaches and examples of testing hypotheses about the importance and mechanisms of adaptation in natural systems. Natural systems currently experiencing changes in their environment are ideal. Climate change, invasive species, and new migration events provide natural experiments for answering this question. These situations have been exploited in squirrels[76] and juncos[70] to empirically show strong positive effects of plastic traits (date of parturition and breeding season length, respectively) on reproductive success and fixing of selected traits. In bacteria, a pair of recent papers on the evolution of *P. aeruginosa* beautifully demonstrates the changes in regulatory networks and in plasticity during the evolution of a clinical strain within a cystic fibrosis (CF) lung infection[77,78]. In these papers, the mutational trajectory of a clinical *P. aeruginosa* lineage was tracked for 200,000 bacterial generations. Four of these mutations, found in transcriptional regulators, were then constructed into a non-clinical laboratory model strain. Their phenotypic and regulatory effects on the model strain were compared with the late-stage clinical isolate containing these four mutations. Ultimately, the clinical phenotypes were reproducible in the lab strain, and the mutations conferred plasticity in alginate production lacking in the lab strain. Remodeling of the gene regulatory network through a small number of genetic changes was found to be necessary for effective colonization of the CF lung. This work demonstrates the utility of combining natural systems (CF lung) with laboratory strains and powerful molecular tools for mechanistic studies of evolutionary processes.

In addition to using natural systems experience environmental change in real time, natural systems in which both ancestral and derived populations are extant allow the comparison of reaction norms in both ancestral and new environments[70,76]. Similarly, lineages that include derived, divergent taxa inhabiting two different environments can enable inferences about the consequences of plasticity for diversification. This approach was taken in a study of *Drosophila*, which explored the influence of gene expression patterns on sequence variation in divergent populations[79]. This "cross sectional" experimental design can provide evidence of association

between environmental characteristics and the presence of new mutations or plastic phenotypes[68,70]. To establish causality in these systems, additional experiments are needed to test the influence of specific mechanisms.

Osmotolerance in prokaryotes

Changes in salinity (external osmolality) are a nearly-ubiquitous facet of microbial environments. Whether freshwater, marine or hypersaline, cell membranes do not permit effective transport of dissolved ions. Therefore, water moves in and out of cells, changing the turgor pressure and concentration of intracellular solutes. Because this experience is so universal, osmotolerance mechanisms are strikingly similar among bacteria and archaea. In all groups, cells must employ strategies to balance intracellular and extracellular osmolalities, in order to maintain protein hydration and membrane pressure within physiological limits[80,81].

Among all life, two fundamentally different strategies are employed to cope with high salt: 1) a "Low salt-in" strategy in which salts are excluded from the cytoplasm and osmotic balance is maintained by the energetically expensive production of compatible organic solutes (glycerol, ectoine, betaine, sucrose, etc), and 2) a "High salt-in" strategy in which salts accumulate in the cytoplasm at concentrations equal to the external concentration[82,83]. The low-salt strategy is used by prokaryotes and eukaryotes alike, while the high-salt strategy is limited to halophilic bacteria and archaea[40]. I will focus on the high-salt strategy in halophilic archaea for the remainder of this discussion. This strategy is used by all archaea in the family Halobacteriaceae, which dominate microbial communities (5:1 archaea:bacteria) in the north GSL and give it its characteristic red color[84]. Although it is energetically cheaper than the low salt strategy 1, the high salt strategy 2 requires deeply ingrained adaptations to the intracellular presence of high salts. These adaptations include an acidic proteome, a high GC content and low numbers of hydrophobic amino acids[85,86]. The intracellular ion composition is also tightly controlled by more than 8 different types ion channels, active transport proteins and proton pumps. KCl is imported as the main intracellular salt and must be pumped in because most hypersaline environments are dominated by NaCl. The maintenance of the intracellular ionic environment despite external salinity changes requires coordination of ion channels, pumps, signaling and regulatory networks, and therefore is an ideal "complex phenotype" for exploring microbial adaptation to physicochemically linked environmental stresses.

Characteristics of hypersaline ecosystems and their halophilic prokaryotes

Extremely halophilic microbes belong to both the bacteria and the archaea, and they have evolved to persist in salinities above 1.5-2 M[40]. In addition, hypersaline environments require adaptation to extremely low oxygen and large doses of UV radiation, together with large seasonal and diurnal changes in temperature. One halophilic archaeon, *Halobacterium salinarum,* has become a simple yet powerful model organism for gaining systems-level insights into microbe-environmental interactions. *H. salinarum* NRC-1 has a small genome (~2.6 Mb), grows easily in laboratory culture media, and has been the subject of decades of molecular tool development. In addition to a homologous recombination system for genetics[87], systems biology tools and software have been developed for inferring gene regulatory networks and predicting how the cell might adapt to novel stimuli[88-90]. Although these cells thrive in hypersaline conditions, their home environment precludes the survival of most other life. Studying adaptation and plasticity in halophilic archaea will not only reveal how an extremeophile has adapted to one of the most harsh and unique environments on earth, but will provide fundamental insight into the nature of microbial adaptation to complex environmental change.

Hypersaline environments can be found throughout the world, and are biologically classified as such if their total salinity falls above 1.5 M[40]. Of all the environmental signals and stimuli encountered by halophilic archaea, salinity and ion composition are most striking and least well studied. Here, I will describe on the unique chemical aspects of hypersaline lakes and the microbial communities found in these environments, with a special focus on salinity the Great Salt Lake in Utah, USA.

Hypersaline lakes fall into two general categories, thalassohaline and athalassohaline, meaning that they are either derived from the concentration of ocean water or from other non-marine brine, respectively. Athalassohaline systems, like the Dead Sea, have a distinct ion profile, with divalent cations (Mg^{++} and Ca^{++}) more abundant than monovalent cations[91]. Not surprisingly, these brines support distinct microbial communities compared with thalassohaline brines. Of the thalassohaline systems, the Great Salt Lake (GSL) in Utah is a well-known example.

The GSL is located in the arid, Great Basin region of the west central United States and is the remnant of the much larger, prehistoric freshwater Lake Bonneville. In the last 100,000 years, geological evidence indicates that this lake has cycled from freshwater to hypersaline

twice, with the last freshwater cycle ending 6,000 years ago[75]. The composition of the GSL can be considered in terms of its two "arms": north and south halves of the lake, divided by a railroad causeway in 1959, with only two culverts for the north-south mixing of water. A third breach was installed in 1984, allowing additional mixing at the west side of the causeway (Figure 1-3). The causeway installation caused a vast salinity differential, with the south arm currently at ~16% salt by weight, and the north at >26%[84]. The south arm receives fresh water from three streams, whereas the north arm is all but cut off to fresh water with the exception of precipitation events and crossover of south arm water through the culverts and breach. Salinity in both arms of the lake fluctuates over time and space. It has fluctuated drastically in the last 50 years, responding to long-term climate patterns including major flooding during the 1980s, and seasonal, cyclic changes due to precipitation and evaporation[84,92] (Figure 1-4). Spatially, the lake is stratified vertically as well as horizontally from north to south. Heavy brine and halite precipitation creates a density gradient from the water surface to the bottom of the lake, a maximum of 10 m deep. The south arm is most distinctly stratified by depth, with a sharp interface between the less saline top layer and the deep brine layer underneath[93].

In order to further discuss biologically relevant salinity changes in the GSL, an operational definition of salinity must be established. First, in the GSL, salinity change occurs as gradients in time (evaporation, dilution by rainfall) or space (depth gradients, isolated pools with divergent chemistries), as described above. (Figure 1-5a, b). Further, these gradients may exist in ion composition (ion ratio), total salinity (total dissolved ionic salts), or both. For example, a total salinity gradient exists between the north and south arms of the lake, on either side of the causeway. Ion composition between the two arms is relatively constant, although total salinity in the north is often twice as high as in the south (Figure 1-5c)[84,94]. The Dead Sea, which does not support dense halophilic archaea as the GSL does, has an ion composition distinct from the GSL although its total salinity is comparable[95]. In both places, evaporation and rainfall dilution create gradients in both ion composition and total salinity. During evaporation, ion composition changes as different salts precipitate out in order of solubility (Figure 1-5d)[96]. "Salinity", then, is a high-dimensional EF that belies the often-used [NaCl] salinity proxy in biological research. A systems approach is required to understand how cells perceive, parse and respond to the distinct attributes that comprise salinity.

What microbial species and assemblages inhabit hypersaline environments like the GSL? Although a few select halophiles (*e.g. H. salinarum* NRC-1) have been studied in great detail as model organisms, the papers describing the structure of the microbial ecosystem are few and far between. The first such work, by F. Post in the 1970s and 1980s, described cell densities of 10^7-10^8 cells/mL in the north arm, and about 10^6 in the south[75,97]. This early work isolated a number of organisms from the lake, and all were found to be of the genera *Halobacterium* and *Halococcus,* based on the presence of the membrane carotenoid bacterioruberin. Following the description of the Archaea as a distinct domain of life, several other fingerprinting studies were undertaken in the GSL and other hypersaline lakes, including Lake Chaka in Tibet.

In the Great Salt Lake, Archaea were shown to outnumber bacteria by a factor of five to one[84]. In a more recent survey of diversity in the GSL, phylogenetic and functional gene chips were used to estimate taxonomic and functional diversity, and to relate these back to physicochemical gradients in the lake, including salinity[98]. Collecting from 9 different sites in the GSL at 3 depths in the north and south arms, the authors identified 5,000 unique microbial taxa by 16S rRNA PhyloChip analysis. These ranged from communities of 100 taxa at sites in the north arm, to 2,400 taxa in sediment collected in the south. The distribution of functional gene diversity was similar (few genes in north arm sites, many in south arm sites), although functional and phylogenetic diversity were not significantly correlated at all sites. In general and as expected, higher salt concentrations restrict Cyanobacteria, beta-proteobacteria and Bacteroides, but favor Archaea and Thermotoga. The paper concludes that evidence of horizontal gene transfer within the GSL is demonstrated in biogeographic patterns in functional and 16S genes: across all sites, 16S sequences have a significantly lower similarity than functional genes. Further, functional similarity is positively correlated with selection pressures including salinity and chromium concentration.

In Tibet's Lake Chaka, there exists a vertical salinity gradient within sediments that ranges from nearly freshwater at 9m below lake bottom to saturated hypersaline (>32% salinity) at sediment surface. Although salinity is a known dominant factor constraining microbial communities, Jiang *et al* used both qPCR and culture techniques to explore how communities vary as salinity changes[99]. Briefly, the Archaea identified along the depth gradient represented both Crenarchaeota and Euryarchaeota, with the saline and hypersaline samples (30-45% salinity) dominated by Euryarchaeota. Within this group, 80 of 120 sequences were

Halobacteriales and most of these (49 of 80) matched *Halobacterium salinarum* (AJ496185) and *H. salinarum* NRC-1. In addition to these matches, several other Halobacteriales sequences matched (>99%) additional *Halobacterium* species (Ch12 and AUS-1). The Euryarchaeota were completely absent from the lowest salinity samples.

Taken together, these studies demonstrate decreasing phylogenetic and gene functional diversity as salinity increases. Further, within hypersaline environments, Archaea are strongly dominant and *Halobacterium* spp comprise the majority of prokaryotes, both in the GSL and elsewhere.

Concluding remarks

Current basic research on microbial adaptation to changing environments has far-reaching consequences for some of the world's most intractable challenges. Climate change, antibiotic resistance, and novel sources of sustainable fuel are just three fields in which the ability to predict, prevent and encourage specific physiological processes could impact innumerable lives (human, animal and plant alike). Advances in systems biology and experimental evolution enable understanding the genetic and mechanistic basis of adaptation in real time, and these new insights pave the way for microbial rational reengineering and better decision-making where drug development and climate policy are concerned. Osmotolerance in particular is a critical property of microbes, exploitable in the food industry to prevent spoilage. A deeper understanding of adaptation to external osmotic stress could improve the safety and efficacy of anti-spoilage measures and technologies.

In addition to the myriad practical applications of further work in microbial adaptation, basic unanswered questions about microbial evolution still remain. First, how do microbes perceive and respond to the environment, as a dynamic and complex experience? Just as the human brain cannot possibly process all the external signals it receives, microbes are under strong selection to make sense of the aspects of their environment most critical for survival. This is inherently a systems-level question, for which simultaneous measurements of biological and environmental information are needed. The Baliga lab has begun to tackle this large question by amassing a vast collection of sequencing, phenotypic and microarray experiments that document genome- and transcriptome-wide data in hundreds of experimental conditions. Together with software tools for inferring and predicting microbial response, our group has demonstrated the

power of systems tools and technologies for learning and predicting microbial-environmental response[32,90,100].

Secondly, how can cellular responses to different rates of environmental change be deciphered? A recent paper demonstrated that different evolutionary trajectories, even extinction, are accessible to an evolving microbe depending solely on the rate at which environmental stresses increase[101]. The authors used multiplexed cultures of *E. coli* to test the effects of three different rates of increase in antibiotic concentration. They found that antibiotic resistance required several selected mutations, and that high rates of change blocked the cultures from experiencing the required mutational trajectory and thus promoted extinction. This work is alarming given the current acceleration of climate change and atmospheric CO_2 on Earth, and systems approaches should be leveraged to find the environmental stimuli for which rate of change matters most.

As reviewed earlier, systems biology is extremely powerful and is ushering in a new frontier in the resolution at which biological processes are understood. However, the tradeoff between studying tractable models and urgently relevant natural systems is alive and well in this field. To bridge the gap, research on the human body as a natural system or small microcosms are gaining foothold. The cystic fibrosis lung has become a model system for measuring adaptation in real time, and cannot be replicated artificially[77,78,102]. This "natural environment" is accessible and measureable over time, and the course of the disease is an important covariate, and not noise to be accounted for. The human lung lies somewhere between an open and a closed system.

Apart from disease hosts, the natural environment includes every other microbe-inhabited space on earth, all of which are complex, open systems. Of these environments, an especially important one is agricultural land. Optimization of plant growth is a major goal, and agriculturalists do not have the luxury of a controlled laboratory. Despite this challenge, Atsushi Nagano and colleagues recently published perhaps the most remarkable example of systems biology in the "real world" to date[103]. This group built a model of the transcriptional response of rice grown in the field, and measured climate, developmental age and genotype on various time scales. Using these environmental data they built a linear model that accurately predicts most gene expression changes in the rice, based on atmospheric and developmental data alone. The study required continuous environmental monitoring, and curiously the model did not perform as

well with plants grown under controlled laboratory conditions. These results may have interesting implications for other projects that aim to characterize environmental response in the laboratory, because they raise the possibility that organisms require environmental variability in order to establish stable expression patterns. The paper is a call for a shift toward embracing, and not excluding this environmental variation in systems biology.

Finally, returning to basic biological questions, how do microbes adapt as they transition from a complex natural environment to a controlled laboratory system, and vice versa? At its core, this question asks about the process of adapting to new constant and variable environments, or the adaptation of generalists and specialists in "real world" contexts. The continued intermixing of systems biology tools with natural environments begs an answer to this question; are our models relevant for studying natural systems? How can we ensure that we are answering the questions we think we are, if so much about generalist-specialist adaptation in these contexts is unknown? There is ample evidence for environmental and clinical isolate adaptation after transitioning to model organism status (reviewed in Chapter 5), but halophilic archaea are younger model organisms. While there is reason to suspect adaptation in halophiles, this question has not yet been addressed.

Toward that end, the results presented in this dissertation represent an investigation into the salinity adaptation of a model halophilic archaeon in the context of its conspecific GSL isolate counterparts. Chapter 2 describes the baseline measurements of GSL water chemistry, and forms the foundation for further investigation at the interface of the lab and the lake. Chapter 3 details a full, multifactorial analysis of *H. salinarum* NRC-1 growth response to multiple facets of salinity, a complex environmental factor traditionally reduced to [NaCl]. Chapter 4 is a pilot study of gene expression in hypersaline media that vary in ion composition, but share the same total salinity. This chapter uncovers patterns in gene expression that recapitulate known effects of K^+ but also reveal other ion-specific patterns in gene functional groupings. Chapter 5 is a discussion of the role of a ribonuclease in modulating expression during concomitant changes in oxygen, growth phase and salinity. Chapter 6 is a cross-sectional study exploring the osmotolerance profiles of a decades-old model organism and recently isolated GSL haloarchaea of the same species. We show that despite 50 years of evolutionary opportunity in the lake and in the lab, these strains maintain strikingly similar osmotolerance

profiles. The final chapter is a brief synthesis of the preceeding sections, and considers additional outstanding questions and avenues for future work in this area.

Chapter 1 Figures and Tables

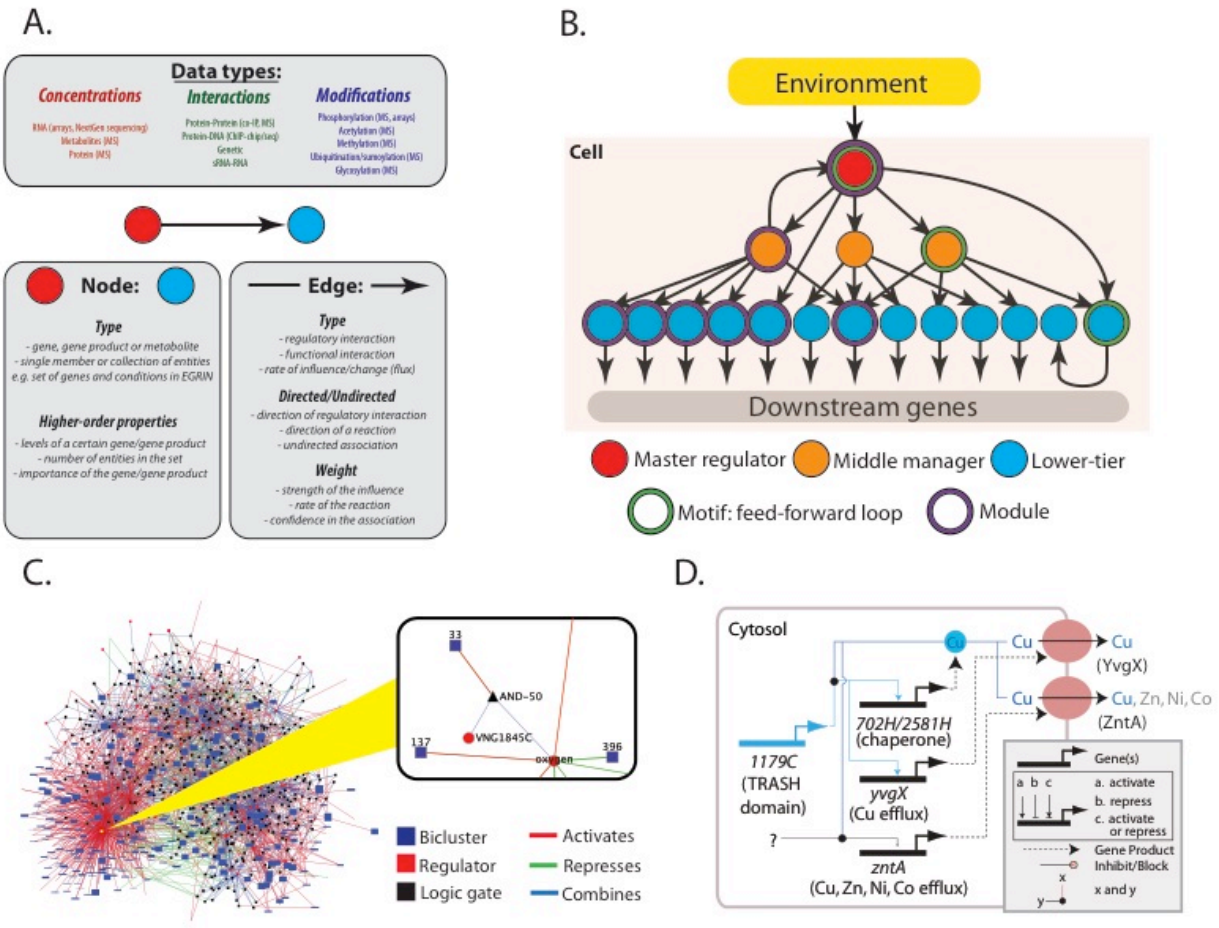

Figure 1-1: Generation and properties of networks. Networks are a useful formalism to represent, catalogue, and analyze biological information. (A) Networks are generic entities that are used to represent many types of biological interactions. The fundamental units of a network (or graph) are nodes, and edges. Three types of biological information are commonly represented as a network: transcriptional, metabolic, and protein-protein interactions. The types of data used to build the network and the features that define nodes and edges vary by application. (B) Biological networks share common features. Here, we represent a transcriptional network. Three features commonly define network topology and have important dynamic consequences for the behavior of the network: (1) hierarchy: transcriptional networks are close to scale-free in the distribution of regulatory connections and exhibit hierarchical arrangement of connections[104]. At the beginning of many biological pathways are few "master regulators" that initiate response to environmental or internal cues. These master regulators propagate information to "middle managers" that have many additional regulatory connections, mostly to "lower-tier" regulators that mediate specific biological functions[105]. (2) modularity: biological networks aggregate pathways and functions into modules, which are defined by groups of genes, regulators, and gene products that are somehow interconnected and interdependent[106]. Here, we denote a transcriptional "module" by a purple ring surrounding the nodes in the module, though we note that biological modules are often composed of a diversity of part types that need not be coregulated. (3) motifs: interesting dynamic behaviors of gene circuits are often mediated by particular wiring of the parts, which defines a motif. The members of a particularly well-studied motif, the feed-forward loop, are depicted in this illustration by green circles surrounding the nodes[107]. (C) Biological networks learned from experimental data, such as the Environmental and Gene Regulatory Influence Network (EGRIN) for *H. salinarum* sp. NRC-1, contain many layers of information that can be mined to aid hypothesis generation[90].

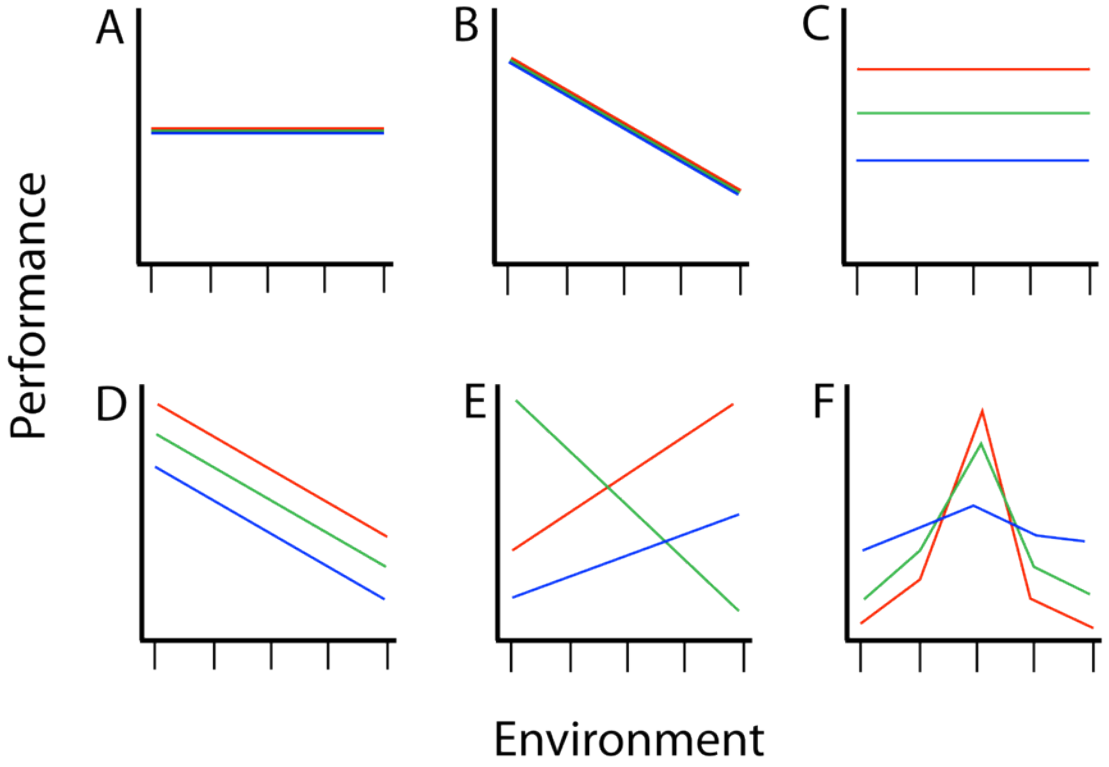

Figure 1-2: Reaction norm plot examples. Reaction norms plot and quantify phenotypic plasticity, the environment-dependent phenotypes produced by a single genotype. Where each colored line represents a genotype, reaction norms convey both environmental and genetic contributions to phenotype (performance). A) Horizontal superimposed lines indicate identical genotypes producing the same environment-independent phenotype. This is called zero plasticity when performance measures a trait that is not a component of fitness. If measuring fitness (*i.e.* survival, fecundity, growth, offspring), these genotypes are generalists, equally adapted in all environments. B) Environmental, but no genetic effect on performance; non-zero plasticity. C) Genetic but no environmental effect on performance. D) Genetic and environmental effects on performance are present, but the genetic effects are not modified by the environment (*i.e.* lines are parallel). E) Non-parallel curves indicate genetic variation for plasticity, and the genotype most fit for an environment depends on the environment. Also called "gene x environment interaction." F) Generalist-specialist curves. When performance measures a component of fitness, reaction norms measure niche width and trade-offs between maximum performance and breadth of fitness across environments. Note that these curves are also non-parallel, indicating that the environment modifies the genetic effect on fitness.

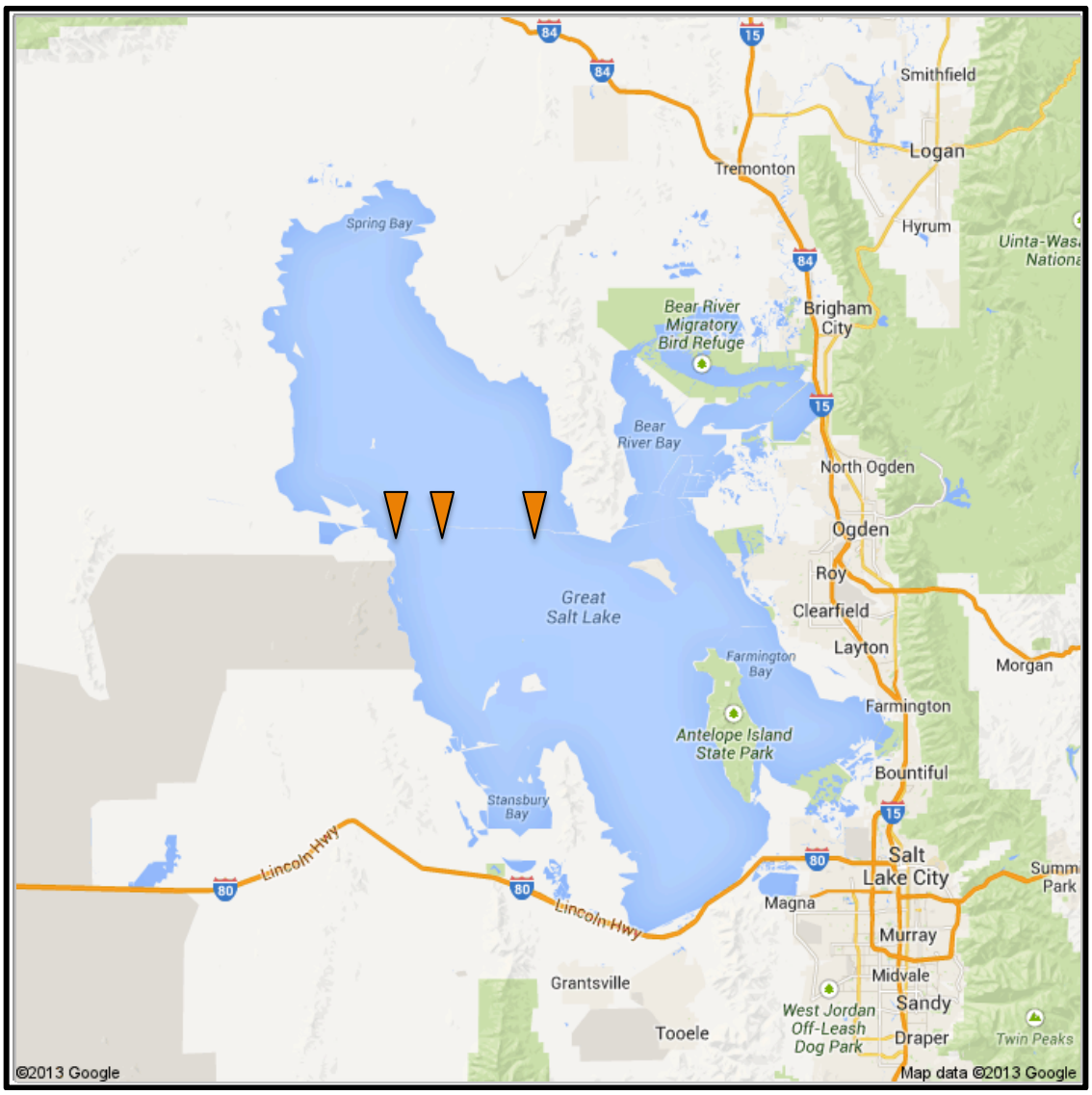

Figure 1-3: Map of the Great Salt Lake, Utah, USA. Map showing the north and south arms of the lake separated by the rock-fill railroad causeway (light line crossing East to West). Orange triangles on the causeway (left, center, and right) denote the 300-ft-wide breach, west culvert and east culvert, respectively. The causeway is nearly impervious to north-south water flow, except for these three inlets.

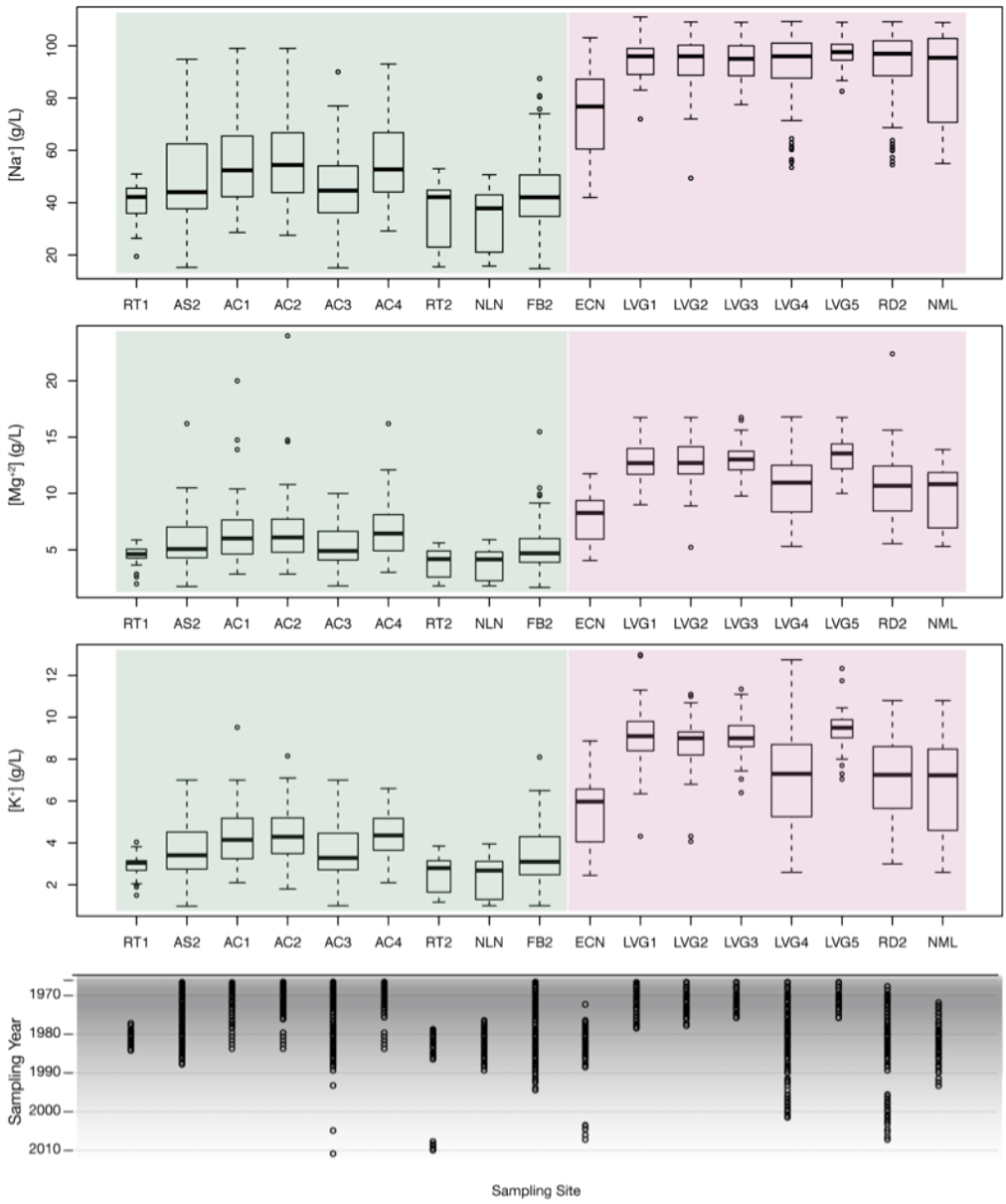

Figure 1-4: Historic variation in Great Salt Lake Salinity.
Longitudinal variation in Na, Mg and K concentrations in 17 different sites within the GSL. Box plots represent ion concentrations measured at times indicated in bottom plot, between 1970 and 2010. Sites are monitored by the Utah Geological Survey and light green background indicates south arm sites. Pink background denotes north arm sites. Site descriptions and locations available in Gwynn 2007[92].

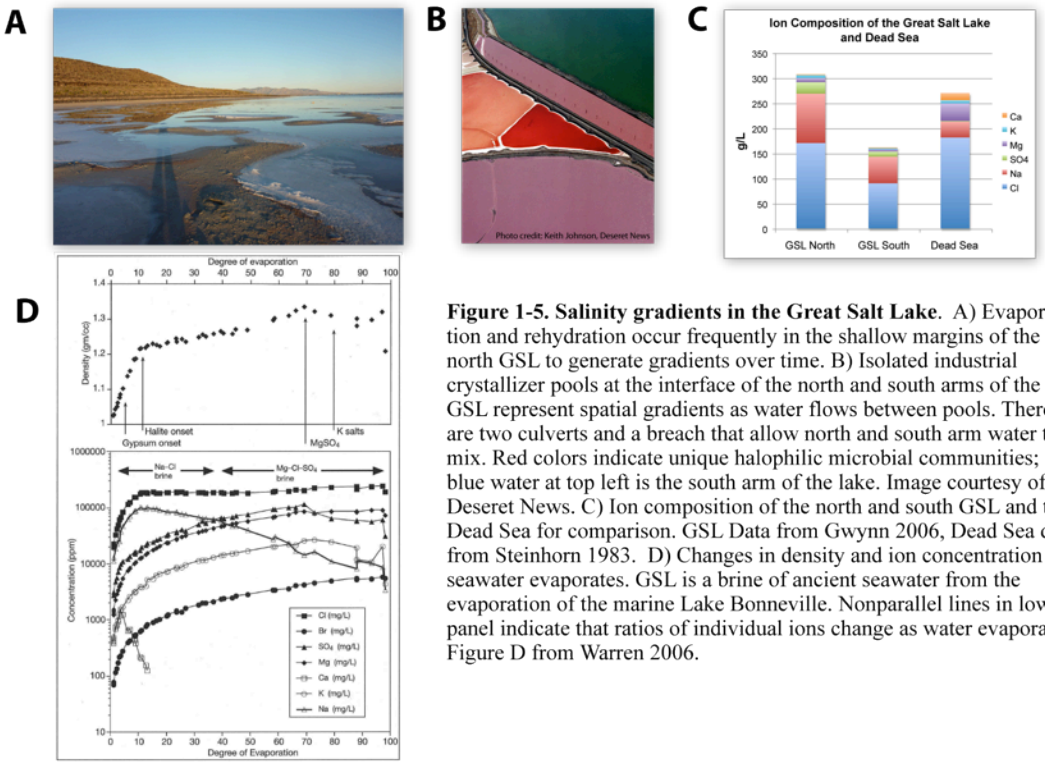

Figure 1-5. Salinity gradients in the Great Salt Lake. A) Evaporation and rehydration occur frequently in the shallow margins of the north GSL to generate gradients over time. B) Isolated industrial crystallizer pools at the interface of the north and south arms of the GSL represent spatial gradients as water flows between pools. There are two culverts and a breach that allow north and south arm water to mix. Red colors indicate unique halophilic microbial communities; blue water at top left is the south arm of the lake. Image courtesy of Deseret News. C) Ion composition of the north and south GSL and the Dead Sea for comparison. GSL Data from Gwynn 2006, Dead Sea data from Steinhorn 1983. D) Changes in density and ion concentration as seawater evaporates. GSL is a brine of ancient seawater from the evaporation of the marine Lake Bonneville. Nonparallel lines in lower panel indicate that ratios of individual ions change as water evaporates. Figure D from Warren 2006.

Chapter 2. Characterization of the physicochemical properties of Great Salt Lake water

Abstract

The Great Salt Lake is a dynamic hypersaline system, in which water properties like temperature, salinity, dissolved oxygen and pH vary over space and time. Here we report the results of a sampling trip in May 2011, in which water was collected and its properties measured at 16 sites within the lake. These measurements inform additional experiments which evaluate haloarchaeal adaptation in GSL-like conditions.

Introduction

Every spring, as temperature and solar irradiance increase, the photosynthetic community of the Great Salt Lake (GSL) blooms. Dominated by the alga *Dunaliella*, these blooms stimulate the growth of the halophylic prokaryotic community in the ecosystem[75]. We are interested in the ecological interactions between *Dunaliella* spp. and halophilic prokaryotes, as well as haloarchaeal adaptation along gradients in salinity. As salinity is physicochemically linked to other environmental factors such as temperature, light, and lake volume, a dataset containing simultaneous measurements of all of these factors will allow detection of microbial anticipatory relationships among them. Further, questions about the evolution of salinity response may be addressed using cells sampled from allopatric lake populations. Thus, we have identified the GSL as a model system for addressing these fundamental questions. The purpose of this work was to collect both water and cell samples from a wide range of ecological conditions within the GSL, and to assess temporal changes in salinity during evaporation in vitro.

Spatially, the salinity in the GSL ranges from near seawater concentration to saturation, with the north arm of the lake nearly 10 times more saline than the south[84,98]. A railroad causeway separates the north from the south, and has served to concentrate the north arm water because of a lack of freshwater inflow. In the south arm, areas adjacent to freshwater stream inflow from streams have lower salinity and this low-density water sits atop a deeper brine layer below. Although the lake experiences temporal fluctuations in salinity, these spatial differences have remained consistent since the installation of the causeway in 1959. Prior to the causeway installation, spatial salinity measurements in the GSL are scarce. However, geological evidence

indicates that in 100,000 years, the lake has cycled from freshwater to hypersaline twice, with the last freshwater cycle ending 6,000 years ago[75].

Temporal variation in the GSL comes from a number of different sources, and has been measured by the Utah Geological Survey since 1966[92] (Fig 1-4, nic's UGS salinity bar plot). Long-term climate cycles as well as major events (drought and flood) have measurable impacts on total salinity in the lake. The most recent nadir in total salinity occurred in the mid-1980s when floods raised the lake's surface elevation by over 10 feet, and diluted the total salinity of the north arm from its pre-flood level (~28% total dissolved solids, by weight) to less than 18%. By 1994, north arm salinity returned to ~26% and continues to hover at that level. The south arm experienced a dip from 16% to 8% and a return to 14% in the same time frame[92].

Salinity change over shorter timescales is less well documented. Annual changes in lake level bring an annual cycle to total salinity in the lake, and wind-driven mixing occurs many times per year. The shallow lake depth combined with its large fetch (length of lake on which the wind acts) contributes to these frequent wind-mixing events. These mixing events are known to keep the top layers of the lake isothermal and isoosmotic, but the deep brine layer rarely mixes, keeping its oxygen, temperature and salinity profiles distinct from the surface layers of the lake[93]. Although the effects on oxygen and temperature have been measured for this differential mixing between the shallow and deep layers, salinity has not been as thoroughly studied.

Nor has salinity been evaluated on shorter temporal scales. In the GSL, much of the lake margin is shallow sediment with precipitated salt and mineral crystals, subject to frequent evaporation and rehydration. Evaporation in these sediments likely causes changes in both total salinity and ion ratio as different salt compounds precipitate out at characteristic concentrations[96]. For example, halite salts (NaCl) precipitate first, followed by magnesium and potassium salts. This sequence is determined by the characteristic concentrations at which solutions saturate with each salt. This temporal variation due to evaporation is distinct from spatial salinity variation, which is due mainly to differences in total salinity without changes in ion ratio[108]. While salinity is a major selective pressure for halophilic microbial communities[98,99], whether total salinity and ion ratio exert independent effects on microbial growth has not been evaluated in hypersaline systems. Therefore, as part of our efforts to answer this question, we sampled from the GSL in order to measure spatial differences within the lake and temporal differences during an in vitro evaporation timecourse.

Materials and methods

Sample sites

Sites were selected to encompass ecological and chemical diversity, and samples were taken where permission was granted and it was safe to do so. GPS coordinates were recorded at each sample site (Table 2-1, Figure 2-1).

In order to sample from geographically and ecologically distinct sites within the GSL, we have established collaboration with Dr. Bonnie Baxter at Westminster College (Salt Lake City, UT), whose group has permissions and experience of regularly sampling the lake. In addition, we have initiated collaboration with the Salt Lake USGS office (Dr. David Naftz and Michael Freeman) and from the Great Salt Lake Ecosystem Project who will include us in their sampling outings, and thus give us access to additional sites within the North and South arm of the GSL.

To coincide with the bloom, we sampled May 2-6, 2011. Sites were selected in both the north and south arms, as well as along the causeway. Causeway sites were selected to sample the effects of proximity to flow-through sites (culverts and breach). The effects of depth on salinity are most pronounced in the south arm, where a sharp interface exists between the less-saline top layer and the dense, deep brine layer that moves in from the north. To capture this variation, we sampled at two depths from one of the south arm sites. In a subset of the measured sites, we collected large volumes of water (10 L) for future laboratory use. Here, we report all data collected from each of the GSL sample sites.

Multi-parameter water and environmental testing

Water chemistry parameters were measured at each site at the time of sampling using an In-Situ Troll 9500 multiparameter water-quality monitor. The high-range specific conductance and standard pH probes were calibrated and verified prior to taking measurements. Measurements from this device include: temperature (°C), pH, specific conductivity (uS/cm at 25°C), practical salinity units[109] (PSU), dissolved oxygen (mg/L), photosynthetically active

radiation (PAR; μE/m^2/s), UVA (W/m^2), and UVB (μW/m^2). The relationship between PSU and parts-per-thousand is approximately 1:1. The conversion from conductivity to salinity is 1000 μS/cm = 0.64 ppt, and is discussed further in Fofonoff, 1985[109] and at this website: http://fermi.jhuapl.edu/denscalc.html. GPS location measurements were taken on a Garmin GPS72 handheld device. GPS accuracy is rated at +/- 5 m. Ultraviolet radiation and PAR were measured using a Solar Light radiometer PMA2100, with detectors PMA1132 (PAR, 400-700 nm), PMA2110 (UVA, 320-400 nm) and PMA2101 (erythemally weighted UVB, 280-320 nm). All light measurements were taken at the water's surface at each site.

Water sampling and processing

Before arrival at the GSL, all glassware, bottles and collection vessels were sterilized with 10% HCl and soaked in a 2% HNO_3 solution to remove any residual metal contamination that might confound downstream ion analyses. Water for rinsing filters and diluting samples was treated with a 0.5% Chelex solution and filter sterilized to chelate any remaining ions in the water. At each site, 1 L of water was collected from the surface. At site GSL01, 1 L was collected from each of 0.30 and 1.00 m, using a weighted shark collector suspended on a rope with depth measurement gauge. An additional 10 L of water were collected from each of sites GSL10, GSL11, GSL14 and GSL16 for future use in culture media.

Within 6 hours of collection, water samples were stored at 4°C before filtering. From each site sample, 500 ml were filtered with a 20 μm nylon filter (to allow prokayrotes to pass through) in 2 x 250 mL replicates, into metal-free dark polypropylene bottles. An additional 100 ml from each sample was filtered in the same way, and inoculated 1:10 into heterotrophic halophilic liquid enrichment media containing algal F/2 media amended with either 0.5% casamino acids or 5mM sodium pyruvate plus 0.05% casamino acids. Duplicate culture tubes containing 2-3 mL of F/2 media[110] with added salts (3.4 M NaCl, 0.07M $MgSO_4$*$7H_2O$, 0.02 M KCl, 0.17 mM NaH_2PO_4, 1 mM $NaHCO_3$) and each of the two amendments were inoculated 1:10 with sample aliquots, and incubated statically at ambient temperature during transport from Salt Lake City (UT) to Seattle (WA), for a total time of 2-3 days. The four 10L samples were filtered by peristaltic pump filter at 0.2 μm upon returning to seattle, and were stored at 4°C. Liquid enrichments were incubated at room temperature in ambient light.

Evaporation and rehydration experiments

The sample taken at site #14 was selected for an evaporation time course, in which GSL water chemistry was measured during the course of near-complete evaporation. This site was selected as it was the most saline of the four sites where 10L samples were collected. For the experiment, six 1000 mL beakers were pre-weighed and then filled with 500 mL each of GSL2011.14 water sample, pre-filtered at 0.2 μm. Beakers were placed in a shaking waterbath incubator set at 37°C and 125 rpm shaking to accelerate the evaporation process. Starting at Time = 0, beakers were weighed, and two 50 mL samples were removed for later chemical analysis. Beakers were weighed again immediately after sampling. A new beaker was weighed and sampled every 15-18 hours, until six samples were taken. After a beaker was sampled, it was sacrificed (no other samples taken from it), and kept until complete evaporation for the rehydration timecourse experiment. Sacrificing beakers after one sample was important so that changes in salinity could be compared with the overall course of evaporation; removing samples changes the volume dramatically and cannot be corrected for post-hoc. Samples were taken at 0, 17.25, 33.5, 46.25, 62.5 and 81.75 hours. The evaporation experiment was carried out for 82 hours (89% evaporation by weight), when total remaining liquid was just sufficient for a final 100 mL sample.

To measure the changes in salinity and ion ratio as precipitated crystals rehydrate in fresh water (as in a precipitation event), experimental beakers were left to evaporate completely, for 7 days, until no visible liquid remained. After total evaporation, the mass of lost water was measured from each beaker. Mass of lost water = (Mass of evaporation T=0 beaker with 500 mL GSL water) – (Mass of 100 mL sample) – (mass of evaporated beaker with salt crystals). The mass of lost fresh water was quickly added back to each beaker, and two 50 mL samples were taken immediately from one beaker. All six re-filled beakers were incubated at 37°C and 125 rpm for 45 hours. One sample per beaker was taken at 0, 0.25, 0.75, 1.33, 4.08 and 44.5 hours, with each beaker sacrificed after one sample. Beaker weights were recorded before and after each sample. Samples were collected in technical duplicate, and the entire evaporation/rehydration experiment was piloted twice in advance without sampling to ensure reproducibility of evaporation course under experimental conditions.

Note that as salt crystals dissolve and rehydrate, evaporation continues to occur. Beakers were not covered to prevent this, in order to recapitulate these competing processes in a

hypersaline lake. If rate of re-solvation is fastest just after water is added, then the first few timepoints (before ~2 hours) should reveal effects of rehydration more than evaporation. Timepoints taken later will be confounded, but will nonetheless indicate how salt chemistry changes after a sudden addition of fresh water.

Chemical analysis of GSL and evaporation water samples

To determine the concentration of the five most abundant ions (Na^+, Cl^-, Mg^{2+}, SO_4^{2-}, K^+) in the GSL, filtered samples from the lake, evaporation timecourse samples, and controls were sent to Hoh Pak Laboratories, Inc (New Iberia, LA) for ion analysis. Because Ca^{2+} ion was measured in the process of titrating Cl^-, it is included here. Note, however, that calcium will not be further discussed in the following chapters. Hoh Pak was selected as it specializes in ion analysis of hypersaline samples. All lake samples except samples 1, 2 and 3 were sent for analysis. These samples were omitted due to the limited collection volume at this site and low likelihood of finding halophilic archaea. In addition, we analyzed samples from time series of evaporation and rehydration of GSL sample 14, including a de-ionized water control.

Several additional control samples of known concentration were also sent for analysis. This basal salts control is a mixture of the basal salt mixture used in standard CM growth media for *H. salinarum*. It contains the following: NaCl (250 g/L, 4.28 M), $MgSO_4*7H_2O$ (20 g/L, 0.081 M $MgSO_4$ alone), KCl (2 g/L, 0.027 M). Other components of the CM growth media were also sent for analysis for the purpose of reference data in additional growth experiments. These included a solution of Oxoid neutralized bacteriological peptone (10 g/L), and a solution containing both Oxoid peptone (10 g/L) and sodium citrate (3 g/L). The peptone and peptone/citrate mixtures were submitted for analysis in triplicate and are labeled a, b and c in Table 2-3.

The following methods were used at HohPak labs, Inc, to measure the concentrations of each ion: Sodium and potassium – atomic absorption spectroscopy; chloride – titration with 0.01N silver nitrate; Magnesium – titration with 0.01M EDTA; Sulfate – gravimetric precipitation with barium chloride. Selected samples were submitted in blind duplicate or triplicate to obtain a measure of technical variability. Coefficients of variation ranged from 0 to 5.5%.

Results and discussion

Great Salt Lake sample chemistry

Ion concentration varied among sites as expected, with the highest salinity in the samples farthest north, and the lowest salinity in the south arm, near freshwater inflow (Table 2-2, Figure 2-2). Interestingly, the sample collected just south of the causeway in the south arm (Sample 11) was half as saline as the most saline north arm sample, but many times more concentrated than samples 4 and 5. This sample was taken approximately 8 km from the nearest culvert where north and south arm water have the opportunity to mix. At these culverts, dense, north arm water is known to sink and flow southward along the bottom of the lake, leaving the lighter south arm water to flow northward at the surface. While this general pattern may be disturbed during wind and storm events, it is a well-known property of GSL hydrology[92,93], (David Nafts, USGS, Personal Communication). No such storm event occurred during sampling at these sites, and thus the intermediate salinity of sample 11 suggests more extensive surface mixing near the causeway or additional surface-level leakage from north to south through the causeway itself. Historic salinity data for this site has not been recorded previously for comparison.

Biologically, the different colors of these sites (ranging from green, to blue, to pink) reflect the different microbial communities found there. Although it is accepted that communities in the south arm are more diverse than those in the north, little is known about how communities vary within the north arm, especially with respect to salinity changes there. Parnell and colleagues showed that salinity was a major selection factor and determinant of functional and phylogenetic diversity[98], but little is known about whether smaller salinity gradients also constrain diversity, or whether wind mixing negates any salinity-based metapopulation structure.

Evaporation and rehydration of GSL water

As GSL water evaporated, ion ratios changed according to solubility limitations as previously observed in seawater. Sodium decreased steadily, reflecting the initial precipitation of halite salts. Magnesium, sulfate and potassium steadily increased as a fraction of dissolved ions (Table 2-3, Figure 2-3). Because evaporation was not carried out to completion in order to allow for sufficient sample volume, we likely did not observe the expected drop in dissolved magnesium, sulfate and potassium. These ions are the very last to precipitate, and are collected last in commercial salt and mineral mining operations. Had we sampled the small volume of

remaining water just before complete evaporation, we may have observed such a reduction in Mg^{2+}, $SO_4^{2-}$and K^+.

During rehydration, there is an apparent exponential increase in total salinity as salts dissolve in the suddenly-added fresh water. The K:Na ratio is initially high, and decreases over time, opposite of the observed dynamics during evaporation. The high rate of increase in ion concentration is likely due to the simultaneous effects of rehydration and evaporation; beakers were not covered in order to capture dynamics that more closely resemble the natural environment.

Such changes in ion ratio likely influence the osmotolerance strategies employed by halophilic archaea that inhabit shallow sediments in hypersaline lakes. Haloarchaea that use a "salt-in" strategy for maintaining osmotic balance sequester potassium in order to balance extracellular sodium concentrations. In *H. salinarum*, a number of potassium uptake systems regulate this sodium-potassium balance, including the low-affinity TrkAH K^+/H^+ symporter, and the high-affinity KdpFABC ATP-driven pump[111,112]. When potassium is limiting, the Kdp system is activated at the level of transcription to pump K^+ into the cell. Intracellular K^+ concentrations also increase when extracellular potassium is limiting. Under conditions of excess K^+, however, little is known about how cells respond at the phenotypic, protein or gene expression levels. Phenotypic and transcriptional response to varying ion ratio will be explored in the following chapters.

Table 2-1: Great Salt Lake sample site locations and water properties

Sample ID	Latitute	Longitude	Sampling date	Sampling time (24h)	Water depth (m)	Water temperature (°C)	pH	Specific conductivity (µS.cm^{-1} at 25 °C)	PSU	DO (mg/L)	PAR (µE/m^2/s^1)	UVA (W/m^2)	UVB (µW/m^{-2})	Site description
2011-GSL01	41°04.017'	112°13.800'	2-May-11	14:30	0.30	11.55	9.43	22900	13.81	13.61	1620	41	16	Farmington Bay outflow, 1 ft below surface to capture FB discharge
2011-GSL02	41°04.017'	112°13.800'	2-May-11	14:30	1.00	11.55	9.43	36000	22.70	13.61	1620	41	16	Farmington Bay outflow, 3 ft below surface to capture south arm deep brine layer
2011-GSL03	41°04.017'	112°13.800'	2-May-11	14:40	0.21	13.6	8.23	145900	115.49	10.79	1620	41	16	approximately 50 m West of Farmington Bay outflow along the Bridger Bay shore (North side of the road) toward Antelope Island
2011-GSL04	41°14.060'	112°20.020'	3-May-11	10:10	0.10	13.2	8.74	25060	15.24	7.68	485	18.5	4	Bear River Bay Outflow at Causeway Bridge near Warren. South side of bridge (south arm of GSL)
2011-GSL05	41°14.050'	112°20.183'	3-May-11	10:10	0.17	11.1	9.06	2379	1.22	9.01	485	18.5	4	Bear River bay outflow: North side of bridge, brackish water
2011-GSL06	41°13.320'	112°50.860'	3-May-11	13:30	0.21	12.7	8.26	120700	91.18	9.44	683.1	15.2	7.7	South causeway at Lakeside breach (far W side of causeway)
2011-GSL07	41°13.350'	112°50.770'	3-May-11	13:30	0.33	13.7	8.17	148533	118.12	8.74	683.1	15.2	7.7	North causeway at Lakeside breach (W side of causeway)
2011-GSL08	41°13.288'	112°47.791'	3-May-11	14:45	0.24	13.6	8.13	164200	134.46	8.98	ND[1]	ND[1]	ND[1]	North causeway, east of Lakeside breach, west of "west crack," where causeway bends slightly.
2011-GSL09	41°13.320'	112°45.390'	3-May-11	15:10	0.25	13.0	7.90	197600	172.18	8.77	ND[1]	ND[1]	ND[1]	North causeway on "west crack"
2011-GSL10	41°13.399'	112°40.108'	3-May-11	15:30	0.17	12.7	7.72	201767	177.18	8.78	ND[1]	ND[1]	ND[1]	North causeway, estimated longitudinal center
2011-GSL11	41°13.400'	112°40.410'	3-May-11	16:00	0.21	11.4	8.31	151900	121.58	9.88	ND[1]	ND[1]	ND[1]	South causeway, estimated longitudinal center
2011-GSL12	41°12.980'	112°30.385'	3-May-11	16:40	0.49	12.6	7.61	221500	201.84	8.29	ND[1]	ND[1]	ND[1]	North causeway btwn "east crack" & promontory point
2011-GSL13	41°26.284'	112°40.100'	4-May-11	19:00	0.26	19.8	7.24	227200	209.27	5.12	530	4.24	0.67	Spiral Jetty: 5 m out from south side of jetty, about halfway between shore and spiral. Site used for timecourse, data here are T=0 hrs.
2011-GSL14	41°26.267'	112°40.133'	5-May-11	11:30	0.265	15.12	7.35	223800	204.82	8.12	ND[2]	ND[2]	ND[2]	Spiral Jetty: center of spiral
2011-GSL15	41°26.267'	112°40.168'	5-May-11	12:00	ND[3]	ND[3]	7.43	263900	260.51	ND	ND[2]	ND[2]	ND[2]	Spiral Jetty: 50 m west of jetty, out in lake.
2011-GSL16	40°44.142'	112°12.575'	6-May-11	11:53	0.18	16.7	8.2	ND	ND	ND	1905	41.9	15.95	Salt lake state marina, along silver sands beach

[1] PAR and UV data not collected due to inclement weather (rain storm).

[2] PAR and UV data not collected due to risk of instrument submersion. Collection required wading in deep water.

[3] Depth and temperature not collected at this site due to deep water and inability to use instrument properly.

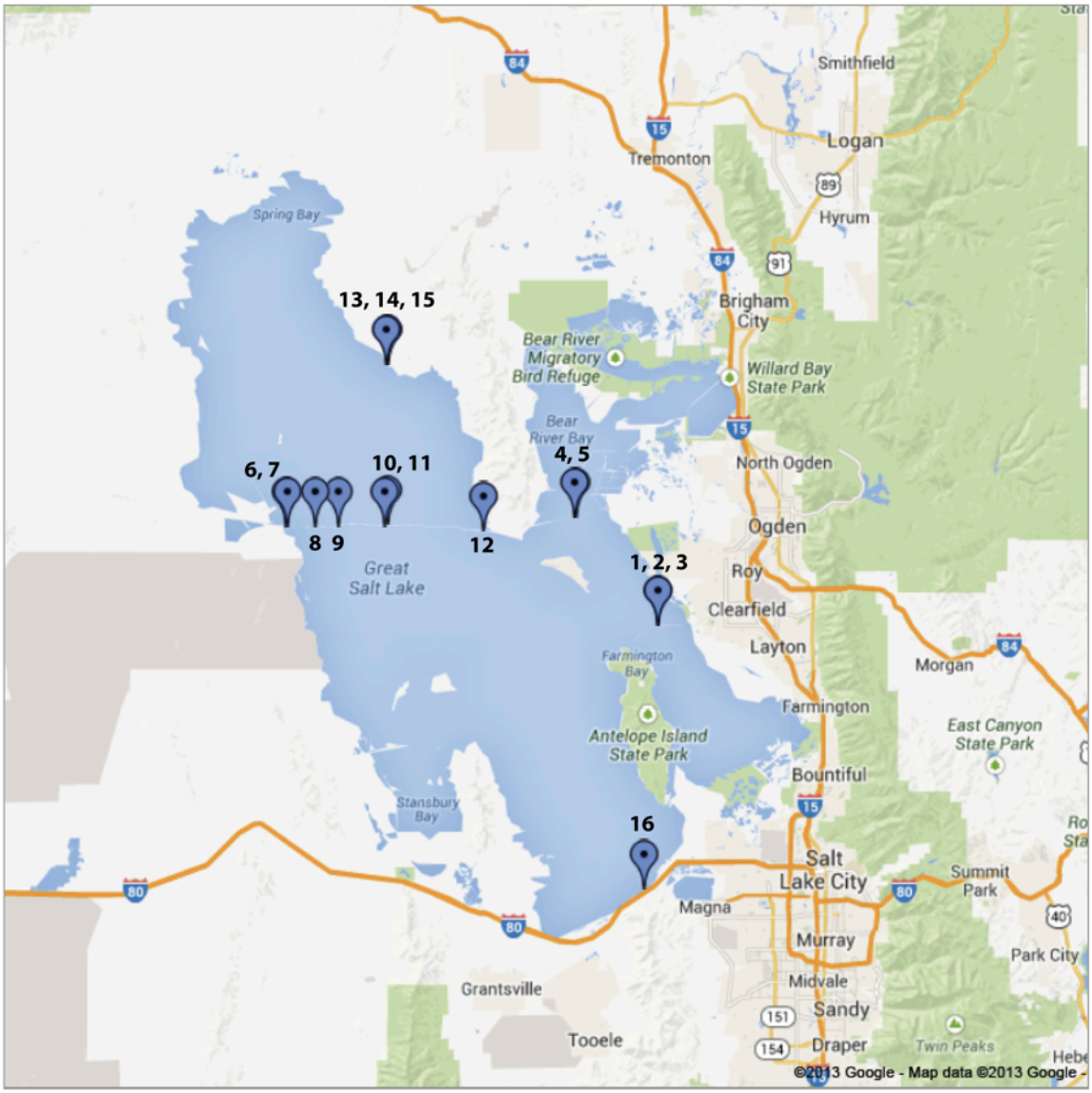

Figure 2-1: Map of all GSL sampling sites. Map of the Great Salt Lake, Utah, USA with blue markers indicating sample sites. Marker locations were made with GPS coordinates in Table 2-1. Water, and metadata were collected and recorded from each site, and sites were selected to capture the chemical and ecological diversity within the GSL.

Table 2-2: Measured ion Composition of Great Salt Lake water samples and basal salt control

g/L

Sample	Chloride	Sodium	Sulfate	Magnesium	Potassium	Calcium	Total Ions
GSL2011.4	6.29	3.33	0.89	0.43	0.27	0.11	11.31
GSL2011.5	0.66	0.33	0.12	0.09	0.04	0.04	1.28
GSL2011.6	69.13	38.16	8.93	3.77	2.41	0.40	122.80
GSL2011.7[1]	79.99	42.25	10.00	4.07	2.72	0.80	139.82
GSL2011.8[1]	83.31	45.36	10.76	4.68	2.81	0.60	147.51
GSL2011.9[1]	117.00	62.35	15.45	6.74	4.21	1.00	206.75
GSL2011.10[1]	143.60	76.81	19.55	8.51	5.21	1.10	254.77
GSL2011.11	73.56	40.50	9.22	4.13	2.58	0.40	130.39
GSL2011.12[1]	153.30	81.39	21.09	9.17	6.06	1.10	272.11
GSL2011.13[1]	179.00	93.69	24.01	11.06	7.31	1.20	316.26
GSL2011.14[1]	186.10	96.67	25.76	11.18	6.31	1.30	327.32
GSL2011.15[1]	179.90	93.85	25.20	11.12	6.53	1.30	317.90
GSL2011.16	60.25	33.77	7.80	3.10	1.96	0.70	107.57
CM basal salts control	164.80	98.33	8.35	1.09	1.11	1.50	275.19

Molar

Sample	Chloride	Sodium	Sulfate	Magnesium	Potassium	Calcium	Total Ions
GSL2011.4	0.177	0.145	0.009	0.018	0.007	0.003	0.36
GSL2011.5	0.019	0.014	0.001	0.004	0.001	0.001	0.04
GSL2011.6	1.950	1.660	0.093	0.155	0.062	0.010	3.93
GSL2011.7[1]	2.256	1.838	0.104	0.167	0.069	0.020	4.46
GSL2011.8[1]	2.350	1.973	0.112	0.192	0.072	0.015	4.71
GSL2011.9[1]	3.300	2.712	0.161	0.277	0.108	0.025	6.58
GSL2011.10[1]	4.051	3.341	0.203	0.350	0.133	0.027	8.11
GSL2011.11	2.075	1.762	0.096	0.170	0.066	0.010	4.18
GSL2011.12[1]	4.324	3.540	0.220	0.377	0.155	0.027	8.64
GSL2011.13[1]	5.049	4.075	0.250	0.455	0.187	0.030	10.05
GSL2011.14[1]	5.250	4.205	0.268	0.460	0.161	0.032	10.38
GSL2011.15[1]	5.075	4.082	0.262	0.457	0.167	0.032	10.08
GSL2011.16	1.700	1.469	0.081	0.127	0.050	0.017	3.44
CM basal salts control	4.649	4.277	0.087	0.045	0.028	0.037	9.12

[1] Indicates north arm sites

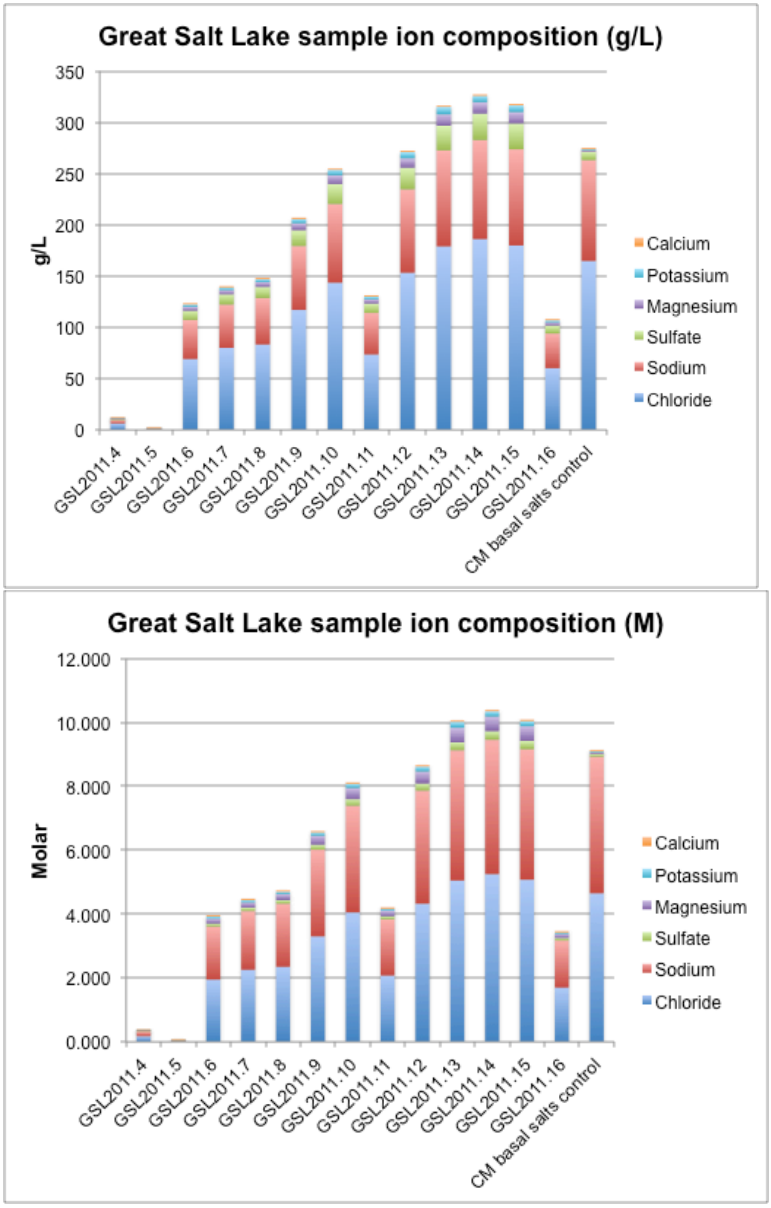

Figure 2-2: Ion concentrations at Great Salt Lake sample sites. Samples were taken from spatially diverse sites within the GSL (Fig 2-1) and ions were measured in g/L and converted to molarity. North arm sites (7, 8, 9, 10, 12, 13, 14, 15) are most saline, with south arm sites near the causeway intermediately saline. CM basal salts is the salt solution that forms the base of the standard rich growth media for *Halobacterium* species. The presence of minimal technical variability in the measurement process was assessed using triplicate controls (Table 2-3).

Table 2-3: Measured ion composition of evaporated and rehydrated Great Salt Lake sample 14.[1]

Evaporation

Sample	% Evaporation	Sample Time (h)	Chloride	Sodium	Sulfate	Magnesium	Potassium	Total Ions	K:Na ratio
GSL 14 Evap 0	0.00%	0	5.100	4.321	0.269	0.494	0.206	10.390	0.048
GSL 14 Evap 1	23.05%	17.25	6.300	5.472	0.354	0.665	0.299	13.089	0.055
GSL 14 Evap 2	42.15%	33.5	6.300	5.271	0.427	0.770	0.334	13.101	0.063
GSL 14 Evap 3	49.13%	46.25	6.500	5.223	0.478	0.879	0.371	13.452	0.071
GSL 14 Evap 4	79.95%	62.5	6.500	4.337	0.750	1.360	0.608	13.554	0.140
GSL 14 Evap 5	89.34%	81.75	6.750	3.898	0.907	1.649	0.728	13.933	0.187

Rehydration

Sample	Sample Time (h)	Chloride	Sodium	Sulfate	Magnesium	Potassium	Total Ions	K:Na ratio
GSL 14 Rehyd 1[2]	0	0.513	0.363	0.023	0.089	0.036	1.023	0.099
GSL 14 Rehyd 2	0.25	0.775	0.534	0.082	0.182	0.077	1.650	0.144
GSL 14 Rehyd 3	0.75	1.150	0.769	0.157	0.306	0.139	2.521	0.181
GSL 14 Rehyd 4	1.33	2.475	2.048	0.156	0.282	0.123	5.085	0.060
GSL 14 Rehyd 5	4.08	2.700	2.219	0.124	0.250	0.103	5.396	0.046
GSL 14 Rehyd 6	44.5	4.150	3.318	0.321	0.597	0.250	8.637	0.075

Controls

Sample[3]	Chloride	Sodium	Sulfate	Magnesium	Potassium	Total Ions	K:Na ratio
Pure deionized water a	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	ND[4]	ND[4]
Pure deionized water b	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	ND[4]	ND[4]
Peptone a	0.011	0.015	0.0006	<0.0001	0.0016	ND[4]	0.107
Peptone b	0.011	0.015	0.0006	<0.0001	0.0017	ND[4]	0.111
Peptone c	0.011	0.015	0.0006	<0.0001	0.0016	ND[4]	0.109
Peptone + citrate a	0.014	0.046	0.0006	<0.0001	0.0017	ND[4]	0.037
Peptone + citrate b	0.014	0.046	0.0006	<0.0001	0.0017	ND[4]	0.038
Peptone + citrate c	0.014	0.046	0.0006	<0.0001	0.0016	ND[4]	0.035

[1] All ion values in molar

[2] Rehydration sample 0 = pure deionized water, and is not included in the timecourse.

[3] letters following sample names indicate technical triplicates

[4] Total not determined; constituent measurements below detection threshold

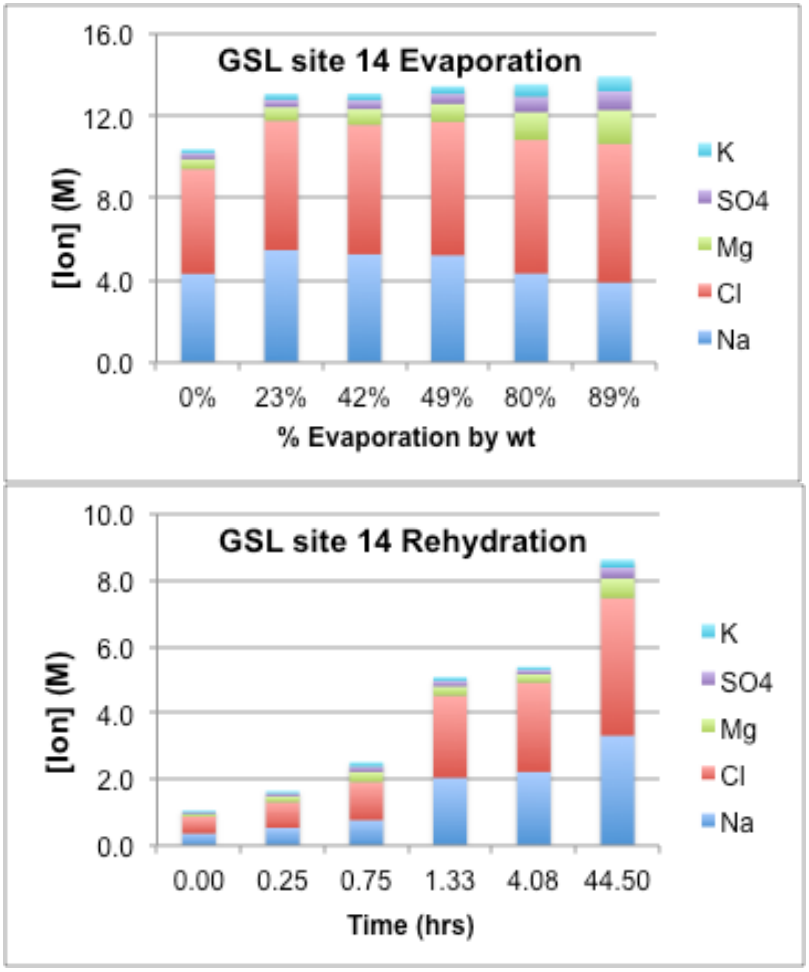

Figure 2-3: Ion concentrations as GSL water evaporates and rehydrates. Measurements were taken on 50 mL samples taken from a starting volume of 500 mL GSL water from site 14. During rehydration, experimental beakers were not covered to prevent additional evaporation. The purpose of this method was to measure dynamics that more closely resemble those in the natural environment. The presence of minimal technical variability in the measurement process was assessed using triplicate controls (Table 2-3).

Chapter 3: Salinity adaptation in *H. salinarum* NRC-1 is multidimensional and reflects natural hypersaline lake chemistry

Abstract

Prokaryotes require the major ions Na, Cl, Mg, SO_4 and K for survival, but imbalances in total ions and in ion ratios produce toxic effects. As a result, cells have evolved mechanisms for maintaining osmotic balance across the membrane in environments ranging from zero dissolved salts to saturation. Because sodium and chloride are the dominant ions in aquatic habitats, most of what is known about prokaryotic osmotolerance is drawn from work on NaCl-based perturbations. This limited scope is surprising, because changes in additional major ion ratios are known to have major impacts on survival and toxicity in invertebrates. Here, we systematically evaluate the effects of both total salinity and NaCl, as well as other ion ratios, on fitness in an extreme halophilic archaeon. Halophilic archaea are uniquely adapted to high salinity, but little is known about their osmotic response to varying ion composition. Here we show that growth rate is determined by total salinity, [NaCl] and K:Mg ratio separately, using a multifactorial set of complex growth media. We define growth optima for multiple aspects of salinity, and finally we show that peaks in growth rate correspond with media compositions that most closely resemble natural hypersaline lakes.

Introduction

Salinity, or the dissolved ionic content of a body of water, is a ubiquitous property that every organism must negotiate and adapt to. Diverse prokaryotes have evolved to thrive along the entire spectrum of salinity, from zero to saturation, and a variety of strategies have emerged for maintaining intracellular salinity at physiologically permissive levels. Two main strategies have emerged for maintaining osmotic balance: 1) a "Low salt-in" strategy in which salts are excluded from the cytoplasm, and osmotic balance is maintained by the energetically expensive production of compatible organic solutes (glycerol, ectoine, betaine, sucrose, etc), and 2) a "High salt-in" strategy in which cytoplasmic K^+ concentrations are adjusted (via active and passive transport) to equal to the external Na^+ concentration[82,83]. The low-salt strategy is used by prokaryotes and eukaryotes alike, while the high-salt strategy is limited to halophilic bacteria and archaea[40]. Despite these osmotolerance mechanisms, all organisms have a limited niche within the salinity spectrum where growth and fitness are optimal. Outside of this range, cells suffer

due to membrane depolarization and loss of electrochemical gradients, excessive or loss of turgor pressure, and systemic metabolic dysregulation[80,113].

Within a given salinity range or niche, however, conditions may vary widely over space and time in the natural environment. Isoosmotic solutions can contain vastly different ion ratios, with consequences for physiology and fitness. These consequences have been explored in the toxicology literature, in order to understand the effects of toxic saline effluent on invertebrate survival and physiology[114-116]. In *Cereodaphnia* (water flea), *Sphaerium* (fingernail clam), *Gyraulus* (snail) and *Tubifex* (worm), toxicity from total dissolved solids (TDS) and chloride is dependent on the ionic composition of the solution. Specifically, increased Ca^{2+} and Mg^{2+} reduce the toxicity (increase the LD50) of high-TDS solutions[115], and this relationship has been recapitulated using both field-collected effluent and effluent-mimicking growth media[116]. Browne, *et al* used a combinatorial approach to examine the effects of total salinity (as increasing concentrations of artificial seawater) and temperature on *Artemia* brine shrimp.

Interestingly, the synergistic effects of ion composition on physiology and fitness have not been examined to the same extent in prokaryotes. Prokaryotic osmotic stress response is critical for cell survival due to the effects of ion imbalance and non-ideal turgor pressure, and osmotolerance has been examined in bacteria and archaea in terms of response to total salinity or changes in ionic salts vs. non-ionic osmolytes (sucrose, polyols) alone. The topic has been thoroughly reviewed elsewhere[80,113]. In osmotolerance experiments in bacteria and archaea alike, strategies generally involve changing the total salinity or concentration of a single salt (with or without constant total salinity) and measuring resultant phenotypes, growth or gene expression[112,117-120]. A few notable exceptions have explored the combinatorial effects of two salts or salt and other environmental effects. An astrobiology study looking at survival of halophilic bacteria (*Bacillus aquimaris, Halomonas salina, Virgibacillus picturae, Salinivibrio costicola*) under martian-like high magnesium conditions found that growth at a given [$MgSO_4$] is modified by [NaCl], and is further influenced by differences in temperature and pH[121].

Halophiles like those in the astrobiology study make attractive models for investigating osmotolerance and complex environmental response in general. *Halobacterium salinarum* in particular is an extreme halophilic archaeon adapted to total salinities upwards of 4 M, and has evolved in hypersaline lakes like the Great Salt Lake where salinity, temperature, and oxygen frequently fluctuate. In addition, *H. salinarum* has been a model organism for the systems

biology of complex microbe-environmental interactions and thus its culturing regime and molecular biology tools are well established. Although survival in extreme high salinities is the most notable aspect of haloarchaeal physiology, it has only been examined in terms of changes in NaCl and KCl independently. In the natural environment, salinity change often occurs as gradients in time (evaporation, dilution by rainfall) or space (depth gradients, isolated pools with divergent chemistries). Further, these gradients may exist in ion composition, total salinity, or both. Thus, a more comprehensive picture of salinity response in halophilic archaea is needed, in order to better characterize the niche that these organisms inhabit. Halophilic archaea hold promise for salinity-dependent hydrocarbon degradation[119,122,123], and may play a role in soil carbon cycling[124,125], making this research additionally pertinent.

Here, we used a multifactorial approach to assay the independent effects of total salinity, [NaCl], and Mg:K ratio on fitness and niche width, in *H. salinarum* NRC-1. We hypothesized that because this model organism has been cultured in the laboratory for over four decades, it will be most successful in standard laboratory media and fitness will suffer in other media compositions. We also investigated salinity-dependent differences in cell morphology. Although rod-shaped *Halobacteria* take on a spherical shape when total salinity falls and when Mg is depleted through chelation[39,126], the extent to which ion composition alone influences cell morphology is not known and may have consequences for ecological interactions within the water column. *H. salinarum* is known to interact with the halophilic alga *Dunaliella salina* for reciprocal carbon sharing[127], and whether changes in osmolality influence this symbiosis is yet unknown.

By measuring maximum growth rate over these three environmental gradients (Total salinity, [NaCl] and K:Mg), we analyzed the contribution of each aspect of salinity to growth and fitness. By viewing the range of tolerated environments in three dimensions, simultaneously, we gain a measure of the degree to which cells are specialized to their chemical environment and more specifically, to laboratory media conditions. This measure of specialization in a laboratory model organism forms a baseline for evolutionary comparisons of generalist vs. specialist strategies in osmotolerance across diverse halophilic archaea.

Materials and methods

Lab strains and culturing conditions

H. salinarum NRC-1 Δ*ura3* was cultured in standard CM complex growth media containing the following: 250g/l (4.28 M) NaCl, 20g/l (0.081 M $MgSO_4$) $MgSO_4$*$7H_2O$, 2g/l (0.027 M) KCl, 3g/l (0.01M) sodium citrate dehydrate, 10g/l Oxoid brand bacteriological peptone, and 50 mg/l uracil, without trace metals[128]. Solid agar plates were made from this same media, with 20 g/L bacteriological agar. In all incubations, cultures were grown at 37°C with continuous light and shaking (220 rpm). The Δ*ura3* mutant is the auxotrophic parent strain of the entire *H. salinarum* NRC-1 mutant library and was selected for use here as an appropriate baseline for future comparisons with mutant strains[87]. All experiments here were performed using Δ*ura3,* hereafter referred to as *H. salinarum.* All cultures were grown at 37 °C with light and shaking, except during growth curve data acquisition when cultures were grown at 37 °C with shaking in the dark.

Growth media and multifactorial design

Media compositions represent a three-factorial design with the following factors: total salinity (5 levels: 100% (8.776 M ionic salts), 90%, 80%, 70% and 60% by volume), [NaCl] (3 levels: high (96% of total M), medium (78%) and low (59%)) and K:Mg molar ratio (5 levels: 0:100, 25:75, 50:50, 75:25 and 100:0), for a total of 75 media (Table 3-1, Figure 3-1). Compositions represent variations on the standard CM media (labeled Media 2: 100% total salinity, 4.28 M NaCl, 25:75 K:Mg), and this media serves as a control for comparison across all *H. salinarum* experiments. [NaCl] and K:Mg ratio were varied by molarity within each level of total salinity. The concentration of the organic ingredients peptone and citrate was held constant across all media, at 10g/L peptone and 3 g/L citrate. To make the total salinity series, the 100% salinity media series was diluted in a solution of peptone and citrate to each of the remaining four factor levels (60-90%, in 10% increments). The pH of all media was adjusted to ~7.20 +/- 0.03, and all were autoclave sterilized. Chemical lot numbers were recorded and all media batches for all experiments were made using the same lot; only two media batches were used to complete all experiments, in order to reduce batch effects.

Bioscreen growth assays and growth rate calculations

Growth assays were performed in two Bioscreen C instruments (Growth Curves USA, Piscataway, NJ), with a maximum throughput of 400 cultures (200 µl each) (Figure 3-2). For all strains and replicates in all experiments, fresh colonies were grown up in 5 mL cultures of Media 2, to OD_{600} = 0.4-0.6 (mid-log phase). These cultures were passaged in Media 2 (Standard CM) at OD 0.05 to synchronize the cultures, and grown to OD_{600} = 0.6. All cultures were then diluted in Media 2 to OD_{600} = 0.55 to ensure identical starting OD. Using a Hamilton Microlab Star robot (Hamilton Robotics, Reno, NV) to pair each of 8 cultures with each of up to 16 different media, we inoculated each well of a Bioscreen honeycomb plate with culture and experimental media to a final OD of 0.05. Hamilton robot program files and protocols are included in the Appendix. All cultures were maintained in Media 2 until transfer to the bioscreen plate, when they were added at <1:10 in experimental media. This method required ~5% differences between media salinity and final experimental salinity, but was ideal because it did not require a stress-inducing centrifugation or rinse step to remove Media 2. Cultures were incubated in the Bioscreen C for 110 hours at 37 °C with maximum shaking. OD_{600} was measured every 15 minutes for the duration of the experiment.

Maximum growth rate was extracted from Bioscreen OD measurements using a custom R package, "Growth Curve Analysis Function," to automate analysis[129]. We use maximium growth rate as a surrogate for fitness, and although it is not the only component of fitness, it has been shown to be the most important[65]. Briefly, cell optical density measurements, relevant metadata and plate layout schemes were combined in a single database, which enabled rapid calculation of growth curve parameters (max growth rate, max OD, area under curve, etc). Maximum growth rate was calculated from smooth spline fitted growth curves, as the inflection point where the second derivative transitioned from positive to negative [130].

Statistical analysis

All analyses were performed in R, versions 2.13 and 2.14. The multifactorial design enabled us to test the hypothesis that mean maximum growth rate differs significantly as a function of total salinity, [NaCl] and K:Mg. To test this hypothesis we used an orthogonal ANOVA model of maximum growth rate on three factors: Total salinity, [NaCl] and K:Mg. Data were means of biological replicates; variance among replicates was low so replicate was not

included as a random effect. No interaction was included because there was no *a priori* hypothesis for what an interaction would mean biologically. In subsequent exploratory models, total salinity × [NaCl] and total salinity × K:Mg interactions were included but models were not originally powered for this analysis, due to lack of *a priori* biological reasons to suspect interactions. Models were generated using the R function lm(); see the Appendix for full R scripts. Residual vs. fit plots and additional diagnostics were performed on each model and revealed satisfactory residual normality and homoscedasticity. Outliers included maximum growth rates in media 11 and 12, which produce low growth rates as expected due to low total and low K:Mg. These outliers were not omitted.

Great Salt Lake chemistry and multifactorial growth media comparisons

Because *H. salinarum* max growth rate was unexpectedly high in non-control experimental media, we wondered if peaks in fitness in certain media were due to those media resembling aspects of natural hypersaline chemistry. To answer this question, we mapped the salinity data from the Great Salt Lake (Chapter 2) onto the matrix of synthetic media compositions. Analyses were performed in R.

We compared pairwise differences between north arm GSL sites and evaporation/rehydration composition, and each growth media in the multifactorial array (20 GSL samples x 75 media). Pairwise differences were calculated for the following aspects separately: Total salinity, Na, Mg and K. Using a nested comparison procedure, we compared media and natural samples according to each ion, in order of ion abundance (total salinity, Na, K, and Mg). To start, we sought the total salinity level that most closely matched each GSL composition. Within this closest-matched group of media with the same total salinity, we then identified closest matches for each additional ion. Often but not always, the media with the best-matched [Mg] or [K] fell within the group of five media with the same best-matched Na level (completely nested). Na- and Mg-level matches are both given in Figure 3-12, with completely nested GSL samples indicated by circles with thick edges. No K-level matches were completely nested, due to the limited number of media (confined to the high sodium group) with low [K] levels. The results of this analysis of salinity differences are presented as heatmaps, with red indicating small differences between media and GSL compositions. White asterisks indicate differences in the smallest 5%, based on empirical cumulative distribution functions. This analysis allows the

identification of a single experimental media that most closely matches each GSL sample and GSL evaporation/rehydration composition, for each major ion.

Cell images and analysis

In addition to maximum growth rate, we also evaluated differences in cell morphology among a subset of experimental media. Specifically, we analyzed images of *H. salinarum* cultures at 400X magnification at six timepoints along the growth curve, in media 1, 2, 5 and 12. These media were selected because they represent both maximal (media 2) and minimal (media 12) growth rate, as well as both extremes in K:Mg ratio (media 1 and 5). Two replicate cultures were grown from colony in 5 mL CM media as described above, and passaged at OD=0.05 after reaching mid-log phase. After reaching mid-log phase once more, cultures were divided into 4 equal aliquots (one per test medium) and pelleted at 8,000 RPM. Pelleting does not visually change cell shape (data not shown). Pellets were resuspended in 5 mL CM, and then diluted further to OD=0.05 in 50 mL of media 1, 2, 5 or 12. T=0 hours was recorded after dilution into test media. At each of six timepoints (0, 19, 27, 40, 64, and 90 hours), ODs were measured, cells were counted in a hemacytometer, and a series of images was taken.

Cultures were diluted in their respective experimental media to OD=0.2 for gel-coated, wet-mount slide preparation. Gel sides were prepared by compressing 2 ul of 2% agarose on slides in advance. Gel slides were dried for 5 minutes in a 60 °C drying oven and allowed to cool just before imaging. These slides immobilize motile cells and allow images to be taken in a single focal plane, making cell area to volume extrapolations more meaningful. Images were taken on a Nikon Eclipse E400 bright field microscope, with the 100X oil immersion objective (phase contrast). Images were captured with an attached Leica camera and analyzed using the Leica Application Suite, version 3.5.0. All images were taken at 2048 x 1536 resolution, at 118.1 ms exposure time. Image analysis of cell circularity ($4\pi(\text{area}/\text{perimeter}^2)$; a value of 1.0 indicates a perfect circle), aspect ratio, Feret's distance (longest distance between any two points), and total area was conducted using ImageJ for Mac, and R. To test the hypothesis that growth media with different ion compositions change cell morphology, Two-way ANOVAs were performed on measured parameters, with media and time as factors. Data were reciprocally transformed for two-way ANOVAs to improve normality and variance equality.

Results and Discussion

Maximum growth rate reflects fitness and niche breadth across a 2D salinity landscape

By growing *H. salinarum* along gradients in total salinity and ion composition, we observed that cells perceive and respond to both components of salinity independently (Figure 3-3). By using a multifactorial approach, the data represent reaction norms (phenotype × environment plots) across three different environmental gradients: total salinity, [NaCl] and K:Mg.

Although it is well-established that *H. salinarum* requires total salinities greater than 2.5 M[40], detailed salinity niche characterizations and fitness curves have not been published. This multifactorial design employed three factor levels to evaluate the two major components of salinity, total salinity (factor: total salinity) and ion composition (factors: [NaCl] and K:Mg). Mean maximum growth rate (μ) (T_d=13.3 h) occurred at 3.949 M total salts (7.959 total ions), or 90% of standard CM media salinity. Higher and lower salinities are less optimal, with doubling times, T_d, reaching 31 hours in the lowest level of total salinity (2.633 M). Along the [NaCl] gradient, mean μ occurred at the highest level, and in the K:Mg gradient, at a ratio of 75:25.

Taken together, the reaction norm plots produced by this design revealed unexpected patterns. First, the growth medium that produced the highest value of μ was not the standard CM media (Media 2), but rather the 90% dilution of Media 2. This is curious because this media was first described in 1960[131,132], and became a standard rich medium for culturing *H. salinarum* in many research groups[128,133]. It is unclear how this media was developed, and whether or not it was formulated to produce the highest growth rate. Regardless of the media formulation 50 years ago, our surprise is borne from the numerous examples of model organisms adapting to their culturing environments[134-137].

A second unexpected result was the intermediate peaks (local maxima) in μ along the ion composition gradients, within each level of total salinity. Within each level of total salinity, three peaks in μ values emerged, one within each level of [NaCl]. The highest of these peaks in mu generally occurred in Media 2 or 3 across all dilutions except 60%, where several patterns broke down due to extremely low growth rates. This is consistent with the conclusion that the ion ratio of CM enables the highest fitness, independent of total salinity. However, additional peaks occurred in media with ion ratios and total salinities that do not resemble CM at all. For example,

Media 9 and all of its dilutions produced fitness peaks nearly as high as those produced by Media 2 and 3 (but still significantly different by Tukey's honest significant difference). Close in composition to Media 9, Media 8 and 10 also produced μ peaks in other total salinity levels. Similarly, Media 14 and 15 also produced local peaks in mu, but at the lowest level of [NaCl]. The highest [NaCl] level tended to produce the highest local peak, and the lowest level produced the lowest local peak.

Common to these intermediate peaks (Media 9, 14 and 15), is the K:Mg ratio of 75:25 or greater. The K:Mg ratio of media 2 (25:75) does not produce peaks within other levels of [NaCl], indicating that increased [NaCl] seems to rescue the effects of low K:Mg. In *H. salinarum*, potassium levels are regulated by a combination of passive transport and active, ATP-driven pumps. Below 5 mM K^+, growth defects were observed in a study of *H. salinarum* potassium transport[112]. Below 200 μM K^+, the Kdp active transport system is expressed and requires ATP to pump potassium into the cell and maintain the cationic balance across the membrane[112]. The converse effects of high relative Mg when K+ is low are unknown, and cannot be disentangled with this study design. It is surprising to see a relieved defect at high [NaCl] nonetheless, because additional intracellular K+ is required to maintain balance.

Maximum growth rate is explained by total salinity and ion composition

We tested the hypothesis that each of total salinity, [NaCl] level and K:Mg ratio influence mean μ independently of one another (Figure 3-4). In ANOVA models, mean μ was significantly influenced by each factor (Total salinity $F_{4,71} = 23.60$, $p<0.0001$; [NaCl] $F_{2,73} = 25.61$, $p<0.0001$; K:Mg $F_{4,71} = 4.97$, $p<0.005$), demonstrating that *H. salinarum* perceives salinity as a complex perturbation, comprised of total salinity and ion ratio components.

Because the major ions Na, Cl, Mg, SO_4 and K are involved in countless cellular processes, their effects on physiology and the regulation of intracellular concentrations must be broad and complex. This opens a number of questions about the coordination of the 8+ membrane-bound ion channels and pumps, and the regulatory networks involved in their regulation. Do ion-specific transcriptional regulators exist, and how are individual osmotolerance proteins regulated at the transcriptional level? For example, in potassium modulation by the KdpFABCQ system, regulation is known to occur by negative regulation upstream of the pKdpF promoter, by KdpQ autoregulation, and by a yet unknown source[111]. Our additional work

suggests a role for post-transcriptional regulation of the Kdp system, and will be discussed in Chapter 4.

Cell shape and morphology are highly plastic across ion composition gradients

Phase-contrast microscopy images revealed that cell shape and timing of gas vesicle biogenesis is affected by changes in ion ratio (Figure 3-5). We imaged cells growing in four media from the ion composition series (Media 1, 2, 5 and 12), selected to represent extreme conditions within the 100% total salinity series. Media 1 and 5 have K:Mg ratios of 0:100 and 100:0, respectively, with identical [NaCl]. Media 2 is the standard laboratory control media, and Media 12 has low [NaCl], 25:75 K:Mg, and produced the second-lowest maximum growth rate in the media series. Media 12 shares a K:Mg ratio with the control media 2, but differs in [NaCl].

As expected[126], Mg depletion causes rod-shaped cells to become irregular spheres. Spherical cells appear at 20 hours of growth and no additional rods are present by 40 hrs. Interestingly, spherical cells were observed to lyse during observation, releasing their gas vesicles. Although cultures were diluted according to OD to achieve uniform density for photographs, later time points in media 5 were more sparse than in other media cultures. This suggests that cell death (perhaps due to lysis) occurred frequently, but the remaining gas vesicles gave the illusion of a high OD. This unique Mg-dependent phenotype has been observed previously and exploited for making competent *Halobacteria*, however the underlying mechanism is still in question. The presence of diplococci suggests that low Mg does not interfere with cell division.

The timing and extent of gas vesicle biogenesis also varied according to media. Media 12 produced cells appearing white with gas vesicles by 40 hours, whereas other media did not produce this phenotype until much later (90 or 137 hrs), if at all. This increase in gas vesicles was also reflected in OD values, which exceeded OD=10.0 in late stationary phase. We ascertained that the relationship between OD and cell count was consistent for all media (Figure 3-6). Interestingly, a previous survey of gene regulation in low (2.9M) and high (5M) salinity did not reveal gas vesicle genes or regulators to be differentially regulated in either of these conditions. In the study, only NaCl (and by extension, total salinity) was varied in the media, and the low salinity condition was 0.3M higher than in our experiments (2.62M). Gas vesicle biogenesis may be a part of an osmotic stress response that is activated at lower [NaCl] levels

than previously tested. In the next chapter, we evaluate gene expression in the same media as were evaluated here.

Finally, we used ImageJ software to quantify the differences in cell shape among media and timepoints (Figure 3-7, Table 3-2). Differences among the media were statistically significant in all cell shape metrics, both linear and area-based, after accounting for growth phase differences in shape. The spheroid cells in Mg-depleted Media 5 appear have the smallest aspect ratio and highest circularity, and are significantly different from all other media by Tukey's honest significant differences. While there was no clear morphological signature of depleted potassium, cells grown in Media 1 exhibit the most rod-like shape, with the lowest circularity and largest aspect ratio across most time points. When K is low, it is actively pumped in across the membrane. If perfect K-Na balance across the membrane is still not achieveable via active transport by the Kdp system, water leakage out of the membrane push the intracellular cation concentration up to its ideal level, thus reducing the volume of the cell[112]. Alternatively, energy-depleted cells from Kdp-associated ATP use may take on a thinner shape.

Differences in cell shape and morphology support the hypothesis that *H. salinarum* NRC-1 employs plasticity as part of its arsenal of osmotolerance strategies. Here we have measured four non-fitness derived phenotypes across an environmental gradient (ion composition), and we observe high plasticity among media varying in ion composition. All cultures shared an identical growth history, and all were started from colony in the same media. History-dependent attributes cannot explain the observed phenotypes. Similarly, plasticity as an adaptive trait itself has been explored along salinity gradients in plants. A 2008 study of salt marsh perennials investigated responses (leaf and root measurements) to salinity treatments, and looked specifically at the effect of previous growth in different saline microhabitats. The study found that all measured traits were highly plastic and did not differ according to microhabitat or growth history. Rather than evidence of high- or low-salt adapted genotypes, the results suggest that plasticity itself is adapted in saline habitats and may be fixed in the plant species under study. This observation in plants is intriguing in light of the eukaryotic characteristics shared within the Archaea. We further explore how plasticity varies among haloarchaea with vastly different recent histories in Chapter 5.

Fitness maxima occur in growth media that most closely resemble Great Salt Lake chemistry

Upon the unexpected observation that *H. salinarum* attains peak maximum growth rates in non-control media, we asked whether any of the non-control media share saline properties with the natural hypersaline environments where *H. salinarum* evolved. If laboratory model organisms reach peak fitness levels in salinity conditions they have never seen in the lab, perhaps they are successful in these conditions because they resemble environments experienced prior to laboratory culturing. To answer this question, we systematically compared the total salinity and ion compositions of each of the 75 experimental media with each of the GSL water samples, including the evaporation series (Table 3-1 (media), Table 3-3 (GSL)).

The first comparison we made was at the level of total salinity (Figure 3-8). As expected, the southern-most GSL sites (7, 8, 9 and 10) matched with 60, 70 or 90% total salinity while the northern samples (12-15) were closest in total salinity to the 100% series. GSL sites that differ from all of the 75 media by 1.6 M or more were omitted from further match analysis due to the decreasing biological relevance of closest matches at this level (*i.e.* Evaporation samples 1-5 and rehydration samples 1-3).

Conditioning on the set of 15 media that comprised the best-matched total salinity level, we then asked which of the three NaCl levels best matched the [Na] concentration of the GSL samples. We performed the same analysis of differences on both [Mg] and [K] as well, conditioning on total salinity (Figures 3-8, 9, 10, 11). In many cases, the single growth media best matched for Mg or K did not fall within the set of 5 media best matched for NaCl. This discordance simply reflects the fact that the 75 multifactorial media do not encompass all possible combinations of ion concentrations, and were not formulated based on GSL chemistry. However, there were eight GSL-growth media pairs for which differences in each ion were fully nested. That is, the media best matched for Mg or K fell within the set of best-matched media for Na, which in turn fell within the set of 15 media best matched for total salinity (Table 3-4).

Six of these eight media, which most closely match GSL conditions in three different aspects of salinity, produced local maxima in *H. salinarum* fitness (Figure 3-12). The six media that produced local maxima represent four different levels of total salinity (100, 90, 70 and 60%), but they share similar sodium and magnesium content. Specifically, regardless of total salinity, the six media were in the "medium" NaCl group (39.2% of total dissolved salts, in Molar). All

six also fell into the K:Mg ratio of 50:50, and contained 5.5% Mg. That this combination of sodium and magnesium produced peak maximum growth rates *and* is most similar to six of the GSL samples is interesting, in light of the fact that the laboratory control media contains 48.9% sodium and 1.2% magnesium. Even if this discordance is only circumstantial, these data suggest that optimal growth in the laboratory does not need to come at the expense of saline conditions that depart vastly from the natural environment.

The observed GSL-media similarities are also potentially significant because they offer that evolutionary growth history is one possible explanation for the peaks in fitness that do not correspond with the laboratory control media (Media 2). Of all possible combinations of total salinity, sodium, magnesium and potassium, I have measured growth in 75, and have chemical measurements in an additional 20. The motivating hypothesis was that perhaps the peculiar growth rate peaks in the 75 media may be attributable media similarity with natural hypersaline chemistry, in which our model organism evolved. The first step toward testing this hypothesis was assessing whether any compositional overlap exists between the group of 75 media and 20 GSL samples. The only *a priori* similarity between these two groups was their shared capacity to support the presence of Halobacteria; although the 75 media were designed to cover *Halobacterium's* wide hypersaline niche, their ability to support growth was not known at the outset. The fact that eight of 20 GSL samples had close media matches in all dimensions of salinity tested, is encouraging in light of the original hypothesis. Even more promising, is the observation that six of those eight GSL-media matches produced local peaks in μ, outside of the control media.

These data are encouraging, but they have merely failed to discount the possibility that laboratory model organisms respond favorably to conditions resembling ancestral hypersaline environments. In order to further explore how well-adapted is a decades-old laboratory culture to its synthetic environment compared with its once-native environment, experiments with natural isolates are required. In chapter 5, I use similar reaction norms to evaluate the extent to which lab models and natural isolates are adapted in synthetic and natural hypersaline conditions.

Chapter 3 Figures and Tables

Table 3-1: Multifactorial media design and compositions with maximum growth rate (mu)

Media composition name	Total salinity dilution series	[NaCl] level	K:Mg (M ratio)	Mean max growth rate (hr^{-1})	SD max growth rate	Total Salinity (M salts)	NaCl (M)	$MgSO_4$ (M)	KCl (M)	NaCl (g/L)	$MgSO_4$*7H2O (g/L)	KCl (g/L)	Na (M)[i]	Cl (M)[i]	Mg (M)[i]	SO_4 (M)[i]	K (M)[i]	Total Salinity (M ions)[i]
CM.Mratio.1	100%	High	0:100	0.061	0.002	4.388	4.28	0.108	0	250	26.6	0.0	4.326	4.294	0.108	0.109	0.002	8.839
CM.Mratio.2	100%	High	25:75	0.067	0.001	4.388	4.28	0.081	0.027	250	20.0	2.0	4.326	4.321	0.081	0.082	0.029	8.839
CM.Mratio.3	100%	High	50:50	0.065	0.003	4.388	4.28	0.054	0.054	250	13.3	4.0	4.326	4.348	0.054	0.055	0.056	8.839
CM.Mratio.4	100%	High	75:25	0.066	0.001	4.388	4.28	0.027	0.081	250	6.7	6.0	4.326	4.375	0.027	0.028	0.083	8.839
CM.Mratio.5	100%	High	100:0	0.055	0.002	4.388	4.28	0	0.108	250	0.0	8.1	4.326	4.402	0.000	0.001	0.110	8.839
CM.Mratio.6	100%	Medium	0:100	0.039	0.010	4.388	3.42	0.968	0	200	238.6	0.0	3.466	3.434	0.968	0.969	0.002	8.839
CM.Mratio.7	100%	Medium	25:75	0.045	0.007	4.388	3.42	0.726	0.242	200	178.9	18.0	3.466	3.676	0.726	0.727	0.244	8.839
CM.Mratio.8	100%	Medium	50:50	0.060	0.002	4.388	3.42	0.484	0.484	200	119.3	36.1	3.466	3.918	0.484	0.485	0.486	8.839
CM.Mratio.9	100%	Medium	75:25	0.062	0.002	4.388	3.42	0.242	0.726	200	59.6	54.1	3.466	4.160	0.242	0.243	0.728	8.839
CM.Mratio.10	100%	Medium	100:0	0.060	0.003	4.388	3.42	0	0.968	200	0.0	72.2	3.466	4.402	0.000	0.001	0.970	8.839
CM.Mratio.11	100%	Low	0:100	0.005	0.001	4.388	2.57	1.818	0	150	448.1	0.0	2.616	2.584	1.818	1.819	0.002	8.839
CM.Mratio.12	100%	Low	25:75	0.017	0.001	4.388	2.57	1.364	0.454	150	336.2	33.8	2.616	3.038	1.364	1.365	0.456	8.839
CM.Mratio.13	100%	Low	50:50	0.041	0.006	4.388	2.57	0.909	0.909	150	224.0	67.8	2.616	3.493	0.909	0.910	0.911	8.839
CM.Mratio.14	100%	Low	75:25	0.051	0.009	4.388	2.57	0.454	1.364	150	111.9	101.7	2.616	3.948	0.454	0.455	1.366	8.839
CM.Mratio.15	100%	Low	100:0	0.054	0.005	4.388	2.57	0	1.818	150	0.0	135.5	2.616	4.402	0.000	0.001	1.820	8.839
CM.Mratio.1	90%	High	0:100	0.063	0.002	3.949	3.85	0.097	0	225	23.9	0.0	3.896	3.864	0.097	0.098	0.002	7.957
CM.Mratio.2	90%	High	25:75	0.070	0.001	3.949	3.85	0.073	0.024	225	18.0	1.8	3.896	3.894	0.073	0.074	0.026	7.963
CM.Mratio.3	90%	High	50:50	0.067	0.001	3.949	3.85	0.049	0.049	225	12.0	3.6	3.896	3.914	0.049	0.050	0.051	7.960
CM.Mratio.4	90%	High	75:25	0.069	0.001	3.949	3.85	0.024	0.073	225	6.0	5.4	3.896	3.934	0.024	0.025	0.075	7.954
CM.Mratio.5	90%	High	100:0	0.056	0.001	3.949	3.85	0	0.097	225	0.0	7.2	3.896	3.964	0.000	0.001	0.099	7.960
CM.Mratio.6	90%	Medium	0:100	0.043	0.002	3.949	3.08	0.871	0	180	214.7	0.0	3.126	3.094	0.871	0.872	0.002	7.965
CM.Mratio.7	90%	Medium	25:75	0.046	0.001	3.949	3.08	0.653	0.218	180	161.0	16.2	3.126	3.314	0.653	0.654	0.220	7.967
CM.Mratio.8	90%	Medium	50:50	0.061	0.002	3.949	3.08	0.436	0.436	180	107.4	32.5	3.126	3.524	0.436	0.437	0.438	7.961
CM.Mratio.9	90%	Medium	75:25	0.064	0.001	3.949	3.08	0.218	0.653	180	53.7	48.7	3.126	3.744	0.218	0.219	0.655	7.962
CM.Mratio.10	90%	Medium	100:0	0.063	0.001	3.949	3.08	0	0.871	180	0.0	64.9	3.126	3.964	0.000	0.001	0.873	7.964
CM.Mratio.11	90%	Low	0:100	0.004	0.001	3.949	2.31	1.636	0	135	403.3	0.0	2.356	2.324	1.636	1.637	0.002	7.955
CM.Mratio.12	90%	Low	25:75	0.023	0.007	3.949	2.31	1.228	0.409	135	302.6	30.5	2.356	2.734	1.228	1.229	0.411	7.958
CM.Mratio.13	90%	Low	50:50	0.042	0.002	3.949	2.31	0.818	0.818	135	201.6	61.0	2.356	3.144	0.818	0.819	0.820	7.957
CM.Mratio.14	90%	Low	75:25	0.053	0.002	3.949	2.31	0.409	1.228	135	100.7	91.5	2.356	3.554	0.409	0.410	1.230	7.959
CM.Mratio.15	90%	Low	100:0	0.060	0.001	3.949	2.31	0	1.636	135	0.0	122.0	2.356	3.964	0.000	0.001	1.638	7.959
CM.Mratio.1	80%	High	0:100	0.063	0.001	3.5104	3.424	0.0864	0	200	21.3	0.0	3.470	3.438	0.086	0.087	0.002	7.084
CM.Mratio.2	80%	High	25:75	0.068	0.001	3.5104	3.424	0.0648	0.0216	200	16.0	1.6	3.470	3.460	0.065	0.066	0.024	7.084
CM.Mratio.3	80%	High	50:50	0.068	0.001	3.5104	3.424	0.0432	0.0432	200	10.6	3.2	3.470	3.481	0.043	0.044	0.045	7.084
CM.Mratio.4	80%	High	75:25	0.064	0.001	3.5104	3.424	0.0216	0.0648	200	5.3	4.8	3.470	3.503	0.022	0.023	0.067	7.084
CM.Mratio.5	80%	High	100:0	0.048	0.001	3.5104	3.424	0	0.0864	200	0.0	6.4	3.470	3.524	0.000	0.001	0.088	7.084
CM.Mratio.6	80%	Medium	0:100	0.052	0.001	3.5104	2.736	0.7744	0	160	190.9	0.0	2.782	2.750	0.774	0.775	0.002	7.084
CM.Mratio.7	80%	Medium	25:75	0.052	0.001	3.5104	2.736	0.5808	0.1936	160	143.1	14.4	2.782	2.944	0.581	0.582	0.196	7.084
CM.Mratio.8	80%	Medium	50:50	0.060	0.001	3.5104	2.736	0.3872	0.3872	160	95.4	28.9	2.782	3.137	0.387	0.388	0.389	7.084
CM.Mratio.9	80%	Medium	75:25	0.059	0.001	3.5104	2.736	0.1936	0.5808	160	47.7	43.3	2.782	3.331	0.194	0.195	0.583	7.084
CM.Mratio.10	80%	Medium	100:0	0.041	0.001	3.5104	2.736	0	0.7744	160	0.0	57.7	2.782	3.524	0.000	0.001	0.776	7.084
CM.Mratio.11	80%	Low	0:100	0.018	0.002	3.5104	2.056	1.4544	0	120	358.5	0.0	2.102	2.070	1.454	1.455	0.002	7.084
CM.Mratio.12	80%	Low	25:75	0.043	0.001	3.5104	2.056	1.0912	0.3632	120	268.9	27.1	2.102	2.433	1.091	1.092	0.365	7.084
CM.Mratio.13	80%	Low	50:50	0.043	0.001	3.5104	2.056	0.7272	0.7272	120	179.2	54.2	2.102	2.797	0.727	0.728	0.729	7.084
CM.Mratio.14	80%	Low	75:25	0.046	0.001	3.5104	2.056	0.3632	1.0912	120	89.5	81.3	2.102	3.161	0.363	0.364	1.093	7.084
CM.Mratio.15	80%	Low	100:0	0.036	0.001	3.5104	2.056	0	1.4544	120	0.0	108.4	2.102	3.524	0.000	0.001	1.456	7.084
CM.Mratio.1	70%	High	0:100	0.053	0.002	3.072	2.996	0.076	0	175	18.6	0.0	3.042	3.010	0.076	0.077	0.002	6.207
CM.Mratio.2	70%	High	25:75	0.056	0.001	3.072	2.996	0.057	0.019	175	14.0	1.4	3.042	3.029	0.057	0.058	0.021	6.207
CM.Mratio.3	70%	High	50:50	0.055	0.003	3.072	2.996	0.038	0.038	175	9.3	2.8	3.042	3.048	0.038	0.039	0.040	6.207
CM.Mratio.4	70%	High	75:25	0.053	0.004	3.072	2.996	0.019	0.057	175	4.7	4.2	3.042	3.067	0.019	0.020	0.059	6.207
CM.Mratio.5	70%	High	100:0	0.035	0.003	3.072	2.996	0	0.076	175	0.0	5.6	3.042	3.086	0.000	0.001	0.078	6.207
CM.Mratio.6	70%	Medium	0:100	0.049	0.001	3.072	2.394	0.678	0	140	167.0	0.0	2.440	2.408	0.678	0.679	0.002	6.207
CM.Mratio.7	70%	Medium	25:75	0.042	0.003	3.072	2.394	0.508	0.169	140	125.3	12.6	2.440	2.577	0.508	0.509	0.171	6.205
CM.Mratio.8	70%	Medium	50:50	0.048	0.002	3.072	2.394	0.339	0.339	140	83.5	25.3	2.440	2.747	0.339	0.340	0.341	6.207
CM.Mratio.9	70%	Medium	75:25	0.047	0.002	3.072	2.394	0.169	0.508	140	41.8	37.9	2.440	2.916	0.169	0.170	0.510	6.205
CM.Mratio.10	70%	Medium	100:0	0.041	0.002	3.072	2.394	0	0.678	140	0.0	50.5	2.440	3.086	0.000	0.001	0.680	6.207
CM.Mratio.11	70%	Low	0:100	0.021	0.002	3.072	1.799	1.273	0	105	313.7	0.0	1.845	1.813	1.273	1.274	0.002	6.207
CM.Mratio.12	70%	Low	25:75	0.038	0.001	3.072	1.799	0.955	0.318	105	235.3	23.7	1.845	2.131	0.955	0.956	0.320	6.207
CM.Mratio.13	70%	Low	50:50	0.036	0.001	3.072	1.799	0.636	0.636	105	156.8	47.4	1.845	2.449	0.636	0.637	0.638	6.205
CM.Mratio.14	70%	Low	75:25	0.043	0.002	3.072	1.799	0.318	0.955	105	78.3	71.2	1.845	2.768	0.318	0.319	0.957	6.207
CM.Mratio.15	70%	Low	100:0	0.038	0.003	3.072	1.799	0	1.273	105	0.0	94.9	1.845	3.086	0.000	0.001	1.275	6.207
CM.Mratio.1	60%	High	0:100	0.026	0.001	2.633	2.568	0.065	0	150	16.0	0.0	2.614	2.582	0.065	0.066	0.002	5.329
CM.Mratio.2	60%	High	25:75	0.027	0.002	2.633	2.568	0.049	0.016	150	12.0	1.2	2.614	2.598	0.049	0.050	0.018	5.329
CM.Mratio.3	60%	High	50:50	0.023	0.002	2.633	2.568	0.032	0.032	150	8.0	2.4	2.614	2.614	0.032	0.033	0.034	5.327
CM.Mratio.4	60%	High	75:25	0.023	0.003	2.633	2.568	0.016	0.049	150	4.0	3.6	2.614	2.631	0.016	0.017	0.051	5.329
CM.Mratio.5	60%	High	100:0	0.022	0.003	2.633	2.568	0	0.065	150	0.0	4.8	2.614	2.647	0.000	0.001	0.067	5.329
CM.Mratio.6	60%	Medium	0:100	0.019	0.001	2.633	2.052	0.581	0	120	143.1	0.0	2.098	2.066	0.581	0.582	0.002	5.329
CM.Mratio.7	60%	Medium	25:75	0.017	0.003	2.633	2.052	0.436	0.145	120	107.4	10.8	2.098	2.211	0.436	0.437	0.147	5.329
CM.Mratio.8	60%	Medium	50:50	0.025	0.002	2.633	2.052	0.29	0.29	120	71.6	21.6	2.098	2.356	0.290	0.291	0.292	5.327
CM.Mratio.9	60%	Medium	75:25	0.027	0.001	2.633	2.052	0.145	0.436	120	35.8	32.5	2.098	2.502	0.145	0.146	0.438	5.329
CM.Mratio.10	60%	Medium	100:0	0.012	0.001	2.633	2.052	0	0.581	120	0.0	43.3	2.098	2.647	0.000	0.001	0.583	5.329
CM.Mratio.11	60%	Low	0:100	0.016	0.005	2.633	1.542	1.091	0	90	268.8	0.0	1.588	1.556	1.091	1.092	0.002	5.329
CM.Mratio.12	60%	Low	25:75	0.016	0.002	2.633	1.542	0.818	0.272	90	201.7	20.3	1.588	1.828	0.818	0.819	0.274	5.327
CM.Mratio.13	60%	Low	50:50	0.028	0.001	2.633	1.542	0.545	0.545	90	134.4	40.7	1.588	2.101	0.545	0.546	0.547	5.327
CM.Mratio.14	60%	Low	75:25	0.023	0.001	2.633	1.542	0.272	0.818	90	67.1	61.0	1.588	2.374	0.272	0.273	0.820	5.327
CM.Mratio.15	60%	Low	100:0	0.025	0.004	2.633	1.542	0	1.091	90	0.0	81.3	1.588	2.647	0.000	0.001	1.093	5.329

[i] Molar ion concentrations include the contribution from organic media components (bacteriological peptone and sodium citrate); all media contain 10 and 3 g/L, respectively, regardless of dilution series

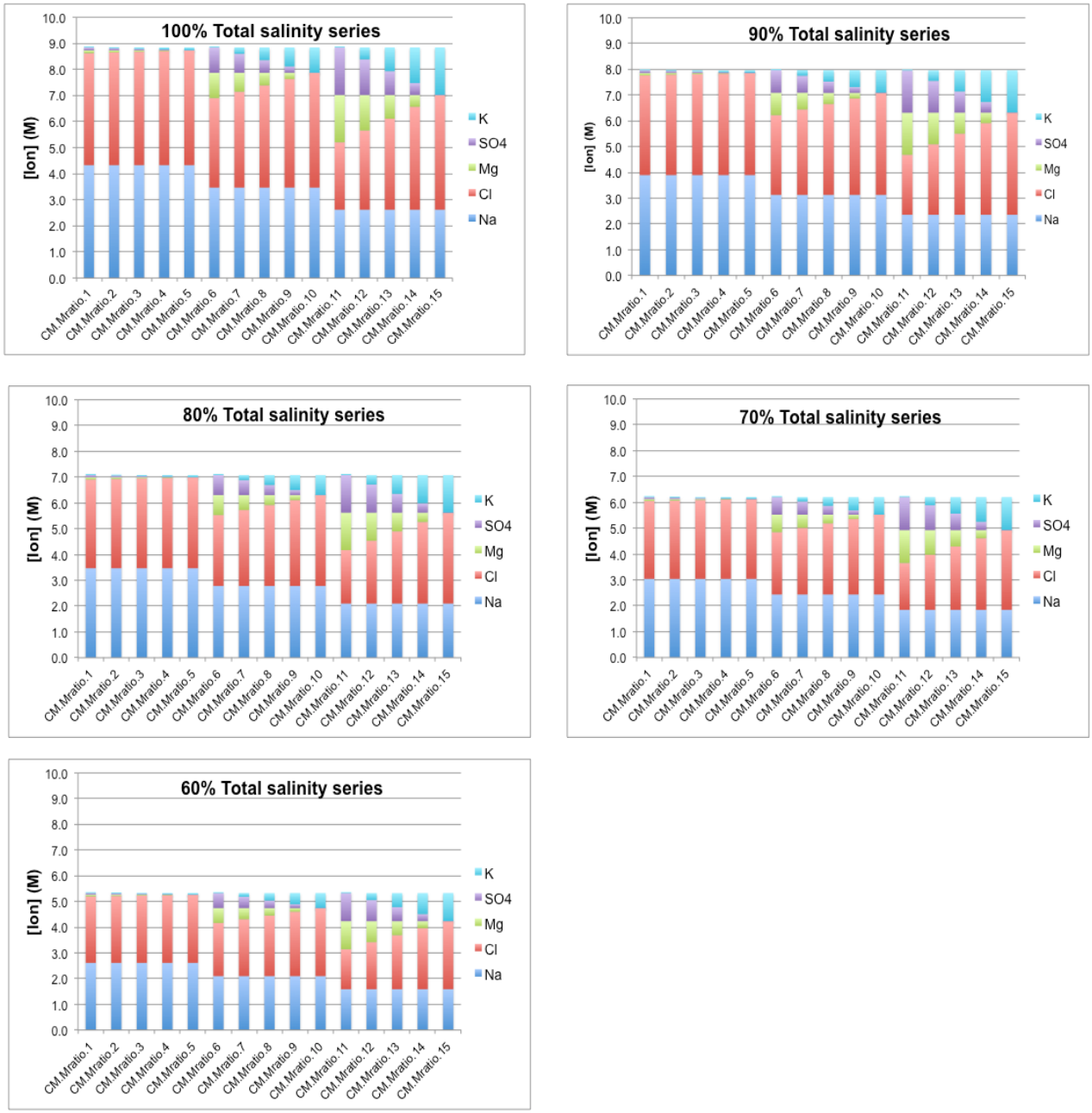

Figure 3-1: Multifactorial media design and composition.
Growth media were designed to encompass all combinations of total salinity, [NaCl] level, and K: Mg ratio. Bars represent the ionic composition of each growth media, not including constant organic components. Compositions are presented in M ions, as converted from the mass of salts added to formulate the media. All media were made with only the salts NaCl, $MgSO_4*7H_2O$, and KCl. Dilution series were generated with v/v dilutions of the 100% total salinity series in a solution of 10 g/L peptone + 3 g/L sodium citrate. Media 2 in the 100% Total salinity series is the control.

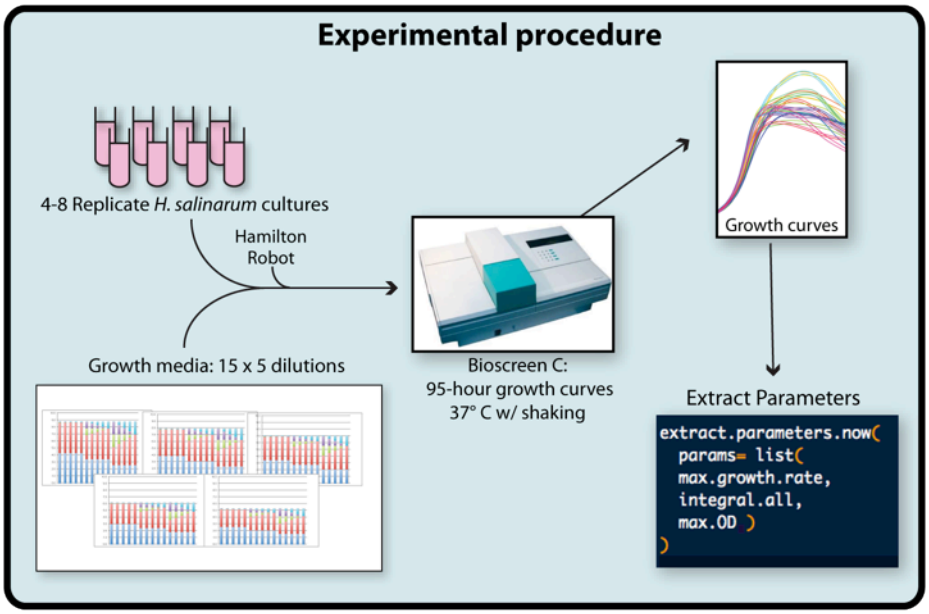

Figure 3-2: Experimental design for salinity reaction norms.
For each of three media series, at least 4 replicate cultures were grown from colony, synchronized by passaging once to OD = 0.05 in control media (Media 2), and then distributed in each media in the series for growth in a Bioscreen C instrument. Optical density was measured every 15 min during growth at 37 °C with shaking for 95 hours. Maximum growth rate was extracted from each growth curve using the GCAF R package developed in the Baliga Lab[129].

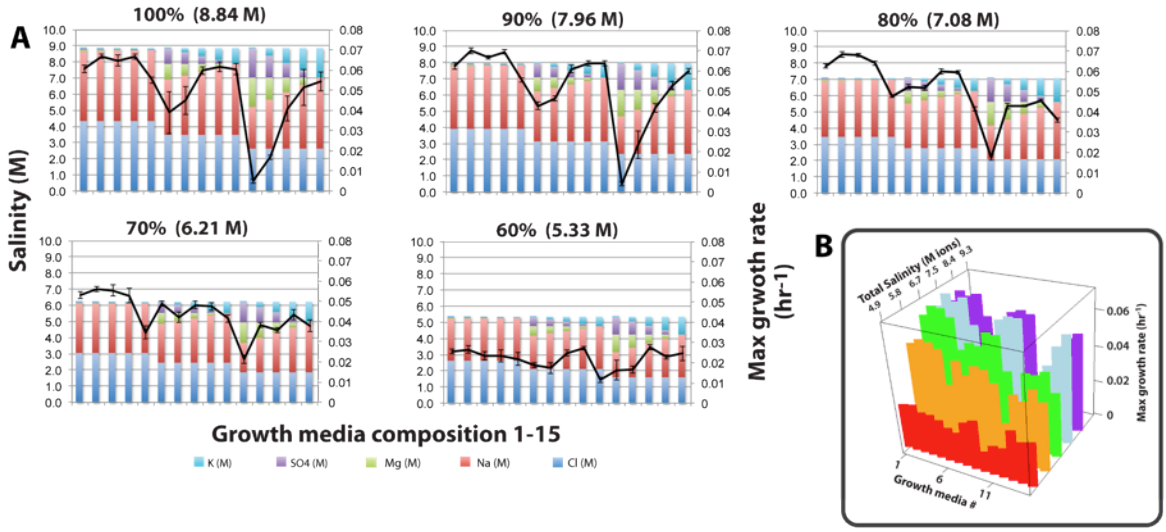

Figure 3-3: Salinity reaction norm plots.
A) Reaction norms plotting maximum growth rate over media composition show how the salinity niche is shaped by three factors (total salinity, [NaCl] and K:Mg). Black lines connect mean max growth rate (μ) from 4-8 replicate experiments in each media. Error bars show standard deviation. **B**) A 3D landscape representation of the same data in **A.** Colors represent total salinity dilution levels (purple: 100%, red: 60%) and bar height is mean μ.

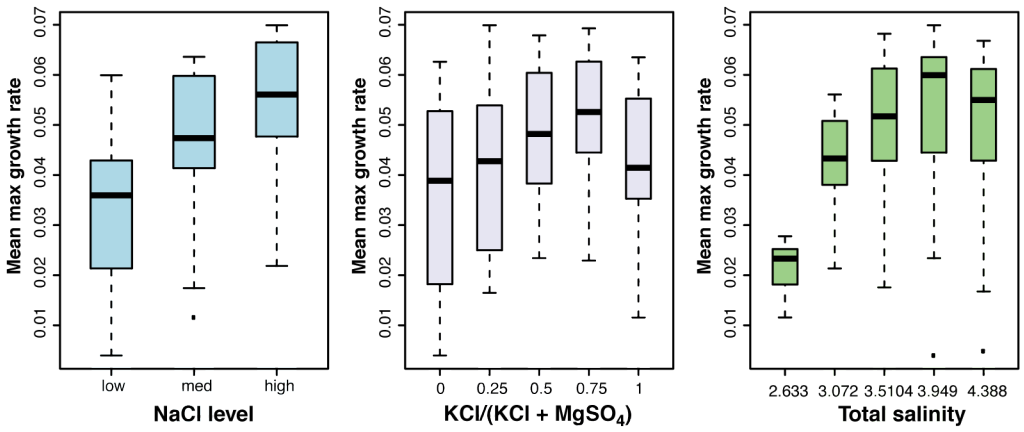

Figure 3-4: Mean max growth rate varies by total salinity, [NaCl] level and K:Mg ratio.
Boxplots indicate median, 1st and 3rd quartiles of mu within each level of each factor. Whiskers extend to 1.5 times the interquartile range, and individual points lie outside of this range. K:Mg is represented as the potassium fraction of total potassium and magnesium salt for ease of plotting.

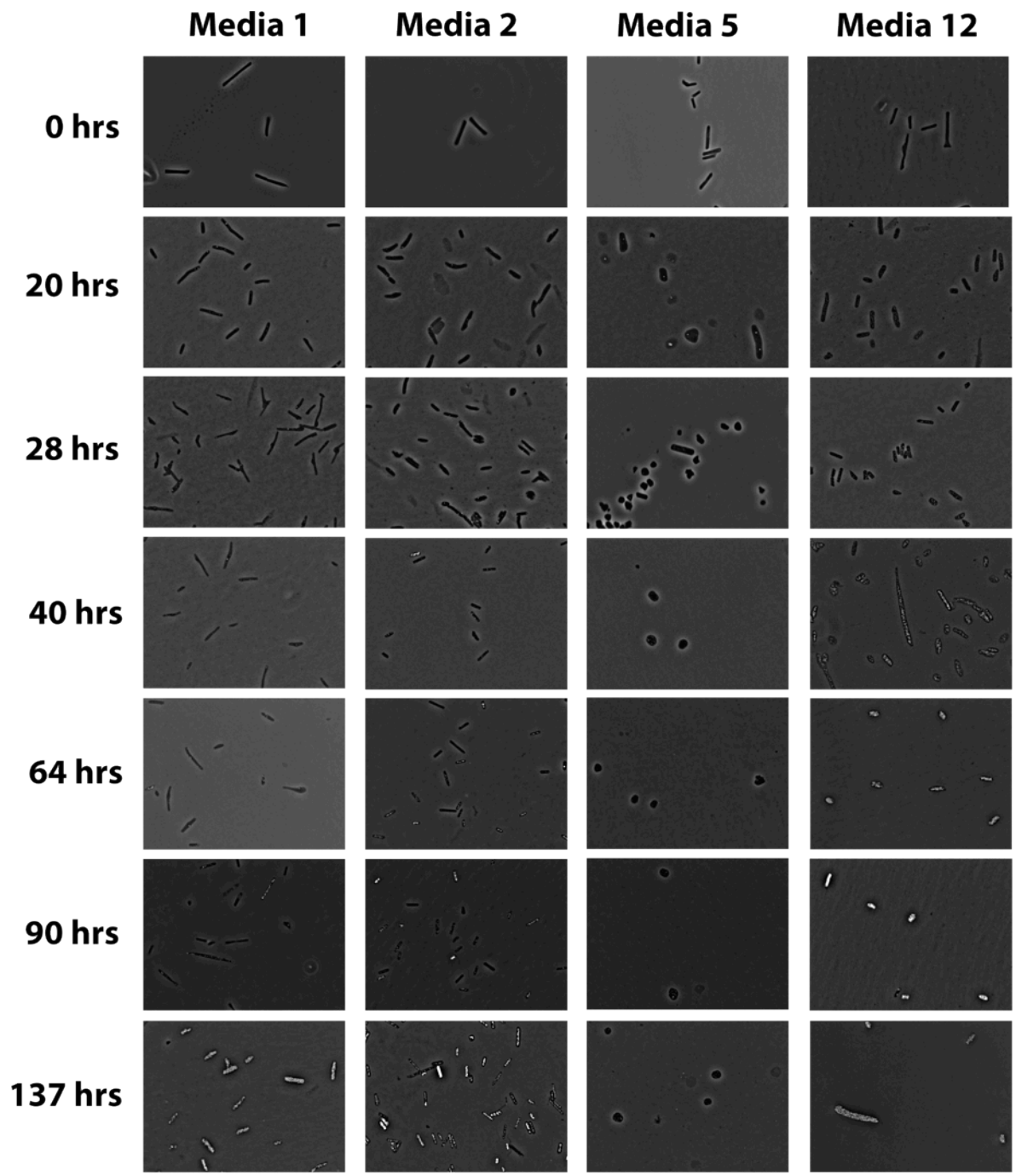

Figure 3-5: Cell and gas vesicle morphology vary with ion composition during batch culturing.
Cells were diluted to OD=0.20 in the respective test media for imaging on gel-coated glass slides. All images were taken using the 100X oil immersion objective. Gas vesicles are the bright, refractile bodies appearing bright white, and may appear as single points or aggregates within cells.

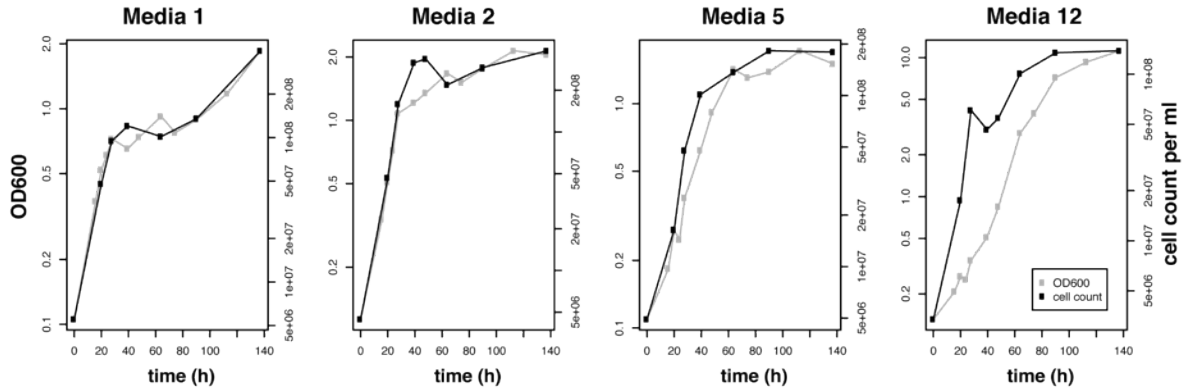

Figure 3-6: OD and cell counts are proportional throughout growth curves in experimental media.
Curves showing timepoints at which OD, cell count and images were taken. OD and cell count were measured at all timepoints, images were taken at 0, 20, 28,40, 64, 90 and 137 hours. Plots show that despite changes in optical properties due to gas vesicle formation at later timepoints, OD and cell count change with similar dynamics throughout the growth curve. In media 12, OD initially lags behind cell count, causing a lag in stationary phase entrance. Reassuringly, cell count follows the OD curve shape even as ODs climb toward 10.0.

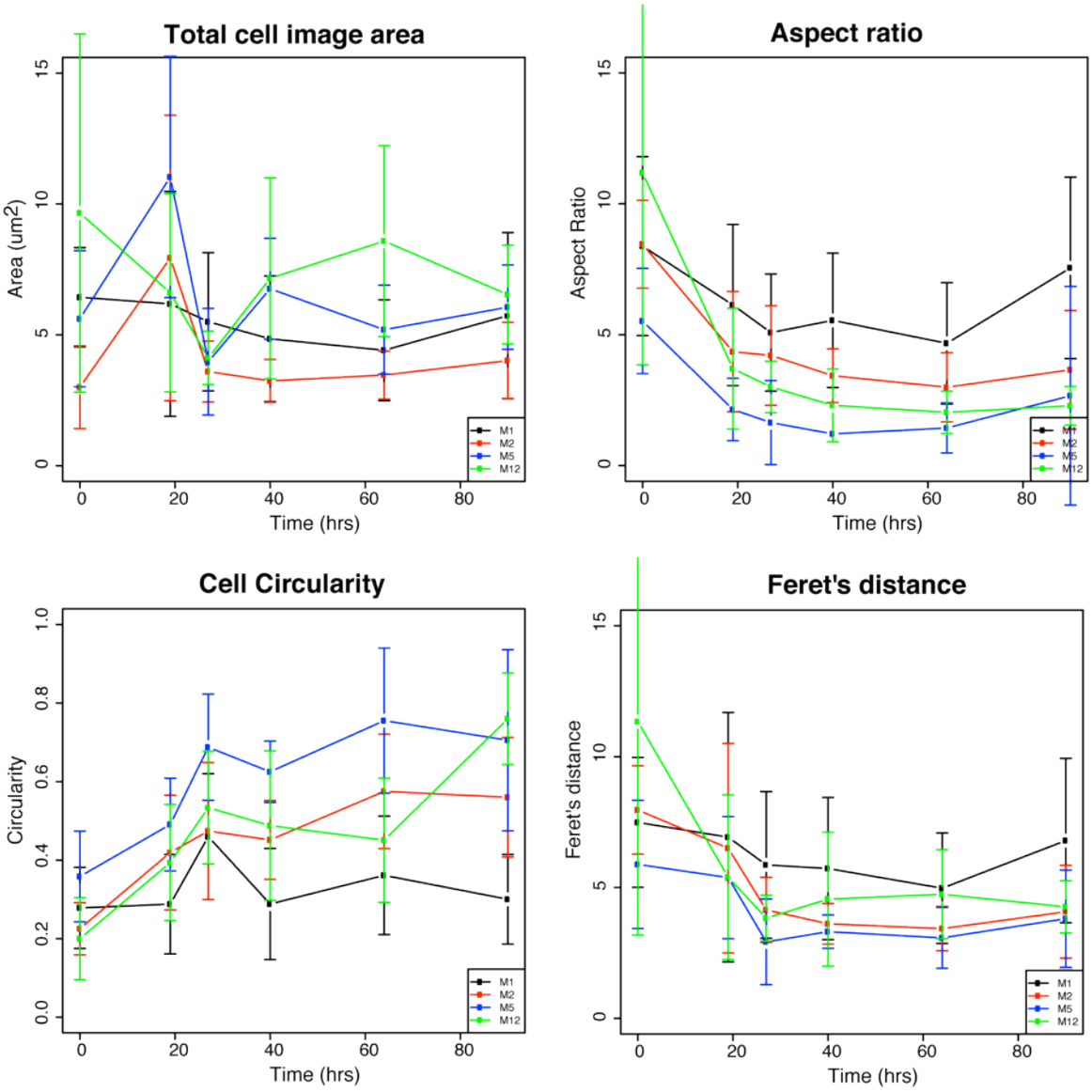

Figure 3-7: Cell morphology varies with ion composition.
Four parameters were used to quantify cell morphology from phase contrast images. Colored lines indicate different media, and error bars represent standard deviation of 5-113 cells. All morphological indices vary significantly with media, by one-way ANOVA (see text). Circularity is calculated using $4\pi(\text{area}/\text{perimeter}^2)$, and a value of 1.0 indicates a perfect circle. Feret's distance is the longest distance between any two points on the 2D cell image. All experiments performed with *H. salinarum* NRC-1 Δ*ura3*. Data in Table 3-2.

Table 3-2: Cell morphology measurements during growth in four different ion compositions

			Cell area (μm²)		Circularity		Feret's distance (μm)		Aspect ratio	
Media	**Time (hrs)**	**# cells**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
CM.Mratio.1	0	7	6.436	1.896	0.278	0.104	7.487	2.482	8.385	3.416
CM.Mratio.1	19	29	6.184	4.294	0.288	0.127	6.923	4.764	6.135	3.081
CM.Mratio.1	27	98	5.500	2.640	0.460	0.160	5.864	2.804	5.081	2.238
CM.Mratio.1	40	37	4.849	2.398	0.288	0.142	5.726	2.714	5.548	2.566
CM.Mratio.1	64	36	4.412	1.927	0.361	0.151	4.973	2.109	4.671	2.318
CM.Mratio.1	90	22	5.734	3.173	0.300	0.114	6.794	3.139	7.555	3.466
CM.Mratio.2	0	5	7.120	1.921	0.225	0.066	7.961	1.685	8.457	1.673
CM.Mratio.2	19	47	7.938	5.453	0.419	0.146	6.504	4.002	4.351	2.299
CM.Mratio.2	27	61	3.599	1.161	0.474	0.174	4.145	1.243	4.207	1.898
CM.Mratio.2	40	26	3.242	0.819	0.451	0.100	3.613	0.776	3.434	1.026
CM.Mratio.2	64	89	3.464	0.916	0.575	0.146	3.428	0.845	2.992	1.318
CM.Mratio.2	90	58	4.015	1.466	0.560	0.152	4.074	1.773	3.659	2.267
CM.Mratio.5	0	13	5.615	2.595	0.358	0.116	5.884	2.448	5.524	2.012
CM.Mratio.5	19	25	11.031	4.609	0.491	0.118	5.380	2.332	2.141	1.194
CM.Mratio.5	27	79	3.972	2.033	0.688	0.135	2.923	1.638	1.642	1.605
CM.Mratio.5	40	9	6.761	1.930	0.625	0.078	3.313	0.639	1.210	0.169
CM.Mratio.5	64	11	5.198	1.700	0.755	0.185	3.077	1.163	1.436	0.957
CM.Mratio.5	90	8	6.058	1.611	0.705	0.231	3.809	1.858	2.668	4.175
CM.Mratio.12	0	10	9.654	6.845	0.200	0.105	11.338	8.156	11.188	7.338
CM.Mratio.12	19	79	6.609	3.783	0.394	0.148	5.395	3.146	3.702	2.306
CM.Mratio.12	27	62	4.116	1.019	0.534	0.143	3.830	0.881	3.005	0.981
CM.Mratio.12	40	113	7.161	3.838	0.488	0.190	4.558	2.558	2.301	1.390
CM.Mratio.12	64	20	8.585	3.647	0.451	0.158	4.742	1.703	2.035	0.809
CM.Mratio.12	90	20	6.540	1.886	0.760	0.117	4.261	1.000	2.284	0.734

Table 3-3: Salinity of GSL samples used for comparison with synthetic media

Sample[1]	Na (M)	Cl (M)	Mg (M)	SO_4 (M)	K (M)	Total salinity (M ions)
GSL2011.7	1.838	2.256	0.167	0.104	0.069	4.435
GSL2011.8	1.973	2.35	0.192	0.112	0.072	4.699
GSL2011.9	2.712	3.3	0.277	0.161	0.108	6.558
GSL2011.10	3.341	4.051	0.35	0.203	0.133	8.078
GSL2011.12	3.54	4.324	0.377	0.22	0.155	8.616
GSL2011.13	4.075	5.049	0.455	0.25	0.187	10.016
GSL2011.14	4.205	5.25	0.46	0.268	0.161	10.344
GSL2011.15	4.082	5.075	0.457	0.262	0.167	10.044
GSL 14 Evap 0	4.321	5.1	0.494	0.269	0.206	10.39
GSL 14 Evap 1	5.472	6.3	0.665	0.354	0.299	13.09
GSL 14 Evap 2	5.271	6.3	0.77	0.427	0.334	13.102
GSL 14 Evap 3	5.223	6.5	0.879	0.478	0.371	13.451
GSL 14 Evap 4	4.337	6.5	1.36	0.75	0.608	13.555
GSL 14 Evap 5	3.898	6.75	1.649	0.907	0.728	13.932
GSL 14 Rehyd 1	0.363	0.513	0.089	0.023	0.036	1.024
GSL 14 Rehyd 2	0.534	0.775	0.182	0.082	0.077	1.65
GSL 14 Rehyd 3	0.769	1.15	0.306	0.157	0.139	2.521
GSL 14 Rehyd 4	2.048	2.475	0.282	0.156	0.123	5.084
GSL 14 Rehyd 5	2.219	2.7	0.25	0.124	0.103	5.396
GSL 14 Rehyd 6	3.318	4.15	0.597	0.321	0.25	8.636

[1] See tables 2-2 and 2-3 for sample descriptions

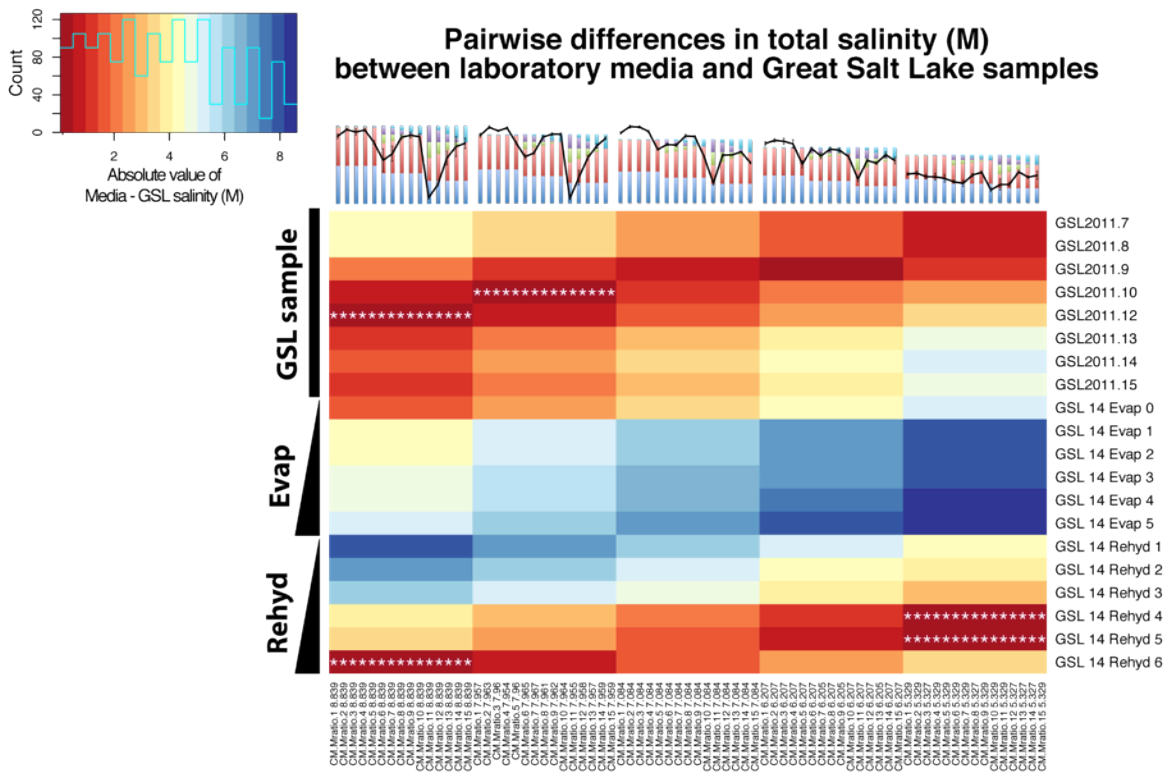

Figure 3-8: Heatmap comparing total salinity in growth media with GSL samples
Step 1, in the nested comparison of multifactorial media compositions with GSL samples and evaporation/rehydration series. Colors represent the absolute value of the difference between each media-GSL pair. Red boxes with white asterisks represent the most similar 5% of pairs in the entire matrix. For each row (GSL site), the single most-similar total salinity media series (set of 15 media) was selected for further comparison of constituent ions. Because the closest media to any of the evaporation series or first three rehydration samples was >1.6 M less saline, this group of samples was omitted from further nested comparisons and ineligible for consideration in the samples most similar to any of the growth media. Labels on bottom indicate ion composition and total salinity level, and correspond with bar plots above. See Figure 3-3 for ion color key. See Figure 2-3 for GSL sample descriptions.

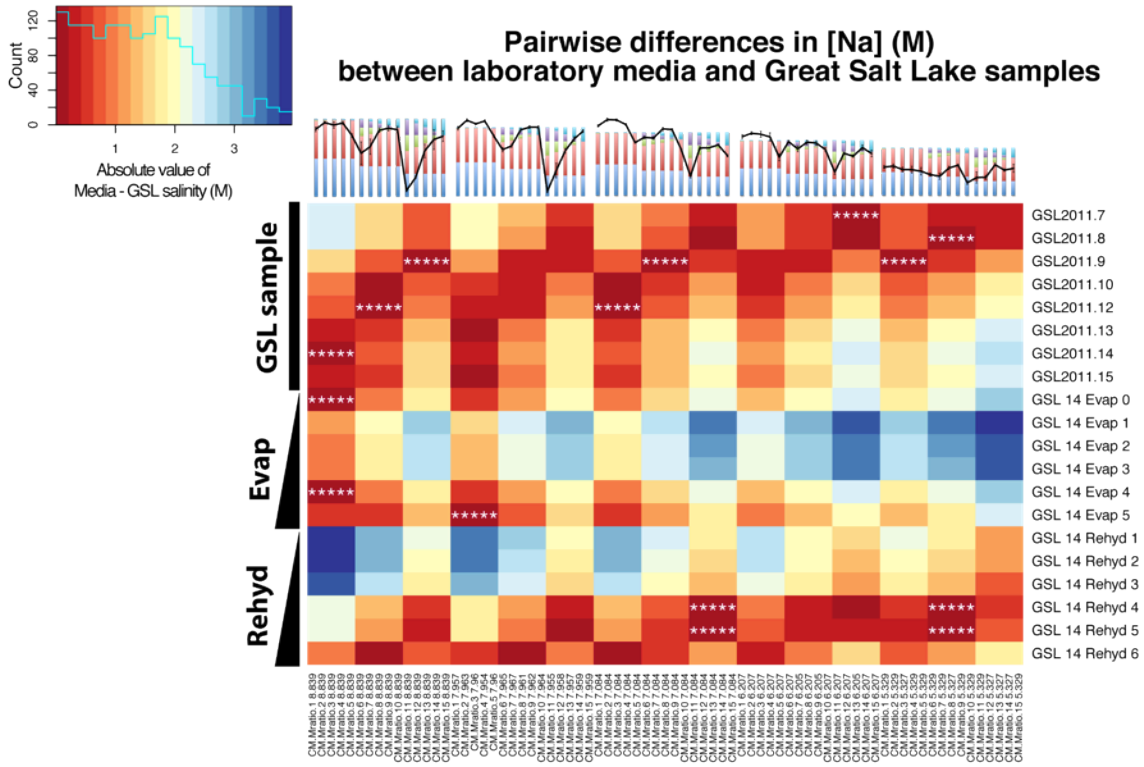

Figure 3-9: Heatmap comparing sodium concentration in growth media with GSL samples
Step 2, in the nested comparison of multifactorial media compositions with GSL samples and evaporation/rehydration series. Colors and asterisks are as in Figure 3-8; color values are described in the top left histogram. For each row (GSL site), within the most similar total salinity media series, the most-similar sodium concentration (set of 5 media) was selected for further comparison of constituent ions. The evaporation series and the first three rehydration were again omitted from further nested comparisons and ineligible for consideration in the samples most similar to any of the growth media. Labels on bottom indicate ion composition and total salinity level, and correspond with bar plots above. See Figure 3-3 for ion color key. See Figure 2-3 for GSL sample descriptions.

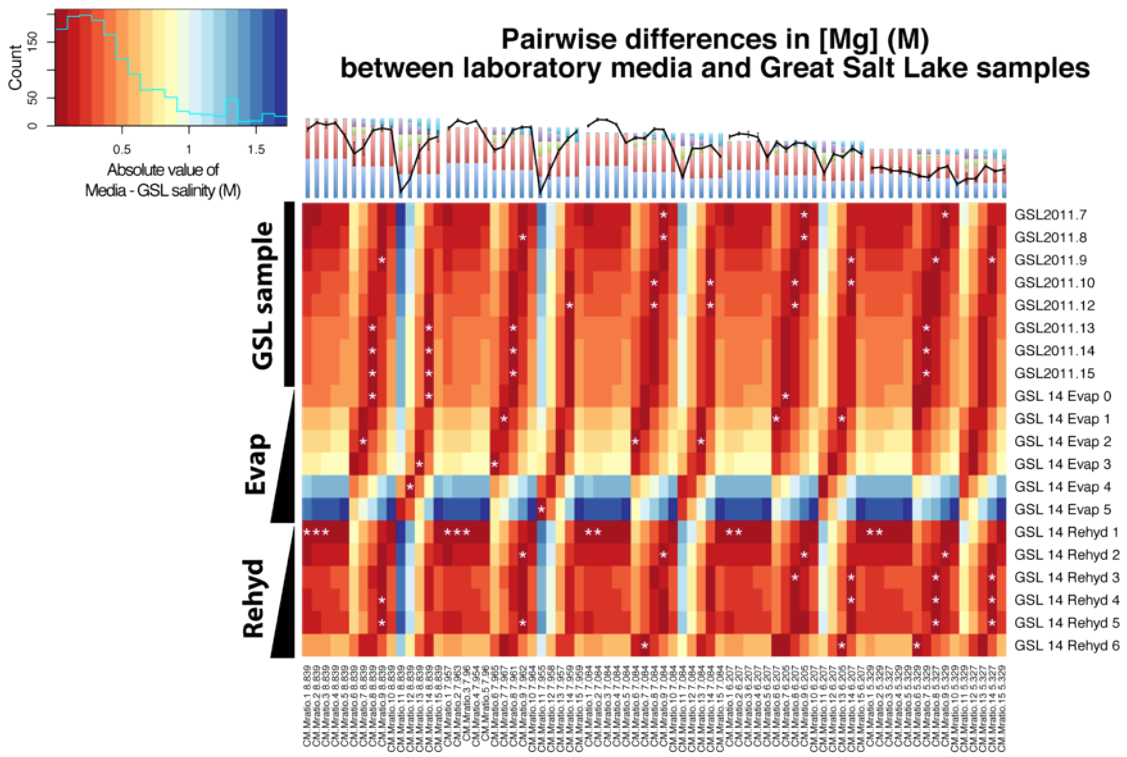

Figure 3-10: Heatmap comparing magnesium concentration in growth media with GSL samples
Step 3, in the nested comparison of multifactorial media compositions with GSL samples. Colors and asterisks are as in Figure 3-8; color values are described in the top left histogram. For each row (GSL site), within the most similar total salinity media series and the nested most similar sodium level, the most-similar single growth medium was selected based on magnesium concentration. The evaporation series and the first three rehydration were again omitted from further nested comparisons and ineligible for consideration in the samples most similar to any of the growth media. Labels on bottom indicate ion composition and total salinity level, and correspond with bar plots above. See Figure 3-3 for ion color key. See Figure 2-3 for GSL sample descriptions.

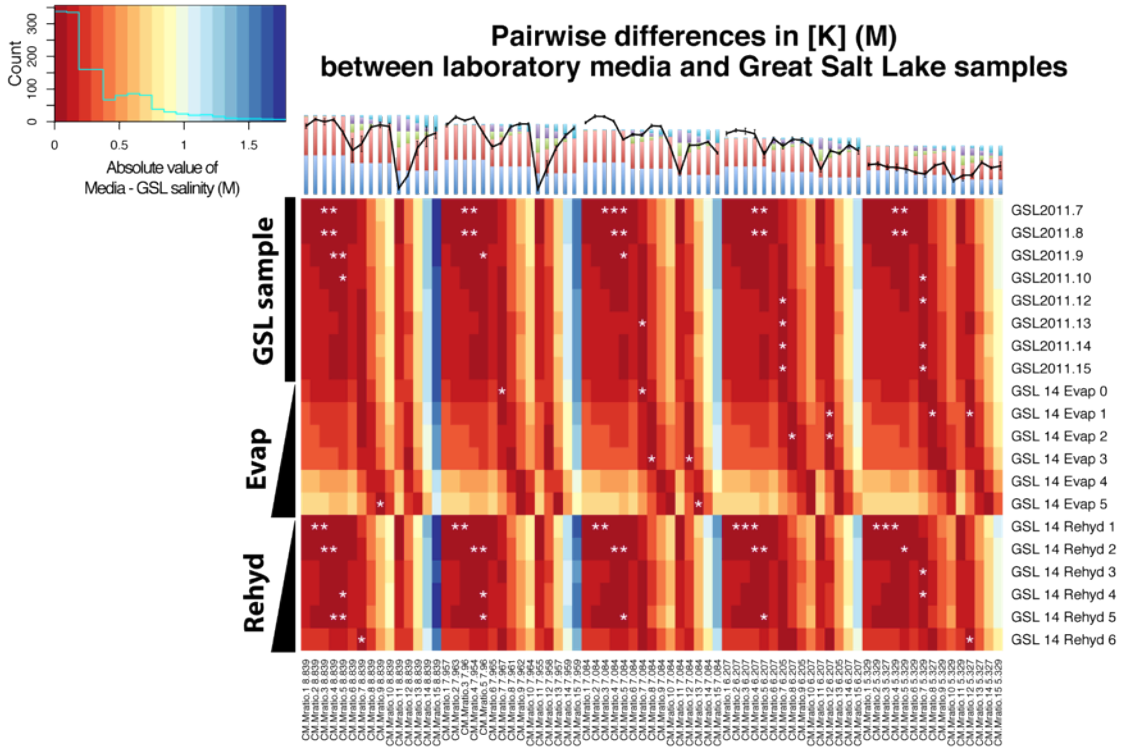

Figure 3-11: Heatmap comparing potassium concentration in growth media with GSL samples
Parallel step 3, in the nested comparison of multifactorial media compositions with GSL samples. Colors and asterisks are as in Figure 3-8; color values are described in the top left histogram. For each row (GSL site), within the most similar total salinity media series and the nested most similar sodium level, the most-similar single growth medium was selected based on potassium concentration. Although potassium and magnesium both comprise the last level of the nested design, potassium is a much less abundant ion in the GSL and therefore the closest-matching media were all in the high sodium series. Few of the closest matches in the two previous steps occurred at high sodium, leaving poorly-matched K levels as remaining choices. Thus, the potassium data are presented, but magnesium concentration determined the best match, because Mg is a more abundant ion in natural hypersaline systems. The evaporation series and the first three rehydration were again omitted from further nested comparisons and ineligible for consideration in the samples most similar to any of the growth media. Labels on bottom indicate ion composition and total salinity level, and correspond with bar plots above. See Figure 3-3 for ion color key. See Figure 2-3 for GSL sample descriptions.

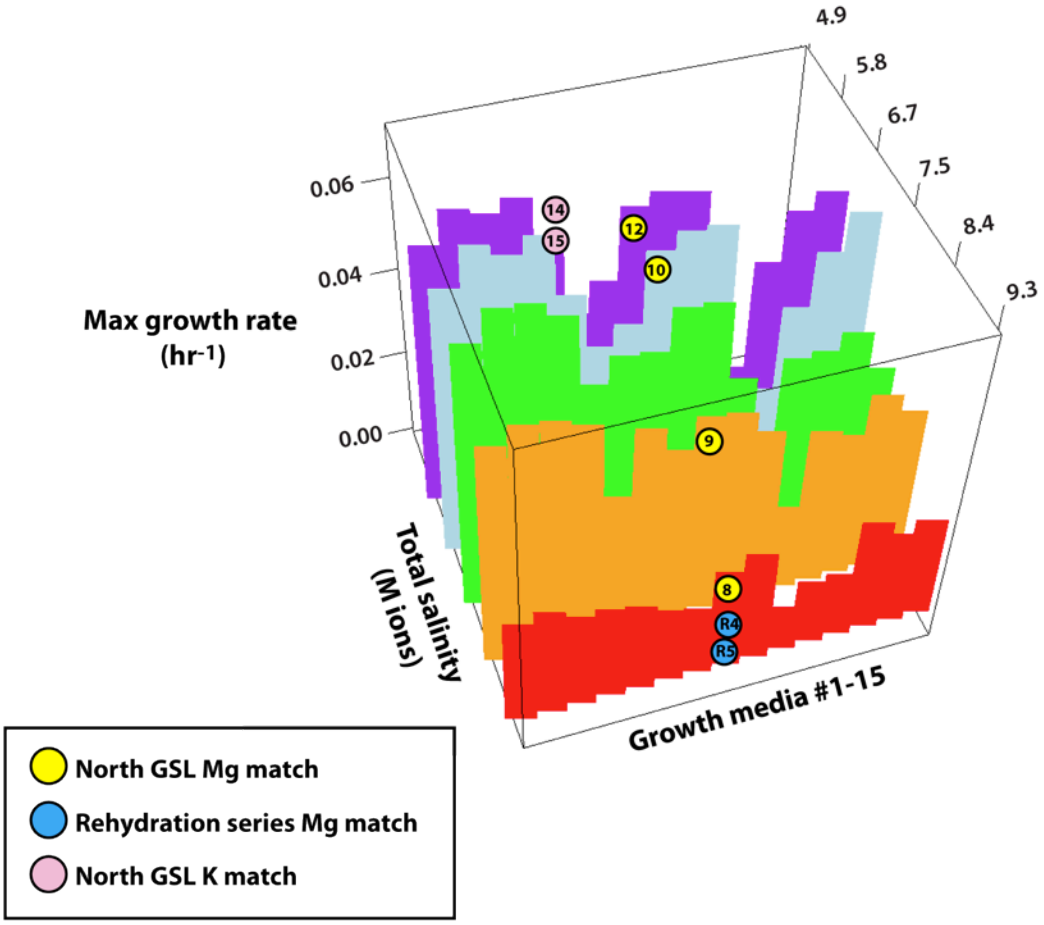

Figure 3-12: GSL sample salinities map to growth media that produce peaks in *H. salinarum* maximum growth rate.
GSL sample salinities (colored dots) are overlaid on the 3D plot of max growth rate in multifactorial growth media. See Figure 3-3 for ion composition of media 1-15. The dots overlaid are GSL samples whose total salinity, [Na], and [Mg] or [K] matched most closely with the underlaid growth media. For example, total salinity in GSL sample 12 was most comparable to the 100% growth media series (purple), and within that series, was most similar to the "medium" [NaCl] level. Within the five media with the same medium [NaCl], the [Mg] of media 12 most closely matched media 8. Heatmap figures (Figures 3-8 – 3-11) indicate the closeness of the matches in each dimension of salinity; data are summarized in Table 3-4. After matching GSL samples with media in total salinity and NaCl, matches for [Mg] (yellow and blue dots) were assessed separately from matches for [K] (pink dot). Yellow dots indicate a [Mg] match between north arm GSL samples and media, and blue dots indicate a match between one of the evaporation or rehydration samples (performed using GSL sample 14) and growth media.

Table 3-4: Summary of ion concentration differences between Great Salt Lake samples and 75 growth media

	Total Salinity		Na			Mg			K		
GSL Sample	Lowest difference level (% total salinity)[1]	Lowest difference level (M)	Lowest difference (% of ions)[2]	Lowest difference (M)[3]	Lowest difference (category)[4]	Lowest difference (% of ions)[2]	Lowest difference (M)[3]	Matching growth media[5]	Lowest difference (% of ions)[2]	Lowest difference (M)[3]	Matching growth media[5]
GSL2011.7	60	5.3	29.80%	1.6	low	2.70%	0.145	9	1.20%	0.11	5
GSL2011.8	60	5.3	39.20%	2.1	med	2.70%	0.145	**9**	1.20%	0.11	5
GSL2011.9	70	6.2	39.20%	2.44	med	5.50%	0.339	**8**	1.20%	0.11	5
GSL2011.10	90	8	39.20%	3.13	med	5.50%	0.436	**8**	1.20%	0.11	5
GSL2011.12	100	8.8	39.20%	3.47	med	5.50%	0.484	**8**	1.20%	0.11	5
GSL2011.13	100	8.8	48.90%	4.33	high	5.10%	0.454	14	2.80%	0.244	7
GSL2011.14	100	8.8	48.90%	4.33	high	5.10%	0.454	14	1.20%	0.11	**5**
GSL2011.15	100	8.8	48.90%	4.33	high	5.10%	0.454	14	1.20%	0.11	**5**
GSL 14 Evap 0	100	8.8	48.90%	4.33	high	5.50%	0.484	8	2.80%	0.244	7
GSL 14 Evap 1	>1.6 M difference in total salinity; omitted										
GSL 14 Evap 2	>1.6 M difference in total salinity; omitted										
GSL 14 Evap 3	>1.6 M difference in total salinity; omitted										
GSL 14 Evap 4	>1.6 M difference in total salinity; omitted										
GSL 14 Evap 5	>1.6 M difference in total salinity; omitted										
GSL 14 Rehyd 1	>1.6 M difference in total salinity; omitted										
GSL 14 Rehyd 2	>1.6 M difference in total salinity; omitted										
GSL 14 Rehyd 3	>1.6 M difference in total salinity; omitted										
GSL 14 Rehyd 4	60	5.3	39.20%	2.1	med	0.055	0.29	**8**	2.80%	0.244	7
GSL 14 Rehyd 5	60	5.3	39.20%	2.1	med	0.055	0.29	**8**	1.20%	0.11	5
GSL 14 Rehyd 6	100	8.8	29.80%	2.62	low	0.055	0.484	8	2.80%	0.244	7

[1] Total salinity series (60-100%) most similar to total salinity of GSL sample
[2] Media ion composition (% of total at corresponding total salinity level) most similar to GSL sample
[3] Media ion composition (M) most similar to GSL sample
[4] Corresponding NaCl category most similar to GSL sample
[5] Single media most similar to GSL sample with respect to [Mg] or [K]; **bold numbers** indicate media matching GSL in all nested dimensions of salinity (total, Na, and Mg or K)

Chapter 4: Roles for transcriptional and post-transcriptional regulation in osmotolerance across ion composition gradients

Abstract

Saline environments are increasing due to industrial waste production, soil salinization, and rising sea levels. Changes in salinity are complex, as they are associated with non-linear fluctuations in both total ionic strength as well as ionic composition. Adaptation to fluctuations in salinity is therefore a systems problem that requires large-scale physiological and regulatory readjustments over short time scales. The objective of this chapter was to characterize the regulatory mechanisms active during salinity adaptation, both at the transcriptional and posttranscriptional levels. Halophilic archaea are ideal organisms in which to conduct these studies because they possess a naturally evolved capability to deal with large and routine fluctuations in salinity. Although gene expression studies have addressed regulation under confounded NaCl/total salinity stress, the effects of ion composition alone have not been explored. Likewise, posttranscriptional regulation of osmotolerance has not been previously investigated in halophilic archaea. Here, we delineate the gene expression signatures of log-phase growth in media that differ in ion composition. We also demonstrate a role for a specific ribonuclease in the posttranscriptional regulation of the Kdp system, a critical operon for maintaining Na/K ion balance across the membrane. This systems-level understanding of complex phenotypes like osmotolerance will allow more proximal, specific analysis of whether and how laboratory organisms have adapted to a new and simpler saline environment compared with their environmental counterparts.

Introduction

The following text and figures contain sections adapted from a manuscript under review: Wurtmann EJ, Ratushny RV, Pan M, **Beer KD**, Aitchison JD, Baliga NS. 2013. RNAse disruption of transcriptional positive autoregulation. Submitted. K. Beer performed experiments on transcriptional regulation and analyzed expression data, while the sections on post-transcriptional regulation are a product of a collaboration with E. Wurtmann.

Laboratory adaptation in natural or clinical isolates is a common phenomenon[134-137], and during evolution experiments, most important mutations are regulatory in nature. In *E. coli*, fitness-enhancing mutations were traced to one of several pleiotropic regulatory genes[138]. Gene regulatory networks (GRNs) coordinate cellular processes at a systems level, and are comprised of interactions between DNA-binding transcription factors (TFs, *e.g. TFBs*) and target genes. GRNs evolve by internalizing environmental factor changes in order to coordinate efficient use of available nutrients, and the activity of TFs in GRNs reflects adaptation in particular environments[18,139]. For example, over half of TFs in bacteria are thought to bind small molecules in order to monitor extracellular chemistry and signals[36]. Thus, differential regulation of genes during diverse environmental changes reveals programs for cell-environmental interactions.

Given the central role of gene regulatory networks as drivers and read-outs of evolution, we expect regulatory patterns during osmotolerance in laboratory model organisms to contain signatures of laboratory evolution, compared with conspecific natural isolates. The first step toward uncovering regulatory signatures of laboratory adaptation is a thorough characterization of salinity response gene regulation in a model organism. In *H. salinarum* NRC-1, a number of genes are known to be differentially regulated in response to low (2.9 M NaCl) and high (5.0 M NaCl) salinity, compared with "optimal" (standard control media, 4.3 M NaCl)[117]. During batch growth in low salt, Coker *et al* found the trkA low-affinity K^+ transporter was upregulated 2.0-fold and the *nhaC3* sodium transporter was 1.3-fold downregulated. Other significantly (2+ fold) differentially regulated genes under low salt in included *dfr* (DinF-related damage-inducible protein, -2.079-fold), *sod2* (superoxide dismutase, 2.023-fold), *hsp1* (small heat shock protein, 2.248-fold), and *cxp* (carboxypeptidase, -2.194-fold).

Under high salt in the Coker study, *H. salinarum* downregulated the *sfuB* ion transporter-like permease and the *phnC* phosphate/phosphonate transport system, suggesting that intake of these species is reduced under osmotic stress[117]. The *nosY* ABC transporter for nitrite/nitrate was upregulated under high salt, along with several oligopeptide/dipeptide/Ni^{2+} transporter genes. Interestingly, compatible solute binding and transporter genes *proX/cosB* and *htr5/cosT* were unchanged. *H. salinarum* is known not to employ a compatible solute strategy to maintain osmotic balance, so these proteins may function differently in this organism. Among transcriptional regulators, one basal transcription factor *(TbpC*) was upregulated 1.7-fold in low-salt, while the *sirR* putative repressor was upregulated in high-salt.

This work was an initial catalog of the gene expression response to two salinity-based perturbations. However, it lacked environmental relevance in its selection of salinity conditions. Addition of NaCl does not reflect a salinity shift encountered in natural systems, and the authors failed to control for changes in total salinity if their main interest was the effect of NaCl on its own. The study also neglected to comment on genes that may play important roles in post-transcriptional regulation (i.e. ribonucleases). There is growing evidence that the relative usage of transcriptional and post-transcriptional regulation is heavily influenced by environmental context.

Post-transcriptional regulation in bacteria, archaea, yeast, and humans appears to be prominent during stress, such as during stationary phase in batch culture and in response to extreme changes in osmotic conditions[140-151]. However, increased regulation at the post-transcriptional level is not a uniform response to all stress conditions. For example, yeast respond to glucose limitation primarily through transcriptional control but respond to nitrogen limitation primarily via post-transcriptional regulation[152]. Similarly, *Bacillus subtilis* adapts to using glucose as a carbon source through a slow, primarily transcriptional response, while the response to malate occurs more rapidly and mostly at the post-transcriptional level[153]. The varied properties associated with transcriptional and post-transcriptional regulation and their non-uniform usage across genes and environments collectively suggest that selective interplay of these mechanisms might be an evolved strategy to generate characteristic response dynamics[154-160]. Yet, there is little known about the architecture of interactions across transcriptional and post-transcriptional regulatory mechanisms, let alone how the topology of these interactions changes with environmental context.

Additional work is necessary in order to delineate the gene regulatory program at work as the model *H. salinarum* NRC-1 responds to environmentally relevant changes in salinity, especially if these data are to provide a baseline for comparison with natural isolates in the future. Given their central location in the hierarchy of genetic information processing, here, we have investigated salinity-response gene regulation at the transcriptional and post-transcriptional levels. Using a systems biology approach that integrates global transcriptome analysis of wildtype and RNase deletion mutants, measurement of ATP levels, and fitness analysis, we have discovered that RNases play a central role in mediating state transitions in specific environmental contexts wherein their action conserves energy and contributes a selective fitness

advantage. We conclude that changes in ion composition (without change in total salinity) invokes a unique gene regulatory response, and that repression of potassium transport by an RNase is an evolutionarily conserved principle of gene regulation that is essential for environment-dependent physiologic state-transitions.

Materials and methods

Strains and culture conditions

The *D2099* strain was created from the *H. salinarum NRC-1 Δura3* parental strain using unmarked, in-frame deletion as previously described[129]. The *Δ2099ΔkdpFABCQ* and the *Δ2099::2099* genomic replacement strain were each created from the *Δ2099* strain with the same technique. The P_{fdx}-2099-CHA plasmid was used to express *VNG2099C* from the ferredoxin gene promoter in the *Δura3* strain. This plasmid was constructed by insertion of the *VNG2099C* gene sequence amplified by forward (5'-GGCCGGCATATGAAGCGCCCGATTGAAACC-3') and reverse (5'-GGCCGGAAGCTTGTCGCGCTGTGCGGCGATGG-3') primers into the pMTF_Pfdx_CHA plasmid digested with NdeI and HindIII. Following transformation, the strain carrying the P_{fdx}-2099-CHA plasmid was cultured with 0.02 mg/ml Mevinolin to select for maintenance of the plasmid.

Strains were cultured at 37°C with continuous shaking (~220 rpm) in standard CM complex media (Media 2: 250 g/l NaCl, 20 g/l $MgSO_4$, 2 g/l KCl, 3 g/l sodium citrate, 10 g/l Oxoid brand bacteriological peptone). This standard CM has a [KCl]:[$MgSO_4$] ratio of 25:75 and was used as the basis for the series of media of differing ion composition as listed in Table 3-1. Media used in this chapter are limited to Media 1-15 in the 100% salinity series. Media were supplemented with 50mg/l uracil to compensate for the uracil deficiency caused by the *Δura3* marker. For K^+-shift experiments, cells were grown to stationary phase in media containing 2 mM KCl before addition of KCl to a final concentration of 100 mM. This media is not listed in Table 3-1 and its composition is as follows: 251.6 g/L NaCl, 20g/L $MgSO_4*7H_2O$, 3g/L Sodium citrate, 10 g/L peptone. pH adjusted to 7.10 with NaOH.

RNA preparation and tiling array analysis

Genome-wide RNA levels were determined for the *Δura3* strain at mid-log phase (OD_{600} = 0.4-0.5), during growth in media 1, 2, 5, 6, 7, 9, 12, 13, 14 and 15 (Table 3-1) in flasks using standard culturing conditions. These 10 media were selected to accommodate array sample size limitations after omitting media where cultures did not reach OD = 0.4 or growth rates were not significantly different from others, by Tukey's honest significant differences. Cells from 2 ml of culture were pelleted (16k x g, 60 s) and flash frozen. Total RNA was isolated using the mirVana RNA kit (Life Technologies) following the manufacturer's instructions and then treated with RNase-free DNase (Promega). Whole-genome tiling arrays were designed using e-Array (Agilent Technologies) for the *H. salinarum NRC-1* main chromosome (NC_002607), and pNRC100(NC_001869) and pNRC200 (NC_002608) plasmids using a 60k feature design of 60-mer strand-specific probes spaced at 24 bp. Arrays included manufacturer's control probes and were printed by Agilent Technologies. RNAs were direct labeled with Alexa547 and Alexa647 dyes (Kreatech), followed by hybridization, washing, slide scanning, spot finding, and normalization as described[129], with dye flip experiments performed for each sample. Raw intensity values were median-normalized and log10 transformed. The matrix of log_{10}-transformed gene expression ratios (media × gene) is included in the appendix.
Statistical analyses were performed in R and Multi-Experiment Viewer (MeV), version 4.8. Gene functional enrichment analysis was performed using DAVID functional enrichment tools[161,162]. For weighted gene coexpression analysis, we used the WGCNA package[163,164] and first limited the gene × media matrix of log10 expression values to the genes with the top 500 coefficients of variation (CV) across all media. The power threshold was set at 10, the soft threshold for scale-free network topology at which all correlations are greater than 0.50. Clustered modules were generated with function blockwiseModules(), and plotted as heatmaps of expression and eigengene expression, a summary measure of expression of all genes in a given cluster.

Transcriptome analysis for comparing Δura3 and VNG2099 strains

Genome-wide RNA levels were determined for the *Δura3* and *Δ2099* strains at four points in the growth curve (Figure 4-7), with strains grown in CM media in flasks using standard culturing conditions. Cells from 2 ml of culture were pelleted (16k x g, 60 s) and flash frozen.

Total RNA was isolated using the mirVana RNA kit (Life Technologies) following the manufacturer's instructions and then treated with RNase-free DNase (Promega). Whole-genome tiling arrays were designed using e-Array (Agilent Technologies) for the *H. salinarum NRC-1* main chromosome (NC_002607) and the pNRC100 (NC_001869) and pNRC200 (NC_002608) plasmids using a 60k feature design of 60-mer strand-specific probes spaced at 24 bp. Arrays included manufacturer's control probes and were printed by Agilent Technologies. RNAs were direct labeled with direct labeled with Alexa547 and Alexa647 dyes (Kreatech), followed by hybridization, washing, slide scanning, spot finding, and normalization as described[129], with dye flip experiments performed for each sample. Data have been deposited in the GEO (GSE45988).

Reverse transcription RT-PCR

Total RNA was isolated from 0.5 OD_{600} units of cells grown to mid-log phase using the Trizol reagent (Life Technologies) according to manufacturer's instructions. RNA pellets were resuspended in H_2O, treated with RNase-free DNase (Promega), and RNA quality and quantity was determined using a Nanodrop spectrophotometer (Thermo Fisher Scientific). Reverse transcription qRT–PCR analyses were performed in 96-well plates with the Power SYBR Green RNA-to-C_t 1-step kit and the Power SYBR Green RT-PCR Mix (Applied Biosystems) in a 7900HT Fast Real-Time PCR instrument (Applied Biosystems) using gene-specific primers for *kdpA*, *kdpQ*, and *VNG2629G*. Fold differences in RNA levels were calculated by the $\Delta\Delta C_t$ method using the glucose kinase gene (*glcK*; *VNG2629G*) as a reference gene. Triplicate biological replicates were used, with each qRT-PCR reaction performed in technical triplicates. Data analysis was performed using the SDS 1.2 software (Applied Biosystems).

ATP measurement

Cultures were grown in standard flask growth conditions to mid-log phase (OD_{600} = 0.5-0.7). Cells (0.3 OD units) were removed, pelleted, washed in basal salt solution (CM without peptone), and lysed by resuspending in 0.3 ml H_2O. ATP was measured using the ATP Bioluminescent Assay Kit (Sigma) by injecting 0.1ml of 1:625 dilution of the bioluminescence assay mix into 0.1 ml of lysate and reading luminescence using a BioTek Synergy H4 Hybrid reader.

Results and discussion

Correlation analysis of gene expression and ion composition

Using whole-genome tiling microarrays, we measured transcript levels for all 2400 genes and 215 noncoding RNAs. Two biological replicate cultures were grown from colony in standard control media 2, before passage into each of 10 media that vary in ion composition but not total salinity, as described in Chapter 3. They represent 10 of the 15 ion composition media in the 100% total salinity series, and were chosen to cover the range of growth phenotypes observed in media 1-15, while omitting redundant media that produced similar phenotypes (Figure 4-1). During batch growth in the experimental media, samples for microarrays were taken at OD_{600} ~ 0.45, to capture mid-log phase and the range where growth rate approaches maximum (Fig 4-2 and 4-3). Total RNA from both samples in each media was compared with a standard RNA reference, extracted from an *H. salinarum* NRC-1 batch culture at OD = 0.471.

We first explored correlations and categorical associations between ion compositions of the ten media and log10-transformed expression of each gene, using Pearson's r correlation tests and one-way ANOVA of expression on media as factor. For a sample size of 10, the absolute value of Pearson's r greater than 0.63 is significant at $p < 0.05$ for a two-sided test. We evaluated correlations between gene expression and each of the following media composition and growth attributes: μ, [Na], [Cl], [Mg], [K], NaCl:Total salt ratio, and K:Mg ratio, with all concentrations in molar. Of 43,290 total correlations (all genes with all media and growth attributes), 5,560 were significant before multiple testing correction. After Benjamini-Hochberg correction, 197 and 1 genes were significantly correlated, at $P < 0.10$ and $P < 0.05$, respectively.

Because changes in growth rate alone affect gene expression patterns in *H. salinarum*[165], we then identified those genes significantly correlated with μ alone. After Benjamini-Hochberg correction for multiple testing, 25 genes were significantly correlated with μ at $p<0.05$. These genes included gas vesicle proteins *gvpA1, gvpC1, gvpA2* and *gvpV1*, bacteriorhodopsin precursor *bop,* an oxidoreductase *inb*, phytoene synthase *crtB1* and a putative thioredoxin VNG0711C. *H. salinarum* is known to modulate production of gas vesicles in order to shift its position in the water column and thus move toward optimal light, oxygen and temperature levels. Gas vesicles are known to vary with growth phase in the lab[166], and are upregulated in media 6

and 12 (low potassium; medium and low sodium, respectively). Gas vesicle genes and others correlated with μ are likely involved in a growth rate-driven stress response, and would not reflect ion-associated regulation. Thus, we identified these genes at the outset and will remove them from analyses of ion and ion ratio correlation.

After removal of μ-correlated genes, 37 genes were significantly correlated with ion composition. Specifically, these 37 genes were correlated at $p < 0.10$ with [Cl] and [Mg] but in opposite directions, as the K:Mg design factor forces anticorrelation between Cl (in the KCl added salt) and Mg (as $MgSO_4$) (Table 4-1). Negative Mg-correlated genes (n = 16) were enriched for porphyrin metabolism, and cobalamin (Vitamin B) biosynthesis and metabolism. KEGG pathways that include these genes are Metabolic Pathways, Porphyrin and chlorophyll metabolism, and ABC transporters (Figure 4-4). Magnesium is required for the synthesis of porphyrin, which when complexed with cobalt or other prosthetic groups, forms cobalamin and its derivatives[167]. Cobalamin is critical for fatty acid and DNA synthesis, and we speculate that reduced extracellular magnesium (and increased Cl) may push these synthesis reactions toward compensatory porphyrin production.

Twenty-one positive Mg-correlated genes were included in the following KEGG pathways: Aminoacyl t-RNA synthesis, Phenylalanine metabolism, Methane metabolism, and Tryptophan metabolism. These genes include iron-sulfur proteins, a peroxidase/catalase, and a transcriptional regulator involved in anaerobic growth on DMSO. Although this experimental design does not allow the deconvolution of Mg, Cl and K gene expression correlations, this correlation pattern suggests that changes in ion composition signal the need for new metabolic states. This hypothesis is supported by the physicochemical relationship between salinity, temperature and oxygen concentration: higher salinity and water activity (as influenced by the ratios of mono- and divalent ions) support lower dissolved oxygen, and this relationship is modified by temperature. In *H. salinarum,* several of the same genes (including *dmsR*, an anaerobic/DMSO growth regulator) are upregulated after quick downshifts in oxygen[168], and allow the cell to quickly change its energy state as oxygen changes. Using these ion-correlated genes, we can establish a set of baseline expression patterns that can be queried in natural isolates.

Salinity effects on ribosomal and global transcription factor gene expression

Next, we took a top-down approach and mined the expression data for patterns in the ribosomal genes and in twelve of the global transcription factors in *H. salinarum.* We first looked at expression of ribosomal subunit genes in order to get an expression-level measurement of the effects of ion composition on active growth (Figure 4-5). As expected for an organism putatively adapted to the control media 2, *rps* and *rpl* genes were most highly expressed in Media 2, followed surprisingly by Media 9. Media 9 contains 3.46 M Na and has a K:Mg ratio of 75:25, and because it contains less NaCl salt, it has contains a larger fraction of K, Mg and SO_4. This media does not represent any synthetic media used in our lab, however in Chapter 3 we reported that Media 8 (which is compositionally similar to Media 9 and growth rate did not differ significantly between the two) is compositionally most comparable to several Great Salt Lake sample compositions. Media in which ribosomal genes were repressed did not correspond perfectly with maximum growth rate (Chapter 3). Media 5 was nearly universally repressive of all *rps* and *rpl* genes, even though media 6 produced the slowest μ values. Similarly, we expected media 12, with its low corresponding growth rates, would have low ribosomal gene expression. These observations call into question the relationship between ribosomal gene expression and growth rate measurements, and the difference may be due to the fact array samples weren't taken at the time of maximum growth rate. Media 12 has highly variable expression among the ribosomal genes and it is possible that this snapshot of expression captured transcripts as their abundance was still decreasing. Nonetheless, the nearly equal ribosomal activity in media 2 and 9 warrant further study of gene expression differences in media that produce the same growth characteristics. What cell processes differ in these environments that still maintain high levels of active growth? These data also argue for a reevaluation of optimal laboratory media.

Next, we looked specifically at the influence of ion composition on the global TFB and TBP regulators (Transcription Factor B and TATA-binding protein, respectively), TfbA, B, C, D, E, F and G, and TbpA, C, D, E, and F. Our group has previously inferred the roles of these top-level regulators across hundreds of environmental conditions[90], and has shown that they interact physically with gene clusters critical for light- and oxygen-induced gene expression[33,168,169]. These global regulators also interact with the promoters of target genes co-regulated during *H. salinarum* response to the oxidative stresses H_2O_2 and paraquat[32]. Differential expression of these global regulators during ion composition stress would be an important first piece in the

puzzle of microbial coordination of osmotolerance together with simultaneously occurring environmental signals.

Indeed, when we look at TFB/TBP expression across the 10 media, several interesting patterns emerge (Figure 4-6). First, *TfbA* is upregulated approximately 10-fold in media 13 compared with media 2. Media 13 has the lowest [NaCl] level and a K:Mg ratio of 50:50, and this combination appears unique even within the set of media with the same low level of [NaCl]. Interestingly, TfbA is also upregulated in media 1, 5 and 7, albeit only 2 to 4-fold. *TfbA* was previously shown to be differentially regulated under transition metal perturbations and is thought to be part of the top level of a stratified regulatory architecture, in which multiple global transcription factors mediate finer control by other transcriptional regulators[170,171]. We also know that *TfbA* is important for fitness in high pH and at 25 °C compared with 37. Taken together, it appears that this global regulator is sensitive to the specific ratios of at least three major ions, in addition to its role in pH and temperature response. The environmental significance of this particular ion ratio, if any, remains an open question.

In addition to *TfbA*, several other global transcription factors show differential expression across the set of ion composition media. These results represent promising new leads into environment-specific gene regulation, especially because ion composition with control for total salinity has not been explored from a systems perspective, to our knowledge. Rather than speculate on the potential connections between TFB and TBP activity in these experiments, and existing data, we plan to follow up with targeted experiments that accomplish two main goals: 1) Establish the salinity EGRIN (Environment and Gene Regulatory Inference Network) and 2) integrate additional salinity experiments into our lab's large database of expression data in more than 1500 conditions. For example, does the gene regulatory network for salinity share topology with the oxygen regulatory network? Because these factors are correlated in the environment, we expect cellular response to be coordinated as well. Salinity response is most interesting in the context of cellular response to physicochemically linked stimuli (*i.e.* light, oxygen and temperature), and this level of understanding requires analysis of this larger data resource. Finally, our data argue for an expansion of osmotolerance research, which is currently limited to studies of total salinity or single-ion changes.

Weighted gene coexpression network analysis of expression data

In order to quantify the genome-wide expression patterns across the ion composition media, we used weighted gene coexpression network analysis (WGCNA) to identify high level patterns among the media composition and for future comparison with natural isolate expression data. WGCNA is a systems biology method for identifying clusters of highly correlated genes, across a set of conditions (*e.g.* salinity media, time courses, case/control)[163,164]. By identifying these clusters of correlated genes, signatures or biomarkers of gene expression lead to additional hypotheses about the mechanistic links between gene regulation and phenotype. As with all correlation-based analyses, additional experimentation will distinguish between co-expression and the elusive conclusion of causal co-regulatory relationships. Here, we took the very first step in this process by identifying clusters of coregulated genes, and generated "fingerprints" of expression for future comparison with natural isolates, using the summary Eigengene measure for each cluster.

After limiting the analysis to the 500 genes with the highest CV values, seven modules (clusters) of correlated genes emerged (Figure 4-7,). To summarize the expression profiles in each module, we also report the eigengene summary metric for each module of correlated genes (Figure 4-8). The eigengene is the first principal component of module gene expression, or in network terms, the most highly connected intramodular hub gene. Eigengenes reveal relationships between modules, and allow the inference of relationships between modules and metadata (*i.e.* ion composition). By reordering the rows and columns in the eigengene heatmap, ion-specific expression patterns became evident. Most notably, downregulated eigenvalues were dominant in the low-[NaCl] media whereas upregulated values were enriched in media 2, 6, 7 and 9 (high and medium-[NaCl] media. Differences within levels of [NaCl] are difficult to distinguish with these data, but overall each media had a unique expression pattern among clusters. Like the correlation and TFB/TBP data above, this analysis provides a systems-level look at expression patterns in 10 distinct chemical environments. These rich data in a laboratory model organism are a platform from which to pursue specific hypotheses about salinity adaptation in the lab and in the Great Salt Lake.

Archaeal RNAse VNG2099 regulates expression of ion transport genes

To investigate the role of post-transcriptional regulation in salinity response, we began by screening deletion mutants of 13 different RNases in *H. salinarum* NRC-1. We discovered a

significant growth defect in one of these mutants, *ΔVNG2099* (*Δ2099* gene henceforth). VNG2099C protein is significantly sequence-similar ($e = 2 \times 10^{-34}$) to the rat liver perchloric acid-soluble protein (L-PSP), a well-characterized endoribonuclease[172]. Key amino acid residues within the active site of L-PSP are also conserved in VNG2099C (data not shown). Interestingly, even the overexpression of VNG2099C resulted in poor growth potentially indicating importance of regulation of its environment-dependent abundance (Figure 4-9a).

We proceeded to identify genes that were dysregulated in the *Δ2099* strain. At four points spanning log and stationary phases of batch culture growth, we harvested total RNA from the *Δura3* (wild-type) and *Δ2099* strains for genome-wide transcriptome analysis (Figure 4-10). Based on the known repressive function of RNases, we predicted that deletion of *VNG2099C* would predominantly result in the upregulation of target genes. Indeed, significance analysis for microarrays (SAM)[173] with an FDR cut-off of 10% revealed that among 24 genes that were differentially regulated upon deletion of *VNG2099C*, the expression of 23 genes was significantly increased (Figure 4-9b). Of the 23 upregulated genes, 5 are associated with monovalent inorganic cation transport (GO:0015672; $P = 0.001$, Benjamini-corrected modified Fisher Exact test). Notably, included in this set are genes from the polycistronic transcript that encodes the positive autoregulator KdpQ and the multi-subunit Kdp potassium (K^+) transport channel (Figure 4-9c)[111]. Using quantitative reverse transcriptase RT-PCR with primers targeting two key genes of this operon, we confirmed that the *kdp* transcript was upregulated 5-fold upon deletion of *VNG2099C* (Figure 4-9d). Finally, we also confirmed that wild type *kdp* transcript level was restored upon complementation with functional *VNG2099C* (i.e. in the *Δ2099::VNG2099* strain) to unequivocally demonstrate that *VNG2099C* mediates repression of the K^+ transport operon.

VNG2099 RNAse is necessary for salinity adaptation in *H. salinarum* NRC-1

Life in a hypersaline environment is very expensive – it requires ATP-consuming production of osmolytes such as proline or the *H. salinarum* strategy of maintaining a high K^+ cytoplasmic content (~3M) to counterbalance the high environmental concentration of Na^+. The salt-in strategy is bioenergetically advantageous if it is managed efficiently[83]. When the K^+ level in the environment is in low mM quantities, *H. salinarum* expends energy to pump K^+ into the cytoplasm using the high-affinity ATP-driven Kdp K^+ transport system. When external K^+

concentration is higher, the cell can use a low-affinity transport system that does not consume as much energy and Kdp is downregulated[111]. External salinity can change abruptly in the natural habitats of *H. salinarum* (hypersaline lakes, oceans, and brines) due to weather events in winter or evaporation in shallow water during summer. Switching on and off the high-affinity transport system is therefore essential for favorable bioenergetics of the salt-in strategy.

We then investigated whether VNG2099C might act to modulate K^+ transport and thereby aid in adaptation to changes in extracellular K^+ levels. Indeed, an increased ratio of K^+:Mg^+ exacerbated the growth defect of the *Δ2099* strain (Figure 4-9e). Genomic replacement of the deleted locus with a functional copy of *VNG2099C* (*Δ2099::VNG2099*) complemented this phenotypic defect in all media conditions and confirmed that VNG2099C is responsible for the salinity-responsive growth defect. Further, as predicted by the model, VNG2099C was required for environment-responsive repression of the *kdp* transcript pursuant to an increase in extracellular K^+ (Figure 4-9f). Interestingly, the transcript level of *VNG2099C* increased transiently in response to the external upshift in K^+ (Figure 4-9g).

Salinity response is regulated in concert with oxygen and growth state transitions

The anti-correlated transcript level changes in *VNG2099C* and its targets (Figure 4-6f and g) suggested that we could use this as a strategy for discovering additional environmental contexts in which VNG2099C coordinately represses target genes. Upon performing this analysis on a compendium of 1,495 transcriptomes, we discovered that *VNG2099C* and *kdp* gene expression are significantly anti-correlated across several environmental transitions associated with a switch from aerobic to anaerobic physiology (e.g. dark>light and log>stationary phase; Figure 4-11). The environmental context for VNG2099C activity inferred from this analysis is consistent with the observation that *VNG2099C* deletion resulted in increased mRNA levels of genes associated with two anaerobic energy transduction modules - the ATP-producing light-driven proton-pump bacteriorhodopsin (*bop*), and an ornithine-arginine antitporter involved in anaerobic arginine fermentation (*yhdG*, Figure 4-6b; Figure 4-8). Moreover, analysis of the global gene regulatory network of *H. salinarum*[90] revealed that *VNG2099C* is co-regulated with aerobic TCA cycle and oxidative phosphorylation genes. This observation suggested that VNG2099C degrades *kdp, bop*, and *yhdG* mRNAs in oxic conditions, when membrane potential generated by the respiration-driven electron transport chain is sufficient for K^+/H^+ symport,

rendering anaerobic energy production and ATP-consuming K^+ uptake unnecessary (Figure 4-12)[83,169]. This is interesting because even though these pathways are under the control of independent transcriptional regulatory programs, the action of VNG2099C accomplishes their coordinated post-transcriptional repression when oxygen tension rises and their functions become simultaneously unnecessary. This scheme also allows other transcriptionally co-regulated genes to be unaffected across environmental conditions; e.g. carotenoid biosynthesis genes are co-regulated with *bop* but are not targeted by the RNase, allowing the pathway to continue to feed intermediates and co-factors to a wide array of cellular processes that operate across both high and low oxygen conditions.

Thus, the slow growth phenotype of the *Δ2099* strain must result from a failure to repress the ATP-dependent Kdp transporter in high oxygen, leading to wasteful hydrolysis of ATP (Figure 4-9). Indeed, intracellular ATP levels were 35% lower in the *Δ2099* strain than in the wild-type *Δura3* strain (Figure 4-13a; $P = 0.004$). This effect could be entirely attributed to *VNG2099C*-repression of the *kdp* operon, as ATP levels were fully restored upon reintroducing *VNG2099C* into the *Δ2099* background ($P = 0.004$ versus the *Δ2099* strain) or upon deletion of the *kdp* operon from the *Δ2099* background (Figure 5A; $P = 0.001$ versus *Δ2099*), both of which also rescued the slow growth phenotype (Figure 4-13b). Thus, these results together demonstrated that VNG2099C-mediated coordination of energy metabolism (oxidative phosphorylation, phototrophy, and arginine fermentation) with on/off switching of energy-dependent K^+ transport is central to maintaining favorable bioenergetics for salinity adaptation using the salt-in strategy.

Discussion

Transcription factors and RNases are known to act in a certain class of network motifs to generate response dynamics that are well matched to certain types of environmental change[107,154]. Here, we have demonstrated specific RNase regulation of an ATP-driven potassium pump, and have shown that environment-specific regulation by this RNase mirrors the physicochemical links among environmental factors. The efficacy with which the VNG2099 RNase is able to turn 'off' positively autoregulated genes such as *Kdp* renders it especially advantageous for the regulation of genes whose unnecessary expression is bioenergetically expensive. Indeed, deleting *VNG2099C* results in untimely expression of the positively

autoregulated ATP-consuming *kdp* operon, unnecessarily draining ATP and ultimately decelerating growth (Figure 4-13a and b). Furthermore, regulation of an RNase with genes of one environmental state while it targets transcripts of an opposing state is another generalized principle uncovered in this work that explains how biological systems achieve efficient environment-dependent state transitions. In *H. salinarum*, by targeting the positively autoregulated *kdp* operon and anaerobic energy-transduction pathways, the transcriptional co-regulation of *VNG2099C* with key oxidative phosphorylation genes accomplishes energy-efficient anaerobic-to-aerobic state transitions. With this systems-level understanding of regulation of a major potassium pump, detailed phylogenetic studies will shed insight into the evolution of osmotolerance.

Chapter 4 Figures and Tables

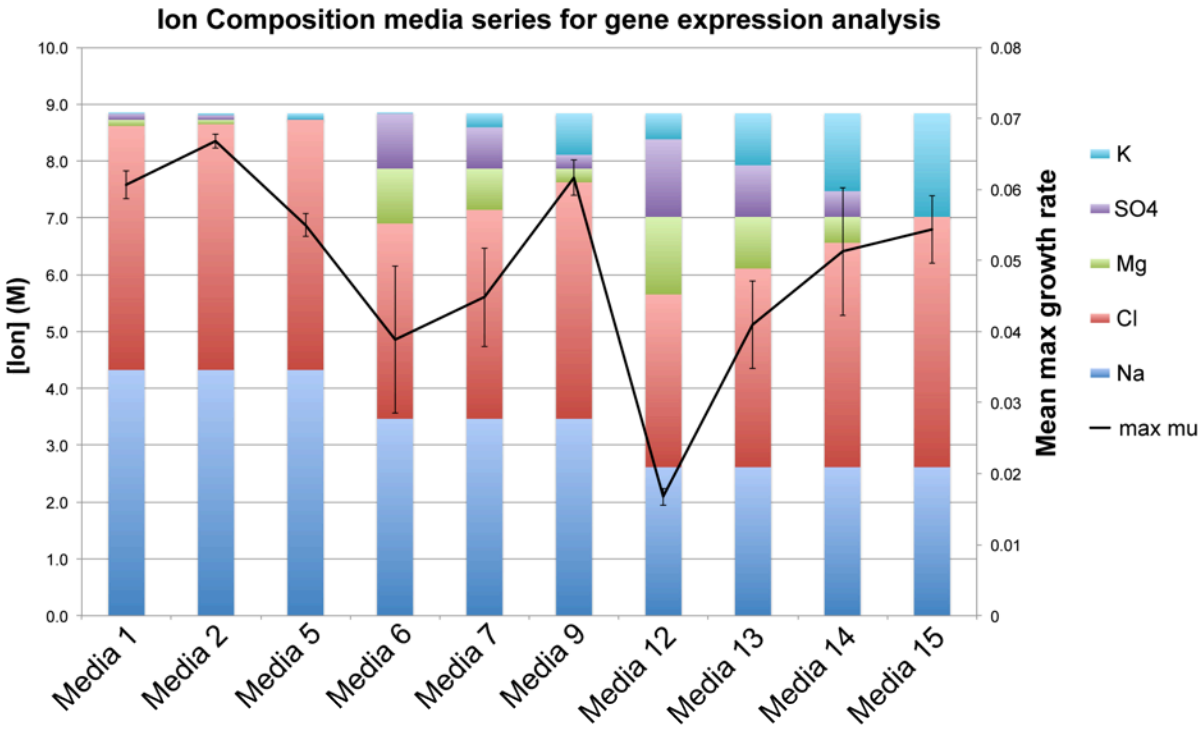

Figure 4-1: Media compositions for expression analysis.
These 10 media vary in ion composition but total salinity is held constant at 8.839 M total ions. These 10 media were selected from the group of 15 media that share this total salinity based on capturing a range of [NaCl] and K:Mg levels, and based on growth rate phenotype. If two media within the same [NaCl] level produced similar values of mu (by Tukey's honest significant differences test, $p > 0.05$), one was omitted. Media 8, 9 and 10 produced similar growth rates, and media 9 was selected because it fell in the center of this optimal range of K:Mg ratios. Maximum growth rate (mu) values are from experiments described in chapter 3, and represent 4-8 biological replicates.

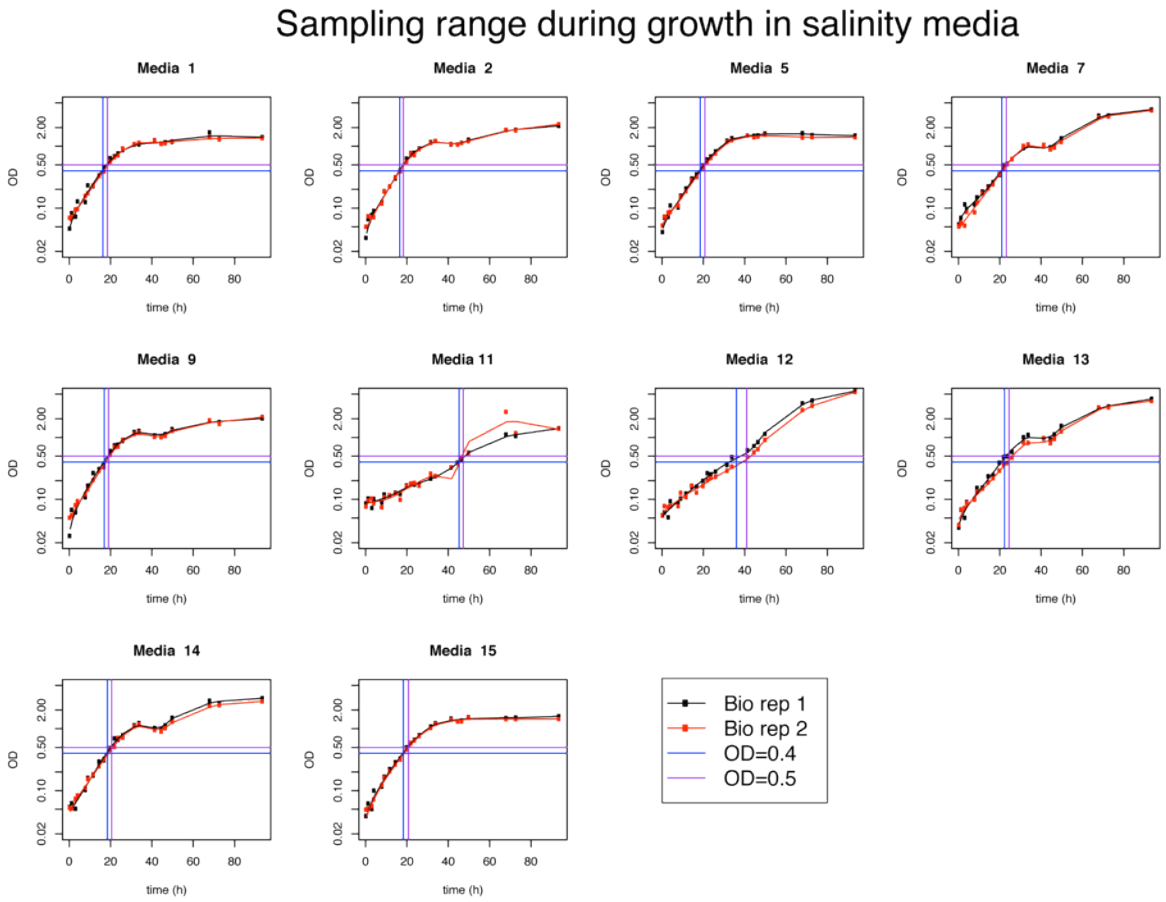

Figure 4-2: Growth curves with planned array sampling range.
Growth curves for 2 biological replicates, collected during an array pilot experiment to determine approximate sample times at which OD 0.4-0.5 occurs. From these data, we opted to sample after ~3 doublings, instead of attempting to sample at the moment of maximum growth rate. Sampling after 3 doublings ensures the seed population has been sufficiently diluted and mitigates growth-phase effects on gene expression as much as possible. In the actual sampling experiment, media 11 was omitted in favor of media 6 due to the slow and erratic growth in media 11, and the proportional levels of K^+ and Mg^+ in these two media.

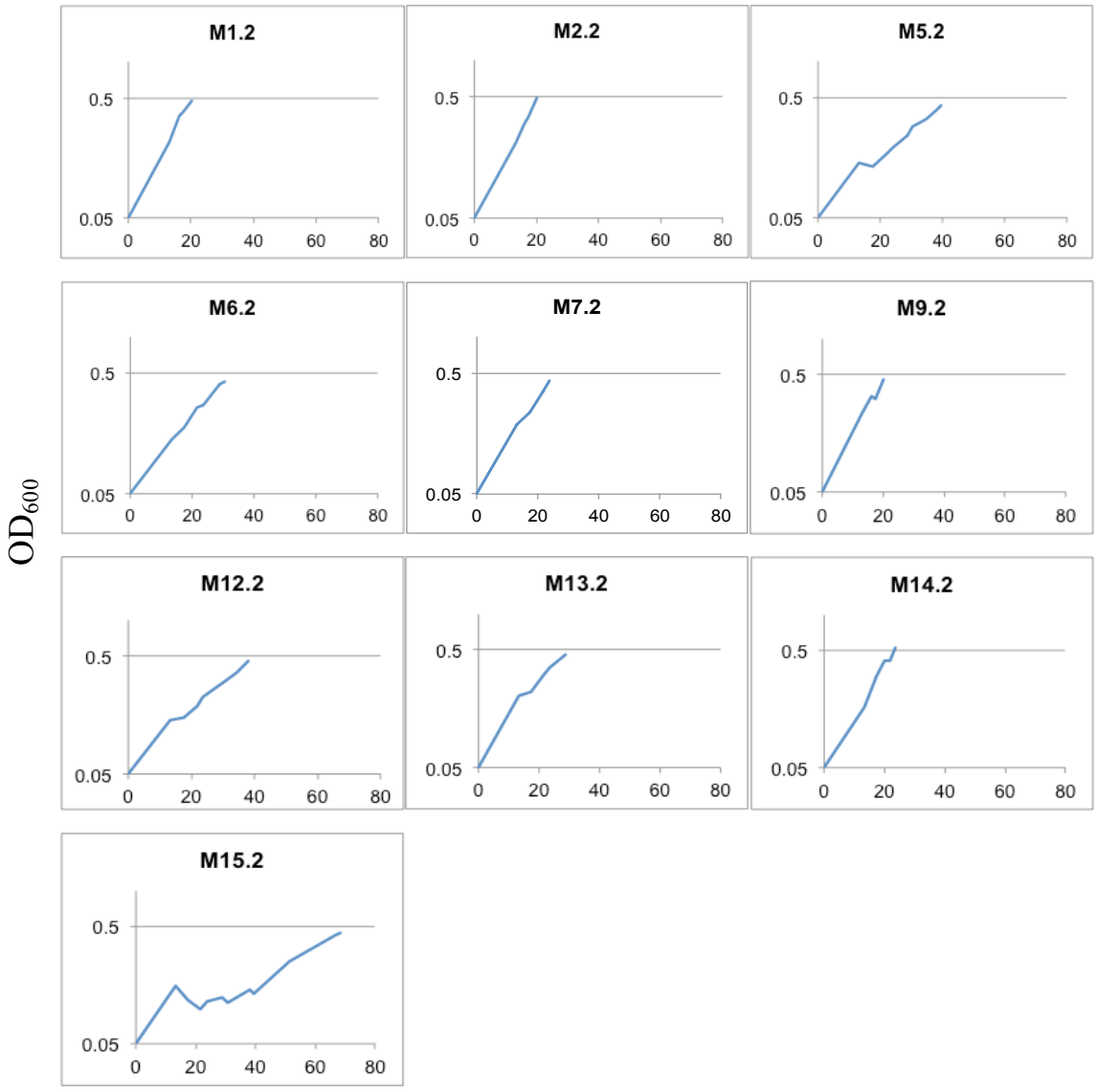

Figure 4-3: Microarray sampling curves
Cultures were grown in 50 mL flasks at 37 °C with shaking, and were passaged through the respective media once before passage to the experimental flask and media, to dilute residual control media 2 and acclimate cells to the new saline environment. Technical quadruplicate samples were taken at the point where each curve terminates. X axes are time in minutes, Y axes are OD_{600} units on a log scale. These plots present one of two biological replicates and plot titles indicate media number followed by replicate 2. (M1.2 is media 1, replicate 2).

Table 4-1: Genes significantly correlated with [Mg] at P < 0.10[1]

Positive			Negative		
ORF	Gene	Annotation	ORF	Gene	Annotation
VNG0044H	VNG0044H		VNG1238C	VNG1238C	Putative ornithine-arginine antiporter
VNG0213H	VNG0213H	transposase	VNG1240G	yhdG	Putative precorrin 8-w decarboxylase (AdoMet methyltransferase).\nInvolved in synthesis of Cobalamin precursors
VNG0826C	dmsR	putative transcription regulator involved in anaerobic growth on DMSO and/or TMAO.	VNG1550G	cbiT	Precorrin-2 C20 methyltransferase\nsynthesis of Cobalamin precursors
VNG1182H	VNG1182H		VNG1551G	cbiL	Precorrin 4-methyltransferase\nCobalamin biosynthesis; CbiF involved in synthesis of precursors
VNG1184Gm	nirJ	predicted Fe-S cluster protein involved in cofactor biosynthesis.	VNG1553G	cbiF	Cobalamin biosynthesis
VNG1486H	VNG1486H		VNG1554G	cbiG	precorrin-3B C17-methyltransferase\nCobalamin biosynthesis
VNG1621H	VNG1621H		VNG1555G	cbiH1	precorrin-3 C-17 methyltransferase\nCobalamin biosynthesis
VNG1664H	VNG1664H		VNG1557G	cbiH2	putative ferredoxin
VNG2302G	yuxL	Acylaminoacyl-peptidase	VNG1558H	VNG1558H	
VNG6085H	VNG6085H		VNG1559H	VNG1559H	
VNG6221H	VNG6221H		VNG1562H	VNG1562H	
VNG6244G	gvpN2	Gas vesicle protein gvpN 2	VNG1564H	VNG1564H	Cobalt transport ATP-binding protein
VNG6251H	VNG6251H		VNG1631G	cbiO2	Cobalt transport protein cbiN
VNG6294G	perA	Peroxidase/catalase	VNG1634G	cbiN	ABC-type cobalt transport system, permease component.\n
VNG6429H	VNG6429H		VNG1635G	cbiM	
NC_001869_20		Non-coding RNA	VNG2656H	VNG2656H	
NC_002607_225		Non-coding RNA			
NC_002607_40		Non-coding RNA			
NC_002608_82		Non-coding RNA			

[1] Significant after removal of mu-correlated genes and Benjamini-Hochberg correction. Genes are anticorrelated with Cl due to experimental design.

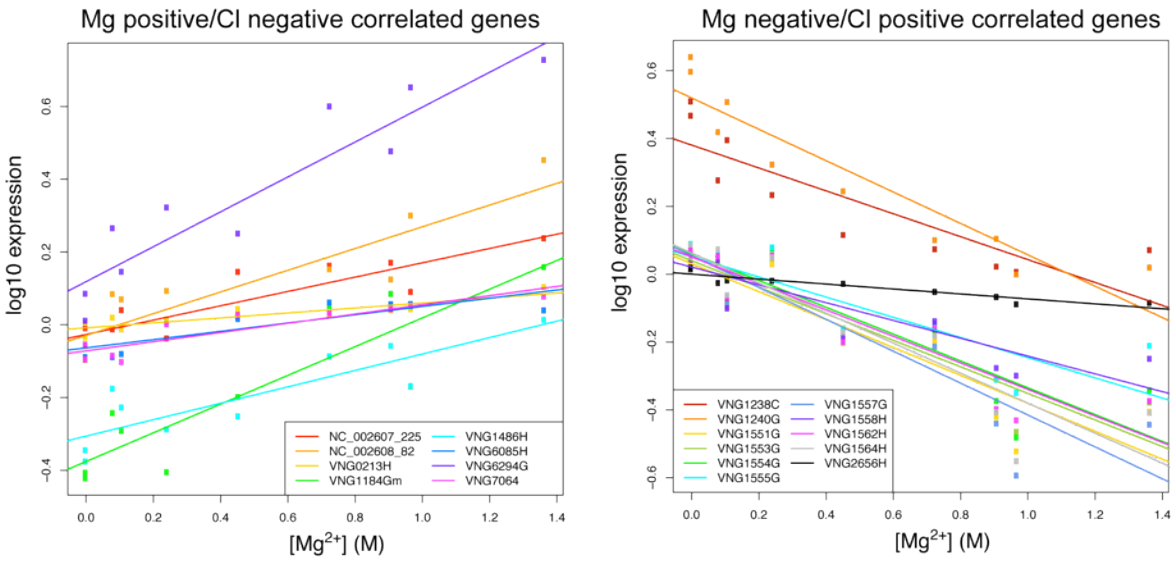

Figure 4-4: Scatterplots of Mg and Cl correlated genes
Plots showing significantly ion-correlated gene expression ($p < 0.05$). Y axes are log10 transformed, normalized ratios of salinity test sample to reference sample. The multifactorial media design necessitates that [Mg] and [Cl] are anticorrelated (Pearson's r = -1.0) among the media. Thus genes positively correlated with Mg are negatively correlated with Cl, and vice versa.

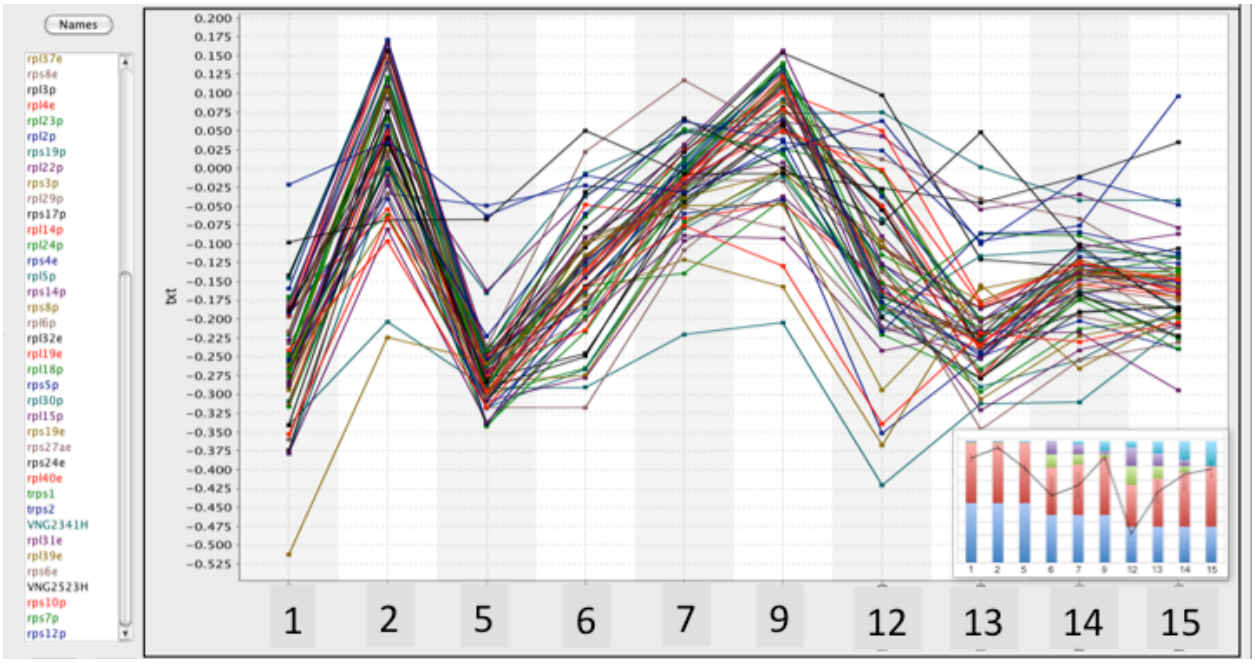

Figure 4-5: Ribosomal gene expression across ion composition media
Coordinated expression of sixty-one *rps* and *rpl* genes encoding ribosomal subunits are shown. tRNA genes were also included to capture the entirety of basal translational response. Y axis shows log10 expression ratios of normalized test to reference spot intensity. Inset shows media compositions for reference; blue: Na; red: Cl; Green: Mg; Purple: SO^4; Turquoise: K.

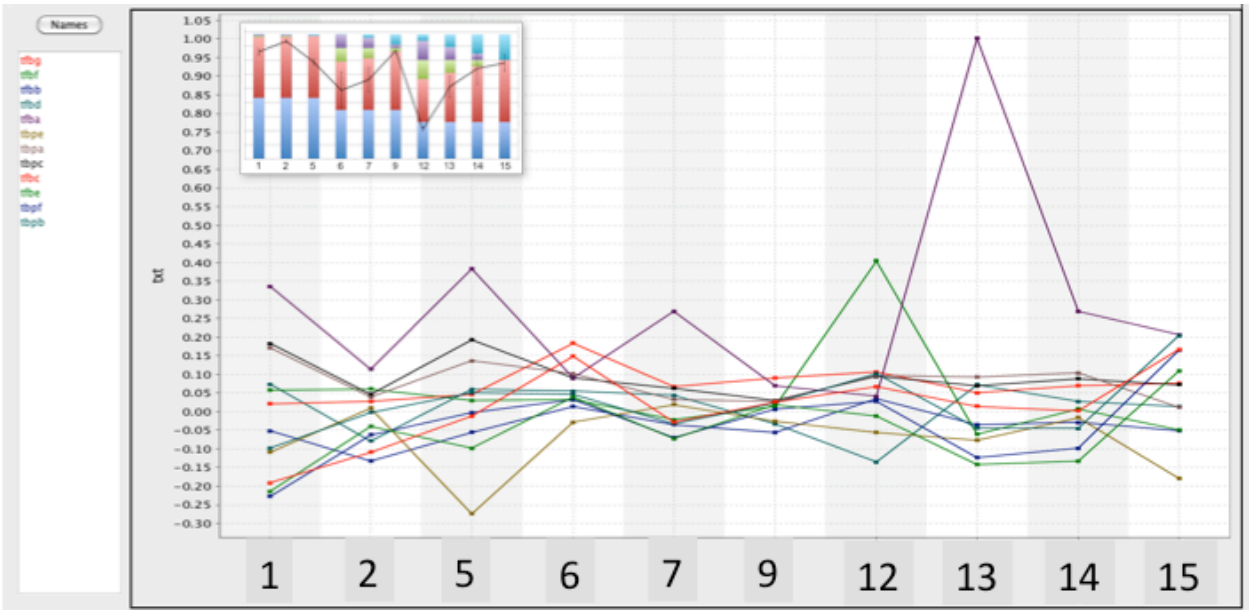

Figure 4-6: General transcription factor expression across ion composition media
Differential expression in 12 general transcription factors (TFBs and TBPs) in new ion composition environments suggest that regulation of osmotolerance is tightly woven into the regulatory network, despite a relatively small number of ion-correlated genes (166). Media numbers are shown at the bottom, and the Y axis is log10 expression ratios of normalized test to reference spot intensity. *TfbA* (plum) shows a 10-fold increase in media 13, compared with control media 2; see text for discussion.

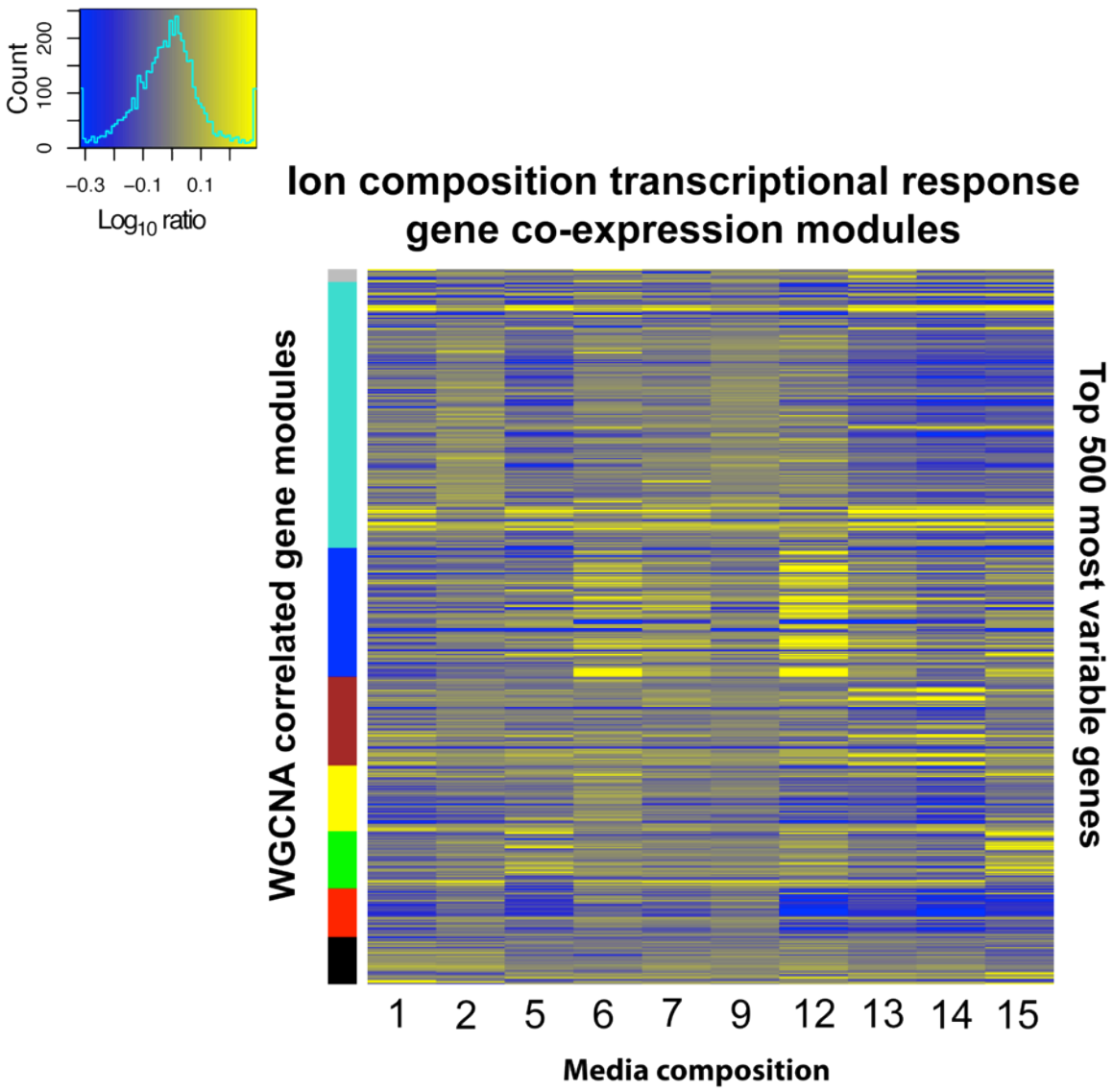

Figure 4-7: Weighted gene expression correlation network analysis (WGCNA) cluster heatmap
Log_{10} expression ratios of the 500 genes with highest coefficients of variation across 10 media. Expression values (blue, down; yellow, up) are means of two biological replicates and each row represents one gene. Vertical banding patterns within clusters indicate coexpression of module genes, and left-side color bars indicate module membership. Co-expression network was generated by measuring the concordance of gene expression with Pearson correlation coefficients to create a topological overlap matrix (or weighted adjacency matrix) showing how strongly each pair of genes is related (correlated)[164]. Because this matrix is a dissimilarity matrix, hierarchical clustering can be applied to produce a tree. Different branch-cutting procedures determine the number and size of modules.

Table 4-3: WGCNA coexpression module cluster membership[/]

Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7
NC_001869_14	NC_002608_34	NC_002607_112	NC_002607_175	NC_002607_90	NC_002607_120	NC_002607_236
NC_001869_17	NC_002607_172	NC_002607_38	NC_002607_216	NC_001869_25	NC_002607_43	VNG0137G
NC_001869_20	NC_002607_201	VNG0008G	NC_002607_225	NC_002607_244	NC_002607_5	VNG0207H
NC_001869_23	NC_002607_212	VNG0032H	NC_002607_232	NC_002607_88	NC_002607_53	VNG0230C
NC_001869_27	NC_002607_230	VNG0035H	NC_002607_42	NC_002608_48	NC_002607_84	VNG0256H
NC_002607_111	NC_002607_54	VNG0064G	NC_002607_45	NC_002608_78	NC_002608_1	VNG0268C
NC_002607_126	NC_002607_57	VNG0106G	VNG0041C	NC_002608_79	NC_002608_73	VNG0282H
NC_002607_129	NC_002607_64	VNG0134G	VNG0045C	VNG0013C	VNG0015H	VNG0511H
NC_002607_177	NC_002607_78	VNG0136G	VNG0255C	VNG0120H	VNG0117H	VNG0656H
NC_002607_189	NC_002608_71	VNG0139H	VNG0301C	VNG0243Cm	VNG0152G	VNG0816G
NC_002607_200	NC_002608_8	VNG0203C	VNG0307G	VNG0459G	VNG0176H	VNG0931G
NC_002607_206	VNG0011C	VNG0251C	VNG0446G	VNG0669H	VNG0267H	VNG1015H
NC_002607_210	VNG0056H	VNG0272H	VNG0572G	VNG0680G	VNG0311H	VNG1151H
NC_002607_242	VNG0063G	VNG0323H	VNG0692C	VNG0738H	VNG0393C	VNG1153G
NC_002607_243	VNG0069H	VNG0360C	VNG0724H	VNG0753G	VNG0398G	VNG1422H
NC_002607_245	VNG0076H	VNG0410G	VNG0736G	VNG0764C	VNG0478C	VNG1451C
NC_002607_251	VNG0148H	VNG0450C	VNG0896G	VNG0778C	VNG0604H	VNG1573G
NC_002607_28	VNG0154G	VNG0469H	VNG0978H	VNG0868H	VNG0971G	VNG1647G
NC_002607_39	VNG0166G	VNG0512G	VNG1071G	VNG0873G	VNG1038C	VNG1666H
NC_002607_58	VNG0196H	VNG0647G	VNG1090H	VNG1064H	VNG1213C	VNG1755G
NC_002607_66	VNG0285C	VNG0686C	VNG1173G	VNG1390a	VNG1562H	VNG1783H
NC_002607_72	VNG0310C	VNG0720G	VNG1301G	VNG1404G	VNG1609C	VNG2196G
NC_002607_80	VNG0383H	VNG0903C	VNG1557G	VNG1417H	VNG1646G	VNG2439H
NC_002607_85	VNG0409C	VNG0993H	VNG1576G	VNG1510C	VNG1688C	VNG2562H
NC_002607_86	VNG0419C	VNG1154H	VNG1674H	VNG1642H	VNG1848H	VNG2593H
NC_002607_99	VNG0507C	VNG1182H	VNG1675H	VNG1779C	VNG2056G	VNG6166H
NC_002608_39	VNG0527C	VNG1249C	VNG1917H	VNG1953C	VNG2150G	VNG6303G
NC_002608_9	VNG0529H	VNG1309G	VNG1965H	VNG2086G	VNG2174H	VNG6339H
VNG0009G	VNG0584H	VNG1379G	VNG1992G	VNG2102G	VNG2378G	VNG6349C
VNG0023H	VNG0676C	VNG1458G	VNG2035H	VNG2217G	VNG2603H	VNG6351G
VNG0037H	VNG0678G	VNG1466H	VNG2162C	VNG2412H	VNG6052H	VNG7053
VNG0052H	VNG0715G	VNG1484H	VNG2167G	VNG2475C	VNG6068C	VNG7130
VNG0072H	VNG0881G	VNG1496G	VNG2323H	VNG2606G	VNG6286H	VNGt39
VNG0084G	VNG0907H	VNG1520G	VNG2324H	VNG2607C	VNG7122	
VNG0160G	VNG0996G	VNG1598a	VNG2419C	VNG6022G		
VNG0163G	VNG1169C	VNG1657H	VNG2443G	VNG6113H		
VNG0198H	VNG1215G	VNG1743C	VNG2490H	VNG7086		
VNG0216H	VNG1259G	VNG1887G	VNG2501C	VNG7115		
VNG0238H	VNG1332G	VNG1912G	VNG2595G	VNG7168		
VNG0254G	VNG1389C	VNG1957G	VNG2629G	VNGt30		
VNG0271C	VNG1390H	VNG1987G	VNG6097C			
VNG0281G	VNG1403H	VNG2138G	VNG6129C			
VNG0286C	VNG1640H	VNG2146H	VNG6233G			
VNG0287a	VNG1665G	VNG2166Cm	VNG6361G			
VNG0291H	VNG1689G	VNG2173G	VNG6432H			
VNG0309C	VNG1693G	VNG2205H	VNG7067			
VNG0319H	VNG1794C	VNG2368G				
VNG0322H	VNG1856G	VNG2381G				
VNG0342G	VNG1914G	VNG2386C				
VNG0406C	VNG1934H	VNG2387H				
VNG0412G	VNG2008H	VNG2390G				
VNG0430H	VNG2025G	VNG2432C				
VNG0458G	VNG2032Gm	VNG2511H				
VNG0473G	VNG2034H	VNG2534C				
VNG0485H	VNG2094G	VNG2554H				
VNG0488H	VNG2131G	VNG2631H				
VNG0502G	VNG2185H	VNG2634H				
VNG0530G	VNG2237G	VNG6074G				
VNG0533H	VNG2307C	VNG6293C				
VNG0568C	VNG2369C	VNG6317G				
VNG0575G	VNG2477H	VNG6340H				
VNG0582C	VNG2555C	VNGt18				
VNG0594H	VNG2563H					
VNG0595H	VNG2582H					
VNG0630G	VNG2599H					
VNG0636G	VNG2624G					
VNG0650C	VNG2648G					
VNG0708H	VNG6010G					
VNG0750C	VNG6027G					
VNG0790G	VNG6081G					
VNG0828H	VNG6125H					
VNG0849C	VNG6126H					
VNG0863H	VNG6131H					
VNG0906H	VNG6148H					
VNG0921G	VNG6159H					
VNG0928G	VNG6244G					
VNG0940Gm	VNG6329H					
VNG0942G	VNG6379C					
VNG0945H	VNG6389G					
VNG0955G	VNG6393H					
VNG0987H	VNG6400H					
VNG1036H	VNG6429H					
VNG1042H	VNG7004					

VNG1187G	VNG7005
VNG1193C	VNG7012
VNG1198C	VNG7024
VNG1202C	VNG7042
VNG1230G	VNG7162
VNG1244C	VNG7164
VNG1255C	VNG7165
VNG1320G	
VNG1343C	
VNG1365C	
VNG1367G	
VNG1464G	
VNG1473H	
VNG1475C	
VNG1515G	
VNG1533H	
VNG1551G	
VNG1580H	
VNG1644G	
VNG1651H	
VNG1687C	
VNG1765G	
VNG1837G	
VNG1851G	
VNG1852H	
VNG1864G	
VNG1882G	
VNG1940H	
VNG1983H	
VNG2041H	
VNG2067H	
VNG2078G	
VNG2128C	
VNG2139G	
VNG2156C	
VNG2165H	
VNG2171G	
VNG2176H	
VNG2181G	
VNG2184G	
VNG2187H	
VNG2197H	
VNG2203G	
VNG2214G	
VNG2223G	
VNG2232G	
VNG2241H	
VNG2251G	
VNG2296C	
VNG2315H	
VNG2349G	
VNG2353H	
VNG2367G	
VNG2404G	
VNG2410G	
VNG2426G	
VNG2472G	
VNG2474C	
VNG2539H	
VNG2576H	
VNG2598G	
VNG2622H	
VNG2628H	
VNG2644C	
VNG2669G	
VNG2674H	
VNG6024G	
VNG6028G	
VNG6031G	
VNG6037G	
VNG6042H	
VNG6088C	
VNG6094H	
VNG6109H	
VNG6127H	
VNG6153G	
VNG6164G	
VNG6168H	
VNG6175G	
VNG6197H	
VNG6210G	
VNG6212G	
VNG6237G	
VNG6250G	
VNG6256G	
VNG6268C	
VNG6283H	
VNG6318G	
VNG6345H	
VNG6370H	
VNG6383G	
VNG6401H	
VNG6403H	
VNG6411H	
VNG6486H	
VNG7080	
VNG7084	
VNG7142	
VNG7167	
VNGt04	
VNGt05	
VNGt07	
VNGt19	

† *H. salinarum* NRC-1 ORF names beginning with "NC" indicate non-coding RNAs

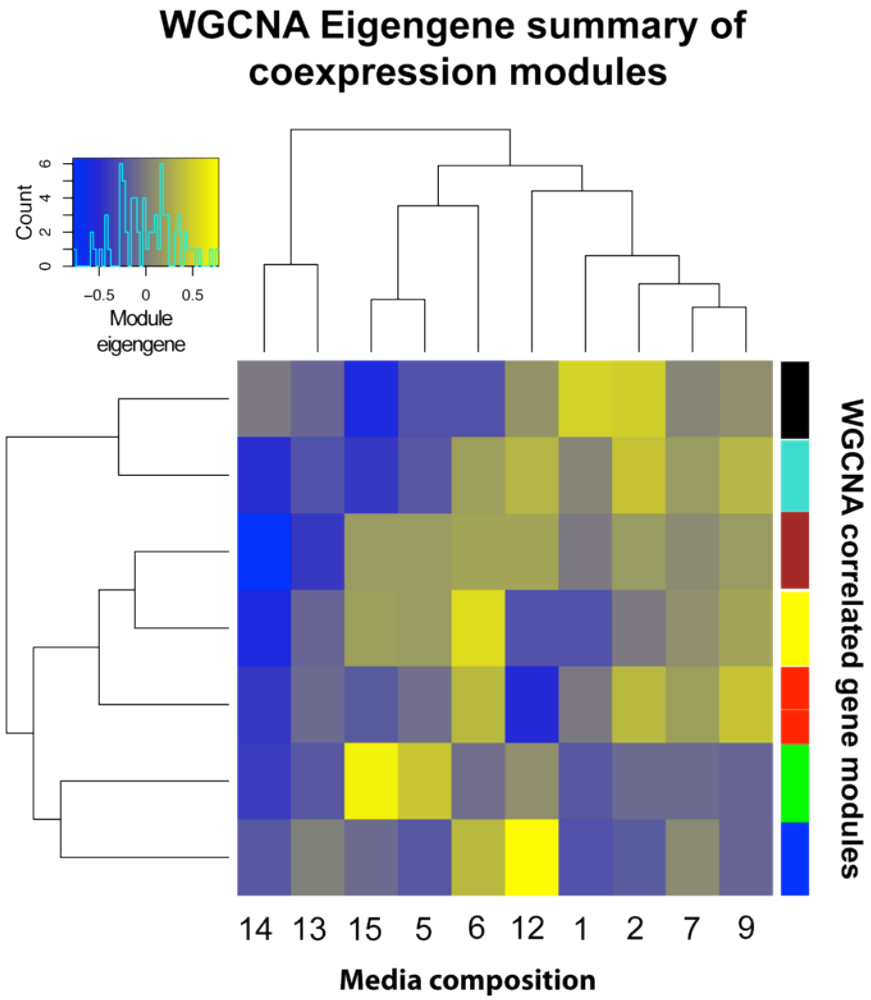

Figure 4-8: WGCNA module eigengene heatmap
Summary of correlated gene expression in each growth media. Positive values (yellow) indicate average upregulation, negative values, average downregulation (blue). Dendrograms show how expression patterns cluster among growth media and WGCNA modules (colored bars correspond with modules in Figure 4-7). Eigengenes summarize the expression profiles in each module, by taking a weighted average of all genes in the module. The eigengene is the first principal component of module gene expression, or in network terms, the most highly connected intramodular hub gene. Eigengenes reveal relationships between modules, and allow the inference of relationships between modules and metadata (*i.e.* ion composition).

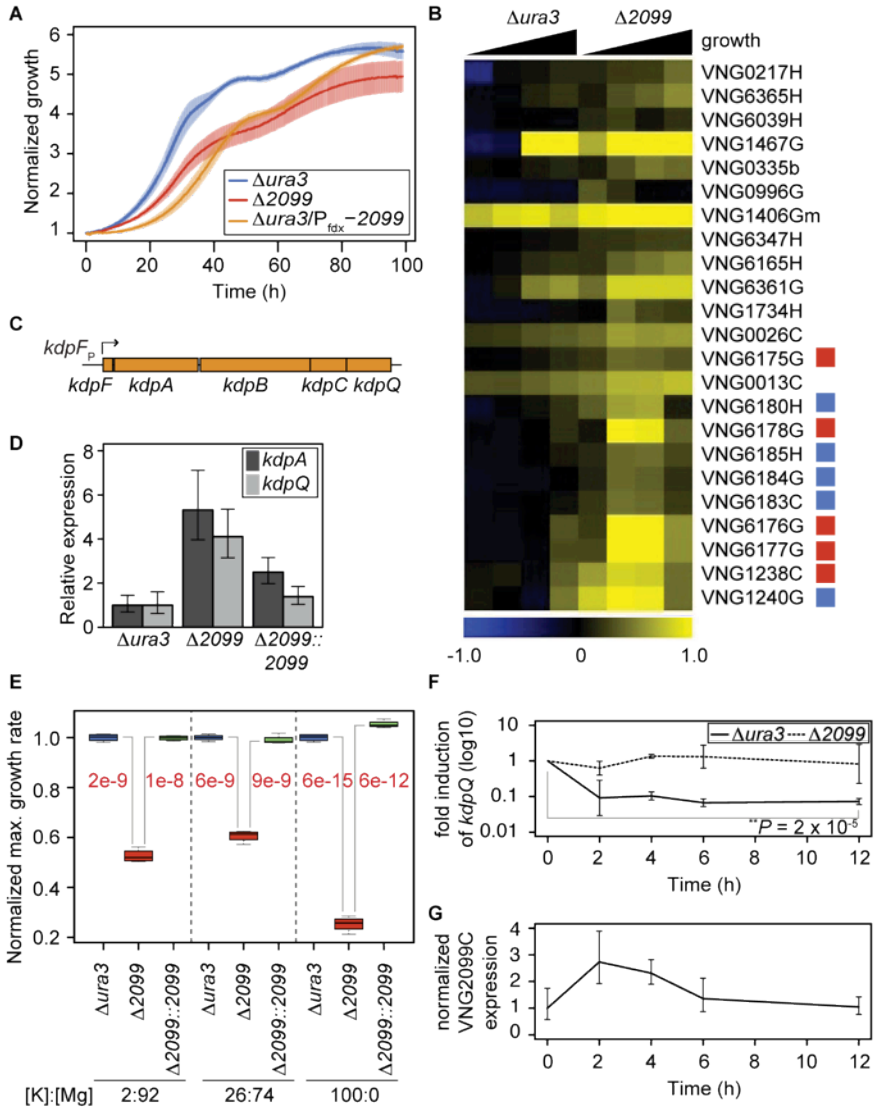

Figure 4-9: Deletion of *VNG2099C* causes a growth defect and transport gene dysregulation.
(A) Perturbations in *VNG2099C* lead to a growth defect. Error bars, s.e.m.; significant *P*-values are listed in red (t-test, n = 4-12 biological replicates per strain). (**B**) Genome-wide mRNA changes were measured at four points spanning the phases of batch culture growth of the *Δura3* and *Δ2099* strains. Log_{10} expression changes relative to a mid-log reference RNA are shown for genes with significantly increased expression in the *Δ2099* strain. Squares mark genes associated with K^+ transport (red) or other transport functions (blue). (**C**) The *kdp* operon structure. The bent arrow denotes the transcription start site. The *kdp* operon includes the positive autoregulator *kdpQ* in *H. salinarum*; in *E. coli*, the operon is controlled by a sensor kinase/response regulator system[111]. (**D**) Gene expression levels for *kdpA* and *kpdQ* mRNAs measured by qRT-PCR, normalized to the *Δura3* strain (error bars, s.e.m; n = 3 biological replicates). (**E**) The growth defect of the *Δ2099* strain is sensitive to growth media ion composition. The three media with listed [K]:[Mg] ratios are Media 1, 2 and 5 and are described in Table 3-1. Ratios include contributions from peptone and citrate. Error bars, s.e.m.; significant *P*-values are listed in red (t-test, n = 4-12 biological replicates per strain). (**F**) Fold induction of *kdpQ* transcript levels before and after an increase in extracellular KCl concentration, calculated relative to replicate cultures without KCl addition as measured by qRT-PCR. *kdpQ* induction is statistically significant in the *Δura3* strain at 12 hours ($P = 2 \times 10^{-5}$, t-test, n = 3 biological replicates; error bars, s.e.m). (**G**) Gene expression levels for *VNG2099C* mRNA measured by qRT-PCR at time points before and after an increase in extracellular K+ level (n = 3 biological replicates).

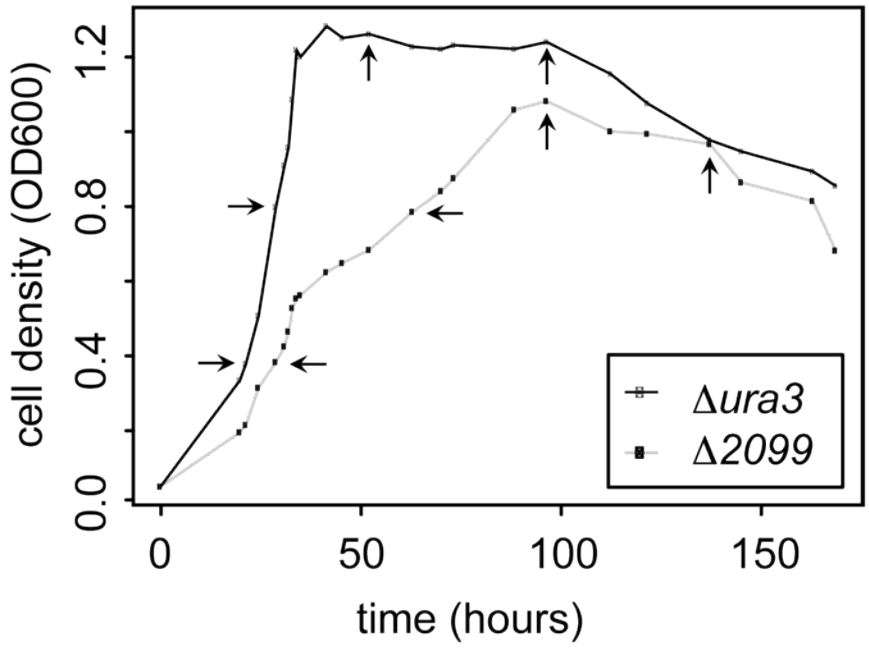

Figure 4-10: Sampling for *Δura3* and *Δ2099* genome-wide expression analysis.
Sampling for total RNA extraction and gene expression analysis (Figure 4-9c) occurred at four points of batch culture growth from the *Δura3* and *Δ2099* strains (mid-log, late-log, early stationary, and late stationary; arrows). The log phase samples were taken at the equivalent OD_{600} values for the *Δura3* and *Δ2099* strains. As the carrying capacity differs for the two strains, stationary phase points were determined as occurring after an equivalent proportion of the duration of stationary phase.

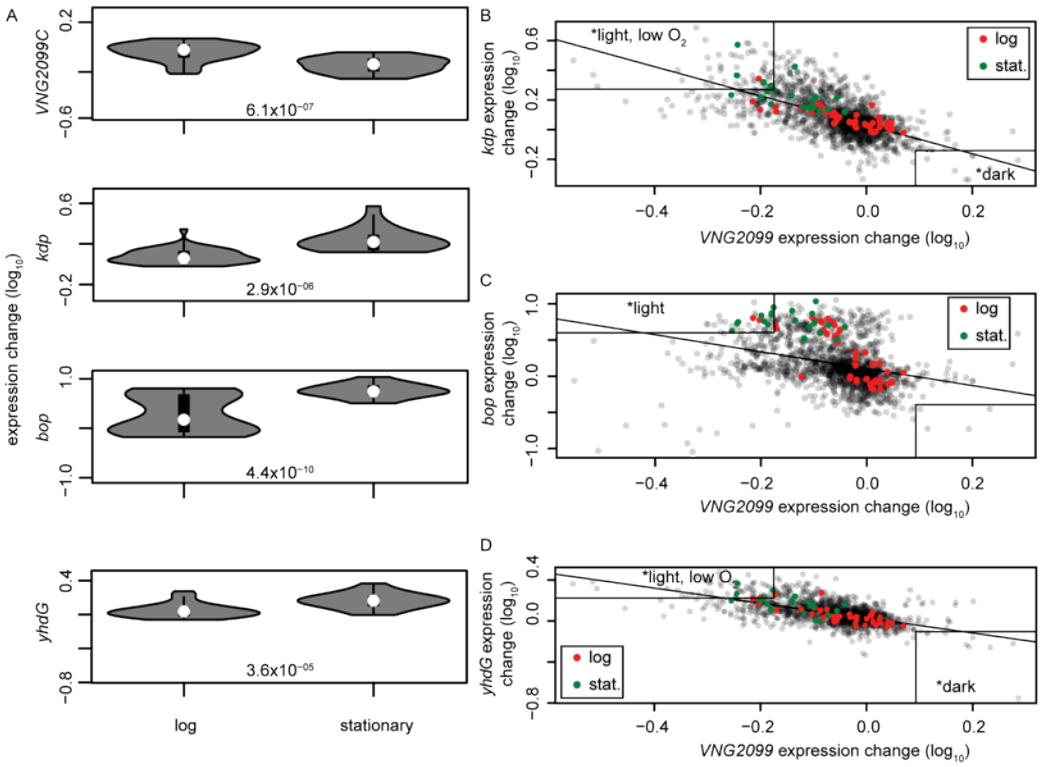

Figure 4-11: Expression levels of *VNG2099C* are anti-correlated to corresponding levels of *kdp*, *bop*, and *yhdG*.

(A) VNG2099C expression decreases from log to stationary phase while expression of *kdp*, the bacteriorhodopsin gene *bop*, and the ornithine-arginine transport gene *yhdG* increase. For log phase (OD_{600} < 1.0; n = 44) and stationary phase points (OD_{600} > 1.0; n = 20) of growth curve experiments (Reiss DJ, Brooks AN, & Baliga NS, unpublished data)[90], expression values are given as $\log_{10}$ ratios relative to a mid-log reference sample. Expression values of *kdp* genes are the median of the values for *kdpA, kdpB, kdpC,* and *kdpQ*. Significance of expression change differences between conditions was determined using two-sample t-tests, and *P*-values are listed. (**B-D**) The relative changes in mRNA expression level of the *kdp, bop,* and *yhdG* genes are plotted against relative changes in expression level of *VNG2099C* from a previous dataset of 1,495 environmental conditions (Reiss DJ, Brooks AN, & Baliga NS, unpublished data)[90]. Points in log phase are colored red; points in stationary phase are colored green. The boxes indicate the subset of points where both *VNG2099C* and target gene expression > 1.5 standard deviations from the median. (**B)** Correlation coefficient: -0.63, R^2 = 0.40, $P < 2.2 \times 10^{-16}$. Within the low *VNG2099C*, high *kdp* subset, the environmental ontology terms light and low oxygen are significantly enriched (P = 0.001, P = 0.001). Within the high *VNG2099C*, low *kdp* subset, the term darkness is significantly enriched (P = 0.01). (**C**) Correlation coefficient: -0.32, R^2 = 0.10, $P < 2.2 \times 10^{-16}$. Within the low *VNG2099C*, high *bop* subset, the term light is significantly enriched (*P*-value: 0.0004). (**D**) Correlation coefficient: -0.60, R^2 = 0.36, $P < 2.2 \times 10^{-16}$. Within the low *VNG2099C*, high *yhdG* subset, the terms light and low oxygen are significantly enriched (P = 0.001, P = 0.001). Within the high *VNG2099C*, low *yhdG* subset, the term darkness is significantly enriched (P = 0.01).

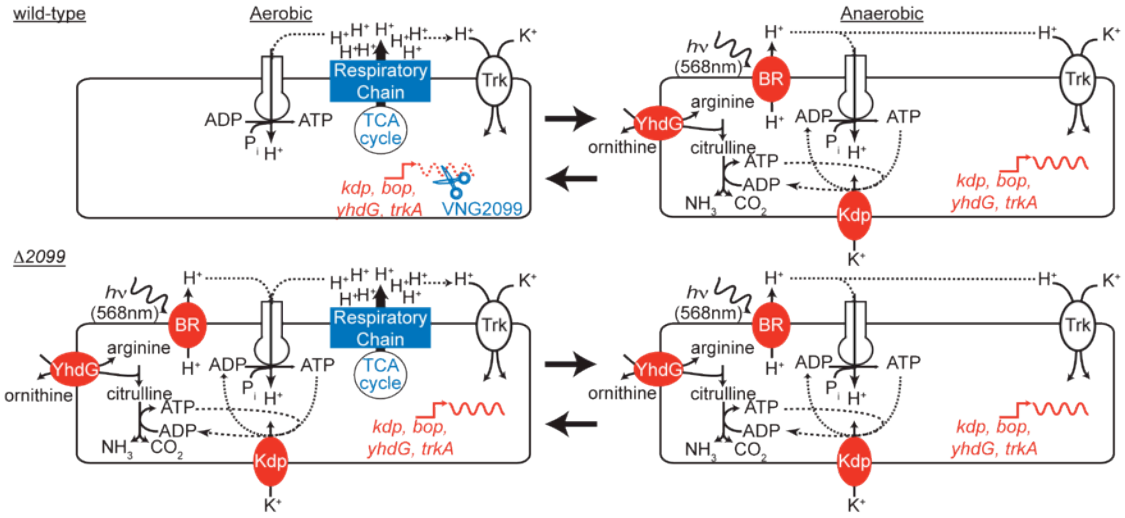

Figure 4-12: RNase-mediated phenotypic state switching.
H. salinarum cells use aerobic metabolism in log phase, dark, and high O_2 conditions[33,168]. In anaerobic conditions such as stationary phase, light, and low O_2, cells switch to alternate metabolic strategies including phototrophy and arginine fermentation. VNG2099C coordinates phenotypic switching in these oxic/anoxic transitions by regulating key genes in K^+ transport (*kdp)*, phototrophy (*bop*), and arginine fermentation (*yhdG*) (Figure 4-9c) and is itself co-regulated with aerobic metabolism genes. In the aerobic state, membrane potential from the respiratory chain may be sufficiently low for passive K^+ transport[83], phototrophy is not used[169], and some arginine fermentation genes are downregulated[174,175] (making *kdp*, *bop*, and *yhdG* expression unnecessary). Another VNG2099C target, *trkA2*, is one of five TrkA homologs in the *H. salinarum* genome. TrkA is the NAD^+/NADH-binding regulatory subunit of the H^+-K^+ symporter Trk and may couple K^+ transport to cellular redox state[170]. In conditions where *kdp* expression is unnecessary, cells may gain a bioenergetic advantage through repression of *kdp* by the VNG2099C RNase. In the *D2099* strain, positive autoregulation of the *kdp* operon leads to unchecked Kdp channel expression and ATP consumption, in turn leading to a growth defect.

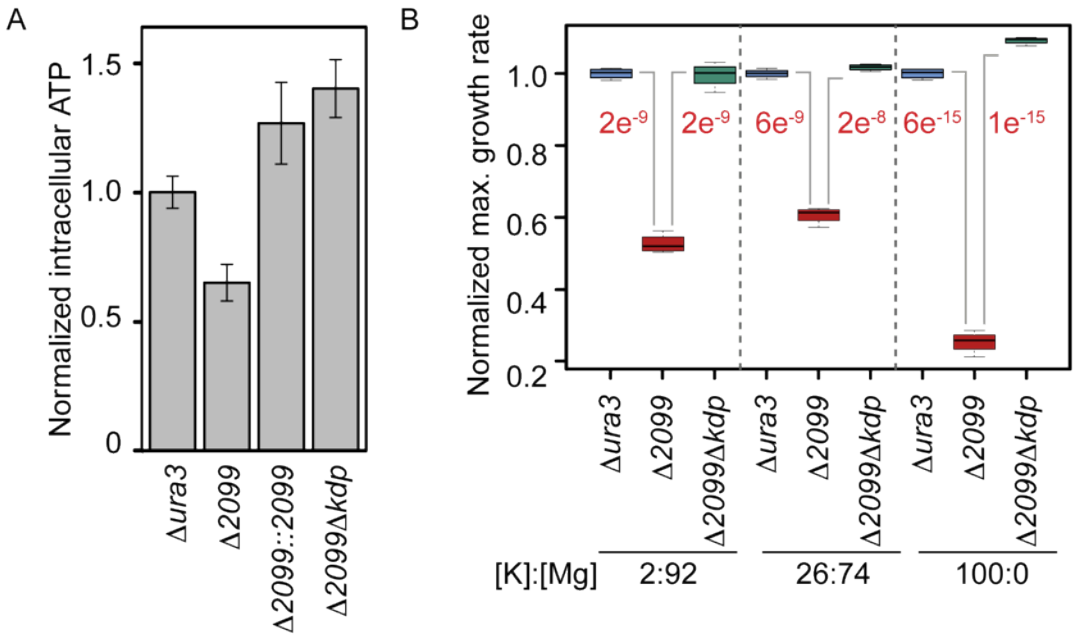

Figure 4-13: VNG2099C-mediated repression of the *kdp* operon conserves cellular ATP level and directly impacts growth rate.
(**A**) Intracellular ATP levels are statistically significant between the *Δura3* and *Δ2099* strains ($P = 0.004$, t-test, $n = 6$ biological replicates; error bars, s.e.m.), the *Δ2099* and the *Δ2099::2099* strains ($P = 0.004$, $n = 6$ biological replicates; error bars, s.e.m.), and the *Δ2099* and the *Δ2099 Δkdp* strains ($P = 0.001$, $n = 3$ biological replicates; error bars, s.e.m.). (**B**) Instantaneous maximum growth rate in media with varying [KCl]:[$MgSO_4$] ratios ($n = 4$ biological replicates; error bars, s.e.m.; the data for the *Δura3* and *Δ2099* strains are repeated from Figure 4-9b).

Chapter 5: Phenotypic and genomic stability in a model haloarchaeon, despite lack of naturally variable environmental forcing

Abstract

Although microbial adaptation to new environments has been studied extensively in laboratory evolution experiments, less is understood about how natural strains adapt to new, constant laboratory conditions after transitioning to become model organisms. Here, we took a cross-sectional approach to query the osmotolerance repertoire of a decades-old model archaeon compared with recently isolated conspecific natural isolates. Using growth rate reaction norms, we found that osmotolerance is highly variable across gradients in total salinity and ion composition for both laboratory and natural strains. Although we expected reaction norms to support evidence of specialization in the model organism, both the model and natural isolates shared similar growth patterns across conditions. Moreover, both model and isolates achieved peak growth rates in salinities departing significantly from the standard halobacterial growth medium, suggesting a generalist, rather than specialist osmotolerance strategy. Whole-genome sequencing and targeted PCR revealed that the strains share high nucleotide similarity but are nonetheless unique strains. Despite at least 50 years of divergence opportunity between the lab and natural isolates, we find that plasticity in osmotolerance is maintained in the laboratory, and that a decades-old model organism remains a relevant study organism for microbial-environmental response.

Introduction

The tremendous utility of laboratory-cultured model microorganisms like *E. coli* and *S. cereviseae* has led to unparalleled foundational insights into core biological processes. When queried to understand biological processes in the context of their natural habitats, however, the utility of such models often diminishes. Best practices for maintaining a model organism necessarily restrict its propagation to selected and controlled growth conditions, intentionally removing the environmental forcing by physicochemical and spatiotemporal variation inherent to the natural environment. As a result, model organisms may be expected to adapt and become

specialists with increased fitness in these narrow laboratory growth conditions, compared with a more generalist ancestor adapted to greater environmental variability[67,69,71,72].

Such adaptation to a new environment can manifest as a change in phenotypic plasticity, or the variation in phenotype produced by a given genotype across an environmental gradient[68,176]. Differences in phenotypic plasticity between a model organism and its natural or clinical counterpart caution against using the model to infer properties of the corresponding natural organism. In *S. cerevisiae* for example, lab strains show unique stress sensitivity and gene expression patterns compared with environmental isolates, in a broad array of lab conditions[134]. Model pathogens often loose critical virulence phenotypes after several passages in growth media[135-137]. These examples motivate the need to understand adaptation after an organism transitions from the environment to the laboratory, and to understand the consequences of such transitions for continued relevance and utility of model microorganisms.

Environment-to-laboratory transitions impose shifts from spatiotemporally variable environments to constant, controlled culturing regimes. Adaptation in new environments, and corresponding changes in phenotypic plasticity, can be quantified using reaction norm plots. Reaction norms show how a trait (e.g. maximum growth rate) varies with an environmental variable[60]. Using these plots, long-term evolution experiments have shown that constant environments favor specialists with high fitness within a narrow niche. For example, in both *E. coli*[64,65] and bacteriophage[50], after just 5,000 and 100 generations, respectively, of directed selection at a constant temperature, laboratory-evolved strains had increased fitness at the selected temperature and reduced fitness above and below it, compared with the ancestor. By contrast, variable environments favor generalists that perform comparably well across a broad niche of multiple conditions [67,69,71].

Using reaction norms to measure phenotypic plasticity in salinity response, we investigated the hypothesis that the laboratory strain of model organism *H. salinarum* NRC-1 has become a specialist, adapted to its narrow laboratory environment, compared with environmental isolates of the same organism (Fig 5-1A). We considered that the physiological repertoire of a cell mirrors the dynamics and complexity of the environment in which it has evolved[18,177], and therefore posit that high fitness across broad environmental gradients that promote success in a variable environment will not be favored after decades of laboratory culture. If so, this would

make *H. salinarum* NRC-1 an unsuitable model for molecular systems biology studies of environmentally-relevant microbe-environmental interactions[90,117,129,133,178,179].

Results and Discussion

To test this hypothesis, we compared reaction norms over salinity gradients for *H. salinarum* NRC-1 (Hsa-lab[85,87]) and three isolates from the Great Salt Lake, Utah, that match Hsa-lab at the 16S rRNA locus (Hsa-GSL, sites 6, 12 and 14). Salinity is a complex and multidimensional factor, defined by a combination of total ions and ion ratios. It remains nearly constant during laboratory culturing, but in hypersaline environments like the GSL, spatiotemporal variation in both total ions and ion ratio due to evaporation, precipitation, and rehydration are defining characteristics of the niche (Figure 5-2). To generate reaction norm plots, we measured the maximum growth rate of the four strains grown in three series of growth media: 1) a series of 15 media with constant total salinity and systematically varying ion composition; 2) a series of five media with constant ion ratio and varying total salinity (from 100% to 60% of control); and 3) a series of seven media that imitate the hypersaline chemistry at each GSL sampling site (Table 5-1). For comparison, the standard CM rich growth media for Halobacteria was included in all three media series as a control (Figure 5-3). Differences in plasticity between Hsa-lab and each GSL strain, and thus evidence of adaptive specialization[180], were evaluated using F-tests of strain × media interactions, Pearson correlations of reaction norms in each media series, and plasticity index (PI_{md})[174] comparisons. PI_{md} = maximum median fitness – minimum median / maxium median[174]. Possible PI_{md} values range from 0 (broad niche) to 1 (narrow niche)

Across each of the three media series, maximum growth rate of all four strains was highly variable (Figure 5-1B) In all media series, Hsa-lab and Hsa-GSL growth curve pairs were highly correlated, with $r > 0.94$ ($P < 0.001$), and all PI_{md} values were > 0.4. Unexpectedly, the Hsa-GSL strains produced reaction norm curves that were nearly identical, or parallel vertical translations of the Hsa-lab curves with few intersections. This is as opposed to intersecting curves, that would indicate local adaptation or an advantage in one condition at the expense of fitness in other conditions. Parallel Y-axis translations of reaction norms are attributable to a strain's genotypic contribution to the measured phenotype, while intersecting curves suggest strain ×

media interactions and differential selection for osmotolerance strategies. With generalist-specialist adaptation specifically, we would expect Hsa-lab to attain the highest maximum growth rate of all strains in laboratory media, with reduced max growth rate in extremity conditions. In the ion ratio series, strain × media interaction terms were statistically significant (Table 5-2), but t-tests comparing Hsa-lab to each Hsa-GSL strain do not point to fitness trade-offs along the gradient. In the total salinity series, however, there is evidence for reduced fitness in the highest and lowest salinity media compared with Hsa-GSL strains from sites 6 and 14, and this difference disappears at intermediate salinities (Bonferroni-corrected t-tests, $p < 0.05$). The strain × media interaction was also significant, however this pattern also does not point to specialist adaptation in Hsa-lab because Hsa-lab did not outgrow Hsa-GSL strains in Media 2. In the GSL imitation media series, no interaction was significant, and t-tests also did not suggest specialist adaptation. The plasticity index (PI_{md}) was highly similar among strains within each media series (CV%: 2.02, 8.35, 2.47 for ion ratio, total salinity and GSL imitation series, respectively), indicating low divergence in overall curve shape among strains. Finally, if Hsa-lab had become a specialist in constant laboratory salinity during its 50+ years as a laboratory model organism, then a comparative fitness advantage in laboratory media along with disadvantages at gradient extremities would be expected. Therefore, the data do not support the hypothesis that Hsa-lab has acquired a specialist strategy to thrive under invariant salinity in the laboratory.

Because of the striking similarity among reaction norms, we then asked whether Hsa-lab osmotolerance phenotypes were preserved because of, or in spite of, genome sequence similarity. To do so, we performed whole genome resequencing on all four strains. Resequencing revealed 95-99% coverage to the *H. salinarum* NRC-1 reference sequence[85], with 21, 24, 16 and 18 SNPs or IS-mediated indels in Hsa-lab, and Hsa-GSL sites 6, 12 and 14, respectively (Table 5-3). The largest differences between Hsa-lab and each GSL strain were several large regions on the two minichromosomes (replicons containing essential genes and large regions of identity) with zero reads mapping to the reference sequence (Figure 5-4, Table 5-4). In particular, Hsa-GSL strains from sites 6 and 14 are missing the 14 *gvp* gas vesicle genes on both minichromosomes, which confer the ability to modulate water column depth. This missingness absence is consistent with the known *gvp*$^-$ translucent colony phenotype, as Hsa-GSL site 6 and 14 exhibit[181](Figure 5-8). Hsa-GSL site 14 is also missing both copies of *TbpC,* a global transcriptional regulator known to influence the expression of potassium transport genes[90]. Despite these large regions of

difference, the nucleotide sequences were too similar to rule out contamination among strains. To do this, we used PCR to amplify three diagnostic loci that distinguish the strains used in the laboratory from each other and from novel *Halobacterium salinarum* strains (Figure 5-5). That the three Hsa-GSL strains are unique and uncontaminated by Hsa-lab, its parent or derivatives was confirmed.

Here we have repeated the process of isolating *H. salinarum* from the natural environment and compared these isolates with a conspecific organism that was collected approximately five decades prior. In those five decades, Hsa-lab has been passed through three generations of investigators, with no central record of strain history. Evolution in the laboratory has most certainly occurred, as demonstrated by the sequence comparison of *H. salinarum* NRC-1 with R1, another *H. salinarum* laboratory strain that differs in genome structure and protein-coding gene set, and is thought to have diverged from NRC-1 after the original isolation[182]. Our data also support evolution in the laboratory in the 12 years since Hsa-lab was first sequenced. We identified 21 SNPs or indels in resequencing Hsa-lab that differ from the reported sequence, not including the deleted *pyrF/ura3* gene, which was knocked out of the sequenced parent strain[87]. Thus, despite best practices for maintaining stocks in the laboratory, model organisms do accumulate mutations and genome rearrangements.

In addition to opportunity for adaptation in the laboratory over the past 50 years, we estimate that the Hsa-GSL strains have had ~10-15,000 generations of continued adaptation to fluctuating salinity in the GSL, given an average doubling time of 30-40 hours over 50 years[183]. In that time, the GSL has experienced drastic salinity fluctuations, including a major flood event that reduced total salinity by almost half for a period of 10 years before salinity returned to pre-flood levels[84]. Nevertheless, resequencing revealed that we had isolated three strains with the same genome structure (a main chromosome and two minichromosomes) and minimal sequence differences.

Despite the small number of sequence differences, we could still expect to see fitness differences among strains reflected in the reaction norms. With *TbpC* missing from Hsa-GSL site 14, a growth defect across the salinity media series would not be surprising given the reported 1.75-fold upregulation of this gene in low salinity[117] and the known involvement of TbpC in potassium ion transport regulation. Data from experimental evolution also support the

expectation of a growth defect in light of similar sequences. Recently, Barrick, *et al* demonstrated that while mutations accumulated linearly over time during long-term evolution experiments, most fitness improvements occurred early on, within 2500 generations and after 5-20 accumulated mutations[54,55]. Although this discrepancy between rates of mutation and adaptation could explain any observed fitness differences in our strains, it may not be apparent from our data for two reasons.

First, the Barrick study is a longitudinal one with fitness measured relative to an ancestor. By contrast, this work is a cross sectional study that sought to measure the functional differences between a model and current isolates. No confirmed ancestor was available. By measuring absolute fitness vs. relative fitness, we may be observing the similar phenotypic end products of two different adaptive trajectories. Long-term experiments that measure evolution of Hsa-lab and Hsa-GSL across salinity media are needed to reveal relative fitness and mutational trajectories after a variable-to-constant environmental transition. Transcriptomic analyses comparing Hsa-lab and Hsa-GSL gene expression during salinity changes would also reveal differences not apparent at the levels of nucleotide sequence or growth rate. Secondly, several of the early mutations in the Barrick study occurred in regulatory loci, and likely had pleiotropic effects on numerous target genes[55]. That critical mutations are often found in regulatory genes is well-established (Reviewed in [69,182] and [184]). But, in our data, the SNPs and indels that differentiated Hsa-lab from Hsa-GSL occurred in non-regulatory or hypothetical proteins (Table 5-3), except for *TbpC*. Thus, even though it is possible that the mutations separating the Hsa strains were non-pleiotropic and insufficient to confer a specialist osmotolerance strategy in Hsa-lab, they may explain the significant differences we did observe in reaction norm shape.

At the outset, our goal was to quantify the phenotypic and genomic divergence between a powerful model organism and conspecific natural isolates adapted to the harsh and fluctuating hypersaline niche, for which numerous biotechnological applications exist[92,119,185,186]. Monitoring the divergence of laboratory and natural isolates is critical, as demonstrated by laboratory pathogens that have lost critical virulence properties. To paraphrase statistician George E.P. Box: All models are false, but some are useful. Applied to experimental biology, the useful models are those that recapitulate the phenomena of interest. Given the opportunity for the Hsa-lab model and Hsa-GSL strain to diverge during their ~50 year parallel tenures in the laboratory and lake, it might appear striking that their osmotolerance phenotypes and genome sequences are so similar.

In reality, however, the surprise may be unwarranted. Compared with its entire evolutionary history in the GSL, selective forces in the last five decades may not have changed substantially enough to eliminate the ancestral genotype from the lake. For Hsa-lab, the similarities should not be surprising given meticulous storage and handling of master culture stocks. Nevertheless, Hsa-lab has ultimately retained its utility as a model for systems-level investigations of osmotolerance in critical hypersaline environments, such as salinized soils and hypersaline industrial waste[129,185].

Supporting Information

Lab strain and growth conditions

Halobacterium salinarum NRC-1 (Hsa-lab) has been maintained the Baliga lab since its inception (12 years). The history of the strain is not well-documented, however it was passed to N. Baliga from S. DasSarma and from F. Doolittle to S. DasSarma. F. Doolittle received the strain from A. Matheson (deceased) and its isolation year is estimated between 1960 and 1970. The Hsa-lab Δ*ura3* strain used here is derived from the original parent isolate,[87] and was selected for its central role in molecular systems biology research.

Hsa-lab and isolated Hsa-GSL strains were cultured identically in the following conditions. The standard control media (CM, or Media 2) consisted of 250g/l (4.28 M) NaCl, 20g/l (0.081 M $MgSO_4$) $MgSO_4*7H_2O$, 2g/l (0.027 M) KCl, 3g/l (0.01M) sodium citrate dehydrate, 10g/l Oxoid brand bacteriological peptone, and 50 mg/l uracil, without trace metals[128]. Ion composition and total salinity media series were based on this media (Table 5-1). All media were formulated from the same chemical lots of all ingredients. In all incubations, cultures were grown at 37°C with continuous light and shaking (220 rpm).

Growth media formulation

Ion composition series media: Media compositions represent a bifactorial design in which the effects of Na and Mg:K ratio on growth rate could be evaluated. The 15 media are divided into 3 groups, with high, medium or low Na (Table 5-1). Within each level of Na, five levels of Mg:K ratio were evaluated: 0:100%, 25:75%, 50:50%, 75:25% and 100:0%.

Total salinity series media: All media in the Ion composition series were diluted in a solution of 3 g/l sodium citrate and 10 g/l Oxoid peptone to maintain a uniform concentration of organic salts and complex peptone components. Dilutions represent 90, 80, 70, and 60% (v/v) total salinity. Hsa-lab was grown in the entire series of 75 media (15 x 5 dilutions) to achieve a full factorial design with factors total salinity, Na level and Mg:K ratio. Hsa-GSL strains were grown in the diluted series of Media 2 only.

GSL imitation series media: Samples from GSL sites were collected and filtered as described below and sent for analysis of the 5 dominant ions in the lake (Cl, Na, Mg, SO_4, and K, in that order). Based on chemical analysis, growth media were formulated to imitate the sample composition as closely as possible. Because ion concentrations were measured but media must be made using a limited combination of ionic salts (NaCl, $MgSO_4$, and KCl), we used the following guidelines to convert between molarity of measured ions and added salts. 1) NaCl molarity = measured Na molarity; 2) KCl molarity = measured K molarity; 3) $MgSO_4$ molarity = average of measured Mg and SO_4 molarity. This formulation resulted in artificially equivalent molarities of Mg and SO_4, but was unavoidable given that Mg and SO_4 were added as a single ionic salt.

GSL water and Haloarchea sample collection

Samples were collected from each of 7 north-arm GSL sites, selected for chemical and ecological diversity (Figure 5-6, Table 5-5). Water samples were collected by hand into sterile glass bottles, and prefiltered through a 20 μm plankton sieve to remove *Artemia* cysts and other macroscopic debris. Samples for water chemistry were further filtered through 5 μm and 0.2 μm membranes, and stored in acid-washed, dark polycarbonate bottles.

Chemical analysis of GSL water samples

To determine the concentration of the five most abundant ions (Na, Cl, Mg, SO_4, K) in the GSL, filtered samples were sent to Hoh Pak Laboratories, Inc (New Iberia, LA) for ion analysis. This analytical lab was selected as it specializes in ion analysis of hypersaline samples. All GSL samples were sent for analysis. In addition, we analyzed samples from time series of evaporation and rehydration of GSL sample 14. We also analyzed the invariant rich media base

(3 g/l sodium citrate + 10 g/l Oxoid peptone) for all series of growth media so that the contribution of peptone and citrate to the major ions could be accounted for.

The following methods were used at HohPak labs, Inc, to measure the concentrations of each ion: Sodium and potassium – atomic absorption spectroscopy; chloride – titration with 0.01N silver nitrate; Magnesium – titration with 0.01M EDTA; Sulfate – gravimetric precipitation with barium chloride. Selected samples were submitted in blind triplicate to obtain a measure of technical variability. Coefficients of variation ranged from 0 to 5.5%.

Hsa-GSL enrichment and isolation

Samples pre-filtered through a 20 μm plankton sieve were used as inocula for enrichments within a day of collection from each GSL site (Figure 5-7). These samples were part of a separate project aimed at isolating halophilic algae and their co-inhabiting heterotrophic prokaryotes, which were targeted simultaneously by amending algal F/2 media with either 0.5% casamino acids or 5mM sodium pyruvate plus 0.05% casamino acids. Duplicate culture tubes containing 2-3 mL of F/2 media[110] with added salts (3.4 M NaCl, 0.07M $MgSO_4*7H_2O$, 0.02 M KCl, 0.17 mM NaH_2PO_4, 1 mM $NaHCO_3$) and each of the two amendments were inoculated 1:10 with sample aliquots, and incubated statically at ambient temperature during transport from Salt Lake City (UT) to Seattle (WA), for a total time of 2-3 days.

Subsequent isolation from heterotrophic enrichments ocurred on 1% (w/v) agar plates containing the same amended F/2 media above, with incubation in the dark at room temperature (21-25°C). Colonies from these plates were selected for isolation, screening and sequencing for all sites, except site 6. Enrichment plates from site 6 crystallized early due to a failed plate seal, leaving no recoverable colonies at the time of this work.

Before the plates crystallized, colonies from enrichment plates were isolated and purified in media containing the following, based on historic GSL salinity data [92]: 4 M NaCl, 0.16 M $MgSO_4$, 0.08 M KCl, 2 mM sodium pyruvate, 0.05% casamino acids (w/v), 0.005% yeast extract (w/v), and 1% (w/v) agar. The colonies subsequently isolated and sequenced from site 6 came from this isolation plate.

Hsa-GSL isolation and 16S sequencing

A total of 94 pink or whitish colonies on enrichment plates (or for site 6, the re-streaked plate) representing all sampled sites in the GSL were screened for membership in the halophilic archaea by PCR of 16S rDNA, using primers HK1F and H589R[99], which amplify the first 524 bp of the 16S gene Qiagen HotStar Taq and accompanying protocol were used for PCR conditions, with an annealing temperature of 60°C. Thirty-five of 94 colonies, representing all sample sites except sites 13 and 15, were PCR-confirmed halophilic archaea and these colonies were cultured in Media 2. Tested colonies from sites 13 and 15 were PCR-negative. Of the 35 PCR-positive colonies, 31 reached $OD_{600} > 0.10$ after three weeks or less of incubation in Media 2. These 31 cultures were streaked on Media 2 agar plates to achieve single colonies for pure culture; two colonies from each plate with growth were selected for PCR re-screen, and sequencing (Genewiz) for pure culture confirmation. After plating, re-screening and sequencing, 21 colonies returned quality sequences, and all but one aligned with *Halobacterium salinarum* NRC-1 with 100% coverage. The remaining sequence aligned with *Halorubrum* spp, another member of the Halobacteriaceae. The 21 colonies that produced quality sequence were grown to OD_{600} ~0.6 in liquid Media 2 and frozen at -80°C.

From the sequences that matched *H.* salinarum, we selected three isolates, one each from sites 6, 12 and 14 (Hsl-GSL 6, 12 and 14) using the criteria that they come from geographically and chemically diverse sites, and that their 16S sequences match GSL-lab. The three selected strains (6, 12 and 14) met all required criteria and we note here the media amendments included in the original enrichments of these strains as described above: 0.5% casamino acids (Hsa-GSL 6) and 5 mM sodium pyruvate plus 0.05% casamino acids (Hsa-GSL 12 and 14). See appendix for 16S sequence alignments of Hsa-GSL strains, Hsa-lab and the reference *H. salinarum* NRC-1. 16S sequences confirmed by whole genome sequencing, described below.

Bioscreen growth assays and growth rate calculations

Growth assays were performed in two Bioscreen C instruments (Growth Curves USA, Piscataway, NJ), with a maximum throughput of 400 cultures (200 µl each). For all strains and replicates in all experiments, fresh colonies were grown up in 5 mL cultures of Media 2, to $OD_{600} = 0.4\text{-}0.6$ (mid-log phase). These cultures were passaged in Media 2 at OD 0.05 to synchronize the cultures, and grown to $OD_{600} = 0.6$. All cultures were then diluted in Media 2 to

OD_{600} = 0.55 to ensure identical starting OD. Using a Hamilton Microlab Star robot (Hamilton Robotics, Reno, NV) to pair each of 8 cultures with each of up to 16 different media, we inoculated each well of a Bioscreen honeycomb plate with culture and experimental media to a final OD of 0.05. All cultures were maintained in Media 2 until transfer to the bioscreen plate, when they were added at <1:10 in experimental media. This method required ~5% differences between media salinity and final experimental salinity, but was ideal because it did not require a stress-inducing centrifugation or rinse step to remove Media 2. Cultures were incubated in the Bioscreen C for 110 hours at 37 °C with maximum shaking. OD_{600} was measured every 15 minutes for the duration of the experiment.

Maximum growth rate was extracted from Bioscreen OD measurements using a custom R package, "Growth Curve Analysis Function," to automate analysis[129]. Briefly, cell optical density measurements, relevant metadata and plate layout schemes were combined in a single database, which enabled rapid calculation of growth curve parameters (max growth rate, max OD, area under curve, etc). Maximum growth rate was calculated from smooth spline fitted growth curves, as the inflection point where the second derivative transitioned from positive to negative[130].

Whole genome sequencing

Purified genomic DNA from *H. salinarum* NRC-1:Δ*ura3* (Hsa-lab) and GSL isolates (Hsa-GSL) from sites 6, 12 and 14 were sent to Beijing Genomics Institute for library preparation and sequencing using the Illumina HiSeq 2000 platform. Two libraries, with 500bp and 5000bp insert sizes, respectively, were paired-end sequenced, with raw read lengths of 90bp. The large library was created in order to capture the lengths and locations of the insertion sequences in the *H. salinarum* genome. Basic bioinformatics analysis, including SNP and InDel calling were performed by BGI as well.

Whole genome sequence analysis and assembly

Raw reads were processed prior to assembly and/or mapping as follows: The first five basepairs at the 5'-end were clipped with *fastx_trimmer* from the **fastx** toolkit (http://hannonlab.cshl.edu/fastx_toolkit/index.html). The 3'-end was trimmed with *fastq_quality_trimmer* (**fastx**) to remove positions with quality values (QV) below 13, discarding

reads trimmed to a length below 70 bp. The remaining reads were filtered at a minimum average QV of 30 with the *trim-fastq.pl* script from the **popoolation** suite[187]. *De novo* assemblies were conducted with IDBA_UD[188] at *kmer* values between 35-69 with steps of 2, and with the *pre-correction* option selected. Unique regions were identified as assembled regions without blast matches against the *Halobacterium salinarum* NRC-1 genome sequence, and were present in Hsa-GSL site 6 only. These regions had BLAST matches to the *H. salinarum* R1 genome. Read mapping against the *H. salinarum* NRC-1 genome sequence was conducted with *bowtie2*[189], with the pre-defined *very-sensitive-local* option. All read libraries were downsampled for mapping to a minimum common value to allow direct coverage comparisons across libraries. Coverage by position was extracted using the *mpileup* utility from **samtools**[190]. Missing regions were defined as regions with coverage 2 standard deviations below mean values of the Freeman-Tukey transformed coverage data. Circular coverage diagrams were constructed with Circos[191] (Krzywinski et al., 2009).

To identify specific mutations, indels and new sequence junctions, we used Breseq[55], a computational pipeline for analysis of re-sequencing data. Breseq uses a reference sequence (*Halobacterium salinarum NRC-1*; RefSeqs for all three chromosomes: NC_002607.1, NC_001869.1, NC_002608.1) to align short reads from related strains. Filtered, quality-trimmed reads (described above). Breseq was run on the 500bp libraries (Paired end F and R reads), using command *breseq -r* with no additional arguments after references and reads. A manual and key to output tables notation (Table 5-3) is available at http://barricklab.org/twiki/pub/Lab/ToolsBacterialGenomeResequencing/documentation/index.html. Hand-verifying of SNP/InDel calls from BGI and Breseq was done using the Integrative Genomics Viewer (IGV)[192].

PCR exclusion of cross-contamination in whole genome sequences

Because the resequencing data revealed highly similar nucleotide sequences, we designed a set of three diagnostic primer pairs in order to ascertain that sequence similarity was not due to sample contamination by any of the laboratory strains in our lab. *H. salinarum* NRC-1, its daughter strain *H. salinarum* NRC-1:Δ*ura3* (Hsa-lab), and a different but related strain, *H. salinarum* R1 are used in the lab and although similar, each has unique identifying sequences that can be exploited by PCR to identify each one (Figure 5-5).

Strain R1 and its derivatives can be distinguished from the NRC-1 strains at the *hcpB* locus (VNG2196 in NRC-1, OE4037R in R1). NRC-1 has one halocyanin domain, while R1 has two[182]. Primers GSLWGS_hcpB_F (GCATCACGGTCGCAATAACC) and GSLWGS_hcpB_R (TGTCTGGCTTCGACGTACAC) produced amplicons of 1342 and 1765 bp in NRC-1 and R1, respectively. After ruling out contamination by R1, contamination in Hsa-GSL strains was ruled out by amplifying a 300-bp fragment within the *ura3* gene, using primers ura3_fwd (ATCACCGTCAACCCGTACCTC) and ura3_rev (CCGTGTTCGCCCGCGAAGATG). Finally, contamination by the parent NRC-1 strain was ruled out by amplifying an IS-mediated insertion/duplication present in the Hsa-GSL strains but not Hsa-lab (Δ*ura3*) or the NRC-1 reference strain. Primers GSLWGS_1125686_WT_F () and GSLWGS_1125686_WT_F () were designed to flank a newly-inserted ISH2 element, and amplified a 489 bp fragment in NRC-1 and Hsa-lab, and a 1010 bp fragment or no fragment in all Hsa-GSL strains. Because mobile element transposition can occur spontaneously, and frequency increases during stationary phase, we followed up on the NRC-1 diagnostic PCR by screening 48 NRC-1 colonies to measure the frequency of ISH2 insertion at this locus. We found that none of the screened colonies had an insertion at this locus matching the Hsa-GSL strains. Specifically, 47 of 48 colonies did not have the inserted ISH2, and the 48th colony had no band. Although Hsa-GSL from site 12 also did not amplify, we conclude that it is not identical to Hsa-lab, or NRC-1. Any similarity between Hsa-GSL site 12 and the 48th colony was not investigated, due to the low probability ($P<0.025$) of a match between NRC-1 and this strain.

Chapter 5 Figures and Tables

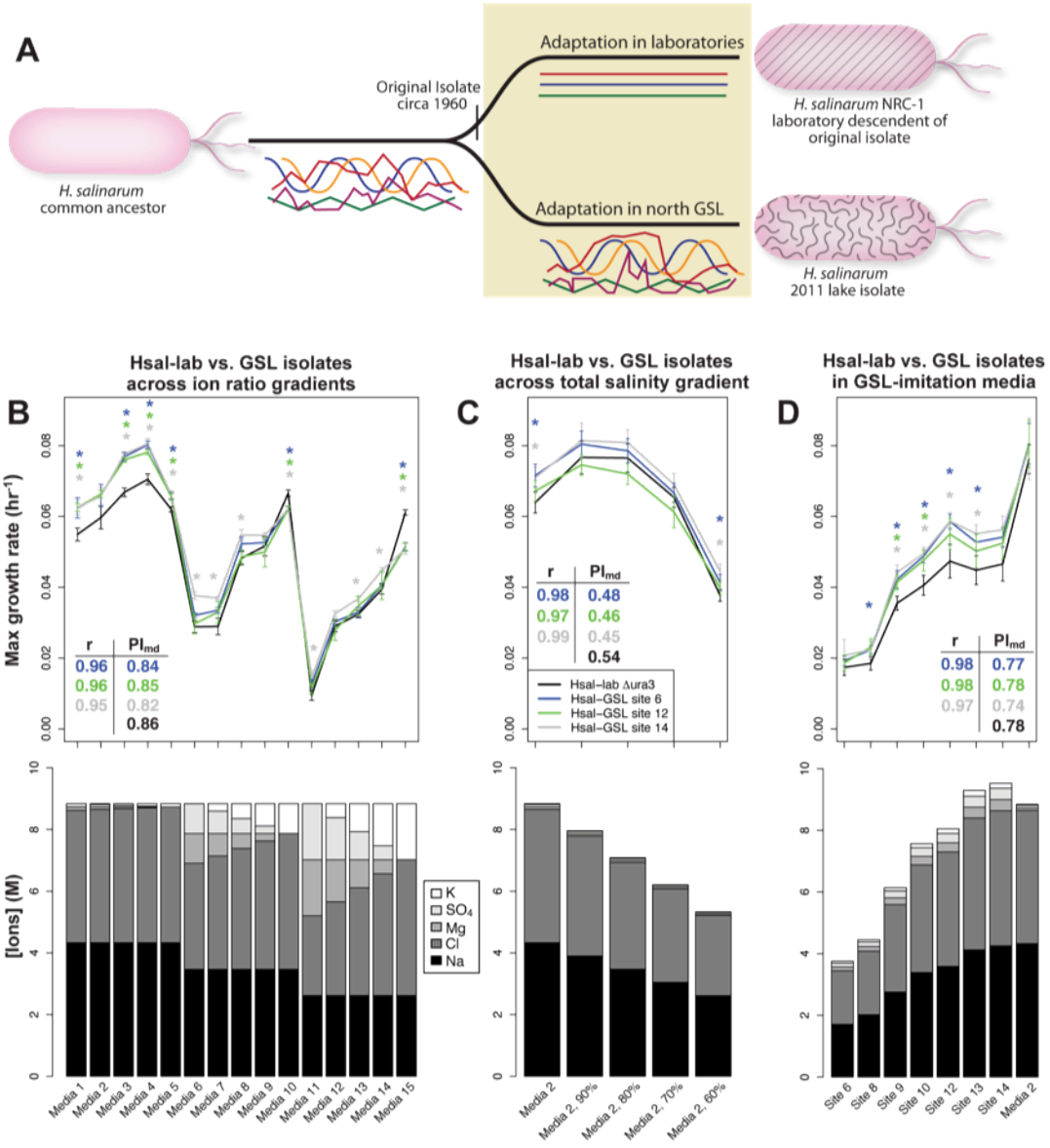

Figure 5-1: Salinity response in H. salinarum model organism and natural isolates.
A. Hypothesis that differential environmental complexity between the lab and the Great Salt Lake have led to adaptive specialization in the laboratory model organism. Colored sparklines within the yellow box denote a larger number and higher complexity of environmental factors in the GSL vs. the laboratory. **B**. Reaction norm (upper plot) showing maximum growth rate of the H. salinarum NRC-1 lab parent strain, Δura3, with three GSL isolates from sites 6, 12 and 14, across 15 growth media with identical total salinity. This is the predominant strain used for molecular systems biology of H. salinarum and is most relevant for comparison. The relatively lower max growth rate of Hsa-lab in some media is explained by the known growth defect of the Δura3 in these media (ΔpyrF ortholog) strain – a counter-selectable genetic background used routinely to make precise chromosomal manipulations with a two-step homologous recombination procedure (data not shown)[87]. Error bars are standard deviations of 6-8 biological replicates. Asterisks, colored according to the strain key, indicate Bonferroni-corrected $P < 0.05$ for t tests comparing Hsal-lab with each Hsal-GSL strain in each growth medium. Pearson's r values indicate correlations between max growth rate in Hsal-lab and each Hsal-GSL strain across all 15 media; all tests are significant at $P < 0.0001$ and text color corresponds with the GSL strain paired with Hsal-lab, as in the plot key. PI_{md}

Plasticity index measures the niche size for each strain within each media series; 0 = an infinitely broad niche (a horizontal line) and 1 = minimum niche size. Text color matches strains in plot key. Media compositions (lower plot) were designed to test the effects of ion ratio independent of total salinity on max growth rate. Media 1-5: high NaCl, Media 6-10: medium NaCl, Media 11-15: low NaCl. Within each level of NaCl, the ratio of K:Mg varied: 0:100%, 25:75%, 50:50%, 75:25% and 100:0%. All media contained the same concentration of a rich media base: Oxoid peptone and sodium citrate. Contributions from the base to the five major ions have been included in the graph. Media 2 is the control rich growth medium, and has an ion profile that reflects the GSL ($[Cl^-] > [Na^+] > [Mg^{2+}] > [SO_4^{2-}] > [K^+]$). **C**. Reaction norm (upper) and media composition (lower) plots of max growth rate in five media that differ only in total salinity. Media 2, the control, was diluted to 90, 80, 70 and 60% total salinity in a solution of rich media base. **D**. Reaction norm and media composition plots of max growth rate in media formulated to reproduce the ion compositions of seven sample sites in the GSL. Media 2 is the control.

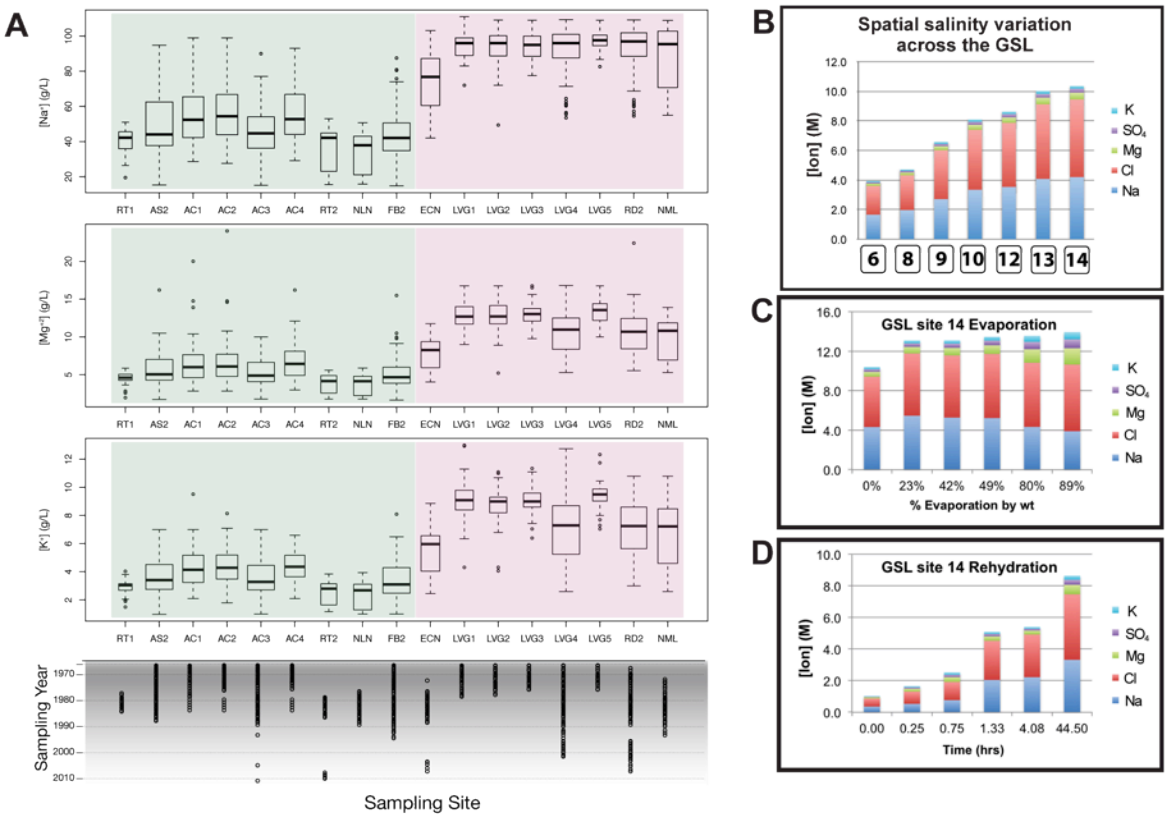

Figure 5-2: Great Salt Lake water ion composition varies temporally and spatially
A. Historical variation in Na, Mg and K concentrations in 17 different sites within the Great Salt Lake. Box plots represent ion concentrations measured at times indicated in bottom plot, between 1970 and 2010. Sites are monitored by the Utah Geological Survey and light green background indicates south arm sites. Pink background denotes north arm sites. Site descriptions and locations available in Gwynn 2007[92].
B. Samples were taken from seven spatially diverse sites within the north arm of the GSL, and ions were measured in g/L before conversion to molarity. Lower salinity sites (6 and 8) are located near the causeway, at the south end of the north arm of the GSL. **C.** Ion composition measurements during evaporation were taken on duplicate 50 mL samples taken from a starting volume of 500 mL GSL water from site 14. Control samples were analyzed in blind triplicate to gain a measure of reproducibility during the ion measurement process (Hoh-Pak Laboratories, Inc. New Iberia, LA). **D.** Rehydration measurements were performed by allowing the evaporation samples to fully crystallize, and then quickly adding back the mass of lost fresh water. Rehydration at T_0 was taken immediately after adding the water. During rehydration, experimental beakers were not covered to prevent additional evaporation. The purpose of this method was to measure dynamics that more closely resemble those in the natural environment. The presence of minimal technical variability in the measurement process was assessed using triplicate controls (Table 2-3).

Table 5-1: Composition and maximum growth rates in three media series

	Media composition name	salinity dilution series	[NaCl] level	K:Mg (M ratio)[1]	Strain	Mean max growth rate (hr^{-1})[3]	SD max growth rate[3]	Total Salinity (M salts)[1]	NaCl (M)[1]	$MgSO_4$ (M)[1]	KCl (M)[1]	NaCl (g/L)	$MgSO_4$*7H2O (g/L)	KCl (g/L)	Na (M)[2]	Cl (M)[2]	Mg (M)[2]	SO_4 (M)[2]	K (M)[2]	Total Salinity (M ions)[2]
Ion composition series																				
	CM.Mratio.1	100%	High	0:100	Hsa-lab	0.055	0.002	4.388	4.28	0.108	0	250	26.6	0.0	4.326	4.294	0.108	0.109	0.002	8.839
	CM.Mratio.2	100%	High	25:75	Hsa-lab	0.060	0.003	4.388	4.28	0.081	0.027	250	20.0	2.0	4.326	4.321	0.081	0.082	0.029	8.839
	CM.Mratio.3	100%	High	50:50	Hsa-lab	0.067	0.001	4.388	4.28	0.054	0.054	250	13.3	4.0	4.326	4.348	0.054	0.055	0.056	8.839
	CM.Mratio.4	100%	High	75:25	Hsa-lab	0.070	0.001	4.388	4.28	0.027	0.081	250	6.7	6.0	4.326	4.375	0.027	0.028	0.083	8.839
	CM.Mratio.5	100%	High	100:0	Hsa-lab	0.062	0.001	4.388	4.28	0	0.108	250	0.0	8.1	4.326	4.402	0.000	0.001	0.110	8.839
	CM.Mratio.6	100%	Medium	0:100	Hsa-lab	0.029	0.002	4.388	3.42	0.968	0	200	238.6	0.0	3.466	3.434	0.968	0.969	0.002	8.839
	CM.Mratio.7	100%	Medium	25:75	Hsa-lab	0.029	0.002	4.388	3.42	0.726	0.242	200	178.9	18.0	3.466	3.676	0.726	0.727	0.244	8.839
	CM.Mratio.8	100%	Medium	50:50	Hsa-lab	0.048	0.002	4.388	3.42	0.484	0.484	200	119.3	36.1	3.466	3.918	0.484	0.485	0.486	8.839
	CM.Mratio.9	100%	Medium	75:25	Hsa-lab	0.052	0.003	4.388	3.42	0.242	0.726	200	59.6	54.1	3.466	4.160	0.242	0.243	0.728	8.839
	CM.Mratio.10	100%	Medium	100:0	Hsa-lab	0.067	0.001	4.388	3.42	0	0.968	200	0.0	72.2	3.466	4.402	0.000	0.001	0.970	8.839
	CM.Mratio.11	100%	Low	0:100	Hsa-lab	0.009	0.001	4.388	2.57	1.818	0	150	448.1	0.0	2.616	2.584	1.818	1.819	0.002	8.839
	CM.Mratio.12	100%	Low	25:75	Hsa-lab	0.029	0.002	4.388	2.57	1.364	0.454	150	336.2	33.8	2.616	3.038	1.364	1.365	0.456	8.839
	CM.Mratio.13	100%	Low	50:50	Hsa-lab	0.032	0.001	4.388	2.57	0.909	0.909	150	224.0	67.8	2.616	3.493	0.909	0.910	0.911	8.839
	CM.Mratio.14	100%	Low	75:25	Hsa-lab	0.039	0.001	4.388	2.57	0.454	1.364	150	111.9	101.7	2.616	3.948	0.454	0.455	1.366	8.839
	CM.Mratio.15	100%	Low	100:0	Hsa-lab	0.061	0.001	4.388	2.57	0	1.818	150	0.0	135.5	2.616	4.402	0.000	0.001	1.820	8.839
	CM.Mratio.1	100%	High	0:100	Hsa-GSL site 6	0.062	0.003	4.388	4.28	0.108	0	250	26.6	0.0	4.326	4.294	0.108	0.109	0.002	8.839
	CM.Mratio.2	100%	High	25:75	Hsa-GSL site 6	0.066	0.003	4.388	4.28	0.081	0.027	250	20.0	2.0	4.326	4.321	0.081	0.082	0.029	8.839
	CM.Mratio.3	100%	High	50:50	Hsa-GSL site 6	0.077	0.001	4.388	4.28	0.054	0.054	250	13.3	4.0	4.326	4.348	0.054	0.055	0.056	8.839
	CM.Mratio.4	100%	High	75:25	Hsa-GSL site 6	0.080	0.001	4.388	4.28	0.027	0.081	250	6.7	6.0	4.326	4.375	0.027	0.028	0.083	8.839
	CM.Mratio.5	100%	High	100:0	Hsa-GSL site 6	0.066	0.001	4.388	4.28	0	0.108	250	0.0	8.1	4.326	4.402	0.000	0.001	0.110	8.839
	CM.Mratio.6	100%	Medium	0:100	Hsa-GSL site 6	0.032	0.001	4.388	3.42	0.968	0	200	238.6	0.0	3.466	3.434	0.968	0.969	0.002	8.839
	CM.Mratio.7	100%	Medium	25:75	Hsa-GSL site 6	0.033	0.001	4.388	3.42	0.726	0.242	200	178.9	18.0	3.466	3.676	0.726	0.727	0.244	8.839
	CM.Mratio.8	100%	Medium	50:50	Hsa-GSL site 6	0.052	0.002	4.388	3.42	0.484	0.484	200	119.3	36.1	3.466	3.918	0.484	0.485	0.486	8.839
	CM.Mratio.9	100%	Medium	75:25	Hsa-GSL site 6	0.053	0.001	4.388	3.42	0.242	0.726	200	59.6	54.1	3.466	4.160	0.242	0.243	0.728	8.839
	CM.Mratio.10	100%	Medium	100:0	Hsa-GSL site 6	0.062	0.001	4.388	3.42	0	0.968	200	0.0	72.2	3.466	4.402	0.000	0.001	0.970	8.839
	CM.Mratio.11	100%	Low	0:100	Hsa-GSL site 6	0.013	0.001	4.388	2.57	1.818	0	150	448.1	0.0	2.616	2.584	1.818	1.819	0.002	8.839
	CM.Mratio.12	100%	Low	25:75	Hsa-GSL site 6	0.030	0.001	4.388	2.57	1.364	0.454	150	336.2	33.8	2.616	3.038	1.364	1.365	0.456	8.839
	CM.Mratio.13	100%	Low	50:50	Hsa-GSL site 6	0.033	0.001	4.388	2.57	0.909	0.909	150	224.0	67.8	2.616	3.493	0.909	0.910	0.911	8.839
	CM.Mratio.14	100%	Low	75:25	Hsa-GSL site 6	0.041	0.001	4.388	2.57	0.454	1.364	150	111.9	101.7	2.616	3.948	0.454	0.455	1.366	8.839
	CM.Mratio.15	100%	Low	100:0	Hsa-GSL site 6	0.051	0.001	4.388	2.57	0	1.818	150	0.0	135.5	2.616	4.402	0.000	0.001	1.820	8.839
	CM.Mratio.1	100%	High	0:100	Hsa-GSL site 12	0.063	0.001	4.388	4.28	0.108	0	250	26.6	0.0	4.326	4.294	0.108	0.109	0.002	8.839
	CM.Mratio.2	100%	High	25:75	Hsa-GSL site 12	0.066	0.001	4.388	4.28	0.081	0.027	250	20.0	2.0	4.326	4.321	0.081	0.082	0.029	8.839
	CM.Mratio.3	100%	High	50:50	Hsa-GSL site 12	0.076	0.001	4.388	4.28	0.054	0.054	250	13.3	4.0	4.326	4.348	0.054	0.055	0.056	8.839
	CM.Mratio.4	100%	High	75:25	Hsa-GSL site 12	0.078	0.001	4.388	4.28	0.027	0.081	250	6.7	6.0	4.326	4.375	0.027	0.028	0.083	8.839
	CM.Mratio.5	100%	High	100:0	Hsa-GSL site 12	0.066	0.001	4.388	4.28	0	0.108	250	0.0	8.1	4.326	4.402	0.000	0.001	0.110	8.839
	CM.Mratio.6	100%	Medium	0:100	Hsa-GSL site 12	0.030	0.003	4.388	3.42	0.968	0	200	238.6	0.0	3.466	3.434	0.968	0.969	0.002	8.839
	CM.Mratio.7	100%	Medium	25:75	Hsa-GSL site 12	0.033	0.003	4.388	3.42	0.726	0.242	200	178.9	18.0	3.466	3.676	0.726	0.727	0.244	8.839
	CM.Mratio.8	100%	Medium	50:50	Hsa-GSL site 12	0.049	0.002	4.388	3.42	0.484	0.484	200	119.3	36.1	3.466	3.918	0.484	0.485	0.486	8.839
	CM.Mratio.9	100%	Medium	75:25	Hsa-GSL site 12	0.050	0.004	4.388	3.42	0.242	0.726	200	59.6	54.1	3.466	4.160	0.242	0.243	0.728	8.839
	CM.Mratio.10	100%	Medium	100:0	Hsa-GSL site 12	0.062	0.001	4.388	3.42	0	0.968	200	0.0	72.2	3.466	4.402	0.000	0.001	0.970	8.839
	CM.Mratio.11	100%	Low	0:100	Hsa-GSL site 12	0.011	0.001	4.388	2.57	1.818	0	150	448.1	0.0	2.616	2.584	1.818	1.819	0.002	8.839
	CM.Mratio.12	100%	Low	25:75	Hsa-GSL site 12	0.028	0.003	4.388	2.57	1.364	0.454	150	336.2	33.8	2.616	3.038	1.364	1.365	0.456	8.839
	CM.Mratio.13	100%	Low	50:50	Hsa-GSL site 12	0.035	0.003	4.388	2.57	0.909	0.909	150	224.0	67.8	2.616	3.493	0.909	0.910	0.911	8.839
	CM.Mratio.14	100%	Low	75:25	Hsa-GSL site 12	0.041	0.004	4.388	2.57	0.454	1.364	150	111.9	101.7	2.616	3.948	0.454	0.455	1.366	8.839
	CM.Mratio.15	100%	Low	100:0	Hsa-GSL site 12	0.051	0.001	4.388	2.57	0	1.818	150	0.0	135.5	2.616	4.402	0.000	0.001	1.820	8.839
	CM.Mratio.1	100%	High	0:100	Hsa-GSL site 14	0.063	0.002	4.388	4.28	0.108	0	250	26.6	0.0	4.326	4.294	0.108	0.109	0.002	8.839
	CM.Mratio.2	100%	High	25:75	Hsa-GSL site 14	0.066	0.002	4.388	4.28	0.081	0.027	250	20.0	2.0	4.326	4.321	0.081	0.082	0.029	8.839
	CM.Mratio.3	100%	High	50:50	Hsa-GSL site 14	0.078	0.001	4.388	4.28	0.054	0.054	250	13.3	4.0	4.326	4.348	0.054	0.055	0.056	8.839
	CM.Mratio.4	100%	High	75:25	Hsa-GSL site 14	0.081	0.001	4.388	4.28	0.027	0.081	250	6.7	6.0	4.326	4.375	0.027	0.028	0.083	8.839
	CM.Mratio.5	100%	High	100:0	Hsa-GSL site 14	0.066	0.001	4.388	4.28	0	0.108	250	0.0	8.1	4.326	4.402	0.000	0.001	0.110	8.839
	CM.Mratio.6	100%	Medium	0:100	Hsa-GSL site 14	0.038	0.001	4.388	3.42	0.968	0	200	238.6	0.0	3.466	3.434	0.968	0.969	0.002	8.839
	CM.Mratio.7	100%	Medium	25:75	Hsa-GSL site 14	0.037	0.000	4.388	3.42	0.726	0.242	200	178.9	18.0	3.466	3.676	0.726	0.727	0.244	8.839
	CM.Mratio.8	100%	Medium	50:50	Hsa-GSL site 14	0.055	0.002	4.388	3.42	0.484	0.484	200	119.3	36.1	3.466	3.918	0.484	0.485	0.486	8.839
	CM.Mratio.9	100%	Medium	75:25	Hsa-GSL site 14	0.055	0.001	4.388	3.42	0.242	0.726	200	59.6	54.1	3.466	4.160	0.242	0.243	0.728	8.839
	CM.Mratio.10	100%	Medium	100:0	Hsa-GSL site 14	0.062	0.001	4.388	3.42	0	0.968	200	0.0	72.2	3.466	4.402	0.000	0.001	0.970	8.839
	CM.Mratio.11	100%	Low	0:100	Hsa-GSL site 14	0.014	0.002	4.388	2.57	1.818	0	150	448.1	0.0	2.616	2.584	1.818	1.819	0.002	8.839
	CM.Mratio.12	100%	Low	25:75	Hsa-GSL site 14	0.033	0.001	4.388	2.57	1.364	0.454	150	336.2	33.8	2.616	3.038	1.364	1.365	0.456	8.839
	CM.Mratio.13	100%	Low	50:50	Hsa-GSL site 14	0.037	0.001	4.388	2.57	0.909	0.909	150	224.0	67.8	2.616	3.493	0.909	0.910	0.911	8.839
	CM.Mratio.14	100%	Low	75:25	Hsa-GSL site 14	0.045	0.001	4.388	2.57	0.454	1.364	150	111.9	101.7	2.616	3.948	0.454	0.455	1.366	8.839
	CM.Mratio.15	100%	Low	100:0	Hsa-GSL site 14	0.051	0.001	4.388	2.57	0	1.818	150	0.0	135.5	2.616	4.402	0.000	0.001	1.820	8.839

Table 5-1, continued: Composition and maximum growth rates in three media series

	Media composition name	salinity dilution series	[NaCl] level	K:Mg (M ratio)[3]	Strain	Mean max growth rate (hr^{-1})[3]	SD max growth rate[3]	Total Salinity (M salts)[1]	NaCl (M)[1]	$MgSO_4$ (M)[1]	KCl (M)[1]	NaCl (g/L)	$MgSO_4$*7H2O (g/L)	KCl (g/L)	Na (M)[2]	Cl (M)[2]	Mg (M)[2]	SO_4 (M)[2]	K (M)[2]	Total Salinity (M ions)[2]
Total salinity series																				
	CM.Mratio.2	100%	High	25:75	Hsa-lab	0.064	0.003	4.388	4.28	0.081	0.027	250	20.0	2.0	4.326	4.321	0.081	0.082	0.029	8.839
	CM.Mratio.2	90%	High	25:75	Hsa-lab	0.077	0.005	3.949	3.85	0.073	0.024	225	18.0	1.8	3.896	3.894	0.073	0.074	0.026	7.963
	CM.Mratio.2	80%	High	25:75	Hsa-lab	0.076	0.004	3.510	3.42	0.065	0.022	200	16.0	1.6	3.470	3.460	0.065	0.066	0.024	7.084
	CM.Mratio.2	70%	High	25:75	Hsa-lab	0.065	0.003	3.072	3.00	0.057	0.019	175	14.0	1.4	3.042	3.029	0.057	0.058	0.021	6.207
	CM.Mratio.2	60%	High	25:75	Hsa-lab	0.038	0.002	2.633	2.57	0.049	0.016	150	12.0	1.2	2.614	2.598	0.049	0.050	0.018	5.329
	CM.Mratio.2	100%	High	25:75	Hsa-GSL site 6	0.072	0.003	4.388	4.28	0.081	0.027	250	20.0	2.0	4.326	4.321	0.081	0.082	0.029	8.839
	CM.Mratio.2	90%	High	25:75	Hsa-GSL site 6	0.080	0.004	3.949	3.85	0.073	0.024	225	18.0	1.8	3.896	3.894	0.073	0.074	0.026	7.963
	CM.Mratio.2	80%	High	25:75	Hsa-GSL site 6	0.079	0.003	3.510	3.42	0.065	0.022	200	16.0	1.6	3.470	3.460	0.065	0.066	0.024	7.084
	CM.Mratio.2	70%	High	25:75	Hsa-GSL site 6	0.067	0.003	3.072	3.00	0.057	0.019	175	14.0	1.4	3.042	3.029	0.057	0.058	0.021	6.207
	CM.Mratio.2	60%	High	25:75	Hsa-GSL site 6	0.041	0.002	2.633	2.57	0.049	0.016	150	12.0	1.2	2.614	2.598	0.049	0.050	0.018	5.329
	CM.Mratio.2	100%	High	25:75	Hsa-GSL site 12	0.067	0.003	4.388	4.28	0.081	0.027	250	20.0	2.0	4.326	4.321	0.081	0.082	0.029	8.839
	CM.Mratio.2	90%	High	25:75	Hsa-GSL site 12	0.075	0.003	3.949	3.85	0.073	0.024	225	18.0	1.8	3.896	3.894	0.073	0.074	0.026	7.963
	CM.Mratio.2	80%	High	25:75	Hsa-GSL site 12	0.072	0.003	3.510	3.42	0.065	0.022	200	16.0	1.6	3.470	3.460	0.065	0.066	0.024	7.084
	CM.Mratio.2	70%	High	25:75	Hsa-GSL site 12	0.061	0.004	3.072	3.00	0.057	0.019	175	14.0	1.4	3.042	3.029	0.057	0.058	0.021	6.207
	CM.Mratio.2	60%	High	25:75	Hsa-GSL site 12	0.040	0.002	2.633	2.57	0.049	0.016	150	12.0	1.2	2.614	2.598	0.049	0.050	0.018	5.329
	CM.Mratio.2	100%	High	25:75	Hsa-GSL site 14	0.071	0.003	4.388	4.28	0.081	0.027	250	20.0	2.0	4.326	4.321	0.081	0.082	0.029	8.839
	CM.Mratio.2	90%	High	25:75	Hsa-GSL site 14	0.081	0.005	3.949	3.85	0.073	0.024	225	18.0	1.8	3.896	3.894	0.073	0.074	0.026	7.963
	CM.Mratio.2	80%	High	25:75	Hsa-GSL site 14	0.081	0.004	3.510	3.42	0.065	0.022	200	16.0	1.6	3.470	3.460	0.065	0.066	0.024	7.084
	CM.Mratio.2	70%	High	25:75	Hsa-GSL site 14	0.069	0.003	3.072	3.00	0.057	0.019	175	14.0	1.4	3.042	3.029	0.057	0.058	0.021	6.207
	CM.Mratio.2	60%	High	25:75	Hsa-GSL site 14	0.044	0.002	2.633	2.57	0.049	0.016	150	12.0	1.2	2.614	2.598	0.049	0.050	0.018	5.329
GSL imitation series																				
	CM.GSL2011.6	NA	NA	0.500	Hsa-lab	0.017	0.002	1.846	1.66	0.124	0.062	97	30.6	4.6	1.706	1.736	0.124	0.125	0.064	3.754
	CM.GSL2011.8	NA	NA	0.471	Hsa-lab	0.018	0.002	2.197	1.97	0.152	0.072	115	37.5	5.3	2.019	2.058	0.152	0.153	0.073	4.456
	CM.GSL2011.9	NA	NA	0.491	Hsa-lab	0.035	0.002	3.039	2.71	0.219	0.108	158	54.0	8.0	2.758	2.833	0.219	0.22	0.109	6.139
	CM.GSL2011.10	NA	NA	0.482	Hsa-lab	0.041	0.003	3.751	3.34	0.277	0.133	195	68.2	9.9	3.387	3.488	0.277	0.277	0.135	7.563
	CM.GSL2011.12	NA	NA	0.519	Hsa-lab	0.047	0.005	3.994	3.54	0.298	0.155	207	73.6	11.5	3.586	3.709	0.298	0.299	0.157	8.049
	CM.GSL2011.13	NA	NA	0.530	Hsa-lab	0.045	0.004	4.614	4.08	0.352	0.187	238	86.9	13.9	4.121	4.276	0.352	0.353	0.188	9.291
	CM.GSL2011.14	NA	NA	0.443	Hsa-lab	0.047	0.005	4.730	4.21	0.364	0.161	246	89.7	12.0	4.251	4.38	0.364	0.365	0.163	9.522
	CM.Mratio.2	100%	High	25:75	Hsa-lab	0.076	0.004	4.388	4.28	0.081	0.027	250	20.0	2.0	4.326	4.321	0.081	0.082	0.029	8.839
	CM.GSL2011.6	NA	NA	0.500	Hsa-GSL site 6	0.019	0.001	1.846	1.66	0.124	0.062	97	30.6	4.6	1.706	1.736	0.124	0.125	0.064	3.754
	CM.GSL2011.8	NA	NA	0.471	Hsa-GSL site 6	0.022	0.001	2.197	1.97	0.152	0.072	115	37.5	5.3	2.019	2.058	0.152	0.153	0.073	4.456
	CM.GSL2011.9	NA	NA	0.491	Hsa-GSL site 6	0.042	0.003	3.039	2.71	0.219	0.108	158	54.0	8.0	2.758	2.833	0.219	0.22	0.109	6.139
	CM.GSL2011.10	NA	NA	0.482	Hsa-GSL site 6	0.049	0.002	3.751	3.34	0.277	0.133	195	68.2	9.9	3.387	3.488	0.277	0.277	0.135	7.563
	CM.GSL2011.12	NA	NA	0.519	Hsa-GSL site 6	0.059	0.002	3.994	3.54	0.298	0.155	207	73.6	11.5	3.586	3.709	0.298	0.299	0.157	8.049
	CM.GSL2011.13	NA	NA	0.530	Hsa-GSL site 6	0.053	0.002	4.614	4.08	0.352	0.187	238	86.9	13.9	4.121	4.276	0.352	0.353	0.188	9.291
	CM.GSL2011.14	NA	NA	0.443	Hsa-GSL site 6	0.054	0.003	4.730	4.21	0.364	0.161	246	89.7	12.0	4.251	4.38	0.364	0.365	0.163	9.522
	CM.Mratio.2	100%	High	25:75	Hsa-GSL site 6	0.081	0.006	4.388	4.28	0.081	0.027	250	20.0	2.0	4.326	4.321	0.081	0.082	0.029	8.839
	CM.GSL2011.6	NA	NA	0.500	Hsa-GSL site 12	0.019	0.002	1.846	1.66	0.124	0.062	97	30.6	4.6	1.706	1.736	0.124	0.125	0.064	3.754
	CM.GSL2011.8	NA	NA	0.471	Hsa-GSL site 12	0.023	0.002	2.197	1.97	0.152	0.072	115	37.5	5.3	2.019	2.058	0.152	0.153	0.073	4.456
	CM.GSL2011.9	NA	NA	0.491	Hsa-GSL site 12	0.041	0.002	3.039	2.71	0.219	0.108	158	54.0	8.0	2.758	2.833	0.219	0.22	0.109	6.139
	CM.GSL2011.10	NA	NA	0.482	Hsa-GSL site 12	0.048	0.003	3.751	3.34	0.277	0.133	195	68.2	9.9	3.387	3.488	0.277	0.277	0.135	7.563
	CM.GSL2011.12	NA	NA	0.519	Hsa-GSL site 12	0.055	0.004	3.994	3.54	0.298	0.155	207	73.6	11.5	3.586	3.709	0.298	0.299	0.157	8.049
	CM.GSL2011.13	NA	NA	0.530	Hsa-GSL site 12	0.050	0.005	4.614	4.08	0.352	0.187	238	86.9	13.9	4.121	4.276	0.352	0.353	0.188	9.291
	CM.GSL2011.14	NA	NA	0.443	Hsa-GSL site 12	0.052	0.005	4.730	4.21	0.364	0.161	246	89.7	12.0	4.251	4.38	0.364	0.365	0.163	9.522
	CM.Mratio.2	100%	High	25:75	Hsa-GSL site 12	0.080	0.007	4.388	4.28	0.081	0.027	250	20.0	2.0	4.326	4.321	0.081	0.082	0.029	8.839
	CM.GSL2011.6	NA	NA	0.500	Hsa-GSL site 14	0.021	0.005	1.846	1.66	0.124	0.062	97	30.6	4.6	1.706	1.736	0.124	0.125	0.064	3.754
	CM.GSL2011.8	NA	NA	0.471	Hsa-GSL site 14	0.022	0.002	2.197	1.97	0.152	0.072	115	37.5	5.3	2.019	2.058	0.152	0.153	0.073	4.456
	CM.GSL2011.9	NA	NA	0.491	Hsa-GSL site 14	0.044	0.002	3.039	2.71	0.219	0.108	158	54.0	8.0	2.758	2.833	0.219	0.22	0.109	6.139
	CM.GSL2011.10	NA	NA	0.482	Hsa-GSL site 14	0.050	0.002	3.751	3.34	0.277	0.133	195	68.2	9.9	3.387	3.488	0.277	0.277	0.135	7.563
	CM.GSL2011.12	NA	NA	0.519	Hsa-GSL site 14	0.059	0.002	3.994	3.54	0.298	0.155	207	73.6	11.5	3.586	3.709	0.298	0.299	0.157	8.049
	CM.GSL2011.13	NA	NA	0.530	Hsa-GSL site 14	0.055	0.002	4.614	4.08	0.352	0.187	238	86.9	13.9	4.121	4.276	0.352	0.353	0.188	9.291
	CM.GSL2011.14	NA	NA	0.443	Hsa-GSL site 14	0.056	0.004	4.730	4.21	0.364	0.161	246	89.7	12.0	4.251	4.38	0.364	0.365	0.163	9.522
	CM.Mratio.2	100%	High	25:75	Hsa-GSL site 14	0.079	0.009	4.388	4.28	0.081	0.027	250	20.0	2.0	4.326	4.321	0.081	0.082	0.029	8.839

[1] Ratios and salt concentrations reflect added salts, and do not include contributions from peptone and citrate

[2] Molar ion concentrations include the contribution from organic media components (bacteriological peptone and sodium citrate); all media contain 10 and 3 g/L, respectively, regardless of dilution series

[3] mean and standard deviation of maximum growth rate (mu) from 6-8 biological replicates

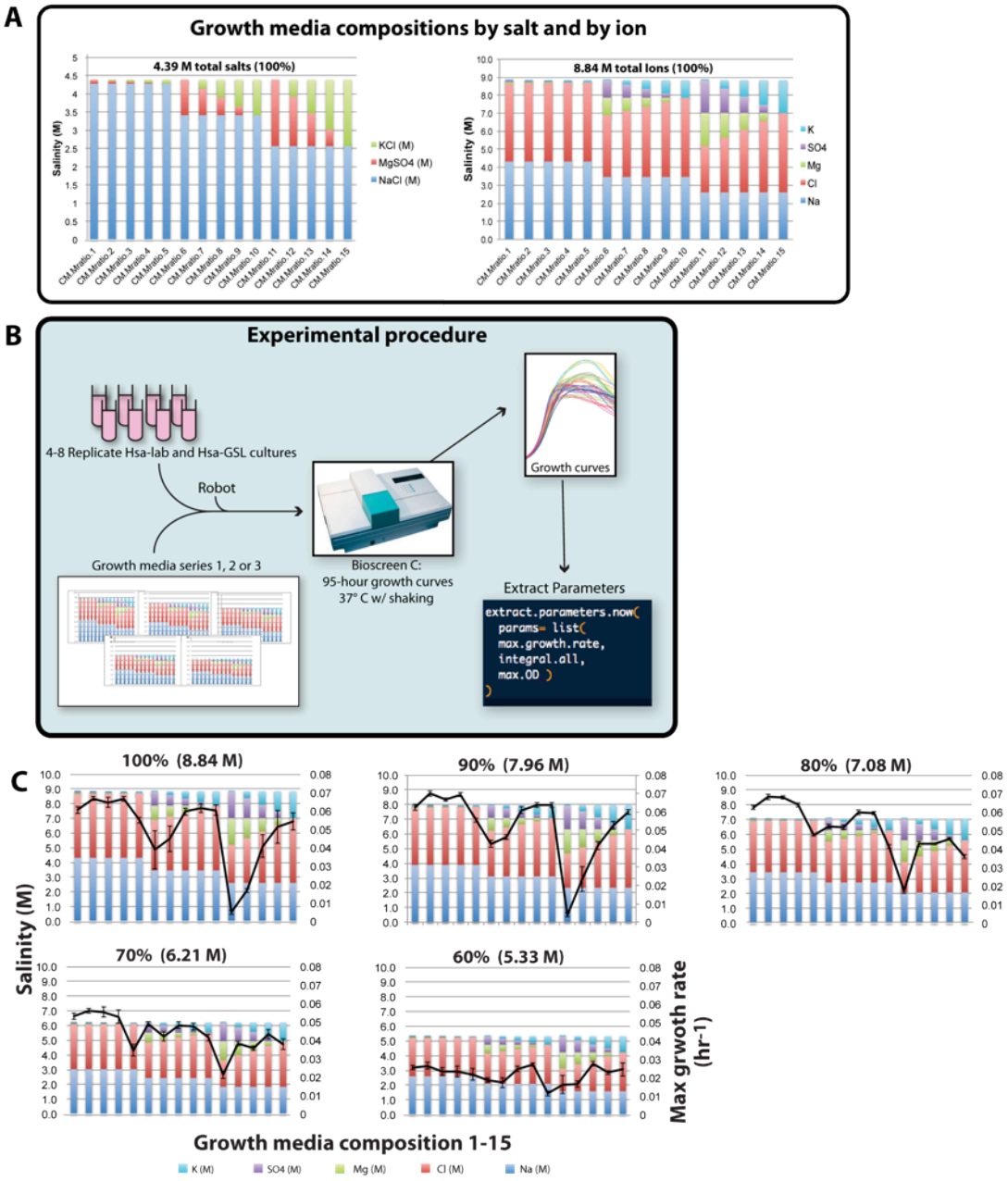

Figure 5-3: Growth media design and experimental pipeline.
A. Graphical representation of the ion composition media series, measured in terms of mol/L of added salts and mol/L individual ions. In this series, CM.Mratio2 media is the control media and all are considered 100% total salinity. The total salinity media series was produced by diluting these media down to 90, 80, 70 and 60% of total, using a stock containing 10 g/L oxoid peptone and 3 g/L sodium citrate to keep organic components constant in all media. GSL imitation media were formulated as described in supplementary methods.
B. For each of three media series, at least 4 replicate cultures were grown from colony, synchronized by passaging once to OD = 0.05 in control media (Media 2), and then distributed in each media in the series for growth in a Bioscreen C instrument. Optical density was measured every 15 min during growth at 37 C with shaking for 95 hours. Maximum growth rate was extracted from each growth curve using the GCAF R package developed in the Baliga Lab[129]. **C**. Reaction norms (black lines) of Hsa-lab in 75 growth media representing all combinations of ion composition and total salinity. All media in the 100% series, which varies only in ion composition, was diluted down to 90, 80, 70 and 60% total salinity in a solution of 10 g/L oxoid peptone and 3 g/L organic sodium citrate to maintain the concentration of rich media components and organic salts across all media. These media represent a full factorial design, testing the independent effects of total salinity, [NaCl]:[Total] ratio and [Mg]:[K] ratio on max growth rate. F tests show that each of these components of salinity contributes significantly to Hsa-lab maximum growth rate (data not shown).

Table 5-2: Analysis of max growth rate variation by media and strain

Media series:	Ion ratio		Total salinity		GSL imitation	
	F	*P*	F	*P*	F	*P*
Maximum growth rate						
Strain	115.9	<0.001	40.2	<0.001	56.2	<0.001
Media	2694.7	<0.001	899.9	<0.001	835.4	<0.001
Strain x Media	15.9	<0.001	2.2	0.013	1.6	0.061

[1] Shown is a two-way ANOVA with interaction, based on type-I sums of squares. Factor order was inconsequential.

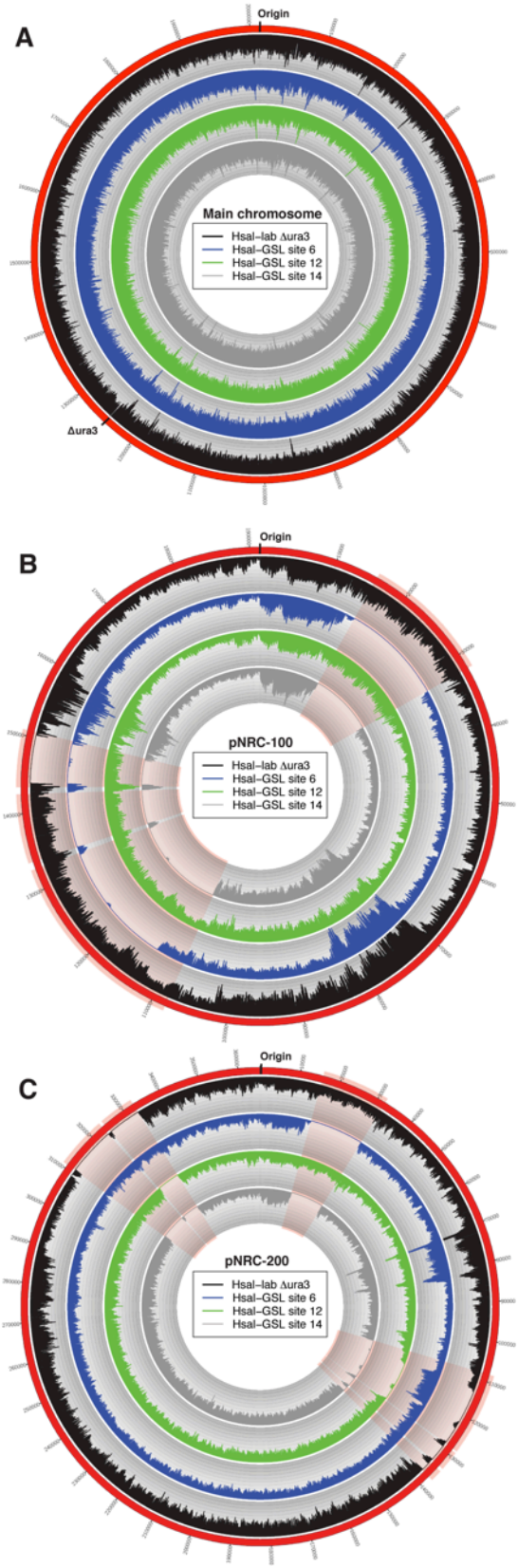

Figure 5-4: Whole genome sequence coverage maps
Depth-of-coverage maps for all aligned reads, in each of four strains, to the *H. salinarum* NRC-1 reference sequence (red outer band). **A.** Main chromosome (RefSeq: NC_002607), **B.** Minichromosome pNRC-100 (RefSeq: NC_001869), **C.** Minichromosome pNRC-200 (RefSeq: NC_002608). In all maps, from outside to inside, strains are: Hsa-lab, Hsa-GSL site 14, Hsa-GSL site 6, and Hsa-GSL site 12. Coverage is plotted in 60 bp windows from the start of each replicon, and the scale of each coverage track is as follows, with ranges given from outside to inside of each strain's coverage track. Main chromosome: 0-150x for all strains; pNRC-100: 0-250x for Δ*ura3*, and 0-400x for all other strains; pNRC-200: 0-250x for Δ*ura3*, and 0-400x for all other strains. The origins of replication on all replicons and the location of the *ura3* deletion on the main chromosome are labeled. Genome coordinates (in # of base pairs) are labeled along the outer edges. Orange-shaded areas indicate regions in which at least one of the four strains does not map to the reference; see table 5-4 for list of ORFs in these coverage-discordant regions.

Table 5-3: SNPs and InDels in Hsa-lab and Hsa-GSL strains[1,2]

							Present/Absent?			
RefSeq	Chromosome	Position (bp)	Mutation[3]	Annotation	Gene	Description	Hsa-lab	Hsa-GSL site 6	Hsa-GSL site 12	Hsa-GSL site 14
NC_002607	Main	30,405	+C	coding (764/813 nt)	VNG0035H ←	hypothetical protein	1	1	1	1
NC_002607	Main	30,417	+G	coding (752/813 nt)	VNG0035H ←	hypothetical protein	1	1	1	1
NC_002607	Main	30,427	+G	coding (742/813 nt)	VNG0035H ←	hypothetical protein	1	1	1	1
NC_002607	Main	30,520	G→T	H217N (CAC→AAC)	VNG0035H ←	hypothetical protein	1	1	1	1
NC_002607	Main	56,173	Δ1,413 bp	IS4-mediated	VNG0059H	VNG0059H	1	1	1	1
NC_002607	Main	425,424	Δ1 bp	coding (1172/1680 nt)	VNG0553C →	hypothetical protein	1	1	1	1
NC_002607	Main	460,879	Δ1 bp	intergenic (+151/-34)	VNG0604H → / → yrhA	hypothetical protein/cysteine synthase	1	1	1	1
NC_002607	Main	586,815	Δ1 bp	coding (1581/1815 nt)	VNG0779C →	hypothetical protein	1	1	1	1
NC_002607	Main	1,023,692	A→C	S208R (AGC→CGC)	kinA1 (VNG1374G) →	signal-transducing histidine kinase-like protein	1	1	1	1
NC_002607	Main	1,230,901	+C	intergenic (-202/-260)	VNG1651H ← / → VNG1653H	hypothetical protein/hypothetical protein	1	1	1	1
NC_002607	Main	749,021	IS4 (+) +9 bp :: +G	intergenic (+170/-363)	trn14 → / → VNG0985H	tRNA-Lys/hypothetical protein	0	1	1	1
NC_002607	Main	749,682	+ACG :: IS4 (–) +7	coding (291-297/2007 nt)	VNG0985H →	hypothetical protein	0	1	1	1
NC_002607	Main	1,125,686	IS4 (+) +10 bp	coding (9-18/2055 nt)	VNG1511C ←	hypothetical protein	0	1	1	1
NC_002607	Main	1,184,714	Δ1,074 bp	IS5-mediated	VNG1587H	VNG1587H	1	1	0	1
NC_002608	pNRC-200	309,347	IS4 (–) +4 bp :: +G	coding (406-409/411 nt)	VNG6393H ←	hypothetical protein	0	1	1	1
NC_002608	pNRC-200	122,251	Δ16,117 bp	between IS4	VNG6147H–[VNG6162H]	VNG6147H, VNG6148H, VNG6149H, orc1, VNG6152H, sojC2, VNG6155H, VNG6156H, VNG6157H, VNG6158H, VNG6159H, VNG6160H, VNG6161H, [VNG6162H]	0	0	1	1
NC_001869	pNRC-100	146,897	T→C	L179L (CTT→CTC)	VNG7133 →	hypothetical protein	0	1	0	0
NC_001869	pNRC-100	147,046	G→T	G229V (GGA→GTA)	VNG7133 →	hypothetical protein	0	1	0	0
NC_001869	pNRC-100	147,059	G→T	W233C (TGG→TGT)	VNG7133 →	hypothetical protein	0	1	0	0
NC_001869	pNRC-100	147,385	A→G	E342G (GAA→GGA)	VNG7133 →	hypothetical protein	0	1	0	0
NC_001869	pNRC-100	147,400	G→A	G347E (GGG→GAG)	VNG7133 →	hypothetical protein	0	1	0	0
NC_001869	pNRC-100	147,462	2 bp→AG	coding (1102-1103/1188 nt)	VNG7133 →	hypothetical protein	0	1	0	0
NC_002607	Main	99,328	Δ1,394 bp	IS4-mediated	yusZ1 → / ← VNG0116H	oxidoreductase/hypothetical protein	1	0	0	0
NC_002607	Main	241,900	A→C	H203P (CAC→CCC)	trpB →	tryptophan synthase subunit beta	1	0	0	0
NC_002607	Main	466,489	IS4 (+) +9 bp	coding (87-95/195 nt)	VNG0612H ←	hypothetical protein	0	1	0	0
NC_002607	Main	673,215	C→T	A347V (GCG→GTG)	gyrB (VNG0887) →	DNA gyrase subunit B	1	0	0	0
NC_002607	Main	750,860	+C :: IS4 (–) +11 bp	coding (1469-1479/2007 nt)	VNG0985H →	hypothetical protein	1	0	0	0
NC_002607	Main	755,304	IS4 (+) +9 bp :: +G	coding (170-178/1023 nt)	VNG0989C ←	hypothetical protein	1	0	0	0
NC_002607	Main	771,564	Δ531 bp		VNG1007H	VNG1007H	1	0	0	0
NC_002607	Main	807,229	IS4 (–) +10 bp	coding (458-467/1197 nt)	VNG1059C ←	hypothetical protein	1	0	0	0
NC_002607	Main	949,727	G→A	G89E (GGG→GAG)	VNG1260C →	hypothetical protein	0	0	1	0
NC_002607	Main	1,197,520	IS4 (+) +9 bp	coding (91-99/591 nt)	cheC2 ←	hypothetical protein	1	0	0	0
NC_002607	Main	1,245,981	Δ1,338 bp		[VNG1672H]–[VNG1674H]	[VNG1672H], pyrF, [VNG1674H]	1	0	0	0
NC_002607	Main	1,855,793	Δ2 bp	coding (839-840/1119 nt)	VNG2480Hm ←	hypothetical protein	0	0	0	1
NC_002608	pNRC-200	108,320	+C :: IS4 (–) +4 bp	coding (1252-1255/1986 nt)	VNG6136H →	hypothetical protein	0	1	0	0
NC_002608	pNRC-200	127,665	+G :: IS5 (–) +7 bp	coding (154-160/2529 nt)	VNG6149H ←	hypothetical protein	0	1	0	0
NC_002608	pNRC-200	142,444	IS4 (–) +9 bp	coding (123-131/252 nt)	VNG6168H ←	hypothetical protein	0	0	0	1
NC_002608	pNRC-200	218,403	G→A	P59S (CCC→TCC)	VNG6283H ←	hypothetical protein	1	0	0	0

[1]Ordered from most common to most unique SNPs/Indels among strains, and then by genome coordinate

[2]Includes SNPs and Indels corroborated by Breseq, BGI and verified by hand in IGV genome viewer.

[3]Mutation descriptions per Breseq manual, see Supplement.

* Indicates locus was used as a diagnostic to rule out contamination of hsa-GSL strains by laboratory strains.

Table 5-4: ORFs in regions of coverage discordance on minichromosomes[1,2]

pNRC-100 discordant reading frames

Contig #	ORF start	ORF end	strand	geneID	annotation[3]	Present/Absent? Hsa-lab	Hsa-GSL site 6	Hsa-GSL site 12	Hsa-GSL site 14
NP_045958.1	15743	16351	+	VNG7013	ORF H0211	1	0	1	0
NP_045959.1	16107	16631	+	VNG7014	ORF 0215	1	0	1	0
NP_045960.1	16705	16451	-	gvpM	GvpM	1	0	1	0
NP_045961.1	17550	16705	-	gvpL	GvpL	1	0	1	0
NP_045962.1	17888	17547	-	gvpK	GvpK	1	0	1	0
NP_045963.1	18234	17890	-	gvpJ	GvpJ	1	0	1	0
NP_045964.1	18661	18227	-	gvpI	GvpI	1	0	1	0
NP_045965.1	19206	18658	-	gvpH	GvpH	1	0	1	0
NP_045966.1	19457	19206	-	gvpG	GvpG	1	0	1	0
NP_045967.1	20102	19461	-	gvpF	GvpF	1	0	1	0
NP_045968.1	20670	20095	-	gvpE	GvpE	1	0	1	0
NP_045969.1	22283	20673	-	gvpD	GvpD	1	0	1	0
NP_045970.1	22484	22714	+	gvpA	GvpA	1	0	1	0
NP_045971.1	22804	23952	+	gvpC	GvpC	1	0	1	0
NP_045972.1	23949	24992	+	gvpN	GvpN	1	0	1	0
NP_045973.1	25017	25376	+	gvpO	GvpO	1	0	1	0
NP_045974.1	25405	26169	+	sojB	sojB	1	0	1	0
NP_045975.1	29436	26683	-	VNG7030	ORF H0337	1	0	1	0
NP_045976.1	30044	30346	+	VNG7031	tbpA Similar to Pyrococcus woesei TATA-box binding proteins; putative TATA-box binding protein	1	0	1	0
NP_045977.1	31896	30622	-	VNG7032	ORF H0386	1	0	1	0
NP_046042.1	107069	109054	+	VNG7097	ORF H1161	1	1	1	0
NP_046043.1	109746	108991	-	VNG7098	ORF H1185	1	1	1	0
NP_046044.1	109051	111990	+	VNG7099	ORF H1186	1	1	1	0
NP_046045.1	111987	112532	+	tbpC	tbpC	1	1	1	0
NP_046045.1	111987	112532	+	tbpC	tbpC	1	0	1	0
NP_046046.1	112973	114127	+	VNG7101	ORF H1224	1	0	1	0
NP_046047.1	114461	114742	+	VNG7102	ORF H1232	1	0	1	0
NP_046048.1	115512	115111	-	VNG7103	ORF H1238	1	0	1	0
NP_046049.1	115884	116498	+	VNG7104	ORF H1243	1	0	1	0
NP_046050.1	117104	120490	+	repJ	repJ	1	0	1	0
NP_046051.1	121064	122392	+	VNG7106	ORF H1292	1	0	1	0
NP_046052.1	123641	122757	-	VNG7107	putative ISH4 transposase	1	0	1	0
NP_046053.1	124238	124582	+	VNG7108	ORF H1313	1	0	1	0
NP_046054.1	124579	125280	+	VNG7109	ORF H1320	1	0	1	0
NP_046055.1	125267	126313	+	VNG7110	ORF H1325	1	0	1	0
NP_046056.1	126334	126504	+	VNG7111	ORF H1332	1	0	1	0
NP_046057.1	126536	126928	+	VNG7112	ORF H1336	1	0	1	0
NP_046058.1	128181	128840	+	VNG7113	ORF H1354	1	0	1	0
NP_046059.1	128906	128346	-	tbpD	tbpD	1	0	1	0
NP_046060.1	129183	129794	+	VNG7115	ORF H1364	1	0	1	0
NP_046061.1	130026	130631	+	VNG7116	ORF H1373	1	0	1	0
NP_046062.1	130628	130840	+	VNG7117	ORF H1378	1	0	1	0
NP_046063.1	131677	132357	+	VNG7118	ORF H1391	1	0	1	0
NP_046064.1	133628	132354	-	VNG7119	putative ISH8 transposase	1	0	1	0
NP_046065.1	134055	135578	+	VNG7120	ORF H1420	1	0	1	0
NP_046066.1	137037	135862	-	VNG7121	ORF H1434	1	0	1	0
NP_046067.1	137059	136700	-	VNG7122	ORF H1442	1	0	1	0
NP_046068.1	139724	137790	-	arsA	arsA	1	0	1	0
NP_046069.1	140051	139725	-	arsD	arsD	1	0	1	0
NP_046070.1	140392	140781	+	arsR	arsR	1	0	1	0
NP_046071.1	140884	141312	+	VNG7126	arsC	1	0	1	0
NP_046072.1	141871	142311	+	VNG7127	ORF H1484	1	0	1	0
NP_046073.1	145108	143915	-	VNG7128	ORF H1502	0	0	1	0
NP_046074.1	145659	145165	-	VNG7129	ORF H1511	0	0	1	0
NP_046075.1	145195	145497	+	VNG7130	ORF H1513	0	0	1	0
NP_046076.1	145498	146010	+	VNG7131	ORF H1517	0	0	1	0
NP_046077.1	146042	146371	+	VNG7132	ORF H1522	0	0	1	0
NP_046078.1	146361	147548	+	VNG7133	ORF H1525	0	0	1	0
NP_046079.1	147005	147997	+	VNG7134	ORF H1529	0	0	1	0
NP_046080.1	148190	148636	+	VNG7135	ORF H1537	0	0	1	0
NP_046081.1	149110	150270	+	VNG7136	ORF H1549	0	0	1	0

Table 5-4, cont'd: ORFs in regions of coverage discordance on minichromosomes[1,2]

pNRC-200 discordant reading frames

Contig #	ORF start	ORF end	strand	geneID	annotation	Present/Absent? Hsa-lab	Hsa-GSL site 6	Hsa-GSL site 12	Hsa-GSL site 14
NP_395580.1	15743	16351	+	VNG6017H	Vng6017h	0	0	1	0
NP_395581.1	16705	16451	-	gvpM1	GvpM protein, cluster A	0	0	1	0
NP_395582.1	17550	16705	-	gvpL1	GvpL protein, cluster A	0	0	1	0
NP_395583.1	17888	17547	-	gvpK1	GvpK protein, cluster A	0	0	1	0
NP_395584.1	18234	17890	-	gvpJ1	GvpJ protein, cluster A	0	0	1	0
NP_395585.1	18661	18227	-	gvpI1	GvpI protein, cluster A	0	0	1	0
NP_395586.1	19206	18658	-	gvpH1	GvpH protein, cluster A	0	0	1	0
NP_395587.1	19457	19206	-	gvpG1	GvpG protein, cluster A	0	0	1	0
NP_395588.1	20102	19461	-	gvpF1	GvpF protein, cluster A	0	0	1	0
NP_395589.1	20670	20095	-	gvpE1	GvpE protein, cluster A	0	0	1	0
NP_395590.1	22283	20673	-	gvpD1	GvpD protein, cluster A	0	0	1	0
NP_395591.1	22484	22714	+	gvpA1	GvpA protein, cluster A	0	0	1	0
NP_395592.1	22804	23952	+	gvpC1	GvpC protein, cluster A	0	0	1	0
NP_395593.1	23949	24992	+	gvpN1	GvpN protein, cluster A	0	0	1	0
NP_395594.1	25017	25376	+	gvpO1	GvpO protein, cluster A	0	0	1	0
NP_395595.1	25405	26169	+	sojB	Spo0A activation inhibitor	0	0	1	0
NP_395596.1	29436	26683	-	htlD	Htr-like protein	0	0	1	0
NP_395597.1	29637	29789	+	VNG6036H	Vng6036h	0	0	1	0
NP_395598.1	30131	30346	+	tbpA	transcription initiation factor IID	0	0	1	0
NP_395599.1	31896	30622	-	VNG6038H	Vng6038h	0	0	1	0
NP_395676.1	107069	109054	+	VNG6136H	Vng6136h	0	1	0	0
NP_395677.1	109051	111990	+	hepA	ATP-dependent RNA helicase	0	1	0	0
NP_395678.1	111987	112532	+	tbpC	transcription initiation factor IID	0	1	0	0
NP_395679.1	113157	112954	-	VNG6141H	Vng6141h	0	1	0	0
NP_395680.1	114211	113327	-	VNG6142C	Vng6142c	0	1	0	0
NP_395681.1	118026	114208	-	VNG6143H	Vng6143h	0	1	0	0
NP_395682.1	120253	118019	-	trsE	transfer complex protein	0	1	0	0
NP_395683.1	121358	120237	-	VNG6145H	Vng6145h	0	1	0	0
NP_395684.1	122090	121887	-	VNG6146H	Vng6146h	0	1	0	0
NP_395685.1	122374	123648	+	VNG6147H	Vng6147h	0	1	0	0
NP_395686.1	124873	123755	-	VNG6148H	Vng6148h	0	1	0	0
NP_395687.1	127824	125296	-	VNG6149H	Vng6149h	0	1	0	0
NP_395688.1	129403	130329	+	orc1	orc / cell division control protein 6	0	1	1	0
NP_395689.1	130295	130819	+	VNG6152H	Vng6152h	0	1	1	0
NP_395690.1	130856	131923	+	sojC2	Spo0A activation inhibitor	0	1	1	0
NP_395691.1	132066	132377	+	VNG6155H	Vng6155h	0	1	1	0
NP_395692.1	132832	133239	+	VNG6156H	Vng6156h	0	1	1	0
NP_395693.1	135987	135571	-	VNG6157H	Vng6157h	0	1	1	0
NP_395694.1	136285	136124	-	VNG6158H	Vng6158h	0	1	1	0
NP_395695.1	136724	136593	-	VNG6159H	Vng6159h	0	1	1	0
NP_395696.1	136952	136761	-	VNG6160H	Vng6160h	0	1	1	0
NP_395697.1	137079	138353	+	VNG6161H	Vng6161h	0	1	1	0
NP_395871.1	311100	309826	-	VNG6395H	Vng6395h	0	1	1	0
NP_395872.1	311246	311617	+	VNG6396H	Vng6396h	0	1	1	0
NP_395873.1	311629	312180	+	VNG6397H	Vng6397h	0	1	1	0
NP_395874.1	313047	312226	-	VNG6400H	Vng6400h	0	1	1	0
NP_395875.1	313502	313038	-	VNG6401H	Vng6401h	0	1	1	0
NP_395876.1	314011	313499	-	VNG6402H	Vng6402h	0	1	1	0
NP_395877.1	315084	314185	-	VNG6403H	Vng6403h	0	1	1	0
NP_395878.1	315651	315154	-	VNG6404H	Vng6404h	0	1	1	0
NP_395879.1	316931	315786	-	VNG6406H	Vng6406h	0	1	1	0
NP_395880.1	317128	316952	-	VNG6407H	Vng6407h	0	1	1	0
NP_395881.1	317427	318326	+	phzF	phenazine biosynthetic protein	0	1	1	0
NP_395882.1	319002	318817	-	VNG6409H	Vng6409h	0	1	1	0
NP_395883.1	319015	319578	+	VNG6411H	Vng6411h	0	1	1	0
NP_395884.1	319817	319575	-	VNG6412H	Vng6412h	0	1	1	0
NP_395885.1	320476	319865	-	VNG6413H	Vng6413h	0	1	1	0
NP_395886.1	321673	320684	-	VNG6416H	Vng6416h	0	1	1	0
NP_395887.1	322067	321729	-	VNG6418H	Vng6418h	0	1	1	0
NP_395888.1	322357	322067	-	VNG6419H	Vng6419h	0	1	1	0
NP_395889.1	323033	322488	-	VNG6420H	Vng6420h	0	1	1	0
NP_395890.1	324345	323341	-	VNG6421H	Vng6421h	0	1	1	0
NP_395890.1	324345	323341	-	VNG6421H	Vng6421h	0	1	0	0
NP_395891.1	325516	325010	-	VNG6424H	Vng6424h	0	1	0	0
NP_395892.1	326050	325607	-	VNG6427H	Vng6427h	0	1	0	0
NP_395893.1	326877	326500	-	VNG6429H	Vng6429h	0	1	0	0
NP_395894.1	327051	327602	+	VNG6430C	Vng6430c	0	1	0	0
NP_395895.1	327929	327699	-	VNG6431H	Vng6431h	0	1	0	0
NP_395896.1	328134	328547	+	VNG6432H	Vng6432h	0	1	0	0
NP_395897.1	328989	329141	+	VNG6434H	Vng6434h	0	1	0	0
NP_395898.1	329196	329609	+	VNG6437C	Vng6437c	0	1	0	0
NP_395899.1	329788	330348	+	tbpF	transcription initiation factor IID	0	1	0	0
NP_395900.1	331546	330770	-	VNG6439H	Vng6439h	0	1	0	0
NP_395901.1	331672	332220	+	VNG6441H	Vng6441h	0	1	0	0
NP_395902.1	334165	332801	-	VNG6442H	Vng6442h	0	1	0	0

[1]Listed in clockwise order of orange-shaded regions in Figure 5-4B and C, starting at the Origin

[2]Horizontal section dividers denote region boundaries for at least one strain

[3]Annotations from NCBI .ptt file and Microbesonline.org

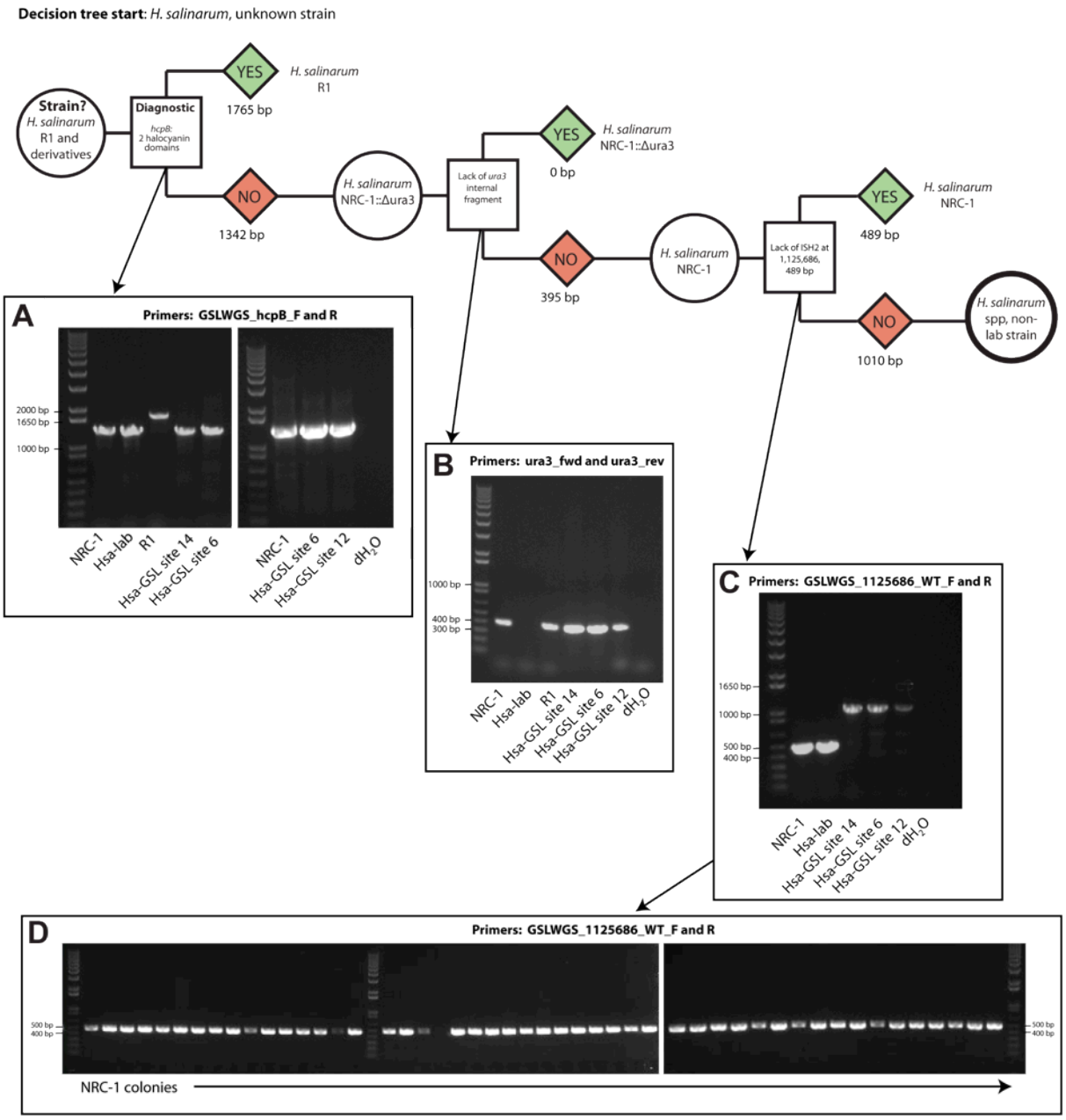

Figure 5-5: Decision tree and gel evidence for lack of contamination among strains
To sequentially rule out the possibility of contamination of hsa-GSL strains by laboratory *H. salinarum,* we performed diagnostic PCRs according to a decision tree. **A.** Beginning at the upper left, we assume an unknown *H. salinarum* strain and ask first whether its genome contains a sequence specific to the R1 strain. **B.** Upon confirming that no strains possess the R1 diagnostic except for the R1 positive control, we then asked whether the strain carried the diagnostic *ura3* deletion of the *H. salinarum* NRC-1 Δ*ura3* (Hsa-lab) strain. **C.** Following this test, we used the presence of a newly-inserted ISH2 element in the Hsa-GSL strains (per whole genome sequencing data, Table 5-3) as a diagnostic marker of a non-NRC-1 laboratory strain. **D.** Because ISH elements present in the genome may move laterally to new locations even in laboratory strains, we checked the frequency of this ISH insertion in a group of 48 NRC-1 colonies. None of these colonies produced a band matching the GSL strains; 47 of 48 produced bands identical to the NRC-1 band in **C**, and one produced no band at all.

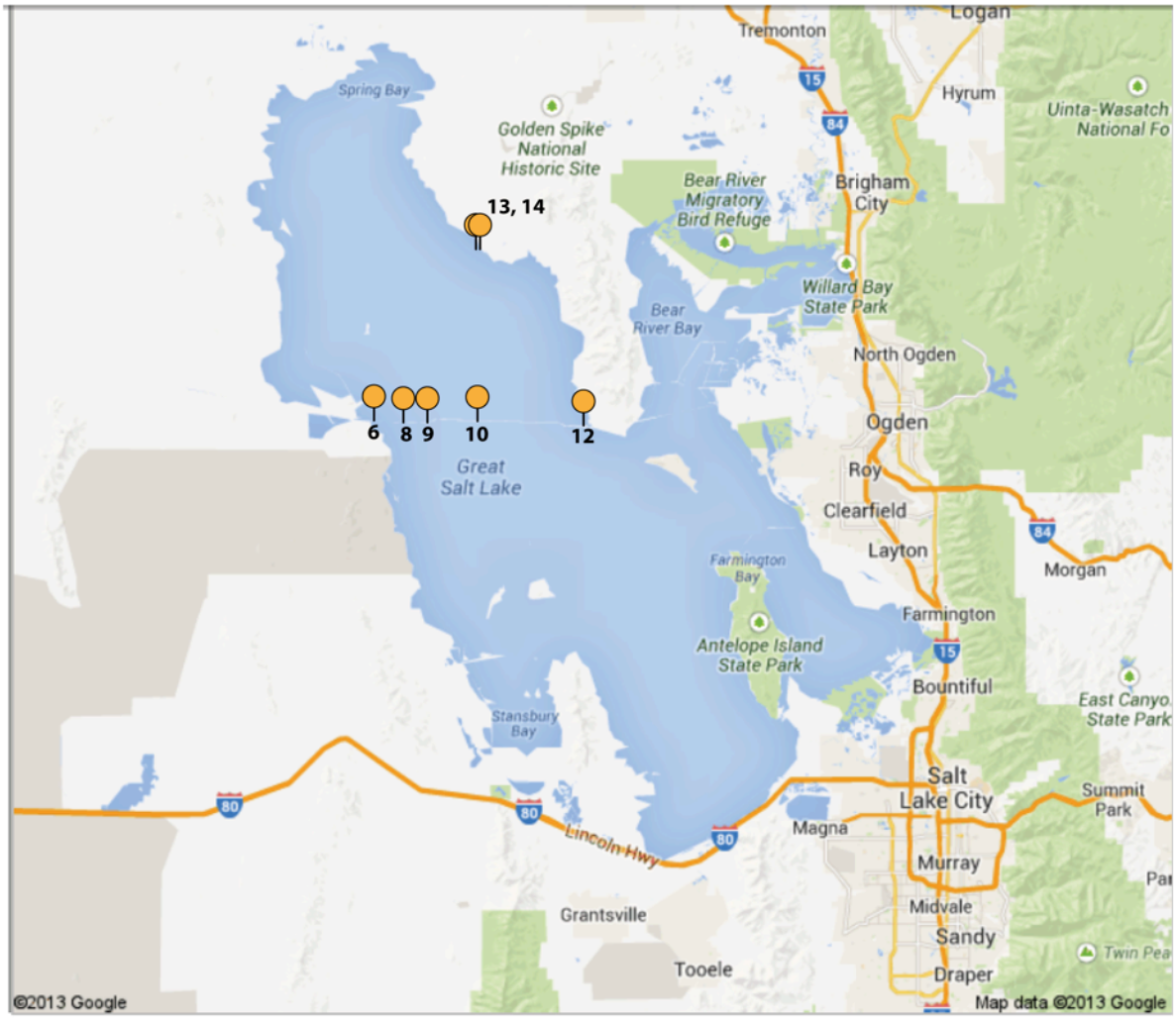

Figure 5-6: Map of the Great Salt Lake and seven sampling sites.
Samples were taken at seven sites in the north arm and along the railroad causeway that crosses the GSL. Sites were selected to capture the chemical and ecological diversity of the lake. Site 6 was collected in the breach that allows water from the north and less-saline south arms to mix, and sites 13 and 14 were taken from sites within and around the Spiral Jetty, a spiral-shaped basalt rock formation in which water concentrates and mixes to varying degrees. GPS locations and metadata for each site are listed in Table 5-5.

Table 5-5: Great Salt Lake sample site locations and water properties

Sample ID	Latitute	Longitude	Sampling date	Sampling time (24h)	Water depth (m)	Water temperature (°C)	pH	Specific conductivity (µS.cm^{-1} at 25 °C)	PSU	DO (mg/L)	PAR (µE/m^2/s^1)	UVA (W/m^2)	UVB (µW/m^{-2})	Site description
2011-GSL06	41°13.320'	112°50.860'	3-May-11	13:30	0.21	12.7	8.26	120700	91.18	9.44	683.1	15.2	7.7	South causeway at Lakeside breach (far W side of causeway)
2011-GSL08	41°13.288'	112°47.791'	3-May-11	14:45	0.24	13.6	8.13	164200	134.46	8.98	ND[1]	ND[1]	ND[1]	North causeway, east of Lakeside breach, west of "west crack," where causeway bends slightly.
2011-GSL09	41°13.320'	112°45.390'	3-May-11	15:10	0.25	13.0	7.90	197600	172.18	8.77	ND[1]	ND[1]	ND[1]	North causeway on "west crack"
2011-GSL10	41°13.399'	112°40.108'	3-May-11	15:30	0.17	12.7	7.72	201767	177.18	8.78	ND[1]	ND[1]	ND[1]	North causeway, estimated longitudinal center
2011-GSL12	41°12.980'	112°30.385'	3-May-11	16:40	0.49	12.6	7.61	221500	201.84	8.29	ND[1]	ND[1]	ND[1]	North causeway btwn "east crack" & promontory point
2011-GSL13	41°26.284'	112°40.100'	4-May-11	19:00	0.26	19.8	7.24	227200	209.27	5.12	530	4.24	0.67	Spiral Jetty: 5 m out from south side of jetty, about halfway between shore and spiral. Site used for timecourse, data here are T=0 hrs.
2011-GSL14	41°26.267'	112°40.133'	5-May-11	11:30	0.265	15.12	7.35	223800	204.82	8.12	ND[2]	ND[2]	ND[2]	Spiral Jetty: center of spiral

[1] PAR and UV data not collected due to inclement weather (rain storm).

[2] PAR and UV data not collected due to risk of instrument submersion. Collection required wading in deep water.

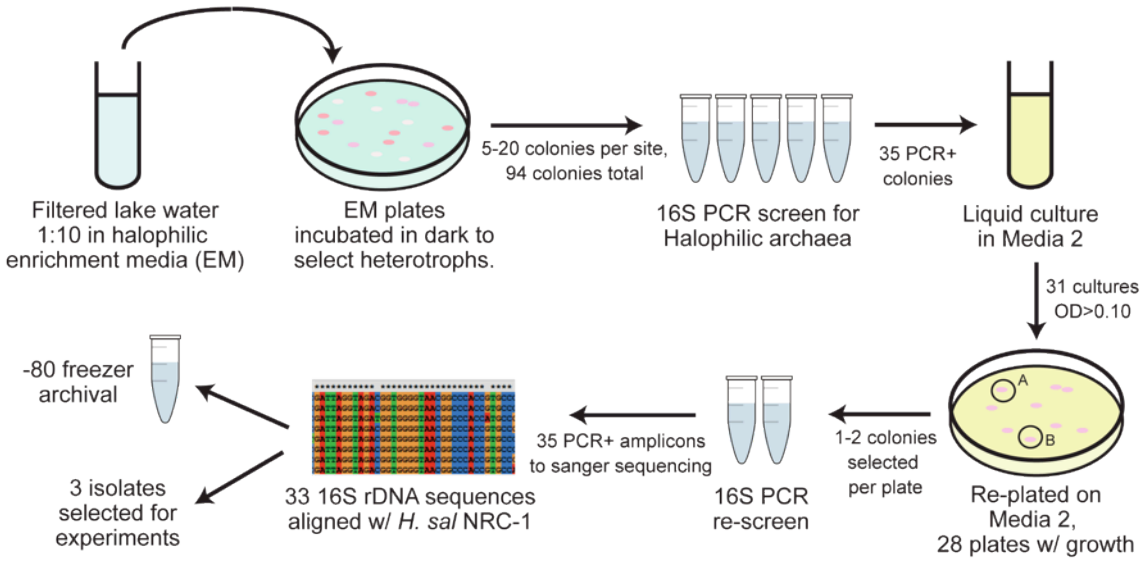

Figure 5-7: Great Salt Lake strain isolation overview.
Lake water samples from each site were 20 µm-filtered, and processed identically, as described in Supporting Materials. Briefly, filtered samples were inoculated 1:10 into halophilc enrichment media containing F/2 algal media amended with ~3.5 M total salts and casamino acids or pyruvate. Enrichments were plated onto the same media made with agar, and incubated in the dark to select for heterotrophs. Roughly equal numbers of colonies from each site were PCR-screened at the 16S rDNA locus for membership in the halophilic archaea. Positively screened colonies were passaged and re-plated to ensure isolated colonies and pure cultures. These new colonies were re-screened using the same 16S PCR, and the colonies that screened positive a second time were frozen in glycerol at -80°C and Sanger sequenced. Three colonies that came from different, chemically diverse GSL sites, *and* had identical 16S sequences, were selected for further experimentation (Sites 6, 12 and 14). See appendix for 16S alignments.

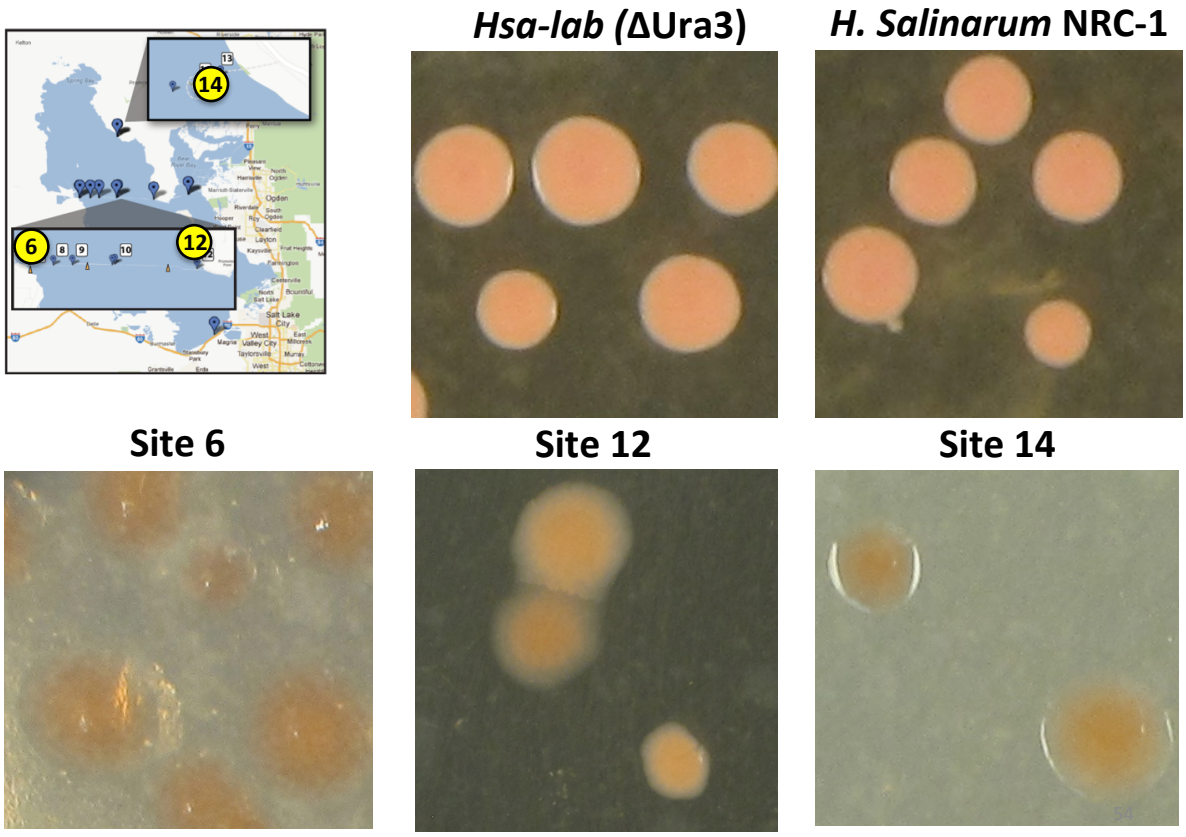

Figure 5-8: Colony phenotypic varability among Hsa-lab and Hsa-GSL strains.
Sites 6, 12 and 14 are indicated on the inset map, and correspond with labeled images. Opacity of *H. salinarum* colonies is attributable to the presence of gas vesicles (gvp^+ phenotype), used to modulate water column depth[181]. Both Hsa-lab and its parent strain *H. salinarum* NRC-1, are gvp^+ and shown for comparison.

Chapter 6: Conclusions and future directions

Novel characterizations of haloarchaeal osmotolerance

This dissertation makes a novel contribution to our collective understanding of archaeal osmotolerance in hypersaline environments. Most importantly, this work is a thorough evaluation of the utility of an old laboratory strain for ongoing investigation of environmental stress response. I began with a broad review of microbial adaptation to new environments, and a discussion of plasticity and niche width as markers of adaptation. The process of quantifying adaptation is a complex challenge itself, and multiple perspectives and data types – a systems approach – is powerful for producing an aggregate analysis of adaptation. I continued by focusing on one specific and timely question within the realm of microbial adaptation. This question surrounds adaptation not often measured, and always unwelcome. This is the question of uncontrolled adaptation during an organism's tenure as a model research organism.

Organisms of human interest and consequence have been brought into the controlled laboratory environment since van Leeuwenhoek first saw them, in order to better understand their peculiar existence. However, the very act of transferring and studying microbes changes the tapestry of selection pressures, and can thwart an investigator's understanding of the organism's true nature. This is a long-recognized paradox, which has created a trade-off between the relevance of recent natural isolates and the utility and depth of insight gained from laboratory models.

While finding a solution or a perfect balance for this trade-off is a much greater endeavor than this dissertation, I am nonetheless fascinated by the challenge, and have begun by asking whether or not *H. salinarum,* a powerful model archaeon, has acquired an altered environmental response profile compared with its conspecific natural isolate counterparts. The remainder of the introduction provides background on halophilic archaea, the hypersaline ecosystem, and osmotolerance, the environmental perturbation selected for this study.

Chapter 2 continues by characterizing the hypersaline ecosystem from which we collected samples. Salinity is a multi-dimensional environmental factor, although it is often reduced to simple dilutions or changes in a single salt for biological research on osmotolerance. I motivate the need to approach salinity as a multifactorial perturbation, and chapter 3 takes a systematic approach to measuring the *H. salinarum* response to a multifactorial set of 75 growth

media. To make these measurements, I opted for a high-throughput growth assay from which maximum growth rate could be extracted as a marker of fitness. By designing a workflow for a state-of-the-art liquid handling robot together with a 200-plex small-volume incubator/shaker, I scaled up a classical microbiology assay in order to capture fitness across all media with 8x replication. This strategy afforded an uncommon power of the test and was a fine example of technology as it enables new and previously difficult questions to be answered.

We found that *H. salinarum* responds independently and significantly to changes in total salinlity, [NaCl] concentration, and [K]:[Mg] ratio. Microbial response was measured as maximum growth rate, a major component of fitness. Our initial expectation was that fitness would be maximized in growth media resembling the media composition used nearly universally in the laboratory. Surprisingly, local maximum growth rates occurred in different media compositions, leading us to ask whether these compositions resembled conditions experienced during earlier evolutionary history. I conclude Chapter 3 by showing that media that produce peak maximum growth rate are more chemically similar to GSL water chemistry than non peak-producing media.

The high-level, fitness-based characterization of osmotolerance in Chapter 3 showed that cells respond to changes in ion ratio, even when total salinity is held constant. This has not been well established in the literature, and has never been addressed for *H. salinarum.* This result led me to ask how cells negotiate changes in ion ratio more proximally. I asked, how is gene expression coordinated when ion ratio but not total salinity is changed? Chapter 4 provides an analysis of the major patterns in gene expression that emerge when ion ratio is varied, and I found that classes of genes important for metal ion transport, gas vesicles and vitamin biosynthesis are most differentially regulated across conditions. These genes emerge even after controlling for the effects of growth rate changes on gene expression, and suggest that osmotolerance is regulated in concert with numerous other cellular stress responses.

Together with my colleague, Dr. Elisabeth Wurtmann, I continued to explore the regulation of osmotolerance by investigating the role of post-transcriptional regulation in potassium uptake. A screen of ribonucleases in *H. salinarum* revealed one RNAse with expression anticorrelated with a potassium transporter operon, KdpFABCQ, suggesting that the RNAse negatively regulates this important membrane-bound transport system. We ultimately demonstrated that the RNAse (VNG2099) is needed for efficient management of the “salt-in”

osmotolerance strategy, whereby high concentrations of K^+ are pumped into the cell to counterbalance external Na^+, in an ATP-dependent fashion. When VNG2099 is missing and extracellular K^+ is high, the Kdp transport system fails to turn off, thus wasting ATP and producing the observed growth defect. The Kdp system is already known to operate under mM [K^+] but not higher concentrations[111]; the RNAse is needed for its efficient regulation. In addition, we found that this regulation is critical during environmental transitions such as dark-to-light and log-to-stationary phase. This systems-analysis of over 1500 array experiments highlights the importance of environmental context for understanding cellular response, and the extent to which relationships among environmental factors are mirrored in gene regulatory networks.

The analysis of transcriptional and post-transcriptional regulation during osmotolerance made it clear that salinity response is a complex phenotype, deeply engrained within the genome and transcriptome. Thus the answer to my original question of laboratory adaptation in a model organism would not be a binary yes or no, and there would be no simple phenotype to score in order to measure adaptation. Unlike some virulence traits (toxins, secretion systems, exopolysaccharides) encoded on plasmids by one or a few genes, osmotolerance is a complex phenotype, under the control of numerous genes and regulators, with redundant mechanisms throughout the cell. The question would also be difficult to answer because no ancestral strain exists, as in controlled laboratory evolution experiments. Pressing forward, I designed a cross-sectional study that would take a snapshot of current osmotolerance phenotypes and genomes in the laboratory model and in freshly isolated *H. salinarum.* Both strains share a common ancestor, at least five decades old, and both have had those intervening years to evolve at the level of phenotype and genome within their respective environments. Examples of laboratory adaptation in microbes (whether simple or complex phenotypes) abound, and I expected to uncover yet another case, in yet another model organism.

Stunted adaptation: explanations for the lack of phenotypic and genomic divergence among laboratory and lake strains

Chapter 5 describes how, counter to my expectations, the lab and lake strains were hardly divergent at all. They shared similar reaction norms, indicating that they thrived in similar niches

along salinity gradients, and this was subsequently explained when I sequenced their genomes and found them highly similar. The least exciting explanation for the similar phenotypes and genomes is that the strains had contaminated one another. This was quickly ruled out using a series of PCR reactions that differentiated the lake strains from each laboratory strain in turn.

Another explanation for the similar genomes is that exceptionally careful strain maintenance over the years permitted too few generations for evolution to occur. Perhaps the strain sequenced here is identical to the organism that was isolated from the lake decades ago. Careful maintenance of freezer stocks combined with a 10+ hour generation time leaves little opportunity for genomic evolution or adaptation of a complex phenotype like osmotolerance; recall that the ability to metabolize citrate, another complex phenotype, required over 30,000 generations in *E. coli*[52]. Nonetheless, I can rule out this explanation using the sequencing data. I re-sequenced the Hsa-lab strain, which was subcultured from the sequenced parent reference strain in 2000 to generate the Δ*ura3* mutant. When this sequence was aligned with the reference, it contained 21 SNPs and ISH-mediated InDels in the main chromosome alone, and several regions on both minichromosomes did not map to the reference. Thus, the Hsa-lab sequence differs from the reference as much as the Hsa-GSL strains do, albeit in different ways. In addition, the published differences between lab strains NRC-1 and R1 imply evolution in the laboratory as well[182]. I was not working with Hsa-lab as a perfect fossil of the original isolate so long ago.

The observed similarity among sequences and reaction norms must then be attributed to two testable factors: initial selection for genomes that produce growth in laboratory media, and minimal spatial and temporal diversity within 16S-identical-*H. salinarum* in the GSL. Few microbes are culturable in laboratory conditions, and the initial transfer from environmental sample to culture media is unavoidably selective for those capable of growth. For this experiment, I sought conspecific *H. salinarum* deliberately, in order to meaningfully compare the model to its environmental counterpart. What I did not know, and what has not been established in the literature, is the relationship between 16S rDNA diversity vs. whole genome sequence-based diversity, in non-laboratory evolved populations. Here, I have shown that despite decades of evolutionary opportunity for both Hsa-lab and Hsa-GSL, 16S identity is nearly recapitulated in the entire genome sequence in natural and laboratory strains. Finally, and perhaps most practically important, I have shown that *H. salinarum* NRC-1, despite 50+ years of laboratory

culture, has not become a specialist and has retained utility as a model for study of microbial response to complex environmental stimuli.

Future directions

This work has unearthed many more questions and hypotheses than it has answered. Although it is reassuring to know that *H. salinarum* NRC-1 research may continue to be applied to questions relevant to its hypersaline environment and natural isolate counterparts, many basic questions about adaptation, plasticity and osmotolerance remain unanswered. Here I elaborate on the most interesting and important lines of inquiry that follow from this work.

The original question posed here asked whether environment-to-laboratory transitions produced transitions in generalist to specialist strategies for fitness along an environmental gradient. Using Hsa-GSL strains in a longitudinal, laboratory evolution experiment, this question may be addressed directly. In such an experiment, the fresh isolate serves as the ancestral strain which will be subcultured and passaged through static and varying salinity conditions. Samples taken every 1000 generations will then be sequenced and assayed for fitness and phenotype, creating a real-time record of adaptation following environmental isolation. Beginning such an experiment with a diverse population (*e.g.* through use of a transposon mutagenized library) may help jumpstart the population on a selective trajectory toward laboratory salinity adaptation.

Continuing to explore the evolutionary basis for plasticity in osmotolerance, a phylogenetic approach is ideal. I have access to a collection of halophilic archaeal stocks, each varying at the level of 16S rRNA. With these strains, I can ask how divergent species solve the problem of tolerating variant saline conditions. The notion of microbial species is contentious and it is possible that divergent genomes that have evolved in a similar hypersaline environment have converged to solve the osmotolerance problem in similar ways. A comparative phylogenomic analysis will lead to testable hypotheses about the genes and networks involved in osmotolerance, and the extent of their conservation within halophilic microbial consortia.

Another line of inquiry that follows from our systems-level analysis of the osmotolerance transcriptome is the mechanistic basis for salinity adaptation. I uncovered several leads within the microarray analysis (Chapter 4), including the involvement of the *cbi* and *cob* cobalamin biosynthesis genes. The connection between osmotolerance, transition metal influx (*i.e.* cobalt), and vitamin biosynthesis is intriguing, and the transduction of an osmotic imbalance signal

through the cell to the level of transcription has not been described. I suspect that cells perceive changes in ionic salts in much the same way they perceive transition metal stresses[171], and that the mechanisms and networks that govern both have significant overlap.

This field also requires a new set of methods for generating and analyzing compositional data. When using media composition as a cellular perturbation, it is impossible to evaluate independently varying components of the media. This is a compositional problem, and this means that when one factor (*i.e.* [NaCl]) is increased, at least one more factor must decrease if the total dissolved concentration is to remain the same. If total is not held constant, then total salinity confounds the effect of [NaCl]. The very nature of compositions forces correlations in the design, which leave few valid statistical analyses without violations of assumptions or collinearity within models. In chapter 3, this problem becomes apparent by the fact that the effects of [K] and [Mg] cannot be disentangled. Similarly, [Cl] and [Mg] are anticorrelated in the array experiments. Compositional analysis exists as a field, and a number of resources, statistical packages and papers have been written on the topic[193-195]. However, these methods, like PCA and MDS, apply mainly to non-experimental compositions, where multivariate data must be reduced for analysis. Systematically varying compositions as experimental treatments will require additional innovation within the field of statistics.

Finally, in the analysis of the whole-genome sequence data in Chapter 5, I have become interested in the implications of different microbial diversity measurements. For over 30 years, Carl Woese's introduction of the 16S rRNA sequence has served as gold standard for assessing evolutionary distance among microbes[196]. It is an ideal molecular chronometer, neither varying too quickly or slowly over evolutionary time. The choice of a single gene as diversity measure, however, was constrained by technological capacity at the time of its development. Today, whole genomes are cheaper than ever to sequence, and soon (if not already), it will be feasible and economical to undertake a comprehensive evaluation of whole genomes as measures of diversity. When 16S analysis declares two genomes identical, what can be inferred about the similarity of the rest of the genome? How does the relationship between 16S and whole-genomic diversity change across prokaryotic taxa? We are at a juncture in microbial genomics, where technology has enabled paradigm shifts in time-tested methodologies. This will be an important answer for microbiology, and will require large computing capacity for alignment of entire genomes.

Reflections

In closing, this has been an exploration of a single idiosyncratic species' response to a multifactorial and ubiquitous environmental pressure. Using this narrow system, however, I have been transformed in my approach to all manner of complex questions in biology. The philosophy that underlies systems analysis is one that keeps the mind from constraining the possibility space of any question's answer. In other words, it encourages investigators to observe and measure first, then form hypotheses, followed by additional, more informed measurements. Clearly this is not a new approach, to biology or to science, even though "Systems Biology" as a field was born in the last 15 years with the advent of advanced measurement technologies. It is not new, but conscious and deliberate follow-through is often not met, and I can attest to the incredible amount of time required to complete one "cycle" of the process.

I look forward to carrying this whole-systems philosophy with me, as I move on to questions and problems that face public health today. Much like the stress response of a single cell, infectious diseases in populations are emergent properties of complex systems, and no single drug or policy decision will make significant impact on disease burden. Reducing the burden of infectious disease, most especially for those who suffer disproportionately through poverty and oppression, is a lifelong commitment I will embark upon following successful completion of my doctorate.

Systems approaches in public health

To close this dissertation and elaborate on systems approaches to public health and infectious disease, I am including an article I wrote on the topic, with its accompanying responses from three researchers and professionals in the field. The article is published as: **Beer, K**. News from the IAEH. Discussion on the role of national public health agencies in the implementation of ecohealth strategies for infectious disease prevention. *EcoHealth* **10,** 111–114 (2013), and is reproduced with permission.

Discussion on the role of national public health agencies in the implementation of ecohealth strategies for infectious disease prevention

Why should ecohealth principles be adopted for infectious disease control by public health agencies in high-income countries? Infectious diseases (IDs) remain the most important contributor to all-ages mortality in low- and middle-income countries. Although high-income countries face more substantial burdens from chronic diseases, emerging infections (e.g., HIV, SARS, Nipah virus, avian influenza) and outbreaks of foodborne pathogens are reminders that advanced wealth, sanitation, and technology do not offer complete protection against IDs.

Consistent with these realities, the U.S. Centers for Disease Control and Prevention (CDC) and the Public Health Agency of Canada (PHAC) have for decades supported ID research, surveillance, vaccination campaigns, and outbreak responses. These agencies have been highly successful at reducing the ID burden in the United States, Canada, and abroad. However, the spread of antibiotic resistance, global pandemics of emerging IDs, increased ID incidence linked to climate change and biodiversity loss (e.g., Lyme disease and hantavirus), and difficulties in vaccine development for IDs (e.g., for HIV and the dengue virus) necessitate a new parallel approach focused on prevention.

Such an alternative approach can be found in the young scientific disciplines of Ecohealth and Conservation Medicine, as well as the One Health Initiative, collectively referred to here as "ecohealth." Each of these fields, despite three different names, recognize human and animal diseases as emergent properties of a complex system and thus integrate many disciplinary factors into ID research; such factors include ecology, biology, land use, demographics, travel, social inequalities, and climate. Already, this "systems approach" to ID research has demonstrated causal associations between ocean chemistry and cholera incidence and between land use and climate change and numerous zoonoses, such as Lyme disease and dengue fever. In the United States, land use changes affect Lyme disease reservoir abundance and diversity, which in turn influence infection prevalence and disease risk. For dengue fever, models of climate-driven changes in mosquito vector abundance have been recapitulated in annually reported cases in Central America. These ecohealth-based insights offer vast opportunities for novel and complementary prevention strategies.

The prospect of an effective systems approach to ID prevention is compelling, especially for students interested in public health practice in the United States and Canada. Yet, why do public agencies in developed countries seem to play a minimal role in strategic ecohealth implementation? The role that major U.S. and Canadian public health agencies should play in

ecohealth project leadership seems unclear. For example, the CDC has an Office of One Health within the Division of High-Consequence Pathogens and Pathology. Its mission is aligned with ecohealth principles, but documentation of related projects is scarce. The PHAC hosted a "One World, One Health" conference in Winnipeg in 2009, where experts convened to discuss how to implement an ecohealth approach to ID surveillance, monitoring, and prevention. However, follow-up on projects that may have emerged from this meeting is not readily available.

Who *is* leading the implementation of ecohealth-based ID prevention? The Canadian International Development Research Center (IDRC), which funds ecohealth research, recently published an analysis of 15 applied ecohealth case studies. These projects were most often initiated by researchers in developing countries with heavy ID burdens. Many projects achieved success through local government support and did not cite outside agency involvement.

There are many potential explanations for the observation that the well-equipped health agencies in the United States and Canada have seemed slow to participate in ecohealth implementation. First, the ecohealth approach is young. While quality research has been done, there is still much to do, and application generally follows after a scientific field matures. Second, despite strong infrastructure, large public health agencies are often bureaucratic and slower to adopt new approaches and paradigms than are academia or the nonprofit sector. Third, IDs that lend themselves to an ecohealth approach are larger problems in developing countries, and thus there is greater urgency for pioneers in these localities to implement ecohealth-based research on their own. Nonetheless, it is possible that the CDC and PHAC are playing a greater role than their websites and publications reveal, and these efforts may or may not carry the "ecohealth" label. Perhaps a central repository, or a concerted effort to publicize ecohealth-related projects, would better represent current and ongoing cross-disciplinary efforts at the CDC and PHAC.

What *should be* the role of U.S. and Canadian public health agencies in strategic ecohealth implementation for ID prevention? One broad suggestion comes from environmental microbiologist and former NSF director Dr. Rita Colwell, who has called for interdisciplinary "Centers of Ecological Health" that will bring all stakeholders together, including national public health agencies, to reduce global ID burdens. Toward this end, what are the CDC and PHAC doing to make ecohealth strategies like this a reality, and what should the next steps be? How could students stay better informed of agencies' ecohealth-focused activities, and how might

students shape their training to be involved in what is certainly an important new direction for public health?

From a student perspective, it is exciting and encouraging to see the recent uptick in ecohealth-specific graduate programs and post-graduate fellowships. Nevertheless, it can be challenging to find a place in public health where ecohealth approaches and skills may be applied. For example, health agencies often have separate divisions for infectious disease and animal health, and agencies involved in land use and planning are separate from public health altogether. What positions are available to graduates interested in coordinating among these divisions? Perhaps this challenge reflects the excitement and chaos of a nascent paradigm shift in infectious disease prevention. At this point, we students cannot expect to find job descriptions that match the forward-thinking, interdisciplinary work we aim to do. Instead, students themselves are part of the paradigm shift, and if we are persistent and innovative, we may carve out the necessary niche for ecohealth within public health. If we are successful, we will be the ones writing the requests for applications, not seeking them.

Public Health Agencies and EcoHealth Strategies: Opportunity Unanswered?

Dr. James Mills, Editor-in-chief, Journal of Wildlife Diseases; Former chief of medical ecology at the CDC's division of high consequence pathogens and pathology; Adjunct faculty in Population Biology, Ecology, and Evolution, Emory University

As an ecologist who spent 25 years working in public health, including 16 years at CDC, I have some insights that span the EcoHealth spectrum. Ms. Beer is correct; CDC, and probably other public health agencies, have not fully and actively embraced the One Health/EcoHealth concept. Unlike academic investigators who have much autonomy, CDC researchers operate within a defined “mission,” dictated by policies that are not independent of political pressures. Projects pursued are generally those that will yield clear, short-term results. Rushing teams of physicians and epidemiologists to stop an outbreak is important. It also makes headlines and gets people promoted. Working with residents to establish livelihoods that protect ecosystem integrity, maintain biodiversity, and prevent outbreaks from occurring is also important. Yet it, goes unnoticed, produces no heroes, is unlikely to win votes or launch careers, and it may not secure continued agency funding.

Public health agencies *are* beginning to accept the One-Health concept as a useful adjunct. Nevertheless, clear ideas of how to apply the concept have not been developed and agencies are unwilling to take money from traditional programs and invest in what is seen as a new and largely unproven approach. My experience at CDC has shown that change from within happens slowly. However, when strong direction comes from the top, change is rapid. Government agencies (including the CDC) answer to Congress and the Administration, who, in turn, answer to the citizens.

What can students do? Obviously, seek collaborations with and jobs within public health agencies. Conduct research that demonstrates the effectiveness and impact of One-Health approaches. Importantly, this generation of scientists must not repeat the mistakes of the past and sequester itself within the ivory tower. Remain involved with local communities, regain the public trust, share knowledge and exchange ideas with the community, and educate our congressional representatives.

Change in culture

Lawrence C. Madoff, MD. Professor of Medicine, University of Massachusetts Medical School; Director, Division of Epidemiology and Immunization, Massachusetts Department of Public Health; Editor, ProMED-mail

The profile of the One Health (or ecohealth) movement has come increasingly to the forefront in discussions of emerging infectious diseases. Given the complex interplay of factors that affect emergent zoonotic diseases (witness such high profile infectious disease events of the past decade as SARS, swine-origin H1N1, monkeypox, and peanut butter salmonellosis), this trend is entirely justified and indeed overdue. How have governmental public health agencies responded? The bureaucratic response to create additional, often siloed entities, may indeed not be the most productive one, as Ms. Beer notes. Also, in an era of shrinking budgets and government stretched thin, we are not likely to see real progress in the creation of new capacity in the near term. Instead, people within existing agencies need to understand the need for this new capacity, and feel sufficiently empowered to reach across agency boundaries to the resources they need.

This in turn requires a change in culture that is in many ways more difficult to affect than merely creating a new organizational structure. Veterinary health, human health, plant health and ecology occupy very different spheres not only in government. These disciplines are taught in different schools, published in distinct journals, discussed at divergent conferences. The one health movement is, gradually, eroding boundaries and bringing these disciplines to the same table. The establishment of journals like EcoHealth, the creation of meetings like One World, One Health and the International Meeting on Emerging Diseases, the endorsement of the movement by professional organizations like the AMA and AVMA, the growth of surveillance systems like ProMED that monitor plant, animal and human health, and even academic programs like Ms. Beer's speak to this change. We are increasingly seeing this transformation within governmental organizations as well, perhaps not leading the cultural shift, but certainly taking part in it.

A vector-borne disease perspective

Charles B. Beard* and Carol Rubin**
*Centers for Disease Control and Prevention, National Center for Emerging and Zoonotic Infectious Diseases, Division of Vector-Borne Diseases, 3150 Rampart Road, Fort Collins, CO 80521 (corresponding)
**Centers for Disease Control and Prevention, National Center for Emerging and Zoonotic Infectious Diseases, One Health Office, 1600 Clifton Rd. Atlanta, GA 30333

Many infectious diseases have critical environmental links. Consequently, changing ecologies can influence disease emergence. A prime example is Lyme disease emergence and the links to reforestation, habitat fragmentation, and decreasing species biodiversity. The One Health initiative promotes the fact that disease emergence occurs at the interface of humans, other animals, and a changing environment.

Despite articulation, the One Health concept struggles for a place in mainstream public health. In part, this may be due to the reality that public health is much broader than it used to be, and any single priority competes with many others. When the U.S Centers for Disease Control and Prevention (CDC) was established in 1946 as the Communicable Diseases Center, infectious diseases were the sole focus, and ecohealth, while not birthed as a term, was the lens

through which entomologists, engineers, and epidemiologists studied and evaluated interventions for diseases like malaria, plague, and polio. Today, however, public health at all levels is pressed by economic constraints and a vastly expanded mission.

Regardless of the challenges, One Health is very much alive within the world of public health. It is seen in the ArboNET surveillance system that maps data on arboviruses detected in humans, birds, mosquitoes, and livestock. It is seen in Lyme disease projects that evaluate upstream interventions such as reservoir-targeted vaccines. It is further demonstrated in novel interdisciplinary training programs such as one that provides cross-training in both public health and climatology and atmospheric sciences, for the purpose of developing young scientists to provide leadership in addressing the challenges that climate disruption and change will bring to the world of public health.

While these programs are small and seminal, they indicate a growing recognition of the importance of a systems approach to public health, one that demonstrates an evolving ecohealth perspective.

References

1. Galhardo, R. S., Hastings, P. J. & Rosenberg, S. M. Mutation as a stress response and the regulation of evolvability. *Critical Reviews in Biochemistry and Molecular Biology* **42,** 399–435 (2007).
2. Kennedy, S. P., Ng, W. V., Salzberg, S. L., Hood, L. & DasSarma, S. Understanding the adaptation of Halobacterium species NRC-1 to its extreme environment through computational analysis of its genome sequence. *Genome Res* **11,** 1641–1650 (2001).
3. Kushner, D. J. & Kushner, D. J. Microbial life in extreme environments. (1978).
4. Woelfle, M. A., Ouyang, Y., Phanvijhitsiri, K. & Johnson, C. H. The adaptive value of circadian clocks: an experimental assessment in cyanobacteria. *Curr Biol* **14,** 1481–1486 (2004).
5. Ouyang, Y., Andersson, C. R., Kondo, T., Golden, S. S. & Johnson, C. H. Resonating circadian clocks enhance fitness in cyanobacteria. *Proc Natl Acad Sci USA* **95,** 8660–8664 (1998).
6. Mitchell, A. *et al.* Adaptive prediction of environmental changes by microorganisms. *Nature* **460,** 220–224 (2009).
7. Huisman, G. W. & Kolter, R. Sensing starvation: a homoserine lactone--dependent signaling pathway in Escherichia coli. *Science* **265,** 537–539 (1994).
8. Ram, R. J. *et al.* Community Proteomics of a Natural Microbial Biofilm. *Science* **308,** 1915–1920 (2005).
9. Hochachka, P. W. & Somero, G. N. *Biochemical adaptation*. xx–537 p. (Princeton University Press, 1984).
10. Schrag, S. J., Perrot, V. & Levin, B. R. Adaptation to the fitness costs of antibiotic resistance in Escherichia coli. *Proc Biol Sci* **264,** 1287–1291 (1997).
11. Spain, J. C. & Van Veld, P. A. Adaptation of natural microbial communities to degradation of xenobiotic compounds: effects of concentration, exposure time, inoculum, and chemical structure. *Appl. Environ. Microbiol* **45,** 428–435 (1983).
12. Finkel, S. E. & Kolter, R. Evolution of microbial diversity during prolonged starvation. *Proc Natl Acad Sci USA* **96,** 4023–4027 (1999).
13. Gribaldo, S. & Brochier-Armanet, C. The origin and evolution of Archaea: a state of the art. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* **361,** 1007–1022 (2006).
14. Soppa, J. *et al.* Genomics and functional genomics with haloarchaea. *Arch Microbiol* **190,** 197–215 (2008).
15. Babic, A., Lindner, AB, Vulic, M., Stewart, E. J. & Radman, M. Direct visualization of horizontal gene transfer. *Science* **319,** 1533–1536 (2008).
16. Lercher, M. J. & Pal, C. Integration of horizontally transferred genes into regulatory interaction networks takes many million years. *Mol Biol Evol* **25,** 559–567 (2008).
17. Rutherford, S. L. From genotype to phenotype: buffering mechanisms and the storage of genetic information. *BioEssays* **22,** 1095–1105 (2000).
18. Tagkopoulos, I., Liu, Y.-C. & Tavazoie, S. Predictive behavior within microbial genetic networks. *Science* **320,** 1313–1317 (2008).
19. Chaffron, S., Rehrauer, H., Pernthaler, J. & Mering, von, C. A global network of coexisting microbes from environmental and whole-genome sequence data. *Genome Res*

20, 947–959 (2010).
20. Barton, N. H. *Evolution*. xiv–833 p. (Cold Spring Harbor Laboratory Press, 2007).
21. Burke, M. K. *et al.* Genome-wide analysis of a long-term evolution experiment with Drosophila. *Nature* **467,** 587–590 (2010).
22. Kassen, R. Toward a General Theory of Adaptive Radiation. *Ann. N. Y. Acad. Sci.* **1168,** 3–22 (2009).
23. Orr, H. A. Theories of adaptation: what they do and don't say. *Genetica* **123,** 3–13 (2005).
24. Papadopoulos, D. *et al.* Genomic evolution during a 10,000-generation experiment with bacteria. *Proc Natl Acad Sci USA* **96,** 3807 (1999).
25. de Visser, J. A. G. M. & Rozen, D. E. Clonal Interference and the Periodic Selection of New Beneficial Mutations in Escherichia coli. *Genetics* **172,** 2093–2100 (2006).
26. Lauro, F. M., Chastain, R. A., Blankenship, L. E., Yayanos, A. A. & Bartlett, D. H. The unique 16S rRNA genes of piezophiles reflect both phylogeny and adaptation. *Appl. Environ. Microbiol* **73,** 838–845 (2007).
27. Hentschel, U. *et al.* Molecular evidence for a uniform microbial community in sponges from different oceans. *Appl. Environ. Microbiol* **68,** 4431–4440 (2002).
28. Gasch, A. P. *et al.* Genomic expression programs in the response of yeast cells to environmental changes. *Mol Biol Cell* **11,** 4241–4257 (2000).
29. Causton, H. C. *et al.* Remodeling of yeast genome expression in response to environmental changes. *Mol Biol Cell* **12,** 323–337 (2001).
30. He, G., He, B., Racey, P. & Cui, J. Positive Selection of the Bat Interferon Alpha Gene Family. *Biochemical Genetics* **48,** 840–846 (2010).
31. Enjalbert, B., Nantel, A. & Whiteway, M. Stress-induced gene expression in Candida albicans: absence of a general stress response. *Mol Biol Cell* **14,** 1460–1467 (2003).
32. Kaur, A. *et al.* Coordination of frontline defense mechanisms under severe oxidative stress. *Mol Syst Biol* **6,** 393 (2010).
33. Whitehead, K., Pan, M., Masumura, K.-I., Bonneau, R. & Baliga, N. S. Diurnally entrained anticipatory behavior in archaea. *PLoS ONE* **4,** e5485 (2009).
34. Wolf, D. M. *et al.* Memory in microbes: quantifying history-dependent behavior in a bacterium. *PLoS ONE* **3,** e1700 (2008).
35. Johnson, C. H., Mori, T. & Xu, Y. A Cyanobacterial Circadian Clockwork. *Current Biology* **18,** R816–R825 (2008).
36. Babu, M. M., Luscombe, N. M., Aravind, L., Gerstein, M. & Teichmann, S. A. Structure and evolution of transcriptional regulatory networks. *Curr Opin Struct Biol* **14,** 283–291 (2004).
37. Whitaker, W. B. *et al.* Modulation of responses of Vibrio parahaemolyticus O3:K6 to pH and temperature stresses by growth at different salt concentrations. *Applied and Environmental Microbiology* **76,** 4720–4729 (2010).
38. Calhoun, L. N. & Kwon, Y. M. The effect of long-term propionate adaptation on the stress resistance of Salmonella Enteritidis. *J Appl Microbiol* (2010). doi:10.1111/j.1365-2672.2010.04750.x
39. Dassarma, S., Berquist, B. R., Coker, J. A., Dassarma, P. & Müller, J. A. Post-genomics of the model haloarchaeon Halobacterium sp. NRC-1. *Saline Syst* **2,** 3 (2006).
40. Oren, A. Microbial life at high salt concentrations: phylogenetic and metabolic diversity. *Saline Syst* **4,** 2 (2008).

41. Sanchez-Perez, G., Mira, A., Nyiro, G., Pasic, L. & Rodriguez-Valera, F. Adapting to environmental changes using specialized paralogs. *Trends Genet* **24,** 154–158 (2008).
42. Feller, G. & Gerday, C. Psychrophilic enzymes: hot topics in cold adaptation. *Nat Rev Micro* **1,** 200–208 (2003).
43. Feller, G. Life at low temperatures: is disorder the driving force? *Extremophiles* **11,** 211–216 (2007).
44. Vieille, C., Epting, K. L., Kelly, R. M. & Zeikus, J. G. Bivalent cations and amino-acid composition contribute to the thermostability of Bacillus licheniformis xylose isomerase. *Eur J Biochem* **268,** 6291–6301 (2001).
45. Jaenicke, R. Protein stability and molecular adaptation to extreme conditons. *European Journal of Biochemistry* **202,** 715–728 (1991).
46. Barák, I. & Wilkinson, A. J. Where asymmetry in gene expression originates. *Mol Microbiol* **57,** 611–620 (2005).
47. Veening, J. W., Smits, W. K. & Kuipers, O. P. Bistability, epigenetics, and bet-hedging in bacteria. *Microbiology* **62,** 193–210 (2008).
48. Bennett, A. F., Dao, K. M. & Lenski, R. E. Rapid evolution in response to high-temperature selection. *Nature* **346,** 79–81 (1990).
49. Bennett, A. F. & Lenski, R. E. An experimental test of evolutionary trade-offs during temperature adaptation. *Proc Natl Acad Sci USA* **104 Suppl 1,** 8649–8654 (2007).
50. Knies, J. L., Izem, R., Supler, K. L., Kingsolver, J. G. & Burch, C. L. The Genetic Basis of Thermal Reaction Norm Evolution in Lab and Natural Phage Populations. *Plos Biol* **4,** e201 (2006).
51. Cooper, T. F. & Lenski, R. E. Experimental evolution with E. coli in diverse resource environments. I. Fluctuating environments promote divergence of replicate populations. *BMC Evol Biol* **10,** 11 (2010).
52. Blount, Z. D., Borland, C. Z. & Lenski, R. E. Historical contingency and the evolution of a key innovation in an experimental population of Escherichia coli. *Proc Natl Acad Sci USA* **105,** 7899–7906 (2008).
53. Zinser, E. R. & Kolter, R. Escherichia coli evolution during stationary phase. *Res Microbiol* **155,** 328–336 (2004).
54. Lenski, R. E. & Travisano, M. Dynamics of adaptation and diversification: a 10,000-generation experiment with bacterial populations. *Proc Natl Acad Sci USA* **91,** 6808–6814 (1994).
55. Barrick, J. E. *et al.* Genome evolution and adaptation in a long-term experiment with Escherichia coli. *Nature* **461,** 1243–1247 (2009).
56. Rosenberg, S. M. Evolving responsively: adaptive mutation. *Nat Rev Genet* **2,** 504–515 (2001).
57. McAdams, H. H., Srinivasan, B. & Arkin, A. P. The evolution of genetic regulatory systems in bacteria. *Nat Rev Genet* **5,** 169–178 (2004).
58. Lozada-Chavez, I., Angarica, V. E., Collado-Vides, J. & Contreras-Moreira, B. The role of DNA-binding specificity in the evolution of bacterial regulatory networks. *J Mol Biol* **379,** 627–643 (2008).
59. López-Maury, L., Marguerat, S. & Bähler, J. Tuning gene expression to changing environments: from rapid responses to evolutionary adaptation. *Nat Rev Genet* **9,** 583–593 (2008).
60. DeWitt, T. J. & Scheiner, S. M. *Phenotypic Plasticity*. (Oxford University Press on

Demand, 2004).
61. Chevin, L.-M., Gallet, R., Gomulkiewicz, R., Holt, R. D. & Fellous, S. Phenotypic plasticity in evolutionary rescue experiments. *Philosophical Transactions of the Royal Society B: Biological Sciences* **368,** 20120089 (2013).
62. Chevin, L.-M., Lande, R. & Mace, G. M. Adaptation, plasticity, and extinction in a changing environment: towards a predictive theory. *Plos Biol* **8,** e1000357 (2010).
63. Leroi, A. M., Lenski, R. E. & Bennett, A. F. Evolutionary adaptation to temperature. III. Adaptation of Escherichia coli to a temporally varying environment. *Evolution* **48,** 1222–1229 (1994).
64. Bennett, A. F. & Lenski, R. E. Evolutionary adaptation to temperature II. Thermal niches of experimental lines of Escherichia coli. *Evolution* **47,** 1–12 (1993).
65. Cooper, V. S., Bennett, A. F. & Lenski, R. E. Evolution of thermal dependence of growth rate of Escherichia coli populations during 20,000 generations in a constant environment. *Evolution* **55,** 889–896 (2001).
66. Kassen, R. & Bell, G. Experimental evolution in Chlamydomonas. IV. Selection in environments that vary through time at different scales. *Heredity* **80,** 732–741 (1998).
67. Reboud, X. & Bell, G. Experimental evolution in Chlamydomonas. III. Evolution of specialist and generalist types in environments that vary in space and time. *Heredity* **78,** 507–514 (1997).
68. Wund, M. A. Assessing the impacts of phenotypic plasticity on evolution. *Integr. Comp. Biol.* **52,** 5–15 (2012).
69. Lenski, R. E. Microbial genetics: Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. *Nature Publishing Group* **4,** 457–469 (2003).
70. Yeh, P. J. & Price, T. D. Adaptive Phenotypic Plasticity and the Successful Colonization of a Novel Environment. *Am Nat* **164,** 531–542 (2004).
71. Buckling, A., Brockhurst, M. A., Travisano, M. & Rainey, P. B. Experimental adaptation to high and low quality environments under different scales of temporal variation. *J Evol Biol* **20,** 296–300 (2007).
72. Kassen, R. The experimental evolution of specialists, generalists, and the maintenance of diversity. *J Evol Biol* **15,** 173–190 (2002).
73. Angilletta, M. J., Jr, Niewiarowski, P. H. & Navas, C. A. The evolution of thermal physiology in ectotherms. *Journal of Thermal Biology* **27,** 249–268 (2002).
74. Bennett, A. F., Lenski, R. E. & Mittler, J. E. Evolutionary adaptation to temperature. I. Fitness responses of Escherichia coli to changes in its thermal environment. *Evolution* **46,** 16–30 (1992).
75. Post, F. J. The microbial ecology of the Great Salt Lake. *Microb Ecol* **3,** 143–165 (1977).
76. Reale, D., McAdam, A. G., Boutin, S. & Berteaux, D. Genetic and plastic responses of a northern mammal to climate change. *Proceedings of the Royal Society B: Biological Sciences* **270,** 591–596 (2003).
77. Damkiær, S., Yang, L., Molin, S. & Jelsbak, L. Evolutionary remodeling of global regulatory networks during long-term bacterial adaptation to human hosts. *Proc Natl Acad Sci USA* **110,** 7766–7771 (2013).
78. Yang, L. *et al.* Evolutionary dynamics of bacteria in a human host environment. *Proc Natl Acad Sci USA* **108,** 7481–7486 (2011).
79. Cruickshank, T. & Wade, M. J. Microevolutionary support for a developmental hourglass: gene expression patterns shape sequence variation and divergence in

Drosophila. *Evol. Dev.* **10,** 583–590 (2008).
80. Csonka, L. N. & Hanson, A. D. Prokaryotic osmoregulation: genetics and physiology. *Annu Rev Microbiol* **45,** 569–606 (1991).
81. Wood, J. M. Bacterial osmoregulation: a paradigm for the study of cellular homeostasis. *Annu Rev Microbiol* **65,** 215–238 (2011).
82. Oren, A. The bioenergetic basis for the decrease in metabolic diversity at increasing salt concentrations: implications for the functioning of salt lake ecosystems. *Hydrobiologia* (2001).
83. Oren, A. Bioenergetic Aspects of Halophilism. *Microbiology and Molecular Biology Reviews* **63,** 334–348 (1999).
84. Baxter, B. K. *et al.* Great Salt Lake Microbial Diversity in Adaptation to Life in High Salt Concentrations in Archaea Bacteria, and Eukarya. 1–17 (2005).
85. Ng, W. V. *et al.* Genome sequence of Halobacterium species NRC-1. *Proc Natl Acad Sci USA* **97,** 12176–12181 (2000).
86. Paul, S., Bag, S. K., Das, S., Harvill, E. T. & Dutta, C. Molecular signature of hypersaline adaptation: insights from genome and proteome composition of halophilic prokaryotes. *Genome Biol* **9,** R70 (2008).
87. Peck, R. F., DasSarma, S. & Krebs, M. P. Homologous gene knockout in the archaeon Halobacterium salinarum with ura3 as a counterselectable marker. *Mol Microbiol* **35,** 667–676 (2000).
88. Reiss, D. J., Baliga, N. S. & Bonneau, R. Integrated biclustering of heterogeneous genome-wide datasets for the inference of global regulatory networks. *BMC Bioinformatics* **7,** 280 (2006).
89. Reiss, D. J., Facciotti, M. T. & Baliga, N. S. Model-based deconvolution of genome-wide DNA binding. *Bioinformatics* **24,** 396–403 (2008).
90. Bonneau, R. *et al.* A predictive model for transcriptional control of physiology in a free living cell. *Cell* **131,** 1354–1365 (2007).
91. Oren, A. The dying Dead Sea: The microbiology of an increasingly extreme environment. *Lakes & Reservoirs: Research & Management* (2010).
92. Gwynn, J. W. *Great Salt Lake brine chemistry databases and reports: 1966-2006*. 1–22 (Utah Geological Survey, 2007).
93. Marcarelli, A. & Wurtsbaugh, W. *Hydrogen Sulfide in Farmington Bay and the Great Salt Lake: A potential odor-causing agent*. 1–30 (2004). doi:Report to Central Davis Sewer Improvement District from Utah State University
94. Gwynn, J. W. *Great Salt Lake: an overview of change*. (Utah Geological Survey, 2002).
95. Steinhorn, I. In situ salt precipitation at the Dead Sea. *Limnology and Oceanography* (1983).
96. Warren, J. K. *Evaporites: Sediments, Resources and Hydrocarbons*. (Springer-Verlag, 2006).
97. Post, F. J. 6. Microbiology of the Great Salt Lake north arm. *Hydrobiologia* **81-82,** 59–69 (1981).
98. Parnell, J. J. *et al.* Functional biogeography as evidence of gene transfer in hypersaline microbial communities. *PLoS ONE* **5,** e12919 (2010).
99. Jiang, H. *et al.* Microbial response to salinity change in Lake Chaka, a hypersaline lake on Tibetan plateau. *Environmental Microbiology* **9,** 2603–2621 (2007).
100. Ashworth, J. *et al.* Genome-wide diel growth state transitions in the diatom Thalassiosira

pseudonana. *Proc Natl Acad Sci USA* **110,** 7518–7523 (2013).
101. Lindsey, H. A., Gallie, J., Taylor, S. & Kerr, B. Evolutionary rescue from extinction is contingent on a lower rate of environmental change. *Nature* **494,** 463–467 (2013).
102. Smith, E. E. *et al.* Genetic adaptation by Pseudomonas aeruginosa to the airways of cystic fibrosis patients. *Proc Natl Acad Sci USA* **103,** 8487–8492 (2006).
103. Nagano, A. J. *et al.* Deciphering and Prediction of Transcriptome Dynamics under Fluctuating Field Conditions. *Cell* **151,** 1358–1369 (2012).
104. Barabási, A.-L. & Oltvai, Z. N. Network biology: understanding the cell's functional organization. *Nat Rev Genet* **5,** 101–113 (2004).
105. Yu, H. & Gerstein, M. Genomic analysis of the hierarchical structure of regulatory networks. *Proc Natl Acad Sci USA* **103,** 14724–14731 (2006).
106. Hartwell, L. H., Hopfield, J. J., Leibler, S. & Murray, A. W. From molecular to modular cell biology. *Nature* **402,** C47–52 (1999).
107. Alon, U. Network motifs: theory and experimental approaches. *Nat Rev Genet* **8,** 450–461 (2007).
108. McCaffrey, M. A., Lazar, B. & Holland, H. D. The evaporation path of seawater and the coprecipitation of Br- and K+ with halite. *J Sediment Petrol* **57,** 928–938 (1987).
109. Fofonoff, N. P. Physical properties of seawater: A new salinity scale and equation of state for seawater. *Journal of Geophysical Research: Oceans (1978 ...* (1985).
110. Guillard, R. in *Culture of Marine Invertebrate Animals* (Smith, W. L. & Chanley, M. H.) (Plenum Press, 1975).
111. Kixmüller, D., Strahl, H., Wende, A. & Greie, J.-C. Archaeal transcriptional regulation of the prokaryotic KdpFABC complex mediating K+ uptake in H. salinarum. *Extremophiles* **15,** 643–652 (2011).
112. Strahl, H. & Greie, J.-C. The extremely halophilic archaeon Halobacterium salinarum R1 responds to potassium limitation by expression of the K+-transporting KdpFABC P-type ATPase and by a decrease in intracellular K. *Extremophiles* **12,** 741–752 (2008).
113. Csonka, L. N. Physiological and genetic responses of bacteria to osmotic stress. *Microbiol Rev* **53,** 121–147 (1989).
114. Tietge, J., Hockett, J. & Evans, J. Major ion toxicity of six produced waters to three freshwater species: Application of ion toxicity models and TIE procedures. *Environmental Toxicology and Chemistry* **16,** 2002–2008 (1997).
115. Soucek, D. J. *et al.* Influence of water hardness and sulfate on the acute toxicity of chloride to sensitive freshwater invertebrates. *Environ. Toxicol. Chem.* **30,** 930–938 (2011).
116. Kennedy, A., Cherry, D. & Zipper, C. Evaluation of ionic contribution to the toxicity of a coal-mine effluent using Ceriodaphnia dubia. *Arch Environ Contam Toxicol* **49,** 155–162 (2005).
117. Coker, J. A., Dassarma, P., Kumar, J., Müller, J. A. & Dassarma, S. Transcriptional profiling of the model Archaeon Halobacterium sp. NRC-1: responses to changes in salinity and temperature. *Saline Syst* **3,** 6 (2007).
118. He, Z. *et al.* Global transcriptional, physiological, and metabolite analyses of the responses of Desulfovibrio vulgaris hildenborough to salt adaptation. *Applied and Environmental Microbiology* **76,** 1574–1586 (2010).
119. Al-Mailem, D. M., Sorkhoh, N. A., Al-Awadhi, H., Eliyas, M. & Radwan, S. S. Biodegradation of crude oil and pure hydrocarbons by extreme halophilic archaea from

hypersaline coasts of the Arabian Gulf. *Extremophiles* **14,** 321–328 (2010).
120. Browne, R. A. & Wanigasekera, G. Combined effects of salinity and temperature on survival and reproduction of five species of Artemia. *Journal of experimental marine biology and …* (2000).
121. Crisler, J. D., Newville, T. M., Chen, F., Clark, B. C. & Schneegurt, M. A. Bacterial growth at the high concentrations of magnesium sulfate found in martian soils. *Astrobiology* **12,** 98–106 (2012).
122. Tapilatu, Y. H. *et al.* Isolation of hydrocarbon-degrading extremely halophilic archaea from an uncontaminated hypersaline pond (Camargue, France). *Extremophiles* **14,** 225–231 (2010).
123. Minai-Tehrani, D., Minoui, S. & Herfatmanesh, A. Effect of salinity on biodegradation of polycyclic aromatic hydrocarbons (PAHs) of heavy crude oil in soil. *Bull Environ Contam Toxicol* **82,** 179–184 (2009).
124. Jarvis, P., Rey, A. & Petsikos, C. Drying and wetting of Mediterranean soils stimulates decomposition and carbon dioxide emission: the 'Birch effect'. *Tree …* (2007).
125. Fierer, N. A proposed mechanism for the pulse in carbon dioxide production commonly observed following the rapid rewetting of a dry soil. *Soil Science Society of America Journal* (2003).
126. BROWN, H. J. & Gibbons, N. E. The effect of magnesium, potassium, and iron on the growth and morphology of red halophilic bacteria. *Can J Microbiol* **1,** 486–494 (1955).
127. Orellana, M. V., Pang, W. L., Durand, P. M., Whitehead, K. & Baliga, N. S. A role for programmed cell death in the microbial loop. *PLoS ONE* **8,** e62595 (2013).
128. DasSarma, S., Fleischmann, E. M. & Rodriguez-Valera, F. *Archaea: A laboratory manual - Halophiles*. (Cold Spring Harbor Laboratory Press, 1995).
129. Turkarslan, S. *et al.* Niche adaptation by expansion and reprogramming of general transcription factors. *Mol Syst Biol* **7,** 554 (2011).
130. Kahm, M., Hasenbrink, G., Lichtenberg-Fraté, H., Ludwig, J. & Kschischo, M. Grofit: Fitting biological growth curves. *Journal of Statistical …* (2010).
131. Onishi, H., McCance, M. & Gibbons, N. A synthetic medium for extremely halophilic bacteria. *Can J Microbiol* **11,** 365–373 (1965).
132. Sehgal, S. N. & Gibbons, N. E. Effect of some metal ions on the growth of Halobacterium cutirubrum. *Can J Microbiol* **6,** 165–169 (1960).
133. Baliga, N. S., Kennedy, S. P., Ng, W. V., Hood, L. & DasSarma, S. Genomic and genetic dissection of an archaeal regulon. *Proc Natl Acad Sci USA* **98,** 2521–2525 (2001).
134. Kvitek, D. J., Will, J. L. & Gasch, A. P. Variations in stress sensitivity and genomic expression in diverse S. cerevisiae isolates. *PLoS Genet* **4,** e1000223 (2008).
135. Cassat, J. *et al.* Transcriptional profiling of a Staphylococcus aureus clinical isolate and its isogenic agr and sarA mutants reveals global differences in comparison to the laboratory strain RN6390. *Microbiology (Reading, Engl.)* **152,** 3075–3090 (2006).
136. Fine, D. H. *et al.* Phenotypic variation in Actinobacillus actionmycetemcomitans during laboratory growth: implications for virulence. *Microbiology* **145,** 1335–1347 (1999).
137. Gaynor, E. C. *et al.* The Genome-Sequenced Variant of Campylobacter jejuni NCTC 11168 and the Original Clonal Clinical Isolate Differ Markedly in Colonization, Gene Expression, and Virulence-Associated Phenotypes. *Journal of …* (2004).
138. Wang, L. *et al.* Divergence involving global regulatory gene mutations in an Escherichia

coli population evolving under phosphate limitation. *Genome Biology and Evolution* **2,** 478–487 (2010).
139. Schmid, A. K., Reiss, D. J., Pan, M., Koide, T. & Baliga, N. S. A single transcription factor regulates evolutionarily diverse but functionally linked metabolic pathways in response to nutrient availability. *Mol Syst Biol* **5,** 282 (2009).
140. Bernstein, J. A., Khodursky, A. B., Lin, P.-H., Lin-Chao, S. & Cohen, S. N. Global analysis of mRNA decay and abundance in Escherichia coli at single-gene resolution using two-color fluorescent DNA microarrays. *Proc Natl Acad Sci USA* **99,** 9697–9702 (2002).
141. Evguenieva-Hackenberg, E. & Klug, G. New aspects of RNA processing in prokaryotes. *Curr Opin Microbiol* **14,** 587–592 (2011).
142. Gierga, G., Voss, B. & Hess, W. R. Non-coding RNAs in marine Synechococcus and their regulation under environmentally relevant stress conditions. *The ISME Journal* **6,** 1544–1557 (2012).
143. Guillier, M. & Gottesman, S. Remodelling of the Escherichia coli outer membrane by two small regulatory RNAs. *Mol Microbiol* **59,** 231–247 (2006).
144. Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. S. & Weissman, J. S. Genome-Wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling. *Science* **324,** 218–223 (2009).
145. Lange, C. *et al.* Genome-wide analysis of growth phase-dependent translational and transcriptional regulation in halophilic archaea. *BMC Genomics* **8,** 415 (2007).
146. Maier, T. *et al.* Quantification of mRNA and protein and integration with protein turnover in a bacterium. *Mol Syst Biol* **7,** (2011).
147. Munchel, S. E., Shultzaberger, R. K., Takizawa, N. & Weis, K. Dynamic profiling of mRNA turnover reveals gene-specific and system-wide regulation of mRNA decay. *Mol Biol Cell* **22,** 2787–2795 (2011).
148. Redon, E., Loubière, P. & Cocaign-Bousquet, M. Role of mRNA Stability during Genome-wide Adaptation of Lactococcus lactis to Carbon Starvation. *Journal of Biological Chemistry* **280,** 36380–36385
149. Romero-Santacreu, L., Moreno, J., Pérez-Ortín, J. E. & Alepuz, P. Specific and global regulation of mRNA stability during osmotic stress in Saccharomyces cerevisiae. *RNA* **15,** 1110–1120 (2009).
150. Storz, G., Vogel, J. & Wassarman, K. M. Regulation by Small RNAs in Bacteria: Expanding Frontiers. *Molecular Cell* **43,** 880–891 (2011).
151. Warringer, J., Hult, M., Regot, S., Posas, F. & Sunnerhagen, P. The HOG Pathway Dictates the Short-term Translational Response after Hyperosmotic Shock. at <http://www.ncbi.nlm.nih.gov.offcampus.lib.washington.edu/pubmed/20587780>
152. Kolkman, A. *et al.* Proteome analysis of yeast response to various nutrient limitations. *Mol Syst Biol* **2,** 2006.0026 (2006).
153. Buescher, J. M. *et al.* Global network reorganization during dynamic adaptations of Bacillus subtilis metabolism. *Science* **335,** 1099–1103 (2012).
154. Beisel, C. L. & Storz, G. Base pairing small RNAs and their roles in global regulatory networks. *FEMS Microbiol Rev* **34,** 866–882 (2010).
155. Beisel, C. L. & Storz, G. The Base-Pairing RNA Spot 42 Participates in a Multioutput Feedforward Loop to Help Enact Catabolite Repression in Escherichia coli. *Molecular Cell* (2011).

156. Gottesman, S. Proteolysis in bacterial regulatory circuits. *Annu. Rev. Cell Dev. Biol.* **19,** 565–587 (2003).
157. Levine, E., Zhang, Z., Kuhlman, T. & Hwa, T. Quantitative characteristics of gene regulation by small RNA. *Plos Biol* **5,** e229 (2007).
158. Mehta, P., Goyal, S. & Wingreen, N. S. A quantitative comparison of sRNA-based and protein-based gene regulation. *Mol Syst Biol* **4,** 221 (2008).
159. Shimoni, Y. *et al.* Regulation of gene expression by small non-coding RNAs: a quantitative view. *Mol Syst Biol* **3,** 138 (2007).
160. Wolf, D. M. & Arkin, A. P. Motifs, modules and games in bacteria. *Curr Opin Microbiol* (2003).
161. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* **37,** 1–13 (2009).
162. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protocols* **4,** 44–57 (2009).
163. Zhao, W. *et al.* Weighted gene coexpression network analysis: state of the art. *J Biopharm Stat* **20,** 281–300 (2010).
164. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9,** 559 (2008).
165. Facciotti, M. T. *et al.* Large scale physiological readjustment during growth enables rapid, comprehensive and inexpensive systems analysis. *BMC systems biology* **4,** 64 (2010).
166. Yao, A. I. & Facciotti, M. T. Regulatory multidimensionality of gas vesicle biogenesis in Halobacterium salinarum NRC-1. *Archaea* **2011,** 716456 (2011).
167. Storbeck, S., Rolfes, S., Raux-Deery, E. & Warren, M. J. A novel pathway for the biosynthesis of heme in Archaea: genome-based bioinformatic predictions and experimental evidence. *Archaea* (2010).
168. Schmid, A. K. *et al.* The anatomy of microbial cell state transitions in response to oxygen. *Genome Res* **17,** 1399–1413 (2007).
169. Baliga, N. S. *et al.* Coordinate regulation of energy transduction modules in Halobacterium sp. analyzed by a global systems approach. *Proc Natl Acad Sci USA* **99,** 14913–14918 (2002).
170. Corratgé-Faillie, C. *et al.* Potassium and sodium transport in non-animal cells: the Trk/Ktr/HKT transporter family. *Cell. Mol. Life Sci.* **67,** 2511–2532 (2010).
171. Kaur, A. *et al.* A systems view of haloarchaeal strategies to withstand stress from transition metals. *Genome Res* **16,** 841–854 (2006).
172. Morishita, R. *et al.* Ribonuclease activity of rat liver perchloric acid-soluble protein, a potent inhibitor of protein synthesis. *J Biol Chem* **274,** 20688–20692
173. Saeed, A. I. *et al.* TM4 Microarray Software Suite. *Meth Enzymol* **411,** 134–193
174. Valladares, F., SANCHEZ GOMEZ, D. & Zavala, M. A. Quantitative estimation of phenotypic plasticity: bridging the gap between the evolutionary concept and its ecological applications. *Journal of Ecology* **94,** 1103–1116 (2006).
175. Ruepp, A. & Soppa, J. Fermentative arginine degradation in Halobacterium salinarium (formerly Halobacterium halobium): genes, gene products, and transcripts of the arcRACB gene cluster. *J Bacteriol* **178,** 4942–4947
176. Pigliucci, M. Evolution of phenotypic plasticity: where are we going now? *Trends Ecol*

Evol (Amst) **20,** 481–486 (2005).
177. Whitehead, K. *et al.* An integrated systems approach for understanding cellular responses to gamma radiation. *Mol Syst Biol* **2,** 47 (2006).
178. Brooks, A. N., Turkarslan, S., Beer, K. D., Yin Lo, F. & Baliga, N. S. Adaptation of cells to new environments. *Wiley Interdiscip Rev Syst Biol Med* n–a–n–a (2010). doi:10.1002/wsbm.136
179. Zillig, W., Stetter, K. O. & Tobien, M. DNA-dependent RNA polymerase from Halobacterium halobium. *Eur J Biochem* **91,** 193–199 (1978).
180. Richards, C. L. *et al.* Plasticity, not adaptation to salt level, explains variation along a salinity gradient in a salt marsh perennial. *Estuaries and coasts* **33,** 840–852 (2010).
181. DasSarma, S. Mechanisms of genetic variability in Halobacterium halobium: the purple membrane and gas vesicle mutations. *Can J Microbiol* **35,** 65–72 (1989).
182. Pfeiffer, F. *et al.* Evolution in the laboratory: the genome of Halobacterium salinarum strain R1 compared to that of strain NRC-1. *Genomics* **91,** 335–346 (2008).
183. Robinson, J. L. *et al.* Growth kinetics of extremely halophilic archaea (family halobacteriaceae) as revealed by arrhenius plots. *J Bacteriol* **187,** 923–929 (2005).
184. Brooks, A. N., Turkarslan, S., Beer, K. D., Lo, F.-Y. & Baliga, N. S. Adaptation of cells to new environments. *Wiley Interdiscip Rev Syst Biol Med* **3,** 544–561 (2011).
185. Potential of halotolerant and halophilic microorganisms for biotechnology. *Extremophiles* **5,** 73–83 (2001).
186. Arulazhagan, P., Vasudevan, N. & Yeom, I. Biodegradation of polycyclic aromatic hydrocarbon by a halotolerant bacterial consortium isolated from marine environment. *International Journal of Environmental Science and Technology* **7,**
187. Kofler, R. *et al.* PoPoolation: a toolbox for population genetic analysis of next generation sequencing data from pooled individuals. *PLoS ONE* **6,** e15925 (2011).
188. Peng, Y., Leung, H. C. M., Yiu, S. M. & Chin, F. Y. L. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* **28,** 1420–1428 (2012).
189. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature Methods* (2012).
190. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).
191. Krzywinski, M. *et al.* Circos: an information aesthetic for comparative genomics. *Genome Res* **19,** 1639–1645 (2009).
192. Thorvaldsdóttir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinformatics* **14,** 178–192 (2013).
193. Aitchison, J. & Greenacre, M. Biplots of compositional data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **51,** 375–392 (2002).
194. Pawlowsky-Glahn, V. & Egozcue, J. Compositional data and their analysis: an introduction. *Geological Society, London, Special Publications* **264,** 1–10 (2006).
195. Hron, K., Filzmoser, P. & Thompson, K. Linear regression with compositional explanatory variables. (2012).
196. Woese, C. R. & Fox, G. E. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *... of the National Academy of Sciences* (1977).