

Large-Scale Effectors of Gene Expression and New Models of Cell Division in the
Haloarchaea.

by

Keely Dulmage

University Program in Genetics and Genomics
Duke University

Date: _____

Approved:

Amy Schmid, Supervisor

Greg Crawford, Chair

Nicolas Buchler

Amy Grunden

David MacAlpine

Dissertation submitted in partial fulfillment of
the requirements for the degree
of Doctor of Philosophy in the
University Program in Genetics and Genomics
in the Graduate School of
Duke University

2015

ABSTRACT

Large-Scale Effectors of Gene Expression and New Models of Cell Division in the
Haloarchaea.

by

Keely Dulmage

University Program in Genetics and Genomics
Duke University

Date: _____

Approved:

Amy Schmid, Supervisor

David MacAlpine

Amy Grunden

Greg Crawford, Chair

Nicolas Buchler

An abstract of a dissertation submitted in partial
fulfillment of the requirements for the degree
of Doctor of Philosophy in the University Program in
Genetics and Genomics in the Graduate School of
Duke University

2015

Copyright by
Keely Dulmage
2015

Abstract

Like most Archaea, the hypersaline-adapted organism *Halobacterium salinarum* exhibits characteristics from all three domains of life, including a eukaryotic histone protein, homologs of bacterial cell division proteins, and the propensity for genetic rearrangement shared by all. Here we investigate the function of histone protein in the Archaea. Transcriptomics, proteomics, and phenotypic assays of histone mutants determine that histone regulates gene expression and cell shape—but not genome compaction—in *H. salinarum*. We further explore the regulation of gene expression on a genome-wide scale through the study of genomic instability. Genomic deletions and duplications are detected through the meta-analysis of 1154 previously published gene expression arrays and 48 chromatin immunoprecipitation arrays. We discover that a 90 kb duplication event in the megaplasmid pNRC100 directly leads to increased gene expression, and find evidence that the chromosome is far more unstable than previously assumed. These events are all linked to the presence of mobile insertion elements. Finally, in response to questions generated by these experiments, we develop a novel time-lapse protocol for *H. salinarum* to answer basic questions about single-cell dynamics during division. Fluorescent labeling of homologs to bacterial cell division proteins confirms their involvement in cell division but localization dynamics contradict the basic bacterial model. The discovery of unusual facets of morphology during cell

division is consistent with these novel protein dynamics and opens up new avenues of inquiry into archaeal cell division.

Dedication

To my other half and co-conspirator.

Contents

Abstract	iv
List of Tables	xiii
List of Figures	xiv
Acknowledgements	xvi
1. Introduction	1
1.1 The Archaea: a diverse domain with characteristics of both Bacteria and Eukarya	1
1.2 <i>Halobacterium salinarum</i>	3
1.3 Chromosome architecture and histones	5
1.3.1 Eukaryotic DNA compaction is mediated by histone protein	5
1.3.2 Bacterial DNA compaction and nucleoid-associated proteins	7
1.3.3 Archaeal DNA compaction is mediated by a diverse set of proteins from all three domains	8
1.4 Genomic instability and gene expression.	12
1.4.1 Genomic instability in the Eukarya and mobile repetitive elements	13
1.4.2 Genomic instability in the Bacteria and mobile IS elements	14
1.4.3 Genomic instability in the Archaea and mobile IS elements	16
1.5. Cell cycle determinants and dynamics	18
1.5.1 Cell division in the Eukarya and the actomyosin ring	20
1.5.2 Cell division in the Bacteria and the Z ring	20
1.5.3 Multi-faceted systems of cell division in the Archaea	21

1.6 Thesis Work.....	23
2. Growth-phase-specific modulation of cell morphology and gene expression by an archaeal histone protein.....	27
2.1 Introduction.....	27
2.2 Methods	31
2.2.1 Strains and growth conditions.....	31
2.2.2 Strain construction	32
2.2.3 Southern blotting and detection.....	33
2.2.4 Sequence and structural analyses	33
2.2.5 Growth rate measurements.....	34
2.2.6 Resistance to UV radiation.....	35
2.2.7 Growth under oxidative stress	35
2.2.8 Light microscopy	36
2.2.9 Gene expression microarrays and analysis.....	36
2.2.10 Quantitation of S-layer glycosylation levels.....	38
2.2.11 Resistance to bacitracin.....	39
2.2.12 Shear stress resistance.....	39
2.2.13 Resistance to novobiocin	39
2.2.14 Identification of nucleic acid components of sucrose gradient profiles	40
2.2.15 Micrococcal nuclease (MNase) digestion of chromatin from <i>Halobacterium salinarum</i>	40
2.2.16 Protein identification and analysis.....	42

2.2.17 Comparison of NAP gene expression and quantitative proteomics data during growth.....	44
2.3 Results	45
2.3.1 HpyA, a fused histone heterodimer, is conserved throughout the halophilic archaea	45
2.3.2 Phenotypic characterization of HpyA function in <i>H. salinarum</i> physiology.....	48
2.3.3 Histone deletion results in growth phase-specific effects on gene expression.....	54
2.3.4 Halophilic histone is not required for genome-wide supercoiling, UV survival, or oxidative stress response	56
2.3.5 Histone regulates genes in cell wall and membrane-associated processes.....	58
2.3.6 Halobacterial DNA is not protected by a regular nucleosome array.....	61
2.3.7 Proteomics analysis of fractionated cells is consistent with open chromatin...	66
2.3.8 Expression of putative haloarchaeal NAPs suggests a dynamic DNA landscape	69
2.4 Discussion.....	71
2.4.1 Unexpected phenotypic consequences of histone gene dosage alteration suggest a unique function for haloarchaeal histones	71
2.4.2 Histone as a regulator of gene expression in open chromatin.....	73
2.5 Acknowledgments.....	76
3. Frequent genomic duplications and deletions are mediated by insertion elements in <i>Halobacterium salinarum</i>	77
3.1 Introduction.....	77
3.2 Methods	81
3.2.1 Strains and growth conditions.....	81

3.2.2 Gene expression microarrays and analysis for strains KAD100 and KAD101	81
3.2.3 Concatenation and normalization of 1154 gene expression arrays	82
3.2.4 Detection and mapping of correlated regions of gene expression in normalized data from 1154 arrays	83
3.2.5 qPCR detection of chromosomal instability	83
3.2.6 Detection of copy number variants in chromatin immunoprecipitation (ChIP) microarray data	85
3.2.7 Statistical analysis of IS element association with copy number variants	86
3.3 Results	86
3.3.1 Genome instability results in large-scale coordinate changes in gene expression	86
3.3.2 Chromosomal copy number variants are associated with mobile elements	94
3.4 Discussion	98
3.4.1 Genomic instability and biodiversity in <i>H. salinarum</i>	98
3.5 Acknowledgments	103
4. Distinct growth patterns and dynamics of the cell division proteins FtsZ1 and MinD2 in the Euryarchaeon <i>Halobacterium salinarum</i>	104
4.1 Introduction	104
4.2 Materials and Methods	109
4.2.1 Comparison of amino acid sequences	109
4.2.2 Growth conditions	109
4.2.3 Strain construction	110
4.2.4 Light microscopy and time-lapse imaging	110

4.3 Results	114
4.3.1 Environment-dependent growth modes of <i>H. salinarum</i>	114
4.3.2 FtsZ1 forms cell division-dependent Z rings at mid-cell.....	117
4.3.3 <i>H. salinarum</i> encodes a diverse and expanded MinD/ParA family	123
4.3.4 MinD2 localizes to the plane of division prior to membrane constriction and remains at the new poles.....	127
4.5 Discussion.....	133
4.5.1 Unique features of growth and cell division in a haloarchaeon.....	133
4.5.2 FtsZ1 and MinD2 participate in cell division in <i>H. salinarum</i> ; observed dynamics are distinct from those of Bacteria	136
4.6 Acknowledgements.....	138
5. Conclusions and Future Directions	140
5.1 Histones in the Archaea and evolution of histone function.....	140
5.2 Genomic instability and fitness	145
5.3 Euryarchaeal shape and cell division	150
5.4 Summary.....	153
Appendix A.....	155
Appendix B	156
Appendix C.....	163
Appendix D.....	171
Appendix E	179
Appendix F	182
References	184

Biography203

List of Tables

Table 1: Distribution of DNA architecture proteins in the three domains of life	9
Table 2: Questions addressed in this study	26
Table 3: Plasmids and strains used in this study.....	31
Table 4: Results of <i>t</i> -tests from quantitative analysis of mutant morphology	54
Table 5: Histone-regulated genes associated with S-layer synthesis and glycosylation ..	60
Table 6: Micrococcal nuclease digestion: tests and conditions.....	63
Table 7: Proportion of total peptides detected for putative nucleoid associated proteins in DNA and control fractions of <i>H. salinarum</i>	68
Table 8: qPCR primer pairs.....	84
Table 9: Plasmids used in this study	113
Table 10: Strains used in this study	113
Table 11: <i>H. salinarum</i> AAA ATPase protein family homology to MinD from <i>Bacillus cereus</i>	124

List of Figures

Figure 1: Haloarchaeal species encode a single conserved histone heterodimer.	47
Figure 2: Histone knockout mutants are readily generated and viable.	50
Figure 3: Histone dosage regulates cell shape and size.	53
Figure 4: Histone-regulated gene expression is growth-phase dependent.	56
Figure 5: Histone is not required for resistance to A) UV radiation, B) reactive oxygen species generated by Paraquat, or C) supercoiling stress generated by novobiocin.	57
Figure 6: Cell wall and cell membrane are unaffected by histone deletion.	61
Figure 7: Micrococcal nuclease digestion of haloarchaeal DNA.	65
Figure 8: Detection of nucleic acids in sucrose gradient fractions.	67
Figure 9: Gene and protein expression of putative NAPs as a function of growth.	70
Figure 10: Large-scale genomic changes influence gene expression.	88
Figure 11: Microarray processing and workflow of gene expression analysis.	90
Figure 12: Regions of correlated gene expression in the genome of <i>Halobacterium salinarum</i>	92
Figure 13: Distribution of chromosomal copy number variants (CNVs).	96
Figure 14: Time-lapse of <i>Halobacterium salinarum</i> cell division.	116
Figure 15: Cell shape comparison between wild-type <i>ura3</i> and smVenus-tagged <i>ftsZ1</i> strains.	118
Figure 16: Average localization of FtsZ1 particles over a population of log phase cells or stationary phase cells.	120
Figure 17: FtsZ1 localizes to the septum prior to constriction of the cell.	123

Figure 18: Alignment of MinD/ParA superfamily proteins from <i>H. salinarum</i> and <i>B. cereus</i>	126
Figure 19: <i>H. salinarum</i> MinD2 localizes to the poles and mid-cell.....	128
Figure 20: smVenus-MinD2 localizes to the cell periphery	129
Figure 21: MinD2 localizes to cell division plane and remains at new poles after division.	132
Figure 22: Comparison of FtsZ and MinD dynamics between <i>Bacillus subtilis</i> and <i>Halobacterium salinarum</i>	136

Acknowledgements

This thesis would not have been possible without support, discussions, and distractions from a host of friends, family, and fellow scientists.

First of all, I would like to thank my mentor, Amy Schmid, for her scientific enthusiasm, years of guidance, hours of editing, and support throughout my scientific wanderings. I'd also like to thank my committee members Amy Grunden, Dave MacAlpine, Greg Crawford, and Nick Buchler for sharing their insights and for all the helpful questions.

My compatriots in the Schmid lab have been invaluable during my graduate studies, not the least because I have had someone else to share my passion for Archaea with! I'd like to thank Kriti Sharma and Adrienne Pittman for buying me stuff, and for providing much needed witticisms and/or biting humor. Thank you to Horia Todor, "the other graduate student," who was always ready to distract with Buzzfeed or serious discussions about science. To all other Schmid lab members, past and present: thanks for making the lab a great place to work and have fun! And thanks to my undergraduate students Morgan Powe, Claudia Ofori-Marfoh, and especially Callan Corcoran, for their patience and hard work.

Much of the science presented here could not have taken place without support from several core facilities and individuals. This includes Sam Johnson and Yasheng

Gao at the Light Microscopy Core Facility; also Heather Hemric at the Duke Microarray Facility, Laura Dubois at the Duke Proteomics & Metabolomics Core Facility, and David Corcoran at the Integrative Genomic Analysis Shared Resource, all within the Duke Center for Genomic & Computational Biology. I'd like to thank the Kiehart Lab in Duke Biology for both their collaborative efforts and the use of equipment. Eli Wilber designed the smVenus fluorophore and was the first to observe localization patterns of FtsZ and MinD proteins. Janice Crawford constructed vectors for testing the functionality of smVenus mutations. Finally, the Keihart Lab kindly allowed us to use their ultracentrifuge and hybridization oven. Most of our research on cell division in *Halobacterium salinarum* was done as part of a collaboration with Ethan Garner's lab at Harvard University. I'd like to thank both Ethan Garner and especially Jenna (Ye-jin) Eun for their support, discussion, efforts, and for their kind hosting.

Much thanks to the communities of UPGG and the Center for Systems Biology for science lunches, retreats, student seminars, and in general, good scientific fun.

Funding was provided from several sources: two years of initial funding by a NIH training grant to UPGG, teaching assistantships from Duke Biology, travel grants from UPGG and the Graduate School, and a grant-in-aid from Duke Biology. Schmid lab funding was through grants to AKS from the National Science Foundation (NSF-MCB-10-52290 and 1417750) and the Duke Arts and Sciences Council Committee on Faculty Research.

1. Introduction

1.1 The Archaea: a diverse domain with characteristics of both Bacteria and Eukarya

The Archaea are a group of single-celled organisms, many of them extremophiles, which were first proposed to be a third domain of life in 1977 by Carl Woese and colleagues (Fox et al., 1977; Woese and Fox, 1977). Previously, these organisms were cataloged in the bacterial domain despite several unique physiological attributes, including distinct ether-linked lipids composed of isoprenoid subunits and cell walls composed of a crystalline proteinaceous S-layer rather than peptidoglycan. Upon the advent of 16S ribosomal RNA sequencing, however, it was clear to Woese and colleagues that these organisms were distinct from Bacteria.

The Archaea share many characteristics and molecular mechanisms with both Bacteria and Eukarya. These include bacterial signatures such as circular genomes, a predominance of transcription factors with helix-turn-helix motifs, and a simple cellular organization without a nucleus (Bell, 2005). On the other hand, the basal transcriptional apparatus and DNA replication machinery resemble the eukaryotic processes (Allers & Mevarech, 2005). Archaeal DNA repair and protein translation complexes share characteristics from both bacterial and eukaryotic organisms (Allers & Mevarech, 2005). The presence of eukaryotic mechanisms in a prokaryotic context is likely due to the origin of these mechanisms in the Archaea and is consistent with the current (albeit

contentious) theory that eukaryotes are derived from an ancient archaeon that engulfed a bacterium (Spang et al., 2015).

The Archaea can be broadly separated into two groups of organisms, the Euryarchaea and the TACK superphylum. The TACK superphylum includes the more basal species on the phylogenetic tree, including the phyla Thaumarchaeota, Aigarchaeota, Crenarchaeota, and Korarchaeota (Guy & Ettema, 2011). While there are few culturable examples of the latter three phyla, many species of Crenarchaea and the Euryarchaea are amenable to laboratory experiments. There are significant physiological differences between organisms in the Crenarchaea and the Euryarchaea. The Crenarchaea are monoploid or diploid, employ the eukaryotic membrane scission ESCRT-III complex in cell division, and generally do not encode histone proteins (Ettema & Bernander, 2009; Zerulla & Soppa, 2014; Zhang et al., 2012). In contrast, the Euryarchaea are polyploid, likely employ bacterial homologs to undergo cell division, and most species encode histone proteins (Allers & Mevarech, 2005; Zerulla & Soppa, 2014; Sandman & Reeve, 2006). Along with many additional molecular differences between these phyla, the diversity of archaeal organisms suggests that research performed in a single phylum or class of archaeal species cannot automatically be extrapolated to other archaea. Indeed, while much is known about cell division in the hyperthermophilic crenarchaeon *Sulfolobus solfataricus*, what little is known about cell division in the Euryarchaea is incompatible with these studies and must be investigated

independently. Because little is known about the Archaea, this is an exciting time to study these organisms; with the aid of high-throughput techniques such as proteomics and transcriptomics, a wealth of basic physiological information may be readily discovered.

1.2 Halobacterium salinarum

Halobacterium salinarum, an organism from the archaeal phylum Euryarchaea, has been studied for almost 100 years (Harrison & Kennedy, 1922), albeit under the impression that it was a bacterial organism. Belonging to the Halobacteria, *H. salinarum* requires high concentrations of Na⁺ for membrane stability, growing optimally at 4.2M NaCl and lysing in concentrations less than 2M NaCl (Steensland & Larsen, 1969). Like all halophilic Archaea, the adaptation of this organism to high salt conditions is not mediated by the exclusion of salt from the cytoplasm as seen in halophilic Bacteria, but by balancing the osmolarity between the environment and the cytoplasm through the import of K⁺ to equimolar concentrations (Perez-Fillol & Rodriguez-Valera, 1986). Because such conditions generally lead to insolubility of proteins, the halophiles have evolved acidic proteomes to maintain protein stability (Kennedy et al., 2001). These organisms are therefore uniquely adapted to high salt environments and are distinct from the halophilic bacteria.

Among the Archaea, the halophiles are excellent model organisms. As many archaea are hyperthermophiles or obligate anaerobes, they can require specialized equipment for cultivation (Horn et al., 1999; Atomi et al., 2012). Most halophiles are mesophiles and many species can be grown in peptone-based media with the simple addition of salt to the required concentration. *H. salinarum* grows rapidly at 37°C and can be cultivated in standard lab conditions. Considerable genetic tools have been developed, from genomic knock-outs and transposon-mediated mutagenesis to a well-annotated genome (Ng et al., 2000) with a significant database of transcriptomics (Brooks et al., 2014; Bonneau et al., 2007; Facciotti et al., 2007; Sharma et al., 2012; Whitehead et al., 2009) and proteomics experiments (Van et al., 2008; Klein et al., 2005; Klein et al., 2006; Blonder et al., 2004; Bisle et al., 2006).

For these reasons, we use the halophilic euryarchaeon *H. salinarum* to answer basic questions about archaeal physiology. We investigate the ancestral, conserved, function of histone protein from the Archaea to the Eukarya. We ask how the organization and stability of genomic DNA influences transcription. Finally, we make seminal forays into the study of euryarchaeal cell division and confront assumptions about the mechanisms involved.

1.3 Chromosome architecture and histones

The genetic material of all organisms must be compacted because the extended length of DNA within the cell exceeds the diameter of the cell itself. This is especially true in humans, where the extended length of the genome is approximately 2 meters (Annunziato, 2008), but is also a feature of Bacteria, whose smaller genomes fit into a correspondingly smaller volume (Phillips, Kondev, & Theriot, 2009).

Diverse organisms employ a variety of DNA packaging strategies. This includes supercoiling, bending, wrapping, and tethering, all of which may be mediated by specialized proteins (Luisterburg et al., 2008; Sandman & Reeve, 2006). However, DNA is not merely restructured: both supercoiling and structural deformation of DNA by proteins can affect gene expression (Luisterburg et al., 2008). This interplay is common to all organisms regardless of the compaction method; we discuss the ramifications below.

1.3.1 Eukaryotic DNA compaction is mediated by histone protein

In eukaryotes, the bulk of DNA compaction is conferred by histone proteins, which share the common structure of a histone fold domain comprised of three alpha helices separated by two loop domains (Sandman & Reeve, 2006; Arents & Moudrianakis, 1995). A histone octamer, consisting of two H3-H4 and two H2A-H2B heterodimers, forms the

core of what is termed a “nucleosome” by wrapping 150 bp of DNA. To stabilize this complex, a cap is formed by histone H1, which, despite the name, does not contain a histone fold domain (Ramakrishnan, 1997). Nucleosomes compact most of the eukaryotic genome, often leaving the regulatory regions unpacked in order to allow for initiation of transcription (Kornberg, 2007).

Nucleosome occupancy in the genome is predictive of transcription in eukaryotic organisms, with post-translational modification (PTM) of histones regulating occupancy. Eukaryotic histone proteins have extended unstructured N-terminal domains, and while the entire protein may be post-translationally modified, the N-terminal domain contains a high density of such sites and it is the regulation of these PTMs that leads to the regulation of gene expression (Tan et al., 2011). Highly chromatinized regions of DNA are generally completely transcriptionally silent (Kornberg, 2007). Because nucleosomes sequester DNA, they must be relocated before transcription can take place. PTMs alter the affinity of histone proteins for DNA and can lead to either tightening or loosening of the DNA wrap (Cosgrove et al., 2004). Therefore, some modifications, such as methylation, may repress or activate gene expression depending on the degree of methylation and genomic context, whereas other modifications, such as acetylation, activate gene expression (Greer & Shi, 2012; Ramakrishnan, 1997). PTMs are not static: rather, the dynamic distribution of N-terminal tail modifications is an important

determinant of regulatory programs for cellular growth, development, and differentiation (Cantone & Fisher, 2013).

1.3.2 Bacterial DNA compaction and nucleoid-associated proteins

Histone proteins have not been observed in Bacteria. Instead, bacterial genomes encode an assortment of proteins that both change the conformation of DNA and regulate gene expression. While the proteinaceous genetic material may still be referred to as chromatin, the bacterial term is “nucleoid” and the associated architectural proteins are collectively referred to as nucleoid associated proteins (NAPs; Luijsterburg et al., 2008). NAP functions range from DNA benders such as HU, IHF, and Fis to DNA bridgers such as Lrp, SMC, and H-NS (Table 1; Luijsterburg et al., 2008). In certain concentrations *in vitro*, HU and Lrp proteins have also been shown to wrap DNA, albeit more loosely than eukaryotic histone (Jafri et al., 1999; Van Noort et al., 2004), suggesting that NAPs may serve multiple architectural roles inside the cell.

As in Eukarya, DNA architecture can directly influence gene expression in Bacteria. *In vivo*, HU stabilizes loop formation, which in turn regulates expression (Becker et al, 2007). Supercoiling in *E. coli* leads to the formations of looped domains with an average size of 10kb, and the expression of genes within each domain is correlated (Postow et al., 2004). In *E. coli*, a double deletion of both IHF homologs led to 477 significant gene expression changes, and a HU double deletion mutant led to 1266

significant gene expression changes (or 56% of all genes in *E. coli*; Prieto et al., 2012).

Therefore, while mechanisms of compaction and the protein effectors may vary between organisms and domains, compaction *per se* appears to be a universal regulator of gene expression.

1.3.3 Archaeal DNA compaction is mediated by a diverse set of proteins from all three domains

The Archaea employ a variety of proteins to compact their genomes (Table 1). The specific set of proteins involved in DNA structure varies between phylum and class and can include histones, bacterial NAPs, and Archaea-specific proteins (Peeters et al., 2015). Most Euryarchaea and several members of early branching archaeal phyla encode histone homologs, suggesting that these proteins have an ancient origin in the archaeal kingdom (Sandman & Reeve, 2006). Some Archaea, such as *Methanothermus fervidus* and *Thermococcus kodakarensis*, package the genome using histone protein to wrap DNA (Pereira et al., 1997; Maruyama et al., 2011). However, chromatin is heterogeneous in *T. kodakarensis* and may be coated with other NAPs. For example, TrmBL2 forms thick chromatin fibers in this organism (Maruyama et al., 2011). In contrast, in some species of the family Methanosarcinaceae, of which all known members encode histone homologs, the archaeal protein MC1 was found to be the primary component of chromatin (Chartier et al., 1985; Laine et al., 1986). Some archaea also encode homologs of bacterial NAPs. For example, *Thermoplasma* species possess a HU-family protein

(DeLange et al., 1981). Similarly, many archaeal genomes encode Lrps, which are environmentally responsive transcription factors (Plaiser et al., 2014). The halophiles encode Dps homologs, which function both as NAPs and as ferritins (Reindel et al., 2002).

Table 1: Distribution of DNA architecture proteins in the three domains of life

Mode of Binding	Eukarya	Bacteria	Archaea
Bending	HMG	Lrp HU Fis IHF	Lrp HU Cren7 Sul7d MC1
Bridging	SMC	SMC Lrp H-NS	SMC
Wrapping	Histone	Lrp	Histone

While genome annotations and proteomics data support the hypothesis that a diversity of proteins in the chromatin environment package DNA in archaeal species, experiments to date have focused on a single protein or time point during growth to the exclusion of all others. Indeed, it is possible that disparities in conclusions between *in vivo* and *in vitro* experiments or between species may simply be explained by a dynamic chromatin environment. In bacterial species, the predominant chromatin protein is strongly growth phase dependent (Luisterburg et al., 2006), and related to the transcriptional activity of the genome (Prieto et al., 2012). Such dynamism is also observed to a lesser extent in the eukaryotes where the relative abundance of histone

variants may change during various phases of growth (Talbert & Henikoff, 2010).

Therefore, it is necessary to determine whether archaeal chromatin is also a dynamic entity whose protein constituency is dependent on the growth phase of an organism.

Although we have some understanding of the proteins involved in archaeal DNA compaction, the role of archaeal histones and NAPs in gene expression is still unclear. Archaeal histone proteins are distinct from their eukaryotic counterparts in that they lack N-terminal tails and thus are unlikely to be post-translationally modified dynamically in order to regulate gene expression (Sandman & Reeve, 2006). *In vitro* experiments suggest that histone may inhibit gene expression, but not to the extent observed in eukaryotes (Wilkenson et al., 2010; Xie & Reeve, 2003). In contrast, in the Class II methanogen *Methanosarcina mazei* Gö1, deletion of the sole histone led to decreased transcription of 25% of annotated transcripts, suggesting that histone is an activator in this species (Weidenbach, 2007). In *T. kodakarensis*, single deletions of either histone led to bi-directional gene expression changes in only a handful of transcripts, despite later evidence that suggests the entirety of the genome contains histone-occupied nucleosome binding sites (Cubanova et al., 2012; Nalabothula et al., 2013). Thus, current transcriptional data are contradictory and suggest that histones may have different functions in different classes of archaea.

The Halobacteria are unusual among Archaea and Eukaryotes in that all sequenced genomes contain a single histone protein homolog whose open reading frame

encodes a tandem fused histone heterodimer. This histone structure has only been observed in one other organism, *Methanopyrus kandleri* (Fahrner et al., 2001). Sharing the conserved histone-encoding operon is a putative histone deacetylase gene. Although the function of the deacetylase has not been investigated, it is known to be essential in *Haloferax volcanii* (Altman-Price & Mevarech, 2008). Electron microscopy (EM) of proteinaceous DNA from *H. salinarum* has shown evidence of the typical “beads-on-a-string” structure of nucleosome-packaged chromosomes, which was found to be more prevalent in stationary phase cells (Takayanagi et al., 1992; Shioda et al., 1989). Interpretation of these results, however, is complicated by the high internal salt concentrations in halophilic archaea (up to 5M K⁺): EM has shown that NaCl concentrations over 150 mM can induce formation of structures resembling nucleosome particles in naked DNA (Perez-Fillol & Rodriguez-Valera, 1986; Eickbush & Moudrianakis, 1978). In 2013, Ammar and colleagues sequenced DNA from MNase “mononucleosomes” and observed phasing around transcriptional start sites as would be expected for histone positioning. However, the protein components of the MNase-digested particles were not identified in this study, leaving unanswered questions. Despite these studies, research into haloarchaeal chromatin still lags behind other archaeal organisms. The function of histone in transcriptional regulation is unknown and few mutant studies have been performed. It is especially critical to understand histone function in the Halobacteria, as the gene encoding histone is in the same operon

as a histone deacetylase homolog in all sequenced haloarchaeal species. This suggests that these proteins might be post-translationally modified, which has not been observed in other archaea. These organisms could prove a crucial link in understanding the evolution of histone, not only as a DNA architecture protein, but also as a regulator of gene expression in the eukaryotic mode. In Chapter 2, we answer several of these questions through investigations into the function of histone in the euryarchaeon *H. salinarum*.

1.4 Genomic instability and gene expression.

Genomic instability is common to all organisms and encompasses both DNA mutations and a wide range of genomic rearrangements, such as amplification, deletion, inversion, translocation, and gene fusions, which we discuss here (Darmon & Leach, 2014).

Genomic rearrangements may be generated by many different factors. Intrinsically, DNA replication may generate recombinogenic events through stalling of replication forks or collision of DNA and RNA polymerase (Kolodner et al., 2002). Many organisms encode multiple copies of small transposable elements throughout their genomes that are prone to mobilization; these randomly interspersed homologous tracts can serve as templates during homologous recombination despite improper genomic location.

Environmental factors, such as desiccation and UV, can lead to the generation of double-stranded DNA breaks (DSBs); improper repair of DSBs can also generate

rearrangements. While these are common methods of inducing genomic instability, this is by no means an exhaustive list. Although most genomic rearrangements are likely to be neutral or deleterious to a cell, a small subset may be beneficial and thus provide a means for organisms to generate diversity and evolve.

1.4.1 Genomic instability in the Eukarya and mobile repetitive elements

Genomic instability in the eukaryotic domain is well understood because most examples of this phenomenon in humans are disease-causing. Most, if not all, forms of cancer are associated with genomic rearrangements (Heyer, 2015). The human genome is dense with transposable elements: nearly 50% of the human genome is composed of repetitive sequences, a large fraction of which are mobile recombinogenic LINE retrotransposons or Alu elements, of which approximately one million copies can be found in a given individual (Hedges & Deininger, 2007). Rearrangements frequently occur when DSBs are erroneously repaired using non-allelic homologous recombination between these repetitive elements, and therefore can be considered a byproduct of DNA repair (Hedges & Deininger, 2007). This is exemplified by an assortment of mutated alleles which may predispose an individual to cancer, such as BRC1, BRC2, RAD51, XRCC2, and PTEN among many others, all of which share the common theme of being required for DNA repair by homologous recombination (Heyer, 2015). With the exception of meiotic crossovers, which are essential for the generation of diversity between

generations (Mancera, 2008), genome instability is relatively infrequent and actively suppressed due to negative fitness consequences.

1.4.2 Genomic instability in the Bacteria and mobile insertion sequence elements

While most bacterial species encode a far smaller proportion of repetitive elements than can be found in eukaryotes, genomic instability and rearrangements are common and can also be mediated by homologous recombination (Darmon & Leach, 2014). Insertion sequences (IS) are small transposable elements flanked by internal repeats, and recombination between these sites lead to genomic rearrangements (Brugger, 2002).

When two such elements are in the same orientation deletion can occur; when in an opposite orientation this can result in inversion (Arber, 2000). Tandem duplications may also be generated and can occur at frequencies of 10^{-3} in a population. These frequencies may be further amplified by homologous recombination; up to 100-fold amplifications have been observed (Darmon & Leach, 2014). While the breakpoints of genomic rearrangements are often located in IS elements, other features may result in erroneous recombination. The minimal sequence homology required for homologous recombination is estimated to be 20–100 bp, and any sites meeting this criterion can be targets; IS elements are merely more numerous and widely distributed throughout the genome (Darmon & Leach 2014; Hedges & Deininger, 2007).

As in eukaryotes, the majority of genomic rearrangements in bacterial species will be deleterious or neutral. In the absence of selection these changes are not normally maintained over time in a population (Arber, 2000). However, such events are important generators of diversity and can aid in rapid response to stress. While beneficial events may be rare, under selective pressure a given rearrangement may provide sufficient fitness benefits to amplify the proportion of a population carrying this genetic change (Darmon & Leach, 2014). This is especially common in pathogenic microbes. For example, gene conversion in the VlsE surface lipoprotein genes of *Borrelia burgdorferi*, the causative agent of Lyme disease, can lead to host immune system invasion (Wisniewski-Dye & Vial, 2008). Inversions resulting in the movement of a promoter in *Campylobacter fetus*, an opportunist human pathogen that also causes ovine abortion, affects the production of surface layer proteins and ultimately the rapidity of host-antibody response (Grogono-Thomas et al., 2003). Examples can also be found in eukaryotic pathogens: *Candida albicans* generates resistance to the fungicide fluconazole by amplifying the left arm of chromosome five, which contains both a cell wall biogenesis gene and a drug efflux pump (Selmecki et al., 2009). Therefore, despite the apparent drawbacks of gross changes to the genome, genomic rearrangements can be an important source of phenotypic diversity in a population.

1.4.3 Genomic instability in the Archaea and mobile IS elements

IS elements are common within the archaeal domain and are generators of genomic rearrangements and phenotypic diversity. Many species of the *Sulfolobus* family encode several classes of active IS elements. Selection for 5-FOA resistance has been shown to yield mutants at a rate of 10^{-4} to 10^{-5} , with each mutant containing IS element insertions at the *pyrEF* locus, rather than individual nucleotide changes (Martusewitsch et al., 2000). In addition, 4/7 of the analyzed mutants had additional large-scale chromosomal rearrangements, indicating that the IS insertion event was recombinogenic (Martusewitsch et al., 2000). These IS-induced recombination events can generate large-scale genomic changes in seemingly isogenic cultures. For example, when two colonies of *Sulfolobus solfataricus* from the same plate were compared genetically, one had a 124 kb deletion (Redder & Garrett, 2006).

Most literature concerning IS element-mediated genetic rearrangements concerns the halophilic archaea (Brugger, 2002). IS elements are common within the genomes of all studied halophile species and are known to be active (Brugger, 2002). In *Halobacterium salinarum*, mobility of IS elements generates gas vesicle null isolates at a frequency of 10^{-2} to 10^{-4} , these isolates are readily identifiable at the plate level as translucent colonies that differ visibly from opaque parent strains (Pfeifer et al., 1988). One such gas vesicle null colony was isolated in 1967 and propagated as a lab reference strain (R1); the gas vesicle positive strain (NRC-1) is equally well-studied (Sapienza et

al., 1982). Comparison of the two strains 41 years after selection in the lab is telling. Although only 12 nucleotide differences between the genomes of the strains are not attributable to rearrangements, the megaplastids contain the highest density of IS elements (approximately 1 element every 7.6 kb in NRC-1), leading to extensive rearrangements including the generation of two additional plasmids in R1 and loss of synteny between strains at these locations (Pfeiffer et al., 2008). In NRC-1, the high frequency of genomic rearrangements as observed on the megaplastids may be attributed both to the density and mobility of the interspersed IS elements (DasSarma et al., 2006) and the highly efficient homologous recombination DNA-repair machinery of this organism (Ng et al., 2000; McCready et al., 2005; Woods & Dyal-Smith, 1997; Kotteman et al., 2005).

While much is known about the activity of IS elements and homologous recombination in *H. salinarum*, the direct effects of genomic instability on gene expression and biodiversity have not yet been explored. Most rearrangements have been observed in the absence of selection. As described above, in organisms from bacteria to humans, genomic instability can directly affect fitness. The organism *H. salinarum* harbors a large family of mobile interspersed homologous tracts; therefore we can expect that genomic instability is a major driver of adaptation. However, the true rate of genomic rearrangements is likely to be underestimated, as no study to date has tracked the changes that occur in the major chromosome or the changes which lack

obvious phenotypic consequences. We investigate these questions in Chapter 3 of this study.

1.5. Cell cycle determinants and dynamics

As described in sections 1.3–1.4 above, the genome is a dynamically regulated entity. From variable gene dosage during DNA replication to chromatin protein components and developmental patterns of gene expression, these phenomena must be considered temporally because they are dynamic processes. In Archaea, much is known about population dynamics, such as the gene regulatory networks during growth (Facciotti et al., 2007; Lundgren & Bernander, 2007), and physiological changes such as cell size (Todor et al., 2014), gas vesicle production (Larsen et al., 1967), and the distribution of flagella (Alam & Oesterhelt, 1984). What is lacking, however, is the understanding of single-cell dynamics. Single cell and single molecule dynamic experiments using microscopy methods have become routine for eukaryotic model organisms, whose larger cell size is easily tracked over time and for whom cell populations may be either chemically synchronized or sorted by age (Banfalvi, 2011). Bacteria are generally too small to track subcellular structures by traditional light-microscopy techniques: the width of an average cell is approximately the size of a wavelength of visible light. However, populations may be separated by age and show characteristic expression profiles and physiologies depending on the phase of growth (Banfalvi, 2011; Madren et

al., 2012). In recent years, microscopy techniques, such as structured illumination (SIM), stochastic optical reconstruction microscopy (STORM), and photo-activated localization microscopy (PALM), have been developed which enable single-molecule imaging at scales smaller than the diffraction limit of light (Gustafsson, 2000; Rust et al., 2006, Betzig et al., 2006). Similar advances have been slow to be adapted to the field of archaeal physiology. This can be partially attributed to the extremophilic nature of many model archaeal species, as microscopes must be specially adapted to resist high temperatures (Horn et al., 1999). In addition, few chemicals used to synchronize the growth of cells within cultures of eukaryotic or bacterial model organisms are suitable for this use in the Archaea. Until recently, many archaea were not genetically tractable (Herrmann & Soppa, 2002; Atomi et al., 2012). Therefore, literature on cell division in the Archaea is largely derived from still images of possibly asynchronous populations and thus the dynamics and mechanisms underlying the archaeal cell cycle remain unclear (Herrmann & Soppa, 2002; Malandrino et al., 1999; Poplawski & Bernander, 1997). Due to our interest in chromatin structure, DNA dosage, and gene expression, it is imperative to develop methods to dissect these phenomena on a single-cell scale and link them to the basic features of cell division. We address these challenges in Chapter 4 of the current study by developing time-lapse growth protocols for *H. salinarum*, which reveal an organism that is flexible in both morphology and placement of the division plane.

1.5.1 Cell division in the Eukarya and the actomyosin ring

Organisms in all three domains of life have similar requirements for coordination and tight regulation of the events of the cell cycle, including DNA segregation, cytokinesis, and formation of new cell wall or membrane at the site of division. In eukaryotes, cell division is mediated by cytoskeletal components. The mitotic spindle, composed of microtubules and molecular motors, segregates DNA (Schweizer et al., 2014). An actomyosin ring, consisting of actin and myosin II, forms at the site of division (Mishra et al., 2014). Constriction of the actomyosin ring both generates the force required for division and recruits additional membrane remodelers to the site (Mishra et al., 2014). In mammals and fission yeast, the ESCRT III complex cuts the membrane once the bud neck is formed by constriction (Bhutta et al., 2014). While the exact mechanisms of cytokinesis show some variability between eukaryotes, the formation of a cytoskeletal ring at the site of division is a common feature.

1.5.2 Cell division in the Bacteria and the Z ring

Bacteria encode homologs of the eukaryotic cytoskeletal protein actin; however, in bacteria actin homologs do not serve as cytoskeleton but are involved in directing the synthesis of new cell wall as the cell elongates (Eun et al., 2015). FtsZ proteins, which function as self-activating GTPases and are members of the tubulin protein family, form a contractile ring at the site of cell division to serve as a scaffold for associated fission

proteins (Erickson et al., 2010). In *Escherichia coli*, the concerted action of the MinCDE proteins inhibits the polymerization of FtsZ subunits and prevents the formation of the Z ring at the cell poles, thus localizing the FtsZ ring to the cell midline (Raskin & de Boer, 1999; Shih & Zheng, 2013). A host of other proteins are required for cell division in *E. coli*, including ZipA, which tethers FtsZ to the periphery of the cell, and ZapA, which positively regulates polymerization and stability. In *Bacillus* species, MinJ and DivIVA, instead of MinE, position MinD (Adams & Errington, 2009; Rowlett & Margolin, 2015). Although different proteins are involved in cell division in the Bacteria and the Eukarya, the model of the cytoskeletal contractile ring is conserved.

1.5.3 Multi-faceted systems of cell division in the Archaea

Relatively little is known about cell division in the Archaea, with many of the current models being derived from protein function predicted by homology to counterparts in other domains. As in all facets of archaeal physiology, these organisms present characteristics of cell division from both the bacterial and eukaryotic domains of life. Crenarchaeal cell division is driven by homologs of the eukaryotic ESCRT-III membrane scission proteins (Cdv), whereas the Euryarchaeal cell division apparatus is predicted to resemble that of Bacteria (Makarova et al., 2010). The hypothesis that Euryarchaeal cell division resembles that of Bacteria is supported by the ubiquitous presence of FtsZ homologs in this phylum, several of which have been characterized. For example,

haloarchaeal FtsZ homologs have been shown to form contractile rings at the site of cell division and to undergo GTPase activity *in vitro* (Duggin et al., 2015; Margolin et al., 1996; Wang & Lutkenhaus, 1996). While little research into the relationship between FtsZ and cell division has been performed in other classes of archaea, heterologous expression of *Thermoplasma* FtsZ leads to a filamentous phenotype in *E. coli* and suggests that this function is conserved (Yaoi et al., 2000). Many archaea encode the FtsZ-positioning protein MinD, but all lack MinD-positioning proteins and MinC, which together are required for inhibition of FtsZ polymerization and thus positioning of the Z ring. The absence of homologs to these proteins precludes the inference that the distribution of MinD leads to negative regulation of FtsZ polymerization in these organisms (Makarova et al., 2010). Additional experiments must be performed in order to understand the relationship between MinD and FtsZ in this context, and to determine whether functional homologs exist for bacterial proteins that are considered to be essential for cell division. We address these questions in Chapter 4 of the current study.

Most FtsZ studies in the euryarchaea have been performed in the Haloarchaea. All sequenced organisms have an expanded FtsZ gene family, and it has been suggested that while *Haloferax volcanii* FtsZ1 and FtsZ2 represent functional cell division proteins, the other 6 homologs are more likely to be structural tubulins – this hypothesis is supported by greater sequence conservation in the tubulin identity sequence, wild-type cell division in knock-out mutants, and differential morphology of an overexpression

strain (Duggin et al., 2015). The implication of this hypothesis is that Haloarchaea, and possibly all Euryarchaea, contain cytoskeletal elements that are involved in cell shape determination. To date, however, only a single study has observed FtsZ-family proteins in live cells.

The Haloarchaea are extremely diverse in shape – species may be rods, cocci, flattened discs, pleiomorphic, or may switch between forms depending on growth phase, growth medium, or nutrient availability (Oren et al., 2009). The Haloarchaea also include the intriguing species *Haloquadratum walsbyi*, which is square and can be found growing together in thin sheets resembling postage stamps (Burns et al., 2007). While it is clear that the proteinaceous cell wall of the Haloarchaea, the S-layer, must be both intact and glycosylated for cells to maintain shape, this does not explain the diversity of forms observed (Mescher & Strominger, 1976). The presence of a diverged tubulin-like FtsZ family that may form a cytoskeleton scaffold provides a starting point for studies into the shape-determining factors behind haloarchaeal morphological diversity, and, ultimately, the morphological diversity of the Archaea.

1.6 Thesis Work

In this study we aim to answer many of the questions posed above using *H. salinarum* as a model organism (see Table 2 for summary). In Chapter 2 we describe an investigation into haloarchaeal chromatin and histone function through the characterization of histone

deletion and overexpression mutants. Transcriptomics and analysis of physical and UV stress response complements experiments in other species, while proteomic analysis of chromatin serves to directly identify the principal DNA architecture proteins in this organism. Quantitative microscopy delves further into histone mutant phenotypes. In Chapter 3, motivated by evidence that genome instability may result in duplication events as large as 90 kb, we perform novel meta-analysis of whole-genome sheared DNA control samples from previously published chromatin immune-precipitation microarrays using a DNA segmenting algorithm. The relationship of the observed genome instability to insertion sequence (IS) element abundance and gene expression is examined. In Chapter 4, we describe the development of methods for single-cell analysis, as well as determining baseline wild-type measurements of archaeal cell cycle and growth for further comparisons with mutant strains. This addresses the necessity of population and single-cell dynamics as metrics for the study of chromatin structure, gene expression regulation, and genome dosage. Specifically, we developed time-lapse microscopy protocols for the live-cell observation of fluorescently tagged FtsZ1 and MinD2 proteins in *H. salinarum*, both of which are homologous to bacterial cell division proteins and thus may serve as landmarks of cell division progress for future studies of growth dynamics. Briefly, we conclude that: a) histones are not required for genome packaging in *H. salinarum* but have likely retained a conserved gene regulation activity, b) that genomic instability as mediated by IS elements extends to the major chromosome

and has direct effects on gene expression, and c) that *H. salinarum* exhibits several unique features of cell division, likely related to the non-canonical dynamics of bacterial cell division protein homologs FtsZ1 and MinD2. A large set of new questions was generated, including questions about the primary method of DNA compaction, the fitness consequences of genomic instability, and the additional protein components and determining factors of euryarchaeal cell division. In the final discussion chapter, we describe the future of this research and the ramifications of our work in a greater context.

Table 2: Questions addressed in this study

Section	Questions	Experiments	Conclusions
Chapter 2	<p>What is the ancestral function of histone?</p> <ul style="list-style-type: none"> Given that other classes of Archaea use histone to package the genome, is this function extended to the Haloarchaea? If not, what other proteins might be involved in DNA compaction? 	<ul style="list-style-type: none"> Knock-out and overexpression phenotypes Proteomics Gene expression dynamics 	<ul style="list-style-type: none"> Histone is likely involved in gene expression but <i>not</i> genome compaction in the Haloarchaea. There are no abundant putative genome compactors in the proteome of <i>H. salinarum</i>.
Chapter 3	<p>What are the effects of genomic instability in the Haloarchaea?</p> <ul style="list-style-type: none"> Where and how often do megaplasmid rearrangements occur? What effects do large-scale duplications and deletions have on gene expression? Does genomic instability extend to the major chromosome? 	<ul style="list-style-type: none"> qPCR and gene expression Meta-analysis of 1154 gene expression microarrays Meta-analysis of 48 ChIP microarrays 	<ul style="list-style-type: none"> A large-scale duplication leads to increased gene expression. pNRC100 may be lost in its entirety. Instability breakpoints may be found at any IS element. Genomic instability is observed at IS elements in the megaplasmid.
Chapter 4	<p>How do euryarchaeal cells divide?</p> <ul style="list-style-type: none"> What does cell division look like in <i>H. salinarum</i>? (This includes timing, features, etc.) Are homologs of bacterial cell division proteins involved in this process? 	<ul style="list-style-type: none"> Time-lapse microscopy Fluorescence labeling of cell division proteins Cell shape quantitation 	<ul style="list-style-type: none"> Cells divide by fission. Cell division may be either symmetric or asymmetric. Cells may divide as rods or flat amorphous cells. FtsZ1, FtsZ2, and MinD2 appear to be involved in division but likely do not interact by the bacterial mechanism.

2. Growth-phase-specific modulation of cell morphology and gene expression by an archaeal histone protein

Chapter 2 was modified from a manuscript of the same title published in mBio (2015).

The authors were Keely Dulmage, Horia Todor, and Amy Schmid. All experiments were conceived of and performed by Keely Dulmage with the exception of the PAS stain experiment, which was performed by Horia Todor.

2.1 Introduction

All organisms must compact their genomes: regardless of the species, the extended length of genetic material greatly exceeds the diameter of the cell or nucleus. Eukaryotic genomes are primarily packaged into nucleosome particles consisting of a histone protein octamer and approximately 150bp of DNA (Clark, 2010). Histone proteins generally contain a core histone fold domain consisting of three α -helices and an N-terminal tail. Covalent modifications of the histone tail, and in some cases the histone core, change the strength of the association of the histone octamer with DNA.

Modification leads to passive regulation of gene expression by modulating accessibility to the basal transcriptional apparatus (Kornberg, 2007; Tan et al., 2011). Like histones, bacterial nucleoid associated proteins (NAPs, e.g. HU, H-NS, Fis) bind DNA non-specifically, causing compaction (Luijsterburg et al., 2006). These DNA-NAP interactions

are dynamic throughout the growth curve (Kahramanoglou et al., 2011; Prieto et al., 2012) and also affect gene expression.

Much of the research on archaeal chromatin focuses on the role of histones in DNA packaging. Proteins homologous to eukaryotic histones H3 and H4 are found in the genomes of most Euryarchaea and some Crenarchaea (Sandman & Reeve, 2006; Zhang et al., 2012). Euryarchaeal histone crystal structures from *Methanothermus fervidus* and *Methanopyrus kandleri* are nearly identical to that of the eukaryotic tetramer, an intermediate in nucleosome assembly (Decanniere et al., 2000; Fahner et al., 2001). *In vitro* experiments have confirmed that these histones compact DNA into nucleosome-like particles. These nucleosomes remain tetrameric and do not go on to form octamers (Decanniere et al., 2000; Pavlov et al., 2002; Tomschik et al., 2001). Although *in vitro* evidence suggests that archaeal nucleosomes may inhibit transcription in a similar manner to eukaryotic tetramers (Soares et al., 1998; Thomm et al., 1992; Wilkenson et al., 2010), *in vivo* gene expression experiments have yielded variable results (Cubonova et al., 2012; Weidenbach et al., 2008).

Several attributes of archaeal histone proteins complicate the interpretation of their function in genome organization and gene expression. First, archaeal histones lack both N-terminal tails and post-translational modifications of the core domain (Sandman & Reeve, 2006; Arents & Moudrianakis, 1995). Second, most archaeal genomes encode a multitude of NAPs, including bacterial packaging proteins such as DpsA and Archaea-

specific packaging proteins such as Alba and the crenarchaeal protein family Sul7d (Zhang et al., 2012). Homologs of these NAPs are often detectable in the genomes of organisms that also encode eukaryotic-type histone proteins. For example, a small non-specific DNA binding protein called MC1 is preferentially used over histone as the principal chromatin packaging protein in *Methanosarcina* species (Weidenbach et al., 2008; Chartier et al., 1988; Laine et al., 1986).

According to whole genome sequence data (Becker et al., 2014; Ng et al., 2000), members of the haloarchaeal clade of Euryarchaeota encode a single histone protein. Histones from these hypersaline-adapted organisms (and a single methanogenic species, *M. kandleri*; Fahrner et al., 2001) contain two histone fold domains fused together with a linker region, which has been suggested as an evolutionary intermediate between archaeal and eukaryotic histones (Sandman & Reeve, 2006; Zhang et al., 2012; Malik & Henikoff, 2003). The halophilic histone protein is primarily comprised of acidic amino acids (pI = 4.4) as an adaptation to high intracellular salt (Kennedy et al., 2001), unlike all other histones and NAPs. For example, histone proteins are typically among the most basic proteins in a eukaryotic cell (pI of *H. sapiens* H3 = 11.71; *S. cerevisiae* H3 = 11.90) and other archaea (*M. fervidus* HMfA = 9.21).

Despite these intriguing differences, the physiological function of histone proteins in halophilic archaea remain to be investigated. The few studies that have been done on haloarchaeal chromatin suggest nucleosome formation. Electron micrographs of

sucrose gradients from *Halobacterium salinarum* have demonstrated the growth-dependent presence of two different populations of DNA: protein-bound fractions with a “beads-on-a-string” appearance predominate stationary phase and protein-free fractions predominate log phase (Shioda et al., 1989; Takayanagi et al., 1992). More recent MNase-seq experiments in *Haloferax volcanii* suggest regular phasing in gene expression and protein occupancy surrounding transcriptional start sites (Ammar et al., 2012). However, the identity of proteins causing these patterns and the physiological function of haloarchaeal histones remain unclear.

In order to shed light on the function of histone protein in the Haloarchaea, we constructed deletion and overexpression mutants of *hpyA* (*VNG0134G*), encoding the sole histone of *H. salinarum* NRC-1. Histone deletion and overexpression mutants are viable and *hpyA* deletion increases the growth rate in batch culture without affecting susceptibility to stress. Unexpectedly, histone dosage appears to be linked to cell shape and size, a function not previously observed in the Archaea. Genome wide expression profiling in histone mutants suggests that histone has a bifunctional, growth phase-specific effect on transcription. Identification of proteins from sucrose-fractionated DNA yielded no evidence to support the existence of eukaryotic-type chromatin, nor was a micrococcal nuclease ladder observed after digestion of chromatin. Instead, proteomics and gene expression analysis supports a model in which a dynamic population of low-abundance NAPs function in open chromatin throughout growth. Together, our

experimental results suggest that the histone protein of *H. salinarum* functions primarily as a modulator of transcription required for wild type morphology.

2.2 Methods

2.2.1 Strains and growth conditions

All strains described are originally derived from *Halobacterium salinarum* NRC-1 (ATCC700922). All other strains used in this study are listed in Table 3. Cultures were routinely incubated at 42°C with agitation in complex medium (CM; 250 g L⁻¹ NaCl, 20 g L⁻¹ MgSO₄•7H₂O, 3 g L⁻¹ sodium citrate, 2 g L⁻¹ KCl, 10 g L⁻¹ peptone). For strains generated in the $\Delta ura3$ background, uracil was added to a final concentration of 50 μ g mL⁻¹. For strains containing the overexpression vector pMTFCHA and its derivatives (Table 3), mevinolin was added to a final concentration of 1 μ g mL⁻¹ to maintain the plasmid.

Table 3: Plasmids and strains used in this study

Plasmid	Description	Source
pMTFCHA	expression vector, P_{fdx} Mev ^R	Wilbanks <i>et al.</i> , 2012
pRSK01	Gateway suicide vector, pNBK07:: <i>attB1,attB2</i> , Mev ^R Ura ⁺	Wilbanks <i>et al.</i> , 2012
pKAD01	pRSK01:: $\Delta hpyA$	This study
pKAD02	pMTFCHA:: <i>hpyA</i>	This study
pKAD03	pMTFCHA:: P_{rpa200} :: <i>hpyA</i>	This study

Strain	Genotype	Description	Source
MDK407	$\Delta ura3$	Parent strain, Ura ⁻	Peck <i>et al.</i> , 2000
KAD100	$\Delta ura3\Delta hpyA$	histone null mutant, Ura ⁻	This study

<i>KAD101</i>	<i>Δura3/pMTFCHA</i>	Ura ^r Mev ^R	This study
<i>KAD102</i>	<i>Δura3/pKAD02</i>	histone overexpression, Ura ^r Mev ^R	This study
<i>KAD103</i>	<i>Δura3ΔhpyA/pKAD03</i>	histone complementation, Ura ^r Mev ^R	This study

2.2.2 Strain construction

Plasmid pKAD01 for in-frame deletion of *hpyA* was constructed by Gateway cloning (Invitrogen, Carlsbad, CA) using PCR SOEing and destination vector pRSK01 as described in (Wilbanks et al., 2012). Knockout clones were transformed and selected in the *Δura3* uracil auxotroph parent strain as described in Peck et al. (2000), and the knockout genotype was confirmed with PCR, Sanger sequencing and Southern blotting. Overexpression vector pKAD02, constructed as described previously (Schmid et al., 2011), contains the *hpyA* (*VNG0134G*) coding sequence driven by the strong constitutive promoter of *VNG2293G* in backbone vector pMTFCHA (Wilbanks et al., 2012). Complementation plasmid pKAD03 was synthesized by cloning the native promoter for the *rpa-hpyA-aup* operon (P_{rpa200}) upstream of the *hpyA* coding sequence; this sequence was inserted into pMTFCHA using Gibson assembly (Gibson, 2011). In order to cure *Δura3/pKAD02* of the plasmid, triplicate cultures were grown in non-selective media for five days and subcultured prior to phenotyping. All plasmid constructs and plasmid-bearing strains were screened using PCR and confirmed by Sanger sequencing. All primers used for strain construction are described in detail in Appendix B.

2.2.3 Southern blotting and detection

Genomic DNA was purified and digested with the restriction enzymes *Bam*HI and *Kpn*I. DNA samples were run on a 0.8% agarose gel prior to blotting on a Biorad Zeta-probe membrane. An alkali-labile DIG-II-dUTP probe of 86bp internal to *hpyA* was synthesized using PCR primers K51 and K52 (Appendix B), gel purified, and prepared according to manufacturer's instructions. Blots were hybridized overnight at 50°C and chemiluminescent probe detection was performed as described by Selmecki et al. (2009).

2.2.4 Sequence and structural analyses

All amino acid sequence alignments were performed using Clustal Omega (Sievers et al., 2011). Histone sequences from the Haloarchaea (Becker et al., 2014) were retrieved from NCBI and manually curated to remove duplicate or spurious entries. Isoelectric point (pI) of the remaining 69 histone protein sequences was calculated using the Sequence Manipulation Suite (Stothard, 2000). Prior to pI analysis, N- and C-terminal histone domains of haloarchaeal histone sequences were separated and linker sequences removed. Significant difference in pI between all N-terminal domains *vs.* all C-terminal domains was determined by *t*-test. The histone protein consensus sequence logo for Haloarchaea was generated using Weblogo3.3 (Crooks et al., 2004). For structural modeling, histones H3 and H4 from *S. cerevisiae* were trimmed of N-terminal tails and

fused *in silico* using the HpyA linker region. Modeling of yeast H3-H4 and *H. salinarum* HpyA onto the HMk scaffold from *Methanopyrus kandleri* (PDB ID: 1f1eA; Fahrner et al., 2001) and calculation of the surface charge distribution was performed using SWISS-MODEL (Biasini et al., 2014). Structural models were visualized with DeepView/PDB-Viewer (Guex et al., 1997). Charge as shown in Figure 1C was calculated for both HpyA and yeast histones at the default dielectric constant of 80.00. Adjusting the constant to 48.4 to account for a high salt environment (Kennedy et al., 2001), slightly affected the magnitude of observed charges but had no detectable effect on charge distribution.

2.2.5 Growth rate measurements

Growth experiments were performed in a high-throughput multiplate reader as follows: at least three biological replicates, each with three technical triplicates, for each of $\Delta ura3$ parent, $\Delta ura3\Delta hpyA$, $\Delta ura3$ empty vector control, and *hpyA* overexpression strains were grown to mid-exponential phase and subcultured to an optical density at 600 nm (OD₆₀₀) of 0.05 in CM with uracil. Subcultures were grown in a multi-well plate at 42°C for 48 hours under continuous shaking (~225 rpm) in a Bioscreen C microbial growth analyzer (Growth Curves USA, Piscataway, NJ). OD₆₀₀ was automatically measured every 30 minutes. Growth rate was calculated from the linear regression of the log₂ optical density during early exponential growth for each replicate. Growth rates from

technical replicate cultures were averaged; the mean and standard deviation were calculated for growth rates of biological replicate cultures.

2.2.6 Resistance to UV radiation

Log phase cultures were subcultured and grown to mid-log phase. 1–2 ml of cells were spun for 1 minute at 13,000 rpm and the supernatant was removed. Cells were resuspended in 1.787 mL of filtered basal salts and transferred to sterile petri dishes (60 × 15 mm). Samples were irradiated without lids at 254 nm in a Stratagene Stratalinker 2400; aliquots were removed after sequential UV doses of 100 J/m². For each of three biological replicates, serial dilutions were spot-plated in duplicate on CM. Spot-plates were incubated in the dark for two weeks and the survival rate of radiated cells over non-irradiated cells was calculated. Error bars represent standard deviation of biological replicates.

2.2.7 Growth under oxidative stress

Three biological replicates of *Δura3* and *Δura3ΔhpyA* were grown to mid-log phase and subcultured to an OD₆₀₀ of 0.05 in CM with uracil. Three technical replicates were grown for each biological replicate. For Paraquat oxidative shock experiments, 195 μl of culture was allotted to each well and 5 μl of diluted Paraquat was added to a final concentration of 0.083, 0.167, and 0.333 mM. 200 μl culture aliquots were grown in a multiwell plate at 42°C for 48 hours under continuous shaking (~225 rpm) in a Bioscreen

C microbial growth analyzer (Growth Curves USA, Piscataway, NJ). OD600 was automatically measured every 30 minutes. Growth rate was calculated by taking the linear regression of the log₂ optical density during early exponential growth for each replicate. Technical replicates were averaged and the mean and standard deviation were calculated for biological replicates.

2.2.8 Light microscopy

Cells were visualized at mid-logarithmic (OD600 ~ 0.3-0.5) and stationary (OD600 ≥ 1.3) phases of growth with a Zeiss Axio Scope A1 microscope under 100x phase contrast.

Images were captured with a Pixelink PL-E421M camera. Cell counting and morphological quantitation was performed with ImageJ and R statistical software.

Ellipses were fit to each cell and measurements of the area, length, and width were collected. *T*-tests were performed on equivalent numbers of data points that had been randomly sampled from pooled biological triplicate images. Resultant *p*-values are listed in Table 4.

2.2.9 Gene expression microarrays and analysis

For strains $\Delta ura3$ and $\Delta ura3\Delta hpyA$, three biological replicate cultures were grown to mid-log phase in CM with uracil and then subcultured to an OD600 of 0.05 for further growth. 4ml aliquots were harvested from 50ml cultures at mid logarithmic (OD600 ~

0.5) and stationary (OD600 ~ 1.5) phase. Culture harvest, RNA extraction, labeling, and hybridization to custom Agilent ORF arrays (6 probes per gene) was performed as described in Sharma et al., (2012). Each sample was hybridized against *H. salinarum* NRC-1 wild type grown under standard conditions (CM to OD600 ~0.4 at 37°C with 225 rpm shaking; Baliga et al., 2004). Dye swaps were performed for each biological replicate. A total of 36 replicate data points were collected per gene in each sample. Spot ratios were determined using Agilent Feature Extraction and all further analysis was performed in the R statistical computing environment. Ratios were normalized within and between arrays using the R package limma (Smyth et al., 2005; Ritchie et al., 2015) in a pipeline adapted from Sharma et al. (2012). Resultant normalized gene expression data was subjected to Student's *t*-test using the TM4 Multiple Experiment Viewer (Saeed et al., 2003). A cutoff of $p < 0.05$ between biological replicates of parent strain *vs.* histone mutant was used to designate those genes that showed significant changes in gene expression in response to histone deletion. Ratios for genes with significant expression changes by *t*-test were calculated by subtracting the mean log₂ ratios of the mutant biological replicates from those of the wild-type. As evidenced by comparing resultant microarray ratios to those obtained by RT-qPCR (Supplementary Figure S2 and Table S2 from Dulmage et al., 2015), the dynamic range of array ratios was artificially narrow. Therefore, prior to setting a 2-fold threshold for significance in the microarray data, a correction factor derived from the linear relationship between the qRT-PCR and

microarray datasets was applied to scale the microarray data appropriately (corrected \log_2 array value = $3.5763 \times \text{array ratio} - 0.3758$). All raw and normalized microarray data have been deposited in NCBI's Gene Expression Omnibus (Edgar et al., 2002) and are accessible through GEO Series accession number GSE54599.

2.2.10 Quantitation of S-layer glycosylation levels

S-layer glycosylation levels were quantified in S-layer enrichments from Δura3 and $\Delta\text{ura3}\Delta\text{hpyA}$. Cultures were grown to stationary phase (OD 600 ~ 1) and approximately 1ml of culture was pelleted by centrifugation at 21,130 r.c.f. for 30 seconds and lysed in 500 μl of water followed by vigorous pipetting. S-layer enrichment was performed using a modified version of the protocol described in Sumper et al. (1990). Approximately 10 ml of culture was pelleted at 4,500 r.c.f. for 10 minutes, resuspended in basal salts buffer (CM lacking peptone) containing EDTA equimolar to the Mg^{2+} concentration (80 mM), and incubated for 30 minutes at 225 r.p.m at 42°C to remove the S-layer. Cells were removed by centrifugation for 5 minutes at 12,000 r.c.f. The supernatant, which contains primarily S-layer protein, was collected. Proteins from aliquots of different volumes (100, 250, 500, and 750 μl) were precipitated using 10% TCA, solubilized in Laemmli buffer containing SDS, and separated on two identically loaded BioRad Mini Protean TGX gels for 1 hr at 125 V (BioRad Laboratories, Hercules, CA). One gel was stained to visualize protein using Colloidal Coomassie Blue (CCB; Invitrogen, Carlsbad, CA), and

the other was stained to visualize carbohydrate using Periodic Acid Schiff (PAS; Sigma, St. Louis, MO).

2.2.11 Resistance to bacitracin

5 μ l of bacitracin solubilized in DMSO was added to media to a final concentration of 0, 1, 5, 10, and 20 μ M. 200 μ l culture aliquots were grown in a multiwell plate at 42°C for 48 hours under continuous shaking (~225 rpm) in a Bioscreen C microbial growth analyzer (Growth Curves USA, Piscataway, NJ). Optical density at 600 nm was automatically measured every 30 minutes. Growth rate was calculated by taking the linear regression of the log₂ optical density during early exponential growth for each replicate. Technical replicates were averaged and the mean and standard deviation were calculated for biological replicates.

2.2.12 Shear stress resistance

Assay was performed essentially as described in Rothfuss et al., 2006. Results of 3 independent experiments, each with 3 biological replicates and at least 2 technical replicates, are summarized in Fig. 5°C. Error bars represent standard deviation of biological replicates.

2.2.13 Resistance to novobiocin

Cultures were grown to mid-log phase; 250 μ l were plated on CM. A 30 μ g novobiocin susceptibility test disc was placed in the center and plates were incubated face-up for 24 hours before being inverted. After several days of growth at 42°C, the zone of clearance was measured.

2.2.14 Identification of nucleic acid components of sucrose gradient profiles

A 5–20% sucrose gradient was run with lysate from an early log phase wild-type culture as described in the main text. Nucleotide content for each fraction was determined by absorbance at 260 nm using a Nanodrop (Thermo Scientific). Aliquots from each fraction were split into three different treatments: a) mock digestion as an intact nucleotide control, b) digestion with RNase A to isolate DNA content, and c) digestion with micrococcal nuclease to isolate RNA content. All treatments were incubated at 37°C for 2 hours. Aliquots were then run on a 1% agarose gel, post-stained with Ethidium Bromide (EtBr) and total band intensity for each lane was quantitated using Image J. Samples 4, 12, and 24 were lost during processing and thus are not represented in Fig. 8B.

2.2.15 Micrococcal nuclease (MNase) digestion of chromatin from Halobacterium salinarum

Our *Halobacterium salinarum* MNase protocol was optimized as described in Table 6. Our final working protocol corresponds to assay #16 (Fig. 7D) and is composed of the following steps: 1.5 OD600 units of stationary phase cells per sample were fixed with 1% formaldehyde for 10 min at room temperature. Fixation was quenched with 100 mM glycine and cells were pelleted at 8000 rpm in a Sorvall rotor for 10 min at 4°C. Cells were washed 3X with cold basal salts buffer and transferred to microfuge tubes. Pellets were resuspended in 800 µl of lysis buffer (50 mM HEPES-KOH pH 7.5, 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% Na-deoxycholate, 1% Pierce Halt protease inhibitor cocktail) and sonicated for 15 min with 30s on/off cycles using a Diagenode bioruptor. Cell debris was removed by centrifugation for 5 min at 10k rpm at 4°C. 500 µl of supernatant was transferred to a new tube and RNase A was added. MNase was diluted in H₂O and added to final concentrations of 0, 10, 50, 70, 100, and 300 units. Digestion was allowed to progress for 20 min at 37°C and then stopped with 30 µl of 500 mM EDTA pH 8.0. Samples were incubated with proteinase K for 1 hour at 55°C and then overnight at 65°C to reverse crosslinks. After crosslink reversal, DNA was extracted twice with phenol/chloroform and precipitated with ethanol, glycogen, and sodium acetate. Pellets were dried and then resuspended in H₂O; equal amounts of purified DNA was run either on a 4–20% Tris-HCl polyacrylamide gel (Biorad) or a 1.5% agarose gel in sample buffer without dye (50 mM EDTA, 25% glycerol). Details for tested alternative methods (as shown in Table 6) are as follows: Spheroplasted cells were

generated by incubation with spheroplast solution (Jarrell & Sprott, 1984: 0.1 M MES, 0.5 M D-sucrose, 0.25 M NaCl, 0.01M MgCl₂•6H₂O). Spermidine was added to a final concentration of 500 μM to stabilize DNA structures prior to MNase digestion. For digestion in MNase buffer, cells were washed 2X in basal salts buffer and then lysed in MNase buffer (50 mM Tris-HCl pH 8.0, 5 mM CaCl₂) for 10 min with vortexing and pipetting until lysate was homogenous. The Ammar et al. (2012) MNase digestion protocol was performed as described with the following modifications: the first two washes were in basal salts to preserve cellular integrity, CaCl₂ was added to the PBS digestion buffer to a final concentration of 50 μM as it is an essential cofactor for MNase activity, and cells were lysed by passage through a needle to prevent fragmentation artifacts from sonication. The DNase digestion protocol was identical to that for MNase digestion but with final concentrations of 0, 0.5, 1, 1.5, 2, 5, and 10 units per sample. Ultracentrifugation and Southern blotting were performed as described elsewhere in Materials and Methods. Yeast DNA was digested with micrococcal nuclease (MNase) as described in Yuan et al. (2005).

2.2.16 Protein identification and analysis

Wild type *H. salinarum* cultures were harvested in log (OD₆₀₀ = 0.5) or stationary (OD₆₀₀ = 1.5) phase and fractionated over a sucrose density gradient as described in Takayanagi et al. (1992) except that cultures were fixed for 30 min in 1% formaldehyde,

lysates were sheared with a needle, and ultracentrifugation was performed at 160,000 r.c.f. for 5 h at 4°C. Nucleic acid concentration in each fraction was determined by absorbance and protein concentration was determined by the Bradford assay (Bradford, 1976). From each phase of growth, six samples were prepared for protein identification: two biological replicates of DNA (fractions 8-13 in log phase, 9-14 in stationary phase) and one of cytosolic components including free RNA (fractions 1-6) as a control. Fractions for either DNA or control samples were pooled and protein was precipitated with 10% trichloroacetic acid. Protein precipitates were separated via SDS-PAGE. All proteins less than 100 kDa were excised to avoid the abundant S-layer protein and subsequently digested with trypsin. Peptide samples were analyzed via LC-MS/MS using a Q Exactive Plus Hybrid Quadrupole-Orbitrap Mass Spectrometer (Thermo Scientific). MS/MS spectra peptide and protein assignment were performed using Mascot (Matrix Science, London, UK; version 2.5.0). Mascot was searched with a fragment ion mass tolerance of 0.020 Da and a parent ion tolerance of 5.0 PPM. O+18 of pyrrolysine and carbamidomethyl of cysteine were specified in Mascot as fixed modifications. Deamidation of asparagine and glutamine, methylation of lysine and arginine, and oxidation of methionine and acetyl of lysine were specified in Mascot as variable modifications. Scaffold (version Scaffold_4.3.2, Proteome Software Inc., Portland, OR) was used to validate MS/MS based peptide and protein identifications. Using the Peptide Prophet algorithm (Keller et al., 2002), peptide identifications were

accepted if they could be established at greater than 7.0% probability to achieve an FDR less than 1.0% with Scaffold delta-mass correction, protein identifications were accepted if they could be established at greater than 53.0% probability to achieve an FDR less than 0.01% and contained at least 2 identified peptides, and protein probabilities were assigned. Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped. Protein identifications are listed in Supplementary Table S3 of Dulmage et al. (2015). Data from biological replicates were added together prior to statistical analysis. Significant enrichment of peptide abundance in control *vs.* DNA fractions was calculated by determining the linear relationship for protein abundance between fractions using the R package robustbase. Peptide abundance per protein was calculated by normalizing to the total number of peptides per sample and protein mass. Residuals of the linear regression were calculated and proteins with residuals at least two standard deviations from the mean were considered enriched. Enriched peptides are listed in Supplementary Table S4 of Dulmage et al. (2015).

2.2.17 Comparison of NAP gene expression and quantitative proteomics data during growth

Gene expression profiles throughout a high-resolution time course of the *H. salinarum* growth curve (Facciotti et al, 2010) were retrieved from the GEO database (GSE14976).

Normalized data points from strains NRC-1 and $\Delta ura3$ were binned into equal intervals

from OD600 = 0.1-1.4, where each interval has at least 2 measurements. Averages of optical densities and expression values from each bin were plotted. For analysis of NAP protein expression profiles, available quantitative iTRAQ proteomics data was analyzed from *H. salinarum* cultures exposed to slow growth conditions (low oxygen) *vs.* rapid growth conditions (high oxygen) in a turbidostat over time (Schmid et al., 2007). Protein level values from high oxygen conditions were averaged to yield rapid growth value, whereas values from low oxygen conditions were averaged to yield no growth values. We focused on five genes encoding putative NAPs within these datasets: *mc1* (VNG2508C), the bacterial-type NAP *dpsA* (VNG2443G), histone (*hypA*) and two homologs to the universally conserved DNA cohesion and condensation protein Smc1, *sph1* (VNG6320C) and *sph2* (VNG6173C).

2.3 Results

2.3.1 HpyA, a fused histone heterodimer, is conserved throughout the halophilic archaea

A single histone homolog, VNG0134G (*hpyA*), is detectable in the genome of *Halobacterium salinarum* (Ng et al., 2000). The protein consists of two histone fold domains arranged in tandem within a single open reading frame. To determine the extent of homology with canonical histones, we aligned the amino acid sequences for each of the N- and C-terminal domains of *hpyA* with well-characterized histone proteins from *Homo sapiens*, *Saccharomyces cerevisiae*, and *Methanothermobacter fermentans* (Fig. 1A).

Because archaeal histones lack N-terminal tails (Sandman & Reeve, 2006), these sequences were removed from the eukaryotic histones prior to alignment. As expected, the HpyA amino acid sequence resembles that of canonical eukaryotic histones (*e.g.* N- and C-terminal domains are 23.6% and 27.9% identical to yeast H3, respectively). Importantly, the majority of core residues thought to be essential for cell viability and for the structure of the histone fold were conserved in both domains of HpyA (Fig. 1A; Decanniere et al., 2000; Matsubara et al., 2007; Nakanishi et al., 2008; Soares et al., 2000). To investigate conservation of histones across the *Halobacteriales*, the amino acid sequences of histones from nearly 70 haloarchaeal genomes currently curated in the NCBI database were aligned. While the linker region between the two domains is highly degenerate, the fused histone heterodimer conformation and essential residues are conserved across all halophile sequences (Fig. 1B).

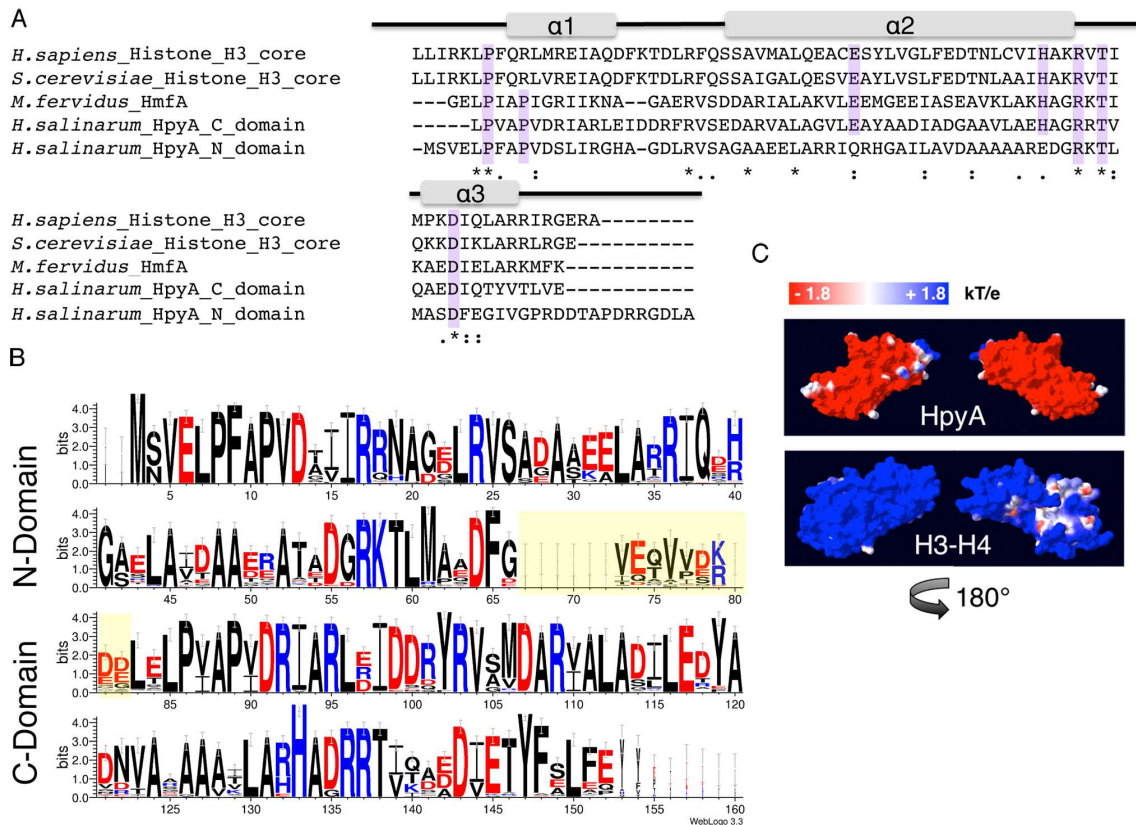


Figure 1: Haloarchaeal species encode a single conserved histone heterodimer.

A) Alignment of histone protein sequences: the eukaryotic H3 core domains from *Homo sapiens* and *Saccharomyces cerevisiae*, HMfA from *Methanothermobacter fervidus*, and each of two histone fold domains of *H. salinarum* HpyA. Residues essential for histone function or cell viability in the Eukarya and/or Archaea are highlighted in purple. Grey bars indicate the locations of the three alpha helical regions that comprise the histone fold domain of the canonical eukaryotic H3. “*”, fully conserved residues; “:”, strongly similar; “.”, weakly similar. **B)** Consensus logo representing histone sequence conservation between haloarchaeal species. The linker region between histone fold domains (yellow shaded residues) is variable in length and sequence. Residue colors are as follows: red, acidic; black, uncharged; blue, basic. **C)** Space-filling representation of the surface charge of *H. salinarum* HpyA and in silico fusion of the *S. cerevisiae* H3-H4 dimer modeled onto the *M. kandleri* histone crystal structure (Fahrner et al., 2001). Blue indicates positive charge, whereas red indicates negative charge.

An unusual feature of haloarchaeal histones is the preponderance of acidic amino acids relative to the histones of eukaryotes and to the histones of other archaea. This acidity is approximately evenly distributed between the two histone fold domains, although the N-terminal domain is significantly more basic on average (4.62 vs. 4.36, $p = 0.005$). In order to predict whether this acidity could affect histone-DNA binding, we projected the amino-acid sequence of HpyA onto the *Methanopyrus kandleri* HMk histone-dimer crystal structure (Fahrner et al., 2001) and calculated the surface electrostatic potential. As a control, *S. cerevisiae* H3 and H4 were also modeled onto the HMk structure (see Materials and Methods). Surprisingly, unlike eukaryotic histones, HpyA is a highly polarized molecule. Our model illustrates that although HpyA is a highly acidic protein, it also contains an isolated basic region largely consisting of arginine residues from both histone fold domains (Fig. 1C). These analyses suggest that, despite its clear conservation with the eukaryotic H3 protein sequence, evolutionary adaptations for high salt have made fundamental structural changes that may affect the function of HpyA.

2.3.2 Phenotypic characterization of HpyA function in *H. salinarum* physiology

To gain insight into the function of the haloarchaeal histone protein, a strain with an in-frame deletion of *hpyA* was constructed (VNG0134G; $\Delta ura3\Delta hpyA$; see Materials and Methods; strain list in Table 3). Out of 110 clones screened, three were heterozygous (*H. salinarum* is highly polyploid), and three were deleted for the *hpyA* ORF as confirmed by

PCR and Southern blotting (Fig. 2A & 2B). This rate of isolation of deletion mutants is consistent with those of non-essential genes (Coker et al., 2009). Clonal isolates of a single clean deletion mutant were used in all subsequent experiments. To explore the effects of altered histone dosage on cell physiology and gene expression, a histone overexpression strain was generated in which a strong constitutive promoter drives ectopic *hpyA* expression in the $\Delta ura3$ parent background ($\Delta ura3/pKAD02$). Histone overexpression strains were obtained for all clones screened.

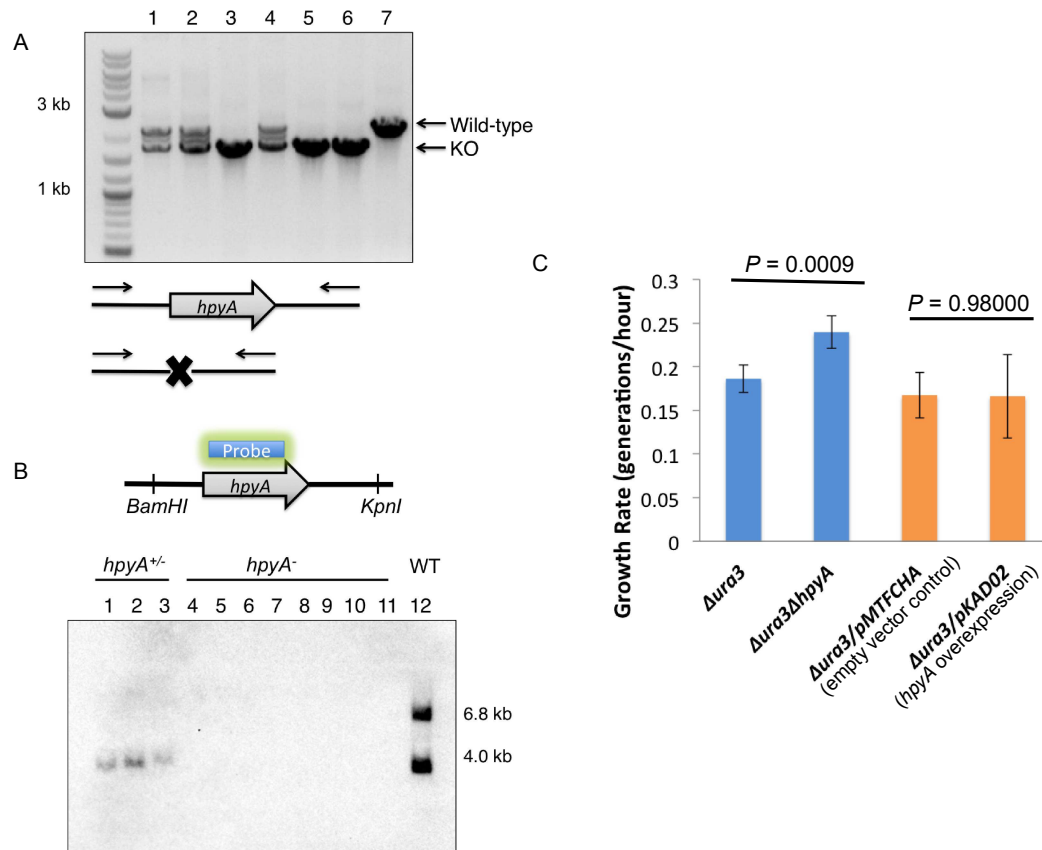


Figure 2: Histone knockout mutants are readily generated and viable. A) PCR validation of transformants: Lanes 1, 2, and 4 are heterozygous at the *hpyA* locus, lanes 3, 5, and 6 are deletion mutants, and Lane 7 corresponds to the Δ *Aura3* parent strain. Wild-type amplicon = 2.1 kb, KO = 1.7 kb. B) Southern blotting of restriction enzyme digested genomic DNA using a probe internal to the *hpyA* ORF confirms that knockouts are homozygous. Lanes 1, 2, and 3 correspond to the heterozygous mutants, lanes 4-9 are biological replicates used in this study, lanes 10 and 11 are independently isolated histone deletion mutants, and lane 12 is a partial digest control of Δ *Aura3* parent strain genomic DNA. Primer sequences for both screens are listed in Appendix B. C) The histone deletion strain Δ *Aura3ΔhpyA* grows significantly faster than the parent *Aura3* ($p = 0.0009$, blue bars), while the *hpyA* overexpression strain *Aura3/pKAD02* is indistinguishable from the *Aura3/pMTFCHA* empty vector control ($p = 0.98$, orange bars). Error bars represent standard deviation from the mean.

To examine the phenotypic effects of deletion and overexpression of the histone-coding gene in *H. salinarum*, growth rate was measured. Unexpectedly, the histone mutant growth rate under optimum growth conditions is significantly faster ($n = 27$, $p = 0.0009$) than that of the parent strain (Fig. 2C). In contrast, the growth rate of the histone overexpression strain $\Delta ura3/pKAD02$ is not significantly different ($n = 3$, $p = 0.98$) than that of $\Delta ura3$ containing the empty vector control (Fig. 2C, right). These results suggest that although *hpyA* encodes the sole histone in this organism, it is not essential for survival.

In order to determine the morphological effects of histone deletion, all four strains described above were visualized using phase contrast microscopy during log ($OD_{600} = 0.2$ to 0.8) and stationary ($OD_{600} \geq 0.8$) phases of growth. As expected, the parent strain $\Delta ura3$ and the $\Delta ura3/pMTFCHA$ empty vector control are rod-shaped regardless of growth phase (Fig. 3). Surprisingly, histone mutant cells are pleiomorphic: cells ranging from wild-type rods to triangular, circular, and irregular forms (Fig. 3A & 3B). As a proxy for the measurement of pleiomorphism in the population, the width/length ratio was calculated for individual cells. The histone knockout mutant cells exhibited a significantly higher width/length ratio than the parent strain in both growth phases tested (log phase mean parent = 0.23; log phase mean mutant = 0.40; $p = 1.3 \times 10^{-13}$, see Table 4 for p -values of t -test comparisons for all strains and growth phases) even though the parent strain exhibited a slightly increased width/length ratio in stationary

phase. This morphological phenotype is significantly complemented by ectopic expression of *hpyA* at near native levels ($\Deltaura3$ parent *vs.* $\DeltahpyA/pKAD03$ complemented strain: $p = 5.04 \times 10^{-2}$, \DeltahpyA *vs.* complemented strain: $p = 1.65 \times 10^{-8}$), ruling out polar effects on surrounding genes or second site mutations (Fig. 3A and 3B, Materials and Methods). In contrast to $\Deltaura3\DeltahpyA$ cells, $\Deltaura3/pKAD02$ cells exhibited a dramatic increase in length in log phase (11.5 microns \pm 1.16 standard error) relative to the empty vector control (5.8 microns \pm 0.26 standard error). Unlike the $\Deltaura3\DeltahpyA$ pleiomorphic phenotype, enlargement of $\Deltaura3/pKAD02$ is largely ameliorated during stationary phase (Figures 3C & 3D). Wild-type dimensions are restored in strains cured of the *hpyA* overexpression plasmid (Figures 3C & 3D), confirming that the observed phenotypes are due to overexpression alone. Taken together, these phenotypic analyses suggest that, although histones are not essential for cell viability, batch culture growth, or stress resistance, they are required to regulate cell morphology in *H. salinarum*.

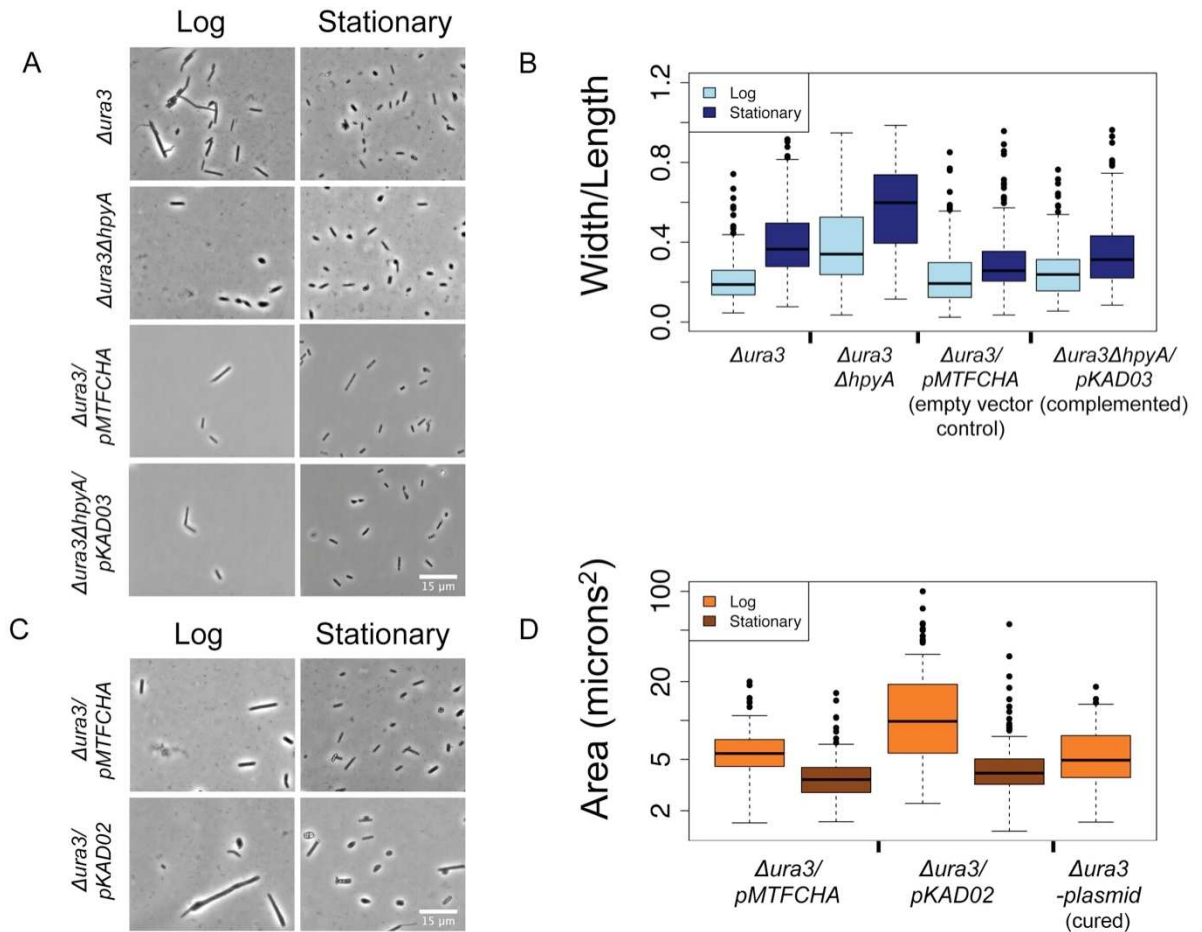


Figure 3: Histone dosage regulates cell shape and size. A) $\Delta ura3\Delta hpyA$ cells are pleiomorphic. Phase contrast light microscopy images are shown for parent $\Delta ura3$ vs. histone deletion mutant $\Delta ura3\Delta hpyA$ and control strain $\Delta ura3/pMTFCHA$ vs. complemented strain $\Delta ura3\Delta hpyA/pKAD03$ in log or stationary phase of growth. B) Individual cells were fit as ellipses and width/length ratio was calculated. Error bars represent the 1st and 4th quartiles. For quantitation of log phase cultures, $n \geq 113$ cells for each biological replicate sample. Stationary phase cultures, $n \geq 227$ cells. Scale bar = 10 μm . C) Histone overexpression cells are larger than isogenic parent cells during logarithmic growth. Phase contrast light microscopy images of $\Delta ura3/pMTFCHA$ vs. $\Delta ura3/pKAD02$ at mid-log or stationary phase of growth are shown. D) Individual cells were fit as ellipses and area (length \times width) was calculated. Error bars are as in (B). For quantitation of log phase cultures, $n \geq 76$ cells. Stationary phase cultures, $n \geq 272$ cells.

Table 4: Results of *t*-tests from quantitative analysis of mutant morphology

Test	Strain A Genotype	Strain B Genotype	Growth Phase	Measure	Number of cells	<i>P</i> -value
Parent vs. histone knock-out	<i>Δura3</i>	<i>Δura3ΔhpyA</i>	Log	Width/Length	136	1.25E-13
	<i>Δura3</i>	<i>Δura3ΔhpyA</i>	Stat	Width/Length	179	4.09E-15
Control vs. histone rescue	<i>Δura3/pMTFCHA</i>	<i>Δura3ΔhpyA/pKAD03</i>	Log	Width/Length	113	5.04E-02
	<i>Δura3/pMTFCHA</i>	<i>Δura3ΔhpyA/pKAD03</i>	Stat	Width/Length	227	2.07E-03
Knock-out vs. rescue	<i>Δura3ΔhpyA</i>	<i>Δura3ΔhpyA/pKAD03</i>	Log	Width/Length	147	1.65E-08
	<i>Δura3ΔhpyA</i>	<i>Δura3ΔhpyA/pKAD03</i>	Stat	Width/Length	220	2.20E-16
Control vs. histone over-expression	<i>Δura3/pMTFCHA</i>	<i>Δura3/pKAD02</i>	Log	Area	76	3.06E-06
	<i>Δura3/pMTFCHA</i>	<i>Δura3/pKAD02</i>	Stat	Area	272	4.08E-03
Over-expression vs. plasmid loss	<i>Δura3/pKAD02</i>	<i>Δura3 -plasmid</i>	Log	Area	76	1.16E-06
Plasmid loss vs. control	<i>Δura3 -plasmid</i>	<i>Δura3/pMTFCHA</i>	Log	Area	76	2.57E-01

2.3.3 Histone deletion results in growth phase-specific effects on gene expression

To investigate whether histone regulates gene expression in *H. salinarum*, we compared genome-wide mRNA expression profiles between the *Δura3* parent strain and

Δura3ΔhpyA at mid-logarithmic and stationary phases of growth. Thirty-seven genes

(1.5% of predicted ORFs in the genome) exhibited significantly altered expression in logarithmic phase, whereas 220 (6.5%) exhibited significantly altered expression in stationary phase (Fig. 4A & 4B). Strikingly, only seven genes were regulated by histone at both time points tested. Expression data were corroborated using RT-qPCR (Pearson correlation of microarray and PCR data = 0.78). In log phase, we observed only downregulation of gene expression in the histone mutant compared to its isogenic parent strain, suggesting that histone activates gene expression during this phase. In contrast, during stationary phase the majority of significantly regulated genes are repressed in the histone deletion relative to the parent (66%), while the remainder is overexpressed. This pattern suggests a bifunctional regulatory role for histone during this phase with a tendency toward activation. Together, these results suggest that histone is required for wild type gene expression levels and that the mechanism of its action is phase-specific.

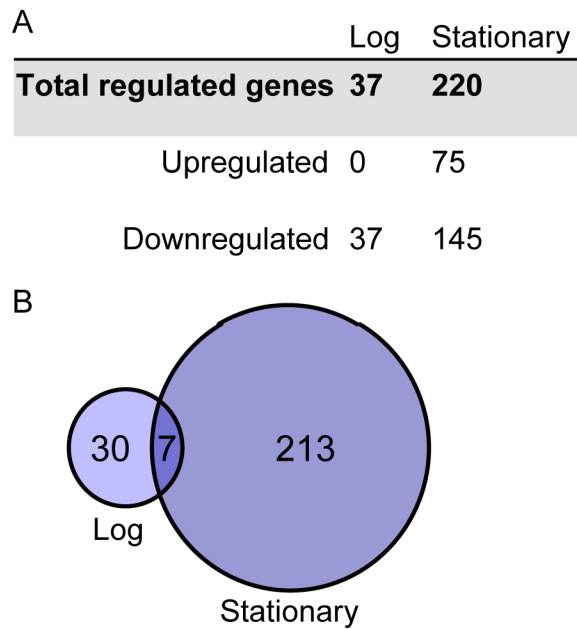


Figure 4: Histone-regulated gene expression is growth-phase dependent. A) The number of genes differentially expressed in $\Delta\text{ura3}\Delta\text{hpyA}$ vs. the Δura3 parent strain is dependent on the phase of growth. B) Of the histone-regulated genes, only 7 are affected by histone deletion in both logarithmic and stationary phase. Circles in the Venn diagram are scaled by gene number.

2.3.4 Halophilic histone is not required for genome-wide supercoiling, UV survival, or oxidative stress response

Previous studies on archaeal and eukaryotic histones have shown impaired stress response survival for histone mutants (Weidenbach et al., 2008). To test the phenotype of ΔhpyA under stress, cells were exposed to lethal amounts of UV (up to 500 J m⁻²) or cultured with the redox-cycling drug Paraquat to generate reactive oxidative species (ROS). No differences in survival or growth from the isogenic parent were observed

(Fig. 5A & 5B), which suggests that histone is not required for light-independent DNA damage repair or ROS-induced damage repair.

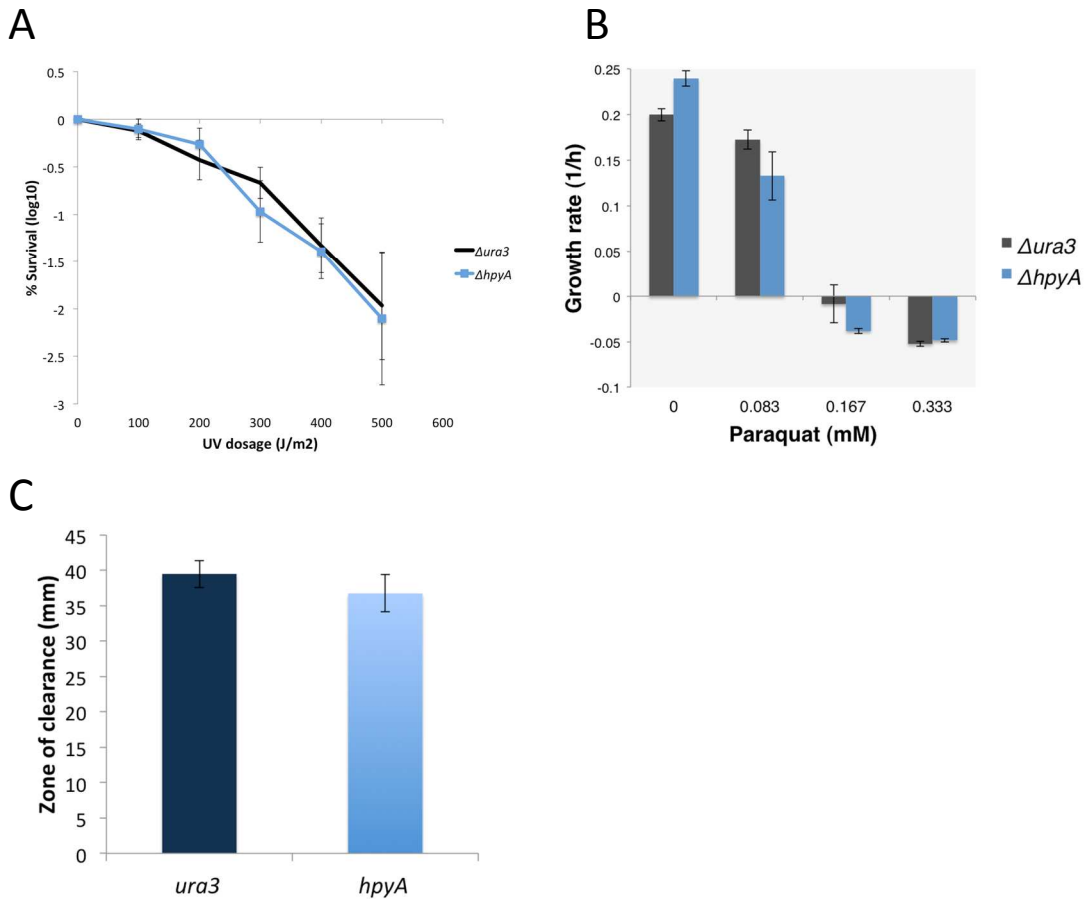


Figure 5: Histone is not required for resistance to A) UV radiation, B) reactive oxygen species generated by Paraquat, or C) supercoiling stress generated by novobiocin. Error bars represent the standard deviation of at least 3 biological replicates.

One of the genes most strongly repressed by histone deletion, *VNG5192H*, may contain a DNA gyrase domain, based on three-dimensional structure prediction

(Bonneau et al., 2004). This was of interest, as both archaeal and eukaryotic histone proteins are known to constrain supercoils (Marc et al., 2002). Since it is well established that the promoter regions of topoisomerases and DNA gyrases are sensitive to changes in supercoiling (Menzel & Gellert, 1983; Tse-Dinh, 1985), we compared the expression of annotated orthologous genes between the histone knockout and the isogenic parent strain. We did not find any significant differences, nor was there a consistent pattern of expression changes among 17 additional proteins with putative supercoiling topoisomerase or DNA gyrase domains. Additionally, tests for novobiocin sensitivity in the histone knock-out yielded wild-type results (Fig. 5C), suggesting that HpyA is not involved in constraining supercoils. Together, these phenotypic tests indicate that (i) the presence of histone is less important for cell viability and ROS stress response in haloarchaea than in other archaea or eukaryotes; and (ii) histone is unlikely to be required for the maintenance of genome-wide supercoiling, though we cannot rule out local effects.

2.3.5 Histone regulates genes in cell wall and membrane-associated processes

No functional enrichment for histone-regulated genes in either gene ontology terms or archaeal clusters of orthologous genes (GO and arCOGs; data not shown) was found, suggesting that histone regulates genes in a wide range of cellular processes. However, out of 220 genes differentially regulated in stationary phase, 3% have been characterized

as being involved in cell wall or membrane biogenesis; all are repressed in $\Delta ura3\Delta hpyA$ relative to the parent strain (Table 5). Further analysis of these proteins revealed strong homology to enzymes required to glycosylate the cell surface S-layer and maintain cell shape in the related halophile *Haloferax volcanii* (Guan et al., 2010; Table 5). Importantly, *csg*, the major protein component of the S-layer, is also downregulated in stationary phase histone deletion mutants. Since cell wall- and membrane-associated genes were found in our dataset, we performed several assays to determine whether the morphological defects were due to instability of the S-layer or membrane. These assays included cell surface layer carbohydrate staining, shear stress survival, and bacitracin sensitivity (bacitracin is known to inhibit the transfer of glycosyl residues from the dolichol phosphate carrier to the S-layer protein, causing rounded cell morphology; Mescher et al., 1976; Mescher & Strominger, 1978). No significant phenotypic differences were observed between the histone deletion mutant and $\Delta ura3$ parent strain (Fig. 6). Therefore, we infer that differential histone expression levels are affecting cell morphology through an uncharacterized mechanism that does not affect the gross integrity of the S-layer scaffold or membrane. Together, these results suggest that while histone is involved in regulation of genes in many different processes, it also regulates key components of cell wall biogenesis.

Table 5: Histone-regulated genes associated with S-layer synthesis and glycosylation

Gene Name	arCOG Name	Log2 Expression, $\Delta hpyA : WT$	Closest Characterized Homolog	E value
VNG0001H	S-layer domain	-1.17		
VNG0062G	Glycosyltransferase	-1.59		
VNG1048G	Predicted UDP- glucose 6- dehydrogenase	-1.64	<i>aglM</i> , <i>HVO1531</i>	2.00E-108
VNG1055G	dTDP-glucose pyrophosphorylase	-1.02	<i>aglF</i> , <i>HVO1527</i>	4.00E-108
VNG1062G	Glycosyltransferase	-1.23	<i>aglI</i> , <i>HVO1528</i>	3.00E-46
VNG2679G	Cell surface protein	-1.05	<i>csg</i> , <i>HVO2072</i>	2.00E-155

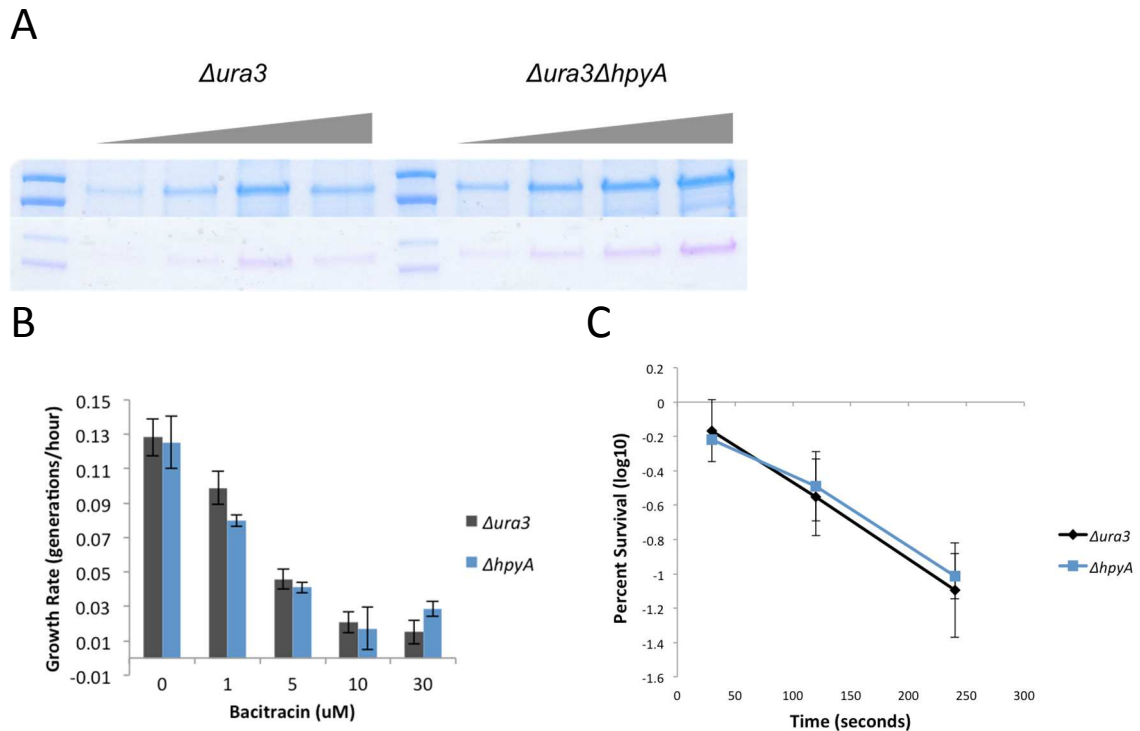


Figure 6: Cell wall and cell membrane are unaffected by histone deletion. A) Comparison of Coomassie-stained S-layer protein (above) with PAS-staining of carbohydrates (below) shows no difference in the proportion of glycosylated S-layer between parent and histone deletion mutant. Shown are technical replicates in which increasing amounts of starting material were used, from left to right. B) *Δura3ΔhpyA* and *Δura3* are equally sensitive to the glycosylation inhibitor bacitracin. C) Mechanical stress was applied to the cell wall and envelope by vortexing cells with beads: *Δura3ΔhpyA* survival is comparable to the parent strain.

2.3.6 Halobacterial DNA is not protected by a regular nucleosome array

A primary characteristic of organisms in which histone is the primary DNA packaging component is the generation of a DNA “ladder” after digestion of chromatin with micrococcal nuclease (MNase). In our first attempt to generate such a ladder in

Halobacterium salinarum, we attempted to permeabilize *H. salinarum* to nucleases by spheroplasting cells in the hopes that this method would not perturb native chromatin structure. However, we found that DNA in spheroplasted cells is not digested by MNase (Fig. 7A & 7B), even at extremely high concentrations of this enzyme. A side-by-side comparison of MNase and DNase digestion in spheroplasts (Fig. 7B) shows that while MNase is not active in these cells, DNase readily cuts DNA, indicating both that spheroplasting has successfully permeabilized the cells and that MNase is specifically inhibited. Intermediate levels of DNA digestion by DNase are suggestive of at least partial protection of DNA by protein *in vivo*. By lysing cells in MNase buffer we were able to restore MNase function, albeit *in vitro*. MNase-digested DNA appeared as smears, without discrete banding (Fig. 7C). Adding crosslinking to preserve protein-DNA interactions in lysate did not appear to change this pattern in MNase-digested samples (Fig. 7D). DNA purified by sucrose gradient was hypersensitive to MNase (Fig 7E); only a single band was observed in the 10 unit sample. This increased activity is likely due to the dilution of salt in the sample. To look at protection of DNA by protein in a specific region, we ran a Southern blot of purified DNA from *Sall*- and MNase-digested chromatin and probed for the chromosomal locus *pykA* (*VNG0324G*). In media without glucose the *pykA* gene is largely silent and therefore we expect this locus to be occupied by nucleosomes if the chromatin environment resembles that of eukaryotes. While a clear ladder is observed in the *Sall* samples, *pykA* bands are only observed in the

no-MNase negative control, despite the DNA smear visualized by EtBr staining (Fig. 7G). This strongly suggests that DNA is not specifically or regularly protected by nucleosome particles.

Table 6: Micrococcal nuclease digestion: tests and conditions.

Assay #	Variables tested	Strain(s)	Method Details or Changes	Micrococcal Nuclease Conditions
1	Rapid assay	NRC-1, OD600 = 1.2	H ₂ O lysis, Ethanol purification of DNA	10 min with 2, 20, 200, and 2000 u
2	Yeast positive control	Yeast, OD600 = 9.83	Lyticase spheroplasting	20 min with 2, 10, 20, 100, 200, and 500 u
3	Crosslink and spheroplast	NRC-1, OD600 = 1.2	All crosslinking with formaldehyde; all spheroplasting with Jarrell & Sprott (1984) solution	200 u for 10, 12, 15, 17, 20, 25, 30, 45, and 60 min
4	Yeast positive control	Yeast, OD600 = 2.7	Crosslink	20 min with 200, 500, 750, 1000, 15000, and 2000u
5	Crosslink and water lysis for lower salt concentration	NRC-1, OD600 = 2.0	Crosslink, 1:10 dilution of lysate in MNase buffer	20 min with 100, 200, 500, 1000, 1500, 2000
6	Spheroplast in pH 9.2, spermidine addition	NRC-1, OD600 = 1.5	With or without spermidine	20 min with 500 or 2000 u
7 (7A)	Decreased DNA concentration, larger MNase range	NRC-1, OD600 = 1.7		20 min with 0, 0.1, 0.5, 1, 5, 10, 50, 100, and 500 u
8 (7B)	MNase versus DNase, stationary phase	NRC-1, OD600 = 1.8	MNase digestion in basal salts buffer to maintain cell integrity; includes overnight digestion test	MNase: 0 u, 50 u 20 min, 50 u O/N DNase: 30 min with 0, 0.5, 1, 1.5, 2, 5, 10
9	MNase versus DNase, log phase	NRC-1, OD600 = 0.51		MNase: 0 u, 50 u 20 min, 50 u O/N DNase: 30 min with 0, 0.5, 1, 1.5, 2, 5, 10

10	Attempt to replicate Shioda et al. (1989) MNase digestion	NRC-1, Stationary phase	DNA isolated with a loop	0, 1, 5, 10 u
11 (7C)	Cell lysis in MNase buffer	NRC-1, OD600 = 1.6	Lysis in micrococcal nuclease buffer	20 min with 0, 5, 10, 50, 100, 500, 1000, and 2000 u
12	Clear lysate	NRC-1, OD600 = 0.38	Preliminary spin clears lysate prior to digestion, lyse in MNase buffer	20 min with 0, 5, 10, 50, 100, 500, 1000, and 2000 u
13	Optimize MNase concentration	NRC-1, OD600 = 1.6	Lyse in MNase buffer	20 min with 0, 5, 10, 30, 50, 70, 100, and 150 u
14	Optimize crosslinking	NRC-1, OD600 = 1.4	Crosslink either 2h or overnight, lyse in MNase buffer	20 min with 0, 5, 10, 30, 50, 70, 100, and 300 u
15 (7G)	Southern blot	NRC-1, OD600 = 1.4	Probe pykA promoter	20 min with 0, 5, 30, 70, and 300 u
16 (7D)	Test ChIP methods	NRC-1, OD600 = 0.8	Protease inhibitors added to ChIP lysis buffer, DNA is sonicated. Crosslink versus no crosslink.	20 min with 0, 5, 50, 70, 100, and 300 u
17 (7E)	Sucrose gradient purification of chromatin	NRC-1, OD600 = 1.5	MNase digestion on sucrose gradient fractions	20 min with 0, 5, 50, 70, 100, and 300 u
18 (7F)	Attempt to replicate Ammar et al. (2012)	NRC-1, OD600 = 2.76	Needle lysis (not sonication), wash in BS 2X (not PBS), but resuspended in PBS + CaCl ₂	20 min with 0, 0.125, 0.25, 0.5, 1, 5 u
19	Attempt to replicate Ammar et al. (2012)	NRC-1, OD600 = 0.37 Hvol, OD600= 0.78	Needle lysis (not sonication), wash in BS 2X (not PBS), but resuspended in PBS + CaCl ₂	20 min with 0, 0.125, 0.25, 0.5, 1, 5 u
20	Attempt to replicate Ammar et al. (2012)	NRC-1, OD600 = 0.5 Hvol, OD600= 0.8	Second wash in 50:50 BS:PBS, digestion in MNase buffer	20 min with 0, 0.125, 0.25, 0.5, 1, 5 u

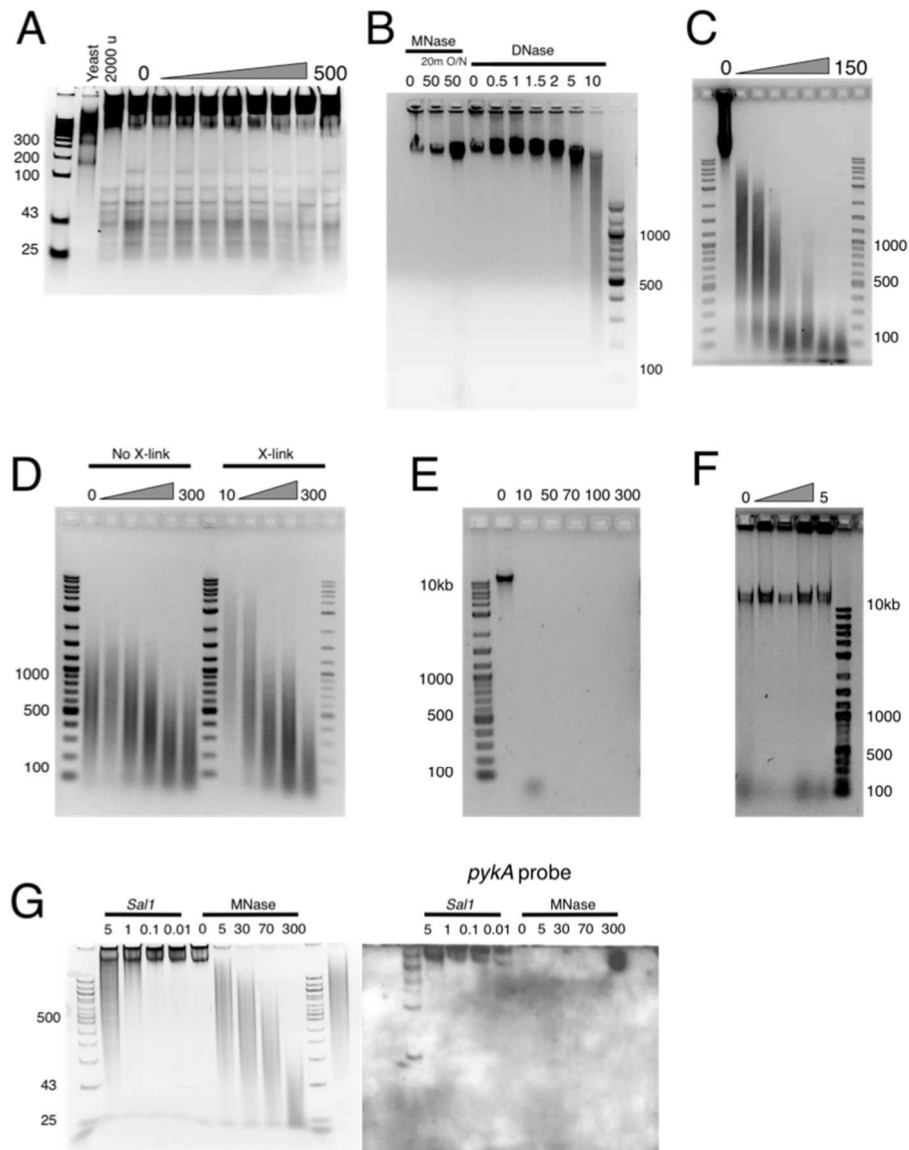


Figure 7: Micrococcal nuclease digestion of haloarchaeal DNA. A) MNase digestion in spheroplasted cells. B) Comparison of MNase and DNase digestion in spheroplasted cells. C) MNase digestion is observed in cells that have been lysed in MNase buffer. D) Crosslinking does not appear to change pattern of MNase-digested DNA; no MNase ladder is observed. E) Rapidly sedimenting DNA from a sucrose gradient is sensitive to MNase digestion. F) Representative attempt (out of 3 independent experiments) to replicate MNase ladder as shown in Ammar et al. (2012); no digestion is observed. G) Southern blot of *SalI*- and MNase-digested DNA. All values for enzyme concentrations are in units of activity. DNA in figures A & G were separated by polyacrylamide gel electrophoresis, for all others agarose was used.

2.3.7 Proteomics analysis of fractionated cells is consistent with open chromatin

While MNase digestion provided no evidence of a regular array of nucleosomes, we reasoned that chromatin could still be composed of a heterogeneous population of architectural DNA-binding proteins. In order to identify the principal protein components of chromatin in *H. salinarum*, DNA was purified *via* sucrose gradient ultracentrifugation. Two major peaks of nucleic acid were detected in sucrose fractions by absorbance at 260 nm. Further analysis by parallel enzymatic digestion of RNA and DNA revealed that the nucleic acid peak in the lower density sucrose fractions consists entirely of RNA, while the secondary peak contains most of the DNA content and includes an additional population of heavier RNA, likely associated with ribosomes (Fig. 8A & 8B). Protein from the DNA fractions, as well as control protein from lighter fractions corresponding to RNA and free protein, was isolated from log (Fig. 8C, left) and stationary phase (Fig. 8C, right) samples.

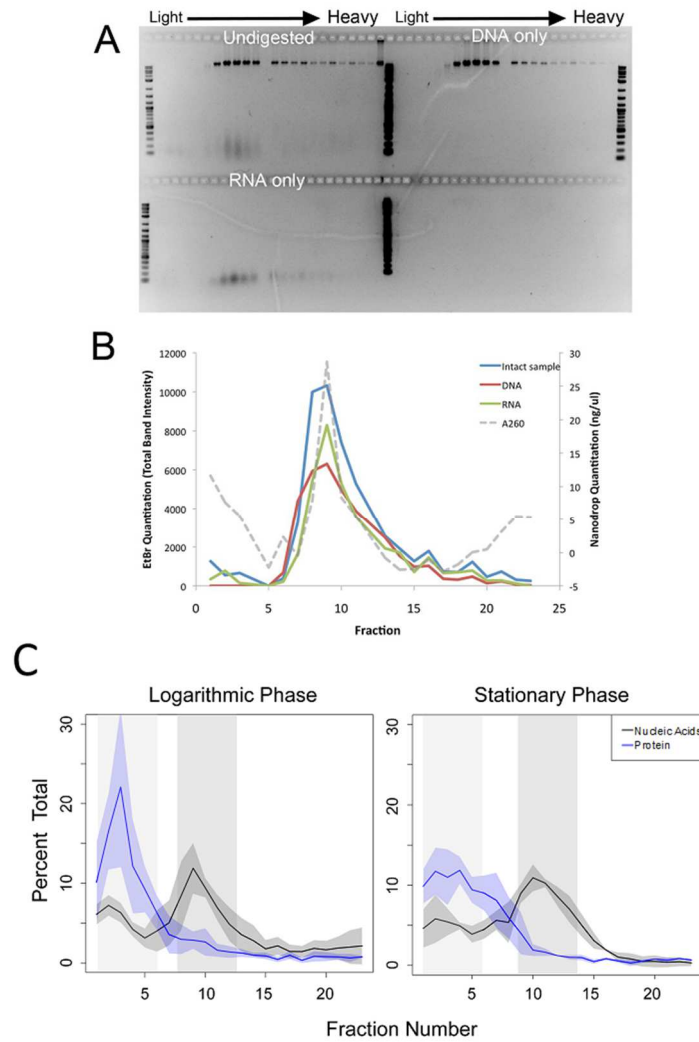


Figure 8: Detection of nucleic acids in sucrose gradient fractions. A) Agarose gel and ethidium bromide (EtBr) staining of nucleic acids from fractions 1-23. Top left: mock digested samples, top right: RNase A-digested samples representing DNA content, bottom left: micrococcal nuclease (MNase)-digested samples representing RNA content. **B)** Image J quantitation of all three treatments from (A) compared to nucleic acid concentration as determined by absorbance at 260 nm (A260; dotted gray line). **C)** Sucrose gradient profiles of DNA and protein from each phase of growth. Proteins present in low-density fractions containing RNA (light gray shaded region) and higher density fractions containing DNA (dark gray shaded region) were identified by tandem mass spectrometry. Nucleic acid profiles (black line) and protein profiles (blue line) represent the mean of 3-4 biological replicate samples; shaded area around each line indicates standard deviation from the mean. Fraction 25, corresponding to unsheared lysate (Takayanagi et al., 1992) is not shown for clarity.

A total of 1,256 unique proteins were identified in all samples using LC-MS/MS (~53% total coverage of the *H. salinarum* proteome). This is a highly sensitive detection level relative to other proteomics experiments conducted thus far for *H. salinarum* (Van et al., 2008). Surprisingly, the majority of proteins detected exclusively in the DNA fractions throughout growth were putatively membrane-associated (43.5%) rather than DNA-associated as expected (3.8% transcription factors and putative NAPs). Furthermore, the five putative NAPs with strong homology to those important for chromatin function in bacteria and other archaea (HpyA, MC1, Sph1, Sph2, DpsA) were present in low abundance in both DNA and control fractions across both log and stationary growth phases: less than 0.41% of total observed peptides was detected for each (Table 7).

Table 7: Proportion of total peptides detected for putative nucleoid associated proteins in DNA and control fractions of *H. salinarum*

Protein Name	Description	Molecular Weight (kDa)	% of Peptides Observed			
			Log Phase		Stationary Phase	
			Control	DNA	Control	DNA
HpyA	Archaeal histone	16	0.06	0.02	0.03	0.02
DpsA	DNA protection during starvation protein	20	0.13	0.12	0.09	0.15
MC1	Archaeal NAP	12	0.06	0.02	0.07	0.02
Sph1	Smc-like protein	71	0.13	0.41	0.20	0.24
Sph2	Smc-like protein	76	0.21	0.28	0.05	0.07

Histone protein (HpyA) accounted for only 0.06% of the observed peptides in any given sample. A linear model was used to determine which proteins common to

both DNA and control populations were significantly enriched over expected frequency in the DNA fractions (see Materials and Methods). Of those proteins enriched in the DNA fractions, 69% of proteins enriched in log phase and 68% of proteins enriched in stationary phase are annotated as ribosome or proteasome subunits. This is consistent with the behavior of rapidly-sedimenting large protein complexes. Therefore, the previous report of rapidly-sedimenting DNA in *H. salinarum* lysate as well as the observations made here are unlikely to be primarily due to the presence of DNA binding proteins. Taken together, these proteomics results are consistent with open, transcriptionally active chromatin throughout growth, rather than chromatin compacted through the binding action of NAP or histone.

2.3.8 Expression of putative haloarchaeal NAPs suggests a dynamic DNA landscape

Gene expression and shape changes in histone mutants are growth phase dependent (Fig. 3 & 4), but NAPs and histone proteins were detected in both growth phases (Table 7). We reasoned that, because the proteomics experiment was only semi-quantitative, this difference might be due to dynamic changes in expression of NAPs throughout growth. To test this, we reanalyzed wild-type and $\Delta ura3$ data from previously published experiments in which gene and protein expression was monitored at several time points throughout growth in batch culture (Facciotti et al., 2010) and during growth rate perturbations in a fermenter (Schmid et al., 2007), respectively.

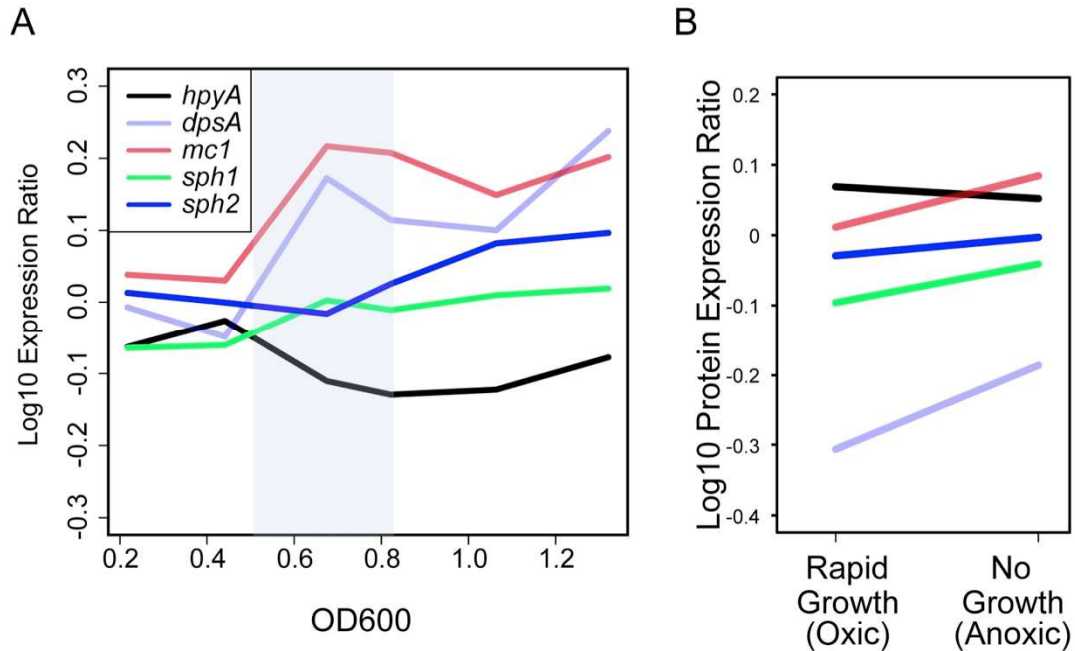


Figure 9: Gene and protein expression of putative NAPs as a function of growth. A) Gene expression profiles as a function of growth phase. Normalized expression ratios for *hpyA* (VNG0134G), *dpsA* (VNG2443G), *mc1* (VNG2508C), *sph1* (VNG6320C), and *sph2* (VNG6173C) from wild-type NRC-1 and Δ *ura3* growth data from Facciotti et al., (2010) were binned and average values for each gene were plotted. The blue shaded region highlights the period of transition from log to stationary phase. Line colors for each gene are explained in the legend. B) Protein expression ratios of NAPs during rapid growth (log phase, oxic conditions) vs. non-growth (stationary phase, anoxic conditions). Each protein profile corresponds to the mean expression in each condition. Line colors are as in (A). Original data was obtained from Schmid et al. (2007).

Clear differences in expression of the five putative NAPs discussed above (Table 7) were observed between log growth and stationary phase (Fig. 9A). Expression of *hpyA* mRNA was higher during rapid growth than slow growth. In contrast, *dpsA*, *mc1*, and *sph1* expression dynamics were anticorrelated with those of *hpyA* (Fig. 9A). Expression of *sph2* was independent of growth. Similar patterns were observed for the proteins

encoded by these genes in quantitative proteomics data (Fig. 9B). In the histone mutant background, *mc1* expression is slightly repressed during stationary phase ($p < 0.05$), suggesting a regulatory interrelationship between these genes. These results are consistent with the hypothesis that, although detected at levels too low to compact chromatin, a dynamic, diverse NAP population may be responsible for growth phase-dependent differential gene expression and phenotypic consequences as observed for *Δura3ΔhpyA*.

2.4 Discussion

2.4.1 Unexpected phenotypic consequences of histone gene dosage alteration suggest a unique function for haloarchaeal histones

Genetics experiments conducted here suggest that histone is not essential for growth under standard or stress conditions, but is required for maintenance of cell shape in *H. salinarum*. These observations are in contrast to those from histone mutants generated in other archaea and the eukaryotes, which require at least one histone fold domain for viability (Cubonova et al., 2012; Matsubara et al., 2007; Nakanishi et al., 2008; Winzeler et al., 1999) and survival under stress response conditions similar to those tested here (Weidenbach et al., 2008). In eukaryotes, histone protein overexpression also leads to cell elongation; however, this is likely related to disruption of cell division dynamics, as growth is severely perturbed in yeast (Sopko et al., 2006). Loss of transformation competence due to histone deletion in the hyperthermophilic archaeon *Thermococcus*

kodakarensis suggests that membrane or cell wall integrity may be compromised (Cubonova et al., 2012), although the morphology of these mutants has not been tested to our knowledge. However, unlike the *T. kodakarensis* histone deletion mutant, *H. salinarum* Δ *ura3* Δ *hpyA* was readily transformed (Figures 3C, Figure 4D) despite gross morphological differences from the parent strain (Figure 4). Deletion mutants also maintained membrane and S-layer integrity (Figure 6). Together, these phenotypic observations point to a novel role for histone as a specific regulator of cell shape in haloarchaea.

Gene expression data presented here suggests potential pathways linking histone, gene expression, and cell shape. Functional analysis of genes perturbed by histone deletion suggest that HpyA is required for wild type expression of several genes whose products are putatively associated with the membrane and the cell wall. This includes transporters, S-layer (archaeal cell wall) glycosylation enzymes, and a putative phospholipase D (Table 5; Guan et al., 2010). Some transport proteins are required for maintenance of wild type cell morphology in *E. coli*, albeit indirectly (Vega & Young, 2014). Further, S-layer glycosylation is required to maintain cell shape in *H. salinarum* (Mescher & Strominger, 1976; Todor et al., 2014). Although the overall integrity of the S-layer of Δ *ura3* Δ *hpyA* mutants remained grossly intact (Figure 6), more sensitive detection methods such as mass spectrometry might be required to detect subtle differences in glycosylation (Cohen-Rosenzweig et al., 2012). Nevertheless, future

investigation into these interesting possibilities could determine mechanisms by which histone affects morphology in *H. salinarum*.

2.4.2 Histone as a regulator of gene expression in open chromatin

In *H. salinarum*, histone and other putative NAPs appear to be present in low abundance in both the cytosolic control and the DNA fractions, according to the semi-quantitative metric of number of peptides observed per protein (Table 7). This is in sharp contrast to eukaryotes, where histone proteins compact the entirety of the genome and are one of the most abundant proteins in the cell. For example, in humans histones comprise ~60% of all proteins associated with chromosomes by mass (Uchiyama et al., 2005). In *E. coli*, despite the large proportion of protein-free DNA, NAPs make up a significant fraction of all cellular protein. For example, Fis proteins make up approximately 12% of all protein molecules in exponential phase (Azam et al., 1999). Therefore, HpyA and NAP proteins in *H. salinarum* are unlikely to play a major role in the sedimentation rate observed here or in “beads-on-a-string” structures and transcription start site phasing observed previously (Shioda et al., 1989; Takayanagi et al., 1992; Ammar et al., 2012).

This is in agreement with the results of micrococcal nuclease digestion of *H. salinarum* chromatin. We did not observe MNase ladders in any of our tests, whether *in vitro* or *in vivo*. This is in contrast to other euryarchaea such as *Thermococcus kodakarensis* (Maruyama et al., 2013), *Methanothermus fervidus* (Pereira et al., 1997), or *Haloferax*

volcanii as described in Ammar et al. (2012), in which two bands were observed in micrococcal nuclease-digested DNA. Attempts to replicate data from Ammar et al. (2012) in both *H. salinarum* and *H. volcanii* using their protocol (with minor modifications, see Materials and Methods) yielded intact undigested DNA in all three experiments (representative gel shown in Fig. 7F). This suggests that while some euryarchaea use nucleosomes to package the entire genome, haloarchaeal genomes are not uniformly compacted by histone as a regular array of nucleosome particles.

While histone may not be the primary compacting agent for DNA, our data suggest that HpyA is required for modulating global gene expression, acting as a repressor or an activator depending on the growth phase (Figure 4). Similarly, many different DNA architectural proteins also regulate gene expression in the eukaryotic and bacterial domains of life. In eukaryotes, nucleosomes inhibit transcription non-specifically by blocking transcription factors and the general transcriptional machinery from accessing DNA. *In vitro* studies suggest that some archaeal histones conform to this paradigm (Soares et al., 1998; Thomm et al., 1992; Wilkinson et al., 2010). The number of genes affected by histone deletion in *H. salinarum* is similar to that observed in *T. kodakarensis* single histone mutants (Cubonova et al., 2012) and *Escherichia coli* single NAP mutants (Prieto et al., 2012). The directions of gene expression changes are also growth phase-dependent in *E. coli* NAP mutants (Prieto et al., 2012; Kahramanoglou et al., 2011). In bacteria, growth phase-dependent effects are commonly attributed to the

dynamic protein composition of the genomic landscape (Prieto et al., 2012; Rimsky & Travers, 2011), consistent with the phase-dependent dynamics of putative *H. salinarum* NAP protein expression observed here (Figure 9). The polar distribution of charges on the surface of HpyA is also reminiscent of halophilic transcription factors (Kennedy et al., 2001). Together, our results suggest that histone in *H. salinarum* functions primarily as a modulator of gene expression. Although this regulatory function is shared with bacterial NAPs, the primary effect of HpyA regulation appears to be specific to cell morphology rather than DNA compaction.

In summary, here we have investigated the phenotypic consequences of histone dosage in the Haloarchaea. Together, the altered morphological and stress resistance phenotypes of histone mutants, strong conservation of the fused dimer sequence throughout the Haloarchaea, and the absence of canonical eukaryotic chromatin suggest specialized functions for histone in the high intracellular salt environment. Histone deletion affects gene expression in a bifunctional, growth phase-dependent manner. Unexpectedly, there is a clear correspondence between histone expression levels and cell morphology. Therefore, haloarchaeal histone protein may serve primarily as a modulator of gene expression regulating cell shape.

2.5 Acknowledgments

I would like to thank Claudia Ofori-Marfoh and Callan Corcoran for technical assistance, Marc Facciotti for the generous gift of plasmids pRSK01 and pMTFCHA, and Duke University School of Medicine Proteomics and Metabolomics Shared Resource for protein identification service. This research was supported by grants from the National Science Foundation (NSF-MCB-10-52290) and Duke Arts and Sciences Council Committee on Faculty Research, both to AKS.

3. Frequent genomic duplications and deletions are mediated by insertion elements in *Halobacterium salinarum*

Data reanalyzed herein was sourced from previously published experiments as described in Appendices E and F with the exception of the histone overexpression data which was generated by the author. Preprocessing of gene expression data was performed by David Corcoran at the Integrative Genomic Analysis Shared Resource in the Duke Center for Genomic and Computational Biology. All figures, analyses, and text within this chapter represent the sole effort of the author.

3.1 Introduction

Genetic rearrangements include DNA amplifications, deletions, inversions and translocations; all such events contribute to genomic instability (Darmon & Leach, 2014). Such rearrangements may occur on a kilobase or megabase scale and can rapidly generate diversity, change the abundance of a given gene, and lead to new regulatory programs (Darmon & Leach, 2014). Genetic rearrangements are common in organisms from all three domains of life and are frequently mediated by illegitimate homologous recombination at interspersed repetitive genomic elements. In humans, this includes retrotransposons and Alu elements that, along with other repetitive elements, comprise approximately half the genome (Hedges & Deininger, 2007). In Bacteria, small repetitive

transposable sequences called insertion sequence (IS) elements are generally far less abundant, comprising no more than 2% of DNA in most species, although they may make up as much as 21% of some genomes (Newton & Bordenstein, 2011). In both Eukaryotes and Bacteria, the presence of repetitive elements leads to genomic instability. While most rearrangements confer neutral fitness changes, they occasionally lead to deleterious or beneficial phenotypes such as cancerous growth or the development of drug resistance (Heyer, 2015; Darmon & Leach, 2014). Active IS elements are also known to facilitate genomic rearrangements in the Archaea; however, the relationship of these changes with fitness on the individual or population level is unknown (Brugger et al., 2002; Martusewitsch et al., 2000; Redder & Garrett, 2006).

IS elements are the most abundant mobile elements in the Archaea and are 0.5–2 kb in length, encode a transposase, and are flanked by terminal inverted repeats (Brugger et al., 2002). These elements are most dense in *Sulfolobus species* and members of the Haloarchaea, with up to 200 copies in a given genome (Brugger et al., 2002). Spontaneous mutants in the uracil biosynthesis pathway conferring resistance to 5-fluoroorotic acid (5-FOA) were solely caused by the mobilization of IS elements in *Sulfolobus solfataricus* and were isolated at a rate of 10^{-4} to 10^{-5} (Martusewitsch et al., 2000). In *Halobacterium salinarum* frequent transposition of IS elements into the promoter or coding regions of the gas vesicle cluster disrupts the formation of these organelles and cause the clear colony phenotype in 10^{-2} to 10^{-4} colonies (Pfeifer et al., 1988).

Because of the high frequency of IS-generated mutants in *H. salinarum*, the mobility of IS elements and the resultant consequences have been most thoroughly investigated in this archaeon.

The chromosomes of the hypersaline euryarchaeon *H. salinarum* sp. NRC-1 are gene-dense with poly- and mono-cistronic operons found on both strands (Grohmann & Werner, 2011). NRC-1 has one large chromosome (~2 Mb) and two smaller megaplasmids (195 & 365 kb). This organism is highly polyploid, with approximately 30 copies of the chromosome in rapidly dividing cells and 15 copies in stationary phase (Breuert et al., 2006). This polyploidy appears to be differentially regulated between chromosomal elements, as two of the three megaplasmids in NRC-1 remained at five copies throughout the growth curve (Breuert et al., 2006). The megaplasmids pNRC100 and pNRC200 of strain NRC-1 (191 kb and 365 kb, respectively) contain several large repeat regions, with ~120 kb shared in common between the two (Ng et al., 2000).

A closely related *H. salinarum* strain R1, which is diverged from NRC-1 by evolution in the laboratory, is no longer able to produce gas vesicles (Sapienza et al., 1982). As evidenced by sequence comparison of the two strains, IS elements may cause genotypic changes not just by transposition of the IS element itself, but by providing interspersed homologous templates by which homologous recombination may lead to rearrangements of large regions of DNA (Darmon and Leach, 2014). Indeed, R1 encodes four megaplasmids representing topological rearrangements, deletions, and

duplications of several regions of the two NRC-1 megaplasms (Pfeiffer et al., 2008). Outside of these genetic rearrangements very few single-nucleotide changes were observed between the two strains despite decades of divergent history, which suggests that rearrangement by IS elements may be a primary generator of biodiversity in this organism (Pfeiffer et al., 2008).

Here we expand the understanding of *H. salinarum* genomic instability by investigating the link between instability and gene expression and employing new methods to detect chromosomal duplications and deletions. We show that DNA amplification can lead directly to increases in gene expression. By applying a chromosome-segmenting algorithm designed to detect DNA copy number variants (CNVs) to the analysis of a large compendium of 1154 mRNA gene expression microarray datasets for *H. salinarum*, we detected the co-expression of large co-linear regions of the megaplasms spanning multiple operons. These regions are delineated by IS elements and suggest variable gene dosage as a causative agent of large-scale co-regulation in the megaplasms. Meta-analysis of microarray data for sheared genomic DNA provides the first evidence that IS-mediated rearrangements extend to the chromosome and are not restricted to the megaplasms. The phenotypic and fitness consequences of these rearrangements in the species *H. salinarum* are discussed. These methods are generalizable to any microbial system for which gene expression and DNA quantification data exists.

3.2 Methods

3.2.1 Strains and growth conditions

Control strain *Δura3/pMTFCHA* (KAD101) and histone overexpression strain *Δura3/pMTFCHA::hpyA* (KAD102) were originally described in Dulmage et al., 2015. All cultures were incubated at 42°C with agitation in complex medium (CM; 250 g L⁻¹ NaCl, 20 g L⁻¹ MgSO₄•7H₂O, 3 g L⁻¹ sodium citrate, 2 g L⁻¹ KCl, 10 g L⁻¹ peptone). The growth medium was supplemented with uracil (50 μg mL⁻¹) and the antibiotic mevinolin (1 μg mL⁻¹).

3.2.2 Gene expression microarrays and analysis for strains KAD100 and KAD101

Three biological replicate cultures of strains *Δura3/pMTFCHA* and *Δura3/pMTFCHA::hpyA* were grown to mid-log phase in CM with uracil and mevinolin and then subcultured to an OD₆₀₀ of 0.05 for further growth prior to harvesting. Culture harvest, RNA extraction, labeling, hybridization to custom Agilent ORF arrays, spot averaging, and downstream analysis were performed as described in section 2.2.9.

3.2.3 Concatenation and normalization of 1154 gene expression arrays

Raw data from 2308 publically available or unpublished gene expression microarrays (1154 arrays with corresponding dye-swap control arrays; see Appendix E for references) was normalized as follows: for each array, probes with a mean low intensity scan value > 0 and an unsaturated mean high intensity scan value were identified. Arrays with $\geq 1,000$ probes fitting this criterion, with an overall $R^2 \geq 0.95$ between the high and low intensity signals for given probes, were included in subsequent analyses. A linear regression model was then used to project the raw low intensity values to raw high intensity values in the instances where the mean high intensity value reached saturation. Data files were read and processed using the R software package *limma* from Bioconductor (Ritchie et al., 2015). Background subtraction and within-array normalization of probe intensities was performed as described in Dulmage et al. (2015). The expression value for each gene per array was then defined as the median probe value for that gene. The expression ratio for each gene per experiment was then calculated as the experimental condition divided by the wild-type control sample. The average of the dye-swap experiments was taken as the final value for each gene within each experiment. Quantile normalization was used to standardize the expression ratios across all of the arrays. The expression ratios across all experiments were then mean-centered and converted to z-scores, where final values correspond to distance from the mean in units of standard deviation.

3.2.4 Detection and mapping of correlated regions of gene expression in normalized data from 1154 arrays

Probes with missing values in any experiment were removed prior to analysis.

Normalized arrays were analyzed and segmented using the R package DNACopy as described by the authors, using default smoothing and segmentation parameters (Venkatraman & Olshen, 2007). Correlated segments of gene expression, as determined by the circular binary segmentation algorithm, were thresholded for a z-score of at least 1. Because the first 113kb of pNRC100 and pNRC200 are identical, representing a large duplicated region, the first 5 genes assigned to pNRC200 on the gene expression arrays were already represented in the pNRC100 probes and thus were removed prior to mapping. Frequency maps (Fig. 12, Fig. 13) were generated by calculating the number of times a probe was detected in segments meeting significance and size thresholds and then dividing this number by the total number of arrays in the analysis set.

3.2.5 qPCR detection of chromosomal instability

Phenol/chloroform extraction was used to harvest DNA from three biological replicates of stationary phase *KAD101 (ura3/pMTFCHA)* and *KAD102 (ura3/pMTFCHA::hpyA)* cultures that had been inoculated from the same plate as the gene expression microarray samples. DNA was amplified using the SsoAdvanced SYBR Green Supermix (Bio-Rad) according to manufacturer's instructions. At least 3 technical replicates were analyzed

for each biological replicate. Relative DNA dosage was calculated using the $\Delta\Delta C_t$ method (Livak & Schmittgen, 2001). Probes in the putatively amplified region (*VNG5097H*, *VNG5102H*, *VNG5148H*) were compared to reference loci *VNG5019G* and *VNG5192H* flanking either side of this region. Primer pairs are listed in Table 8.

Table 8: qPCR primer pairs

Primer Name	Gene	Sequence	Melting Temperature (°C)	Probe Size (bp)	Reference
5019_qPCR_F	<i>VNG5019G</i> , <i>gvpM1</i>	GAGCCAACAAAA GACGAGAC	61	152	This study
5019_qPCR_R	<i>VNG5019G</i> , <i>gvpM1</i>	GTGGTCATGCCA GCAATC	60		This study
5097_qPCR_F	<i>VNG5097H</i> , <i>hypothetical</i>	CAGAATGGCGGT CCGAAG	62	174	This study
5097_qPCR_R	<i>VNG5097H</i> , <i>hypothetical</i>	CCACCAACGGTT GAACTGTC	63		This study
5102_qPCR_F	<i>VNG5102H</i> , <i>hypothetical</i>	GCTTGGACTTGCA CAAGAC	61	208	This study
5102_qPCR_R	<i>VNG5102H</i> , <i>hypothetical</i>	CGCTCGGAGTCC GATAAAG	62		This study
5148_qpcr_F	<i>VNG5148H</i> , <i>hypothetical</i>	CTTCGCCGCATTC GAGAG	62	199	This study
5148_qpcr_R	<i>VNG5148H</i> , <i>hypothetical</i>	CGGAGAGCGAAT CAGTCAATC	62		This study
5192_qpcr_F	<i>VNG5192H</i> , <i>hypothetical</i>	CCGTTATTCAGCG CGTCCTC	65	134	Dulmage <i>et al.</i> , 2015
5192_qpcr_R2	<i>VNG5192H</i> , <i>hypothetical</i>	CATGCTGCGCGTT CAATTCTG	65		Dulmage <i>et al.</i> , 2015

3.2.6 Detection of copy number variants in chromatin immunoprecipitation (ChIP) microarray data

Raw intensity values were extracted from ChIP microarray channels representing randomly sheared genomic DNA from the whole-cell extract ChIP control (also known as “input DNA”). Flagged array spots were removed from analysis and the remaining raw intensities were converted to z-scores. These values were then median normalized so that the majority of DNA ratios centered on zero. A total of 48 arrays were analyzed; the criteria for selection were the availability of single-channel data and the chromosomal coordinates for each probe. Metadata for these experiments can be found in Appendix F. These data were analyzed using the DNACopy segmentation algorithm (Venkatraman & Olshen, 2007) as described above for the analysis of gene expression data. For generation of the composite frequency map, segments were filtered for those fragments that were at least 0.5 standard deviations from the mean and at least 5000 bp in span. Less stringent thresholds were employed here in order to preserve information from our relatively small data set. However, increasing the stringency of the threshold parameters at this step did not change the location of significant peaks in the composite map (data not shown). As these arrays were not designed for the investigation of linearly arranged probes, the large duplicated regions on the megaplasmids were condensed into a single region regardless of their orientation within the genome or relative copy number, and thus we did not analyze this data. Biological replicates were not combined, as genomic rearrangement events in the plasmids are known to occur at a

frequency of up to 1 in 100 cells (Pfeifer & Blaseio, 1989). To control for possible artifacts of array design, segmentation was also applied to input DNA data from high-density ChIP tiling arrays (Tonner et al., 2015); CNVs were observed in similar regions of the genome despite the independent array design (data not shown).

3.2.7 Statistical analysis of IS element association with copy number variants

CNV peak regions were defined as those genomic coordinates in which 10% or more of the arrays contained a fragment in the region, altogether yielding a total genomic fraction of 413,945 bp, or approximately 20.6% of the genome. Here, we use the locations of IS elements as detected by the database IS Finder (<http://www-is.biotoul.fr/>), resulting in total of 72 IS elements in the genome of *H. salinarum*, 17 of which are located in the chromosome. We bootstrapped to generate a *p*-value: 17 IS elements were randomly assigned positions throughout the genome and the number of those elements which fell within the CNV peaks was recorded for each of 100000 iterations.

3.3 Results

3.3.1 Genome instability results in large-scale coordinate changes in gene expression

Megaplasmid instability, including segmental inversions, duplications, and deletions, is well documented in the literature (DasSarma, 1993; Ng et al., 1998; Pfeiffer et al., 2008),

but a direct relationship between instability and large-scale changes in gene expression has not been determined. In the strain *KAD102*, which overexpresses the gene encoding the histone protein of *H. salinarum* (Dulmage et al., 2015), pNRC100 contains a 93.5 kbp region (pNRC100: 38837–132357) that is upregulated across both log and stationary phase growth conditions tested for this strain. In contrast, gene expression in this region remains unchanged in the empty vector control strain, *KAD101* (Fig. 10A). This region is flanked by multiple IS elements (641 bp away from the 5' end of the 93.5 kb region, and directly demarcating the 3' end of the region), which can cause megaplasmid DNA rearrangements under standard laboratory conditions (Pfeifer et al., 1986). To determine whether this 93.5 kb region is regulated at the level of expression or is present in higher copy number than in the wild type, quantitative real-time PCR (qPCR) was performed on DNA purified from strains *KAD101* and *KAD102*. Probes flanking either side of the upregulated region, which also overlaps a region known to be duplicated elsewhere on pNRC100 (pNRC100: 1-112972; Fig. 10B), were compared to three probes internal to the upregulated region (Fig. 10B). All probes inside the upregulated region showed at least a 2-fold increase in DNA copy number (Fig. 10C), suggesting that coordinate expression detected by microarray was due to DNA amplification. While the region of upregulated expression overlaps with the large duplicated region in common between the two megaplasms, it does not share exact breakpoints, suggesting that duplications and deletions may occur at any region encoding IS elements. This result also suggests that

copy number variation in large regions of the genome are associated with significant effects on gene expression, in this case affecting approximately 50% of the genes on megaplasmid pNRC100. Indeed, within this amplified region are two general transcription factors, *tbpB* and *tbpC*. It is possible that by upregulating the copy number and expression of two global regulators this amplification affects gene expression throughout the genome and therefore that gene expression changes are not restricted to the unstable region.

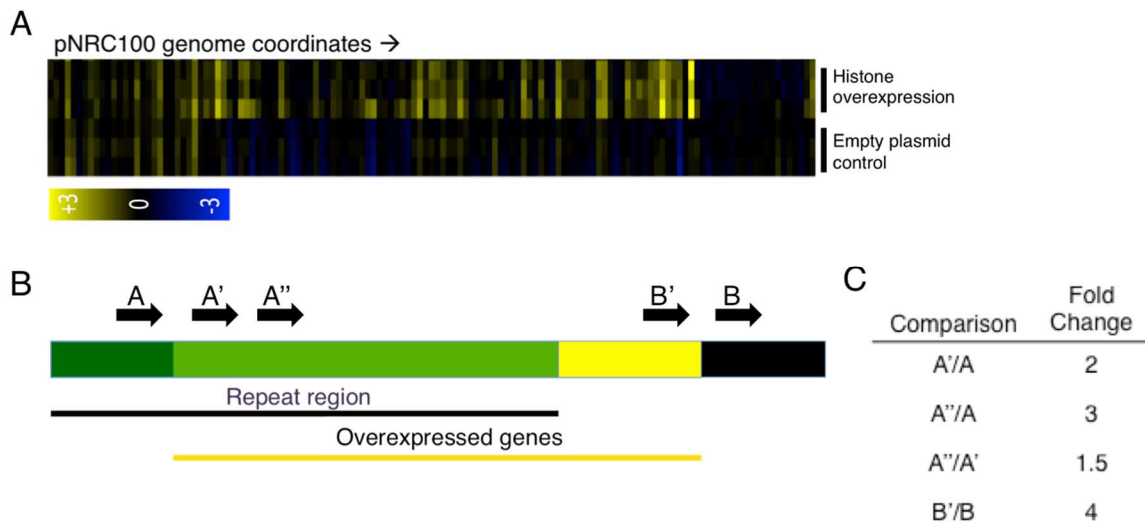


Figure 10: Large-scale genomic changes influence gene expression. A) Gene expression changes between the histone overexpression strain *KAD102* and the empty vector control strain *KAD101* in the megaplasmid pNRC100; shown here are three biological replicates. Arrangement of probes by genomic location shows that a contiguous region is upregulated in *KAD102*. B) Schematic showing location of qPCR probes for confirmation of genomic duplication. In green is the repeat region of this chromosome; this overlaps but does not completely coincide with the region of overexpression as shown in yellow. C) Fold change in DNA content between probes as determined by qPCR.

Having determined that gene expression can be directly affected by duplications or deletions and that these events can be detected by microarray analysis, we developed a workflow to detect other such events in previously published gene expression data. Gene expression arrays were computationally segmented into regions of significant correlation using the R package DNACopy (Venkatraman & Olshen, 2007). This method is commonly used in the detection of genome deletions or amplifications by comparative genome hybridization (CGH) and is readily adaptable to the analysis of any linearly arranged probes. By repeating this process for 1154 microarray experiments performed with a common reference (Brooks et al., 2014; see Appendix E for references), we built a frequency map of position-dependent co-expression for *H. salinarum* on the megaplasmids. See Materials and Methods and Figure 11 for details regarding the normalization and segmentation of these arrays.

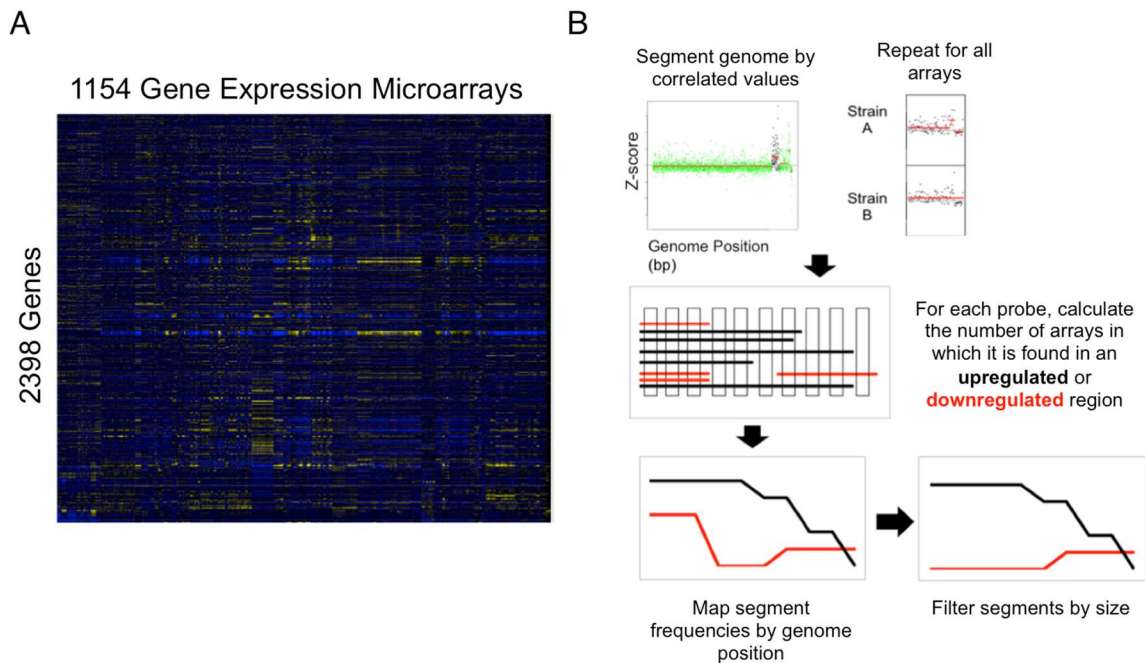


Figure 11: Microarray processing and workflow of gene expression analysis. A) In this analysis, 1154 gene expression arrays were concatenated, normalized, and values were converted to z-scores. B) Correlated regions of gene expression were segmented for each array. All segments were thresholded by z-score, and the frequency of segments in the dataset was plotted by genomic position.

The 1154 microarrays represent data from 56 different mutant and wild-type strains and 72 different experimental conditions (e.g. response to various stressors), 83% of which were conducted in mid-logarithmic phase samples. All experimental RNA samples were hybridized against a common mid-log phase reference and only segments showing significant (1 standard deviation from the mean) gene expression changes from this reference sample were included in our maps. To confirm that this method can accurately detect breakpoints of significantly co-expressed regions, we first generated a

map in which all segment sizes were represented by frequency in the dataset (Fig. 12A). The genome was segmented into size windows of 100 bp to 100 kb and correlation between gene expression values from neighboring segments was calculated (Fig. 12A). The large peak at pNRC100: 16451–25376 corresponds to the gas vesicle cluster. As expected, most of the arrays in which gas vesicle transcription is repressed (10%) correspond to log phase or high oxygen samples, while arrays in which gas vesicle genes are upregulated (22%) include low oxygen or stationary phase samples. Peaks 1, 2, and 3 from the pNRC200 map correspond to previously identified operons involved in siderophore biosynthesis, metal transport, and arginine metabolism (Fig. 12A; Koide et al., 2009). Because operons are typically co-transcribed, the recapitulation of these genomic structures confirms that the algorithm can be robustly applied to gene expression data.

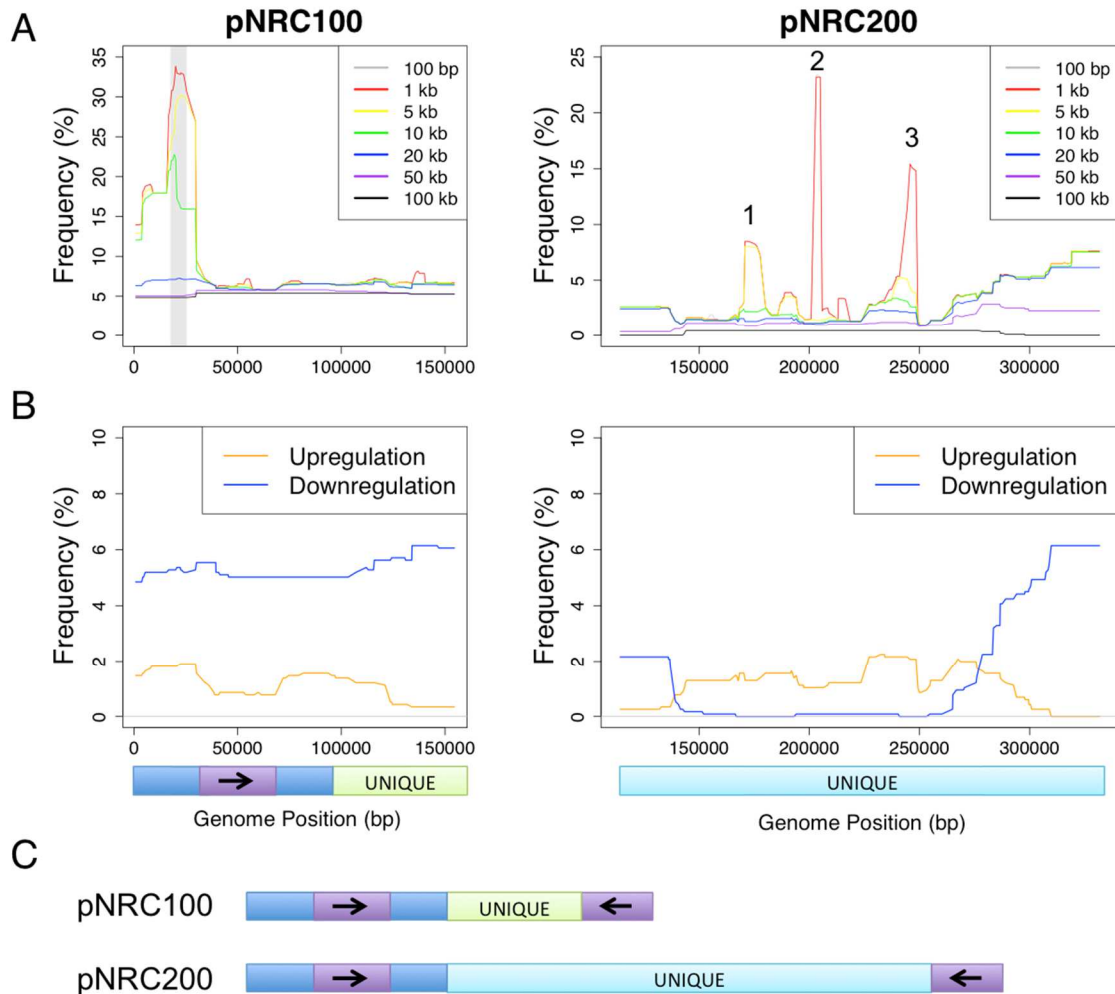


Figure 12: Regions of correlated gene expression in the genome of *Halobacterium salinarum*. A) The frequency of gene expression segments meeting the threshold of one standard deviation from the mean and minimum size criteria as indicated in the legend is plotted against genomic location. Shown here is the distribution of correlated segments from a minimum of 100 bp (gray) to a minimum of 100 kb (black) in length. Vertical gray bar in pNRC100 frequency map represents the location of the gas vesicle cluster. B) Directional regulation of correlated gene expression regions that are at least 20 kb in size. Upregulated gene expression is shown in orange; downregulated gene expression is shown in blue. Bars under pNRC100 and pNRC200 represent the location of repeat and unique regions as represented in the gene expression arrays. Black arrow indicates where a repeat is inverted. C) Diagram of repeat regions (dark blue and violet) and unique regions in the complete genome sequences of the megaplasmids.

To remove operons from analysis and therefore select for those events most likely caused by genomic instability, we looked at only co-expressed segments 20 kb in size or larger. Gene expression follows distinct patterns, suggesting duplication or deletion is the causative agent of these correlated expression agents (Fig. 12B). In pNRC100, gene expression is downregulated over the entirety of the chromosome in about 5% of the arrays. Most of these arrays are from a single large experiment: the long-term tracking of gene expression over diurnally entrained cultures (Whitehead et al., 2009). Because gene expression is downregulated in the megaplasmid over the entire length of the time course, regardless of light conditions, it is possible that the observed changes are not due to repression, but rather due to complete loss of the duplicated region of this plasmid in the original culture. We also find that there are two regions in pNRC200 that are frequently coordinately downregulated in a pattern, suggesting that genetic instability may be the causative agent. Coordinate gene expression observed in the region upstream of the 139 kb breakpoint, in addition to the flanking of this region by the IS element ISH8E, suggests that this region may be unstable and therefore carry an independent gene regulatory program. The region downstream of the gene expression breakpoint at approximately 276 kb corresponds to region T, a segment flanked by an ISH8 element; region T was observed to be one of the many repositioned segments in R1 (Pfeiffer et al., 2008). Contrary to the frequent coordinated downregulation of the aforementioned flanking regions, the center, non-duplicated

region of pNRC100 is only observed to be coordinately upregulated. We suggest that these patterns, rather than representing large transcriptionally linked regions, are due to deletion or amplification of DNA.

3.3.2 Chromosomal copy number variants are associated with mobile elements

While much is known about the instability of haloarchaeal megaplasmids, little is known about stability of the main chromosome, despite the presence of potentially destabilizing mobile elements throughout the genome. In order to investigate this question, we analyzed signal intensities for randomly sheared genomic (input) control DNA from previously published ChIP experiments using microarrays tiled at 500 bp resolution (See Appendix E for references). Scaled raw-intensity signals were used as input into the DNACopy software as described in Methods. Shown in Figure 13 is data from 48 arrays, 36 of which are from stationary phase cultures and the other 12 from log phase cultures. Represented in this analysis is a total of 9 strains, including 2–6 biological replicates each, plus 9 conditions. Segmentation generated a total of 1303 segments for all 48 arrays, 1231 of these being on the chromosome. A total of 368 of the chromosomal segments were above our significance threshold of 0.5 standard deviations in DNA content from the mean and were greater in size than 5 kilobases in length, for an average of about 7 significant chromosomal fragments per microarray.

Putative CNVs were more common than assumed; out of all 48 arrays, only two did not exhibit any CNVs by our thresholding criteria, with some microarrays generating up to 18 of these segments. In order to look for regions with frequent CNVs, we generated a composite map of all fragments. This gives us a total of 17 “hotspots” in the chromosome, with an average size of approximately 24 kb and a frequency of 10-81%. Immediately apparent from this analysis is that DNA amplification is far more common than depletion in our dataset (525 vs. 88 segments, respectively).

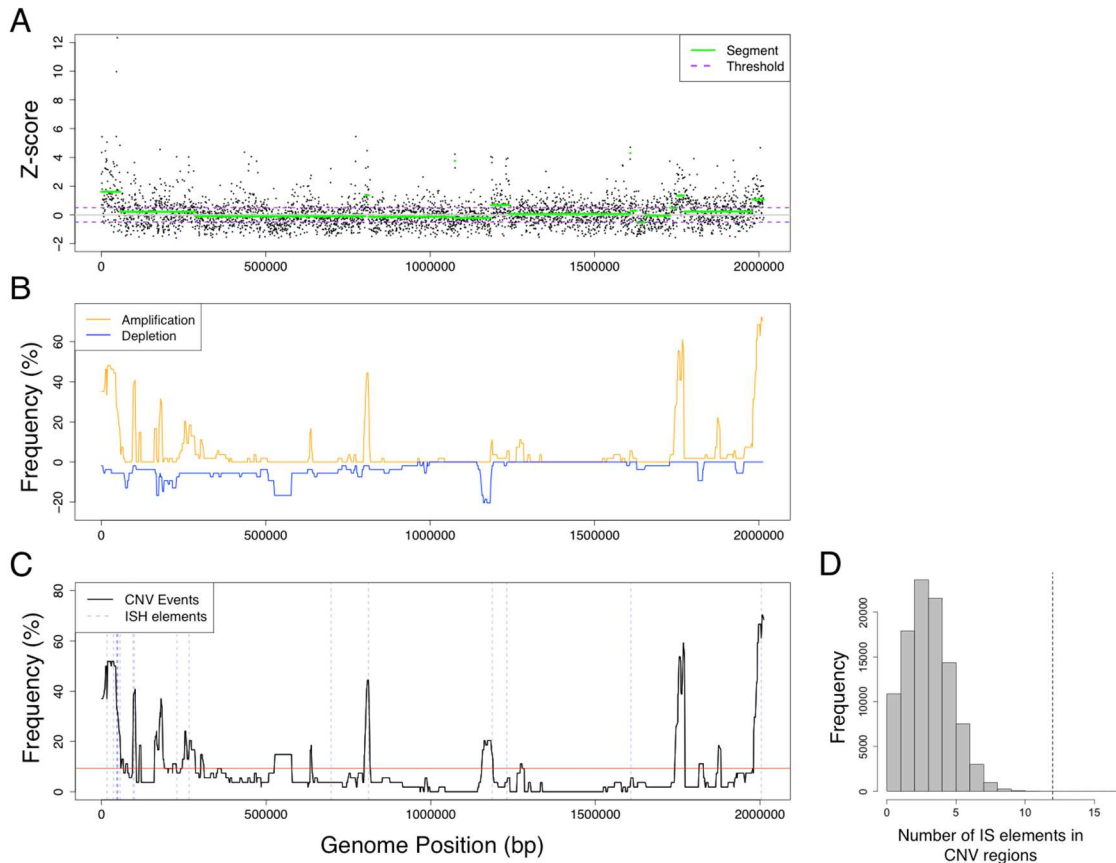


Figure 13: Distribution of chromosomal copy number variants (CNVs). A) Example of segmenting algorithm output. Individual points represent scaled raw intensity values for each microarray probe. In green are the segmented regions according to regions of correlated values; purple represents the threshold for amplified or depleted segments. Shown here is the data for the log phase *trmB* strain grown in 0.1% glucose. **B)** Composite map of CNVs over all 48 arrays. Frequency of events by genomic locus is shown for amplified regions (orange) or depleted regions (blue). **C)** Association of all CNVs (black line) with IS elements (dotted blue lines). IS elements fall within the peak regions (above the red line) more frequently than expected by chance, as shown by bootstrapping analysis (D). Dotted line in (D) corresponds to the observed number of IS elements in the CNV peaks.

Copy number variants were most apparent at the ends of the chromosome (Chr: 1977676-66421); the majority of these CNVs appeared to be amplification events. As the

chromosome is a circular molecule, this is likely to be representative of a single large unstable region rather than two separate peaks. Significantly, this region encodes the S-layer gene, *csg*, as well as several proteins involved in S-layer glycosylation. The genetic coordinates Chr: 1154809–1188524 encompass 29 genes, of which at least 15 are predicted to be involved in cobalamin biosynthesis. Surprisingly, genes involved in cobalamin biosynthesis seem to be subject to only genetic depletion. Many CNVs occurred in regions either spanning one or more IS elements or in those regions flanked by IS elements. Specifically, 12 of the 17 chromosomal IS elements were located within CNV peaks; this association is significantly higher than what would be expected by chance ($p < 1 \times 10^{-5}$, see Methods). Our segment threshold, 5 kb, is approximately twice the size of the largest IS element (~2 kb; Brugger et al., 2002), and therefore the segments are not indicative of the amplification of individual IS elements themselves. This association is consistent with the hypothesis that mobilization of IS elements or recombination between them may lead to CNVs throughout the genome.

Several amplified regions do not contain and are not flanked by any apparent IS elements, but show other encoded evidence of genomic instability. One amplified region from Chr: 1744771–1770599 contains a duplicated and inverted operon. While gene order is conserved, the two operons are degenerate and have only 47% sequence identity. However, this suggests two things: a) that at some point in the past there was a genetic rearrangement, and b) that sequence homology in this region could lead to

further genomic instability through the erroneous use of templates during homologous recombination (Darmon & Leach, 2014). Another frequently amplified region from Chr: 1874018–1881134 contains few genes, with the exception of two rRNA sequences.

Interestingly, directly upstream of this region is a series of 5 short repeats, from 9 to 21 bp of length, with an average distance of 191 bp between them. As mobile IS elements are known to leave short scars in the genome, we suggest that this locus has historically been a target site for IS element integration.

By segmenting DNA signals from the control channels of ChIP microarray experiments, we have shown that DNA can be amplified or depleted in discrete regions of the chromosome of *H. salinarum*. These changes appeared more frequently than predicted in our dataset and seem to be associated with IS elements, as has been shown to be the case for the megaplasmiids.

3.4 Discussion

3.4.1 Genomic instability and biodiversity in *H. salinarum*

Data presented here recapitulate reports of genetic rearrangements in the megaplasmiids of *H. salinarum*. Through the analysis of whole-genome sheared DNA we have also detected CNVs in the major chromosomes of several strains of *H. salinarum* NRC-1 at high frequency, expanding what is known about genomic stability in the Haloarchaea. Pfeiffer et al. (2008) did not report chromosomal CNVs between species R1 and NRC-1

despite the many differences detected between the megaplasmiids. While this would suggest that the chromosome is stable, the presence of 17 IS elements by nature should lead to instability. Indeed, we have found that CNVs detected in the chromosome are significantly associated with IS elements. This is consistent with previous work, as mobile IS elements have been shown to facilitate genomic instability in many archaeal species (Brugger et al., 2002; Martusewitsch et al., 2000). In keeping with the ability of chromosomal IS elements to generate instability, Hawkins et al. (2013) observed the integration of an entire megaplasmiid into the chromosome of *Haloferax volcanii* between two ISH18 loci. Taken together, these data suggest that genomic instability, including insertions, duplications, and deletions, is commonly generated in haloarchaeal chromosomes in addition to megaplasmiids.

All halophilic Archaea tested to date are highly polyploid (Zerulla & Soppa, 2014) and therefore may be especially tolerant of temporary CNVs. Indeed, heterozygosity may be carried in a haloarchaeal population over many generations unless selective pressure is applied to homogenize all genome copies (Lange et al., 2011). While the cobalamin cluster appeared to be a hotspot for deletion events in this study, polyploidy of *H. salinarum* may serve as a buffer such that duplications or deletions in these regions are likely to be partial and do not necessarily extend to all copies of the genome. What then is the relationship between polyploidy, heterozygosity, and gene expression in the Haloarchaea? We have shown that gene amplification due to plasmid

instability causes corresponding changes in gene expression (Fig. 10). This suggests that despite the high frequency of large-scale genomic changes reported for archaeal species with large numbers of IS elements, these organisms do not have mechanisms to titrate gene expression levels. This suggests that heterozygous deletions lead to decreased expression and amplifications lead to increased expression. On a larger scale, the megaplastids and the chromosome have differential ploidy, with the megaplastids typically present in lower copy numbers (Breuert et al., 2006). By extension, it is likely that plasmid genes are typically expressed at lower levels than chromosomal genes. As our transcriptional and ChIP data are derived from independent strains and cultures, future studies should seek to clarify the link between gene expression and CNVs by simultaneous measurement of RNA levels and DNA dosage in a given strain.

Several phenomena could account for the induction of genetic rearrangements, amplifications, and duplications at IS elements. Standard culturing conditions include long-term storage of strains followed by 2–12 weeks on plates and up to 20 generations in liquid media; this may allow for two individual cultures from the same freezer stock to accrue unique CNVs. Indeed, a 124 kb deletion in *Sulfolobus sulfataricus* was observed in just one of two biological replicates from the same plate (Redder & Garrett, 2006). It is also possible that stress induces genome instability: IS element mobility can be induced by long-term incubation of *H. salinarum* plates in the cold (Pfeifer & Blaseio, 1990). However, it is telling that our confirmed duplication event was observed in a histone

overexpression strain in which we were not selecting for homologous recombination events, but merely introducing a plasmid via mevinolin selection. The amplified region did not contain the histone coding gene nor any of the most highly differentially regulated genes as observed in the corresponding histone knock-out mutant (Dulmage et al., 2015). This suggests that selection itself, although it may not directly alter the genome, may result in indirect genetic changes. We also did not observe CNVs in regions of the genome related to the selection itself (e.g., FOA resistance and the *pyrEF* locus, mevinolin resistance and *HMG-CoA reductase*), confirming that the CNVs are not related to the selection pressure itself but are generated by the activity of endogenous IS elements.

The tolerance of genomic instability in the halophiles may reflect the requirement of this phenomenon for the generation of genetic diversity. The Haloarchaea, and *H. salinarum* in particular, are highly effective at DNA damage repair. The Haloarchaea encode homologs from most of the DNA repair pathways found in Bacteria and Eukarya, including base excision repair, nucleotide excision repair, homologous recombination, translesion synthesis, and photoreactivation (Ng et al., 2000). *H. salinarum* is highly resistant to DNA-damage inducing stressors, such as UV and gamma radiation, and is capable of rapidly repairing double-stranded DNA breaks (DSBs) in the absence of light (Kottemann et al., 2005). Gene expression and mutant studies suggest that the predominant method of DSB repair is homologous recombination (McCready et

al., 2005). This suggests that the mutation rate at the nucleotide level is very low in this organism; this has been confirmed in *Haloferax volcanii* (Mackwan et al., 2007). It is the homologous recombination machinery that mediates inversions, transpositions, deletions, and amplifications of DNA at IS elements. Therefore, while an assortment of DNA-repair pathways provides genome stability on the small scale, some of these pathways also predispose *H. salinarum* to gross genomic rearrangement. Homologous recombination also mediates the integration of DNA through horizontal gene transfer (Darmon & Leach, 2014) and the exchange of large fractions of the genome during interspecies mating of halophilic Archaea (Papke et al., 2007; Naor et al., 2012). Therefore, we propose that homologous recombination at the IS elements and other short interspersed regions of homology is the primary method of generating diversity in *H. salinarum* and is likely to extend to all halophilic Archaea.

We have observed frequent CNVs in the chromosome and apparent loss of the megaplasmid pNRC100. Laboratory growth conditions are generally nutrient rich and climate controlled; therefore, it is likely that we are observing events that have no relevance to fitness in the conditions tested and it is unclear whether the CNVs in this study were stably maintained. Further studies should seek to determine the transience of these formations and to relate particular CNVs, such as the amplification of glycosylation enzymes, to phenotypic consequences and selectional pressure.

3.5 Acknowledgments

The authors would like to thank David Corcoran at the Integrative Genomic Analysis Shared Resource in the Duke Center for Genomic and Computational Biology for data preprocessing services.

4. Distinct growth patterns and dynamics of the cell division proteins FtsZ1 and MinD2 in the euryarchaeon *Halobacterium salinarum*

Data presented in Figures 14, 15, 16A, 16B, and 17 were collected in conjunction with Yejin Eun (Ethan Garner Lab, Harvard University). Data presented in Figures 16C, 19, 20, and 21 were collected solely by the author. All figures, analyses, and text within this chapter represent the sole effort of the author.

4.1 Introduction

Faithful division of a mother cell into viable daughter cells is a highly regulated process, requiring that DNA be duplicated and faithfully partitioned, the cell mass be sufficient, the septum be properly placed, and the new cell wall and membrane be synthesized before abscission. The processes involved in cell division are well understood in the Eukaryotes, and to a lesser extent, in the Bacteria. However, cell division is understudied in the Archaea. This is partially due to the only recent development of genetic tools, but can also be attributed to features of archaeal cell biology such as the incompatibility of hyperthermophile culture conditions with standard microscope equipment (Atomi et al., 2012; Horn et al., 1999). In recent years, however, much progress has been made, and we now know that Archaea employ diverse systems of cell division. Species in the phylum Crenarchaeota employ a system homologous to the

ESCRT-III (Endosomal Sorting Complex Required for Transport) machinery of the Eukaryotes, which is responsible for membrane scission at the bud neck in mammals and for formation of membrane vesiculation (Bhutta et al., 2014; Ettema & Bernander, 2009). While lacking ESCRT-III homologs, all sequenced members of the phylum Euryarchaeota encode homologs of the bacterial protein FtsZ, which is required for constriction at the cell division plane (Makarova et al., 2010). While other bacterial cell division factors remain either unidentified or unstudied in the Euryarchaea, FtsZ homologs in several species have been characterized as having functional properties of the bacterial protein, and thus we look to the Bacteria for potential models of euryarchaeal cell division.

Broadly speaking, cell division in Bacteria is driven by a bacterial tubulin homolog called FtsZ. A self-activating GTPase, FtsZ oligomerizes to form filaments (de Boer et al., 1992). Together, many filaments can bundle at mid-cell into a structure called the Z ring (Erickson et al., 2010). It is this ring-shaped structure that serves as a scaffold and initiating factor for septum formation and cell division (Erickson et al., 2010). In *Escherichia coli*, the Min system (named for the miniature size of a series of *E. coli* mutants; Adler et al., 1967) is responsible for the proper localization of the Z ring (Shih & Zheng, 2013). The AAA ATPase MinD, in its ATP-bound form, localizes to the poles of the cell as a dimer (Rowlett & Margolin, 2015). Upon binding by MinE, MinD's ATPase domain is activated and ADP-bound MinD monomerizes and is released into

the cytoplasm (Raskin & de Boer, 1999 [1]). Cycling of these events leads to pole-to-pole oscillation of these molecules in the cell (Raskin & de Boer, 1999 [1]). Coupled to MinD is MinC, which inhibits polymerization of FtsZ (Raskin & de Boer, 1999 [2]). Time-averaged localization of MinCD leads to a zone of depletion in the center of the cell; this is where the Z ring will form (Treuner-Lange & Sogaard-Andersen, 2014). While oscillation of MinCD is characteristic of *E. coli*, it is not observed in *Bacillus subtilis*. MinD in *B. subtilis* is targeted to the poles by MinJ, which, in turn, is recruited by DivIVA, a sensor for negative curvature (Rowlett & Margolin, 2015). This sets up a gradient of MinD and MinC in which protein concentrations are higher at the poles and lower at the mid-cell, inhibiting Z ring formation at the poles and allowing formation at the mid-cell (Treuner-Lange & Sogaard-Andersen, 2014).

The literature on cell division in the Euryarchaea is limited, but a handful of studies have investigated FtsZ function in the Halobacteria. Because they are mesophiles, these organisms can be imaged at their optimal growth temperature without special microscope adaptations. Many species are also genetically tractable, so fluorescently tagged proteins can be introduced either into the genome at the native locus or can be expressed from a stable plasmid (Atomi et al., 2012; Peck et al., 2002; Reuter & Maupin-Furlow, 2004). The species *Halobacterium salinarum* has a sequenced and well-annotated genome, and a wealth of transcriptomics and proteomics experiments make this an attractive model system in which to study cell division

(Dassarma et al., 2006). In addition, it is one of the few members of the Halobacteria that is described as a pure rod-shaped culture and therefore represents a convenient model for comparison of cell division with rod-shaped bacteria such as *E. coli* or *B. subtilis*.

The Haloarchaea encode homologs of FtsZ and MinD (Makarova et al. 2010), and thus can be used as model organisms for the study of bacterial cell division protein function in the Euryarchaea. In fact, several studies have shown that FtsZ proteins in the Haloarchaea behave much like their bacterial counterparts. Immunofluorescence staining against FtsZ1 proteins from *Haloferax volcanii* showed that this protein localizes to the cell division plane in constricted cells; in some of these cells Z rings could be observed (Wang et al., 1996; Poplawski et al., 2000). *In vitro* experiments with purified *H. volcanii* FtsZ1 confirmed that this protein has GTPase activity that is concentration dependent (Wang & Lutkenhaus, 1996). Halophiles have an expanded set of FtsZ proteins: for example, *H. volcanii* encodes 8 homologs. This is unusual, as most bacteria encode only a single FtsZ protein and plastid genomes in the Eukaryotes only encode two (Vitha et al., 2001). Recently it was suggested that haloarchaeal FtsZ proteins group into two families, one that resembles FtsZ in function, and another that possesses functional characteristics resembling both FtsZ and tubulin (Duggin et al., 2014). The crystal structure of *H. volcanii* FtsZ3, a member of the latter group, shows that this protein is capable of filament formation and GTP binding (Duggin et al., 2014). Investigations into the functions of two of the six atypical *H. volcanii* FtsZs suggested

that they are structural proteins, possibly acting as eukaryotic tubulins, and are not involved in cell division (Duggin et al., 2015). Previous studies in *H. salinarum* show that overexpression of one of the conserved FtsZ homologs leads to pleiomorphism of this normally rod-shaped organism, suggesting some role in cell division and/or shape regulation (Margolin et al., 1996). While FtsZ has been studied in the Halobacteria, other factors involved in cell division, such as the negative regulator MinD, have not yet been investigated.

Here we expand the studies of cell division in the halophiles through sequence analysis and time-lapse microscopy of fluorescently labeled cell division proteins in *H. salinarum*. We find that *H. salinarum* is able to undergo rapid growth both as rods and as flat pleiomorphic cells in microcolonies. FtsZ1 forms a Z ring, but rapid dynamics and non-specific localization of puncta outside the Z ring or the midline distinguish it from the bacterial system. Finally, it is apparent that MinD2, although involved in cell division, forms discrete puncta rather than gradients and localizes to the plane of division before invagination of the cell wall is observed. Via observation of cell-cycle dynamics, we conclude the MinD2 is necessary but not sufficient for initiation of cell division. While these results confirm that FtsZ and MinD dynamics are coordinated with cell division, several features are distinct from the bacterial model, opening up new realms of inquiry into the process of euryarchaeal cell division.

4.2 Materials and Methods

4.2.1 Comparison of amino acid sequences

FtsZ, MinD, and ParA sequences were retrieved from NCBI (<http://www.ncbi.nlm.nih.gov/>). Alignments were performed using MegAlign v.11.0.0 (DNASar) to implement the Clustal W method. Phenograms were also generated using MegAlign. FtsZ protein alignment and discussion may be found in Appendix D.

4.2.2 Growth conditions

Cultures were routinely incubated at 42°C with agitation in complex medium (CM; 250 g L⁻¹ NaCl, 20 g L⁻¹ MgSO₄•7H₂O, 3 g L⁻¹ sodium citrate, 2 g L⁻¹ KCl, 10 g L⁻¹ peptone). For strains generated in the *Δura3* background, uracil was added to a final concentration of 50 μg mL⁻¹. For strains containing the overexpression vector pMTFCHA and its derivatives (Tables 9 and 10), mevinolin was added to a final concentration of 1 μg mL⁻¹ to maintain the plasmid. Agarose pads for time-lapse were composed of 80% Complete Defined Medium with uracil (CDM; see Appendix A or Schmid et al., 2009). For all time-lapse experiments, *H. salinarum* strains were subcultured from stationary, grown to log phase, spun at 10k rpm for 5 min, and resuspended to high cell density in CDM with uracil or mevinolin as necessary.

4.2.3 Strain construction

All strains described were originally derived from *Halobacterium salinarum* NRC-1 (ATCC700922). Plasmids and strains used in this study are listed in Tables 9 and 10. smVenus fusion protein plasmids for integration into the genome were transformed and selected in the $\Delta ura3$ uracil auxotroph parent strain as described in Peck et al. (2002), and the fusion genotype was confirmed with PCR and Sanger sequencing. Plasmid constructs were confirmed by Sanger sequencing and transformed into *H. salinarum* as described in Peck et al. (2002) without the secondary selection. Plasmid-bearing strains were screened using PCR. Primers used for strain construction are described in detail in Appendix B. For description of the smVenus and mVenus fluorophores, details can be found in Appendix C. All smVenus or mVenus strains were engineered with a 10 amino acid linker sequence (GGGSGGGSGGGS) between the protein of interest and the fluorophore. Plasmids pMTF-smVenus, pEPW01, pEPW02, and pEPW03 were gifts from Eli Wilber (Dan Keihart lab, Duke University). Plasmids pKAD07, pKAD08, and pKAD09 were designed for integration of smVenus into native loci and were constructed using Gibson assembly (Gibson et al., 2009).

4.2.4 Light microscopy and time-lapse imaging

Cells were imaged using one of the following systems: a) a Nikon Eclipse TE2000-U microscope connected to a Hamamatsu digital camera model ORCA-ER with X-cite 120

illumination, b) an API DeltaVision system with an Olympus IX-71 base connected to a Coolsnap HQ2 CCD Camera with 488 nm 50 mW laser excitation, or c) a Zeiss Axio Observer A1 inverted microscope connected to a Hamamatsu digital camera model ORCA-ER with HBO arc lamp excitation. All cells were imaged under 100X magnification (1000X total) with either phase contrast or DIC. Cell counting and morphological quantitation were performed with ImageJ and R statistical software. Ellipses were fit to each cell and measurements of the area, length, and width were collected. All fluorescence images were background subtracted prior to analysis. Fluorescence localization versus cell length was determined by drawing a line down the length of the cell and measuring intensity along this line. Using the R function `norm_ave_plot3` (<https://github.com/KeelyDulmage>), values were then normalized to maximum intensity and cell length to facilitate comparison between cells. Agarose pads on slides were prepared essentially as described in the HaloHandbook v.7.2 (Dyall-Smith, 2009). Nutrient-rich pads were made with Complete Defined Medium with uracil (CDM; see Appendix A or Schmid et al., 2009) to reduce autofluorescence. Because agarose has low solubility in high salt media, a 1% agarose mixture was prepared in 80% CDM with 20% water for a final concentration of 3.4 M NaCl, still within the optimal salinity range for *H. salinarum* as determined by previous experiments (data not shown). For time-lapse imaging in dishes, 35 mm microwell dishes with No. 1.5 coverglass from MatTek (Ashland, MA) were used. 5 μ l of cells was

pipetted on top of the coverslip and a small pad of agar (approximately $1 \times 1 \times 0.4$ cm) from a CDM+uracil plate was gently placed on top of the culture. The pad of agar was then sealed with molten 1% agarose (80% CDM, 20% H₂O). After the agarose was allowed to cool, excess moisture was removed and the sides were sealed with Parafilm to prevent desiccation. Standard parameters for time-lapse imaging are as follows: 20–30 minute intervals, at least 8 stage points, and 37°C incubation for at least 18 hours. Focus was maintained over this interval with the automated Ultimate Focus system.

Table 9: Plasmids used in this study

Plasmid	Description	Source
pJAM1044	<i>smRS-GFP-ssr1</i> , for expression in <i>Haloferax volcanii</i>	Reuter & Maupin-Furlow, 2004
pMTFCHA	Vector for C-terminal HA tagging and overexpression in <i>Halobacterium salinarum</i> , P _{fdx} , Mev ^R Ura ⁺	Wilbanks et al., 2012
pMTF1044GFP-CHA	pMTFCHA:: <i>smRS-GFP-ssr1</i> , stop codon before the HA sequence	Eli Wilber
pMTF-smVenus	pMTFCHA:: <i>smVenus</i> , P _{fdx} , Mev ^R Ura ⁺	Eli Wilber
pMTF-mVenus	pMTFCHA:: <i>monomeric smVenus (A206K)</i>	This study
pEPW01	pMTFCHA:: <i>ftsZ1-smVenus</i>	Eli Wilber
pEPW02	pMTFCHA:: <i>ftsZ2-smVenus</i>	Eli Wilber
pEPW03	pMTFCHA:: <i>smVenus-minD2</i>	Eli Wilber
pNBK07	Vector for markerless integration of recombinant DNA, Mev ^R Ura ⁺	Nitin Baliga, Institute for Systems Biology, Seattle, WA
pKAD07	pNBK07:: <i>ftsZ1-smVenus</i> plus flanking DNA	This study
pKAD08	pNBK07:: <i>smVenus-minD2</i> plus flanking DNA	This study
pKAD09	pNBK07:: <i>hpyA-smVenus</i> plus flanking DNA	This study

Table 10: Strains used in this study

Strain	Genotype	Description	Source
<i>NRC-1</i>		Wild typed	
<i>MDK407</i>	Δ <i>ura3</i>	Parent strain, Ura ⁻	Peck et al., 2000
<i>KAD104</i>	<i>hpyA-smVenus</i>	C-terminally tagged HpyA, genomic integrant, Ura ⁻	This study
<i>KAD105</i>	<i>ftsZ1-smVenus</i>	C-terminally tagged FtsZ1, genomic integrant, Ura ⁻	This study

KAD106	<i>NRC-1/ftsZ1-smVenus</i>	C-terminally tagged FtsZ1, overexpression vector, Ura ⁻ Mev ^R	This study
KAD107	<i>NRC-1/ftsZ2-mVenus</i>	C-terminally tagged FtsZ2, overexpression vector, Ura ⁻ Mev ^R	This study
KAD108	<i>smVenus-minD2</i>	C-terminally tagged MinD2, genomic integrant, Ura ⁻	This study
KAD109	<i>NRC-1/minD2-smVenus</i>	N-terminally tagged minD2, overexpression vector, Ura ⁻ Mev ^R	This study

4.3 Results

4.3.1 Environment-dependent growth modes of *H. salinarum*

While aliquots of *H. salinarum* cultures have previously been imaged over the growth curve (Herrmann & Soppa, 2002), individual cells have not yet been imaged by time-lapse microscopy, presenting a hindrance to the study of cell division and molecule localization dynamics. Therefore, a protocol to image actively dividing *H. salinarum* cells by light- and fluorescence-microscopy was developed in which cells were incubated in glass-bottom culture dishes with agarose pads at 37°C (see Methods).

Surprisingly, *H. salinarum* is able to undergo two different modes of growth. In liquid culture, *H. salinarum* cells most commonly appear as rods, much like bacterial *Bacillus* species. When cells are allowed to grow without vertical restriction, individual cells will elongate and frequently are seen to divide at the midline. Cell wall constriction is

observed prior to separation of the two daughter cells (Fig. 14A). Surprisingly, amorphous cells can recover the rod shape over growth, yielding rod-shaped daughter cells (Fig. 14A, white arrows). Asymmetric division can also be observed in a subset of cells: in one culture asymmetric division was observed at a rate of 24%. These cells undergo a similar process of constriction and fission (data not shown). However, we found that when cells were restricted vertically (in this case, by agarose pad and coverslip), cells grew outward and lost their stereotypical rod shape (Fig. 14B). In addition to the irregularity of the edge of the flattened cells, small processes extended from the peripheries, a feature that was not observed in rod-shaped cells. Despite the vertical pressure of this environment, and subsequent shape changes, cells continued to grow in this state; lysis was not observed over the course of this experiment (20 hours; Fig. 14B). When pressure was relieved by deformation of the agar, these flattened shapes revealed themselves to be microcolonies of many individual cells who, upon release into a free-swimming environment, immediately began to regain the stereotypical rod shape of this species (Fig. 14B, last panel). The viability of these flattened cell shapes and the ease of their return to rod form suggests that *H. salinarum* readily undergoes either type of growth, depending on signals from the environment.

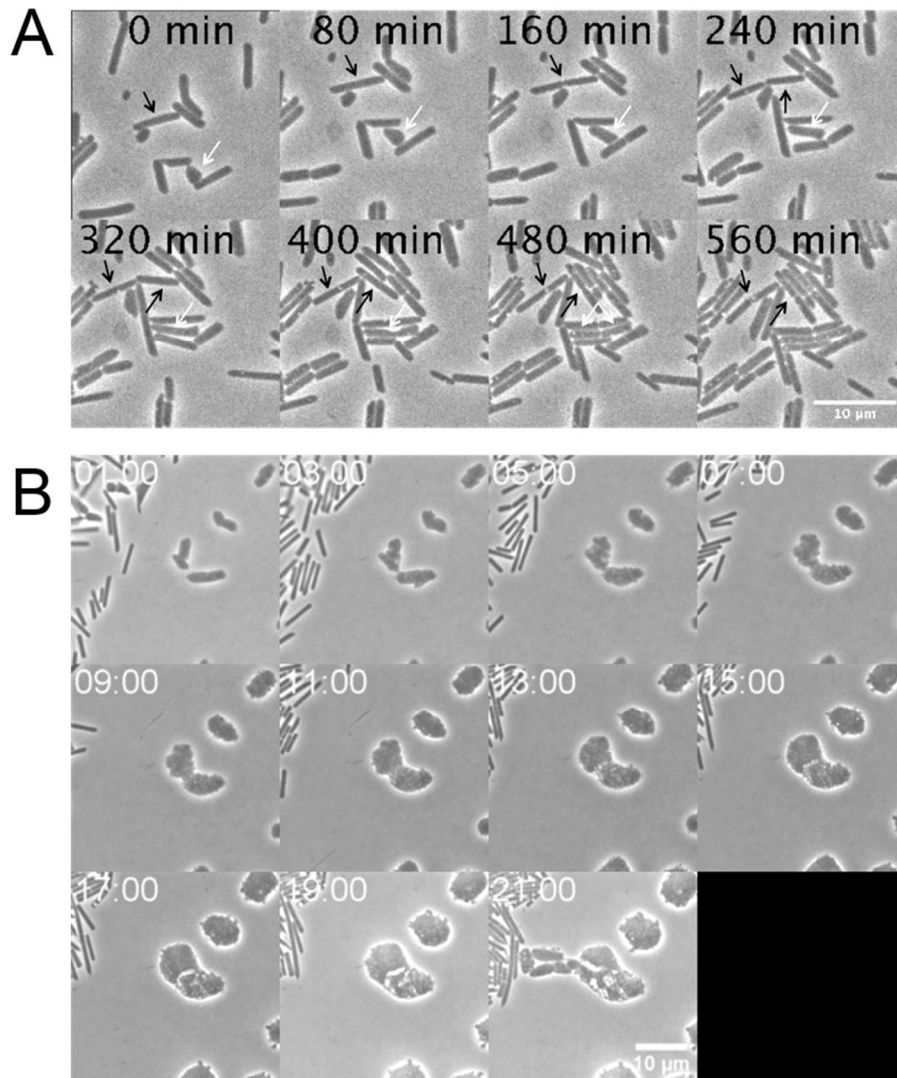


Figure 14: Time-lapse of *Halobacterium salinarum* cell division. A) Example of *H. salinarum* cells undergoing stereotypical growth as rods. Black arrows follow individual rod-shaped cells undergoing division. White arrows indicate an amorphous cell that generates rod-shaped daughter cells. Representative strain shown here is *KAD108* (smVenus-MinD). B) Alternate flat growth mode of *Halobacterium salinarum*. Two isogenic cell populations are shown: free-swimming rods (upper left of images) and sessile cells caught between agarose pad and coverslip (right side of images). When vertically restricted, cells immediately deform from rod shape. Upon release from pressure (final timepoint) cells gradually regain their rod shape. Time elapsed is shown in hours and minutes. Representative strain shown here is *KAD104* (hpyA-c-smVenus).

4.3.2 FtsZ1 forms cell division-dependent Z rings at mid-cell

To investigate whether and/or how the halobacterial homologs of FtsZ are involved in cell division, Ftsz1 was C-terminally tagged with the yellow fluorescent protein smVenus and expressed a) at the native genomic locus, hereafter referred to as the integrant strain *ftsZ1-smVenus*; or b) under control of the strong constitutive promoter P_{fdx} on a plasmid capable of replication in *H. salinarum* (Wilbanks et al., 2012). The latter strain retains the wild-type native copy in addition to the plasmid-born *ftsZ1-smVenus* and is hereafter referred to as the overexpression strain *NRC-1/fts1z-smVenus*. Plasmid and strain details are listed in Table 9 and Table 10.

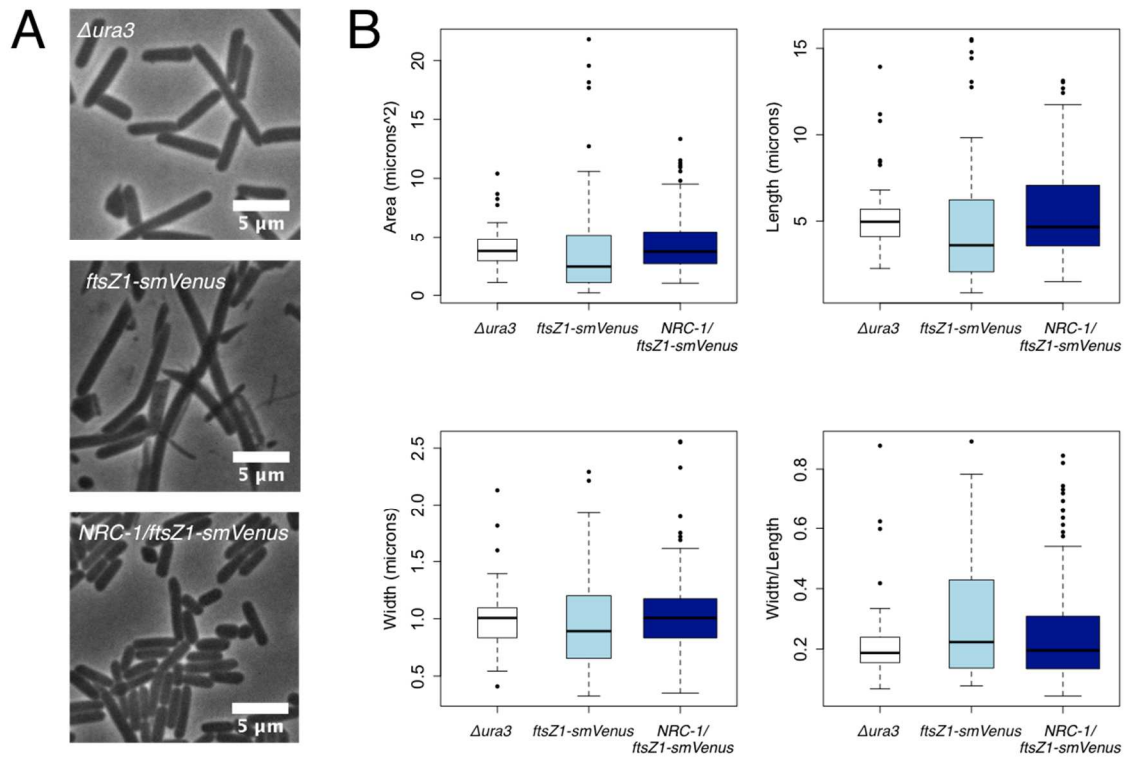


Figure 15: Cell shape comparison between wild-type *ura3* and smVenus-tagged *ftsZ1* strains. A) Phase contrast images of wild-type (top), integrated *ftsZ1-smVenus* (center), and the *ftsZ1-smVenus* overexpression strain (bottom). B) Image J was used to quantitate several parameters of cell shape: area (length × width), length, width, and width/length ratio. C-terminal tagging of FtsZ1 at the native locus disrupts normal cell division, as evidenced by the decreased size of cells with the integrated tag and the increased variability in length and width (light blue boxes). The overexpression strain, *NRC-1/ftsZ1-smVenus*, retains the wild-type copy of *ftsZ1* and therefore the phenotype is not as severe (dark blue boxes). Width of boxplots is proportional to the number of cells quantitated.

To control for cell morphology and size artifacts of non-native expression of cell division proteins, morphological parameters of area, length, width, and roundness were quantitated for smVenus-tagged FtsZ1 mutants and compared with wild-type (Fig. 15).

Morphological variability suggests that the integrated smVenus tag disrupts the native function of the FtsZ1 protein. On average, cells of this strain are smaller than wild-type, as both length and width are decreased (Fig. 15B). While the length and width differences are not statistically significant, *ftsZ1-smVenus* cells exhibit significantly higher variability in both area (Bartlett's test; $p < 1E-10$) and width/length ratio (Bartlett's test; $p < 0.01$) relative to the wild-type. This suggests that FtsZ1 is involved in the determination of both shape and size. The overexpression strain *NRC-1/ftsZ1-smVenus*, which retains a wild-type copy of *ftsZ1* in addition to the plasmid-born *ftsZ1-smVenus*, appears to be wild-type in shape, even though *ftsZ1-smVenus* is under control of the strong promoter P_{fdx} . Therefore, although both strains are used for all further experiments, we assume that FtsZ1 localization and dynamics in the overexpression strain is more representative of wild-type protein behavior.

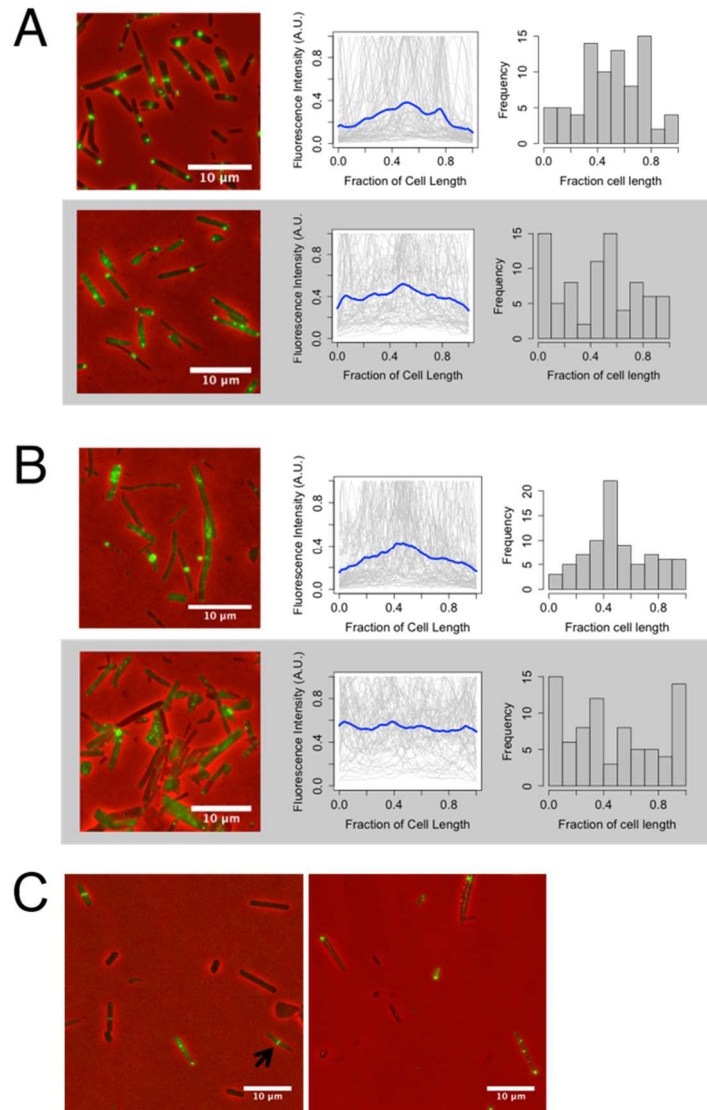


Figure 16: Average localization of FtsZ1 particles over a population of log phase cells (white background) or stationary phase cells (gray background). A) Fluorescence intensity versus length was quantitated for *NRC-1/ftsZ1-smVenus* or B) the *ftsZ1-smVenus* integrant strain. Left: YFP fluorescence overlay onto phase contrast of log phase cultures. Center: Gray lines represent traces for normalized fluorescence intensity versus length in individual cells; the blue line represents average normalized intensity for all cells. Right: Histogram depicting the lengthwise frequency of the location of maximum cellular fluorescence intensity. N for all analyses = 80. C) *FtsZ2* overexpression localization in logarithmic phase (left) and stationary phase (right). Black arrow indicates a cell undergoing septation and division.

Still images of dividing cells demonstrate that although FtsZ1 localizes to the septum it may be observed anywhere in the cell, including the poles, in direct contrast to bacterial FtsZ proteins. A complicating factor in this analysis is that *H. salinarum*, unlike many bacteria, can undergo both symmetric and asymmetric division, which necessitates that if FtsZ1 is involved in division, the Z ring will not always be localized to the midline. Therefore, we reasoned that Z rings can form anywhere in the cell that division can be initiated. To test this, we conducted in-depth analysis of fluorescence signals of individual cells from single snapshots of dividing cultures (Fig. 16A & 16B, center; see Methods). Surprisingly, FtsZ1 does not appear to be excluded from any region along the length of the cell, even though cells in the log phase culture are undergoing division. In contrast, the brightest foci in each cell, putatively representing the bulk of polymerized FtsZ1 protein, have a tendency to be localized to the midline (Fig. 16A & 16B, right), but this is clearly not the case for all cells. This tendency to localize to the midline is characteristic of log phase cells only: upon movement into stationary phase, FtsZ1 appears to be evenly distributed with occasional polar localization (Fig. 16A & 16B, gray boxes). This suggests that FtsZ1 forms Z rings specifically in logarithmic phase cells and Z ring formation is related to cell division. Preliminary fluorescence microscopy imaging of a *FtsZ2-smVenus* overexpression strain suggests a shared function between this protein and FtsZ1. During log phase, FtsZ2 bands can be observed to localize to the plane of division, while localization becomes

less specific in stationary phase (Fig. 16C). Because of the similar localization pattern and conservation of amino acid sequences, it is likely that FtsZ1 and FtsZ2 copolymerize to form the functional Z ring. These fluorescent imaging data suggest that FtsZ1 and FtsZ2 are indeed components of the *H. salinarum* cell division apparatus.

In order to verify that these structures are both dynamic and cell-cycle dependent, *ftsZ1-smVenus* strains were imaged by time-lapse microscopy. Time-lapse of rapidly dividing cells confirms that FtsZ1-dependent Z rings participate in cell division (Fig. 17). Although FtsZ1-smVenus integrant cells are not uniformly wild-type in shape, FtsZ1 can be observed to localize to the septum prior to cell division, as would be expected for a functional FtsZ protein. The Z ring appears to be quite wide, suggesting that FtsZ1 forms bundles *via* lateral interactions and does not only exist as single, independent filaments. Immediately after cell division, the FtsZ1 bundle splits and moves to the center of each daughter cell, where it remains until the next round of division (Fig. 17, double-arrows). Interestingly, this does not appear to be a matter of disassembly and reassembly, but a repositioning of the FtsZ1 bundle itself. The coupling of FtsZ1 dynamics to cell division, along with the apparent Z ring, confirms that FtsZ1 is indeed involved in cell division as in the bacteria, despite the novel observation that FtsZ1 may polymerize throughout the cell. This feature is, to the best of our knowledge, unique to *H. salinarum*, but may extend to other halophilic Archaea.

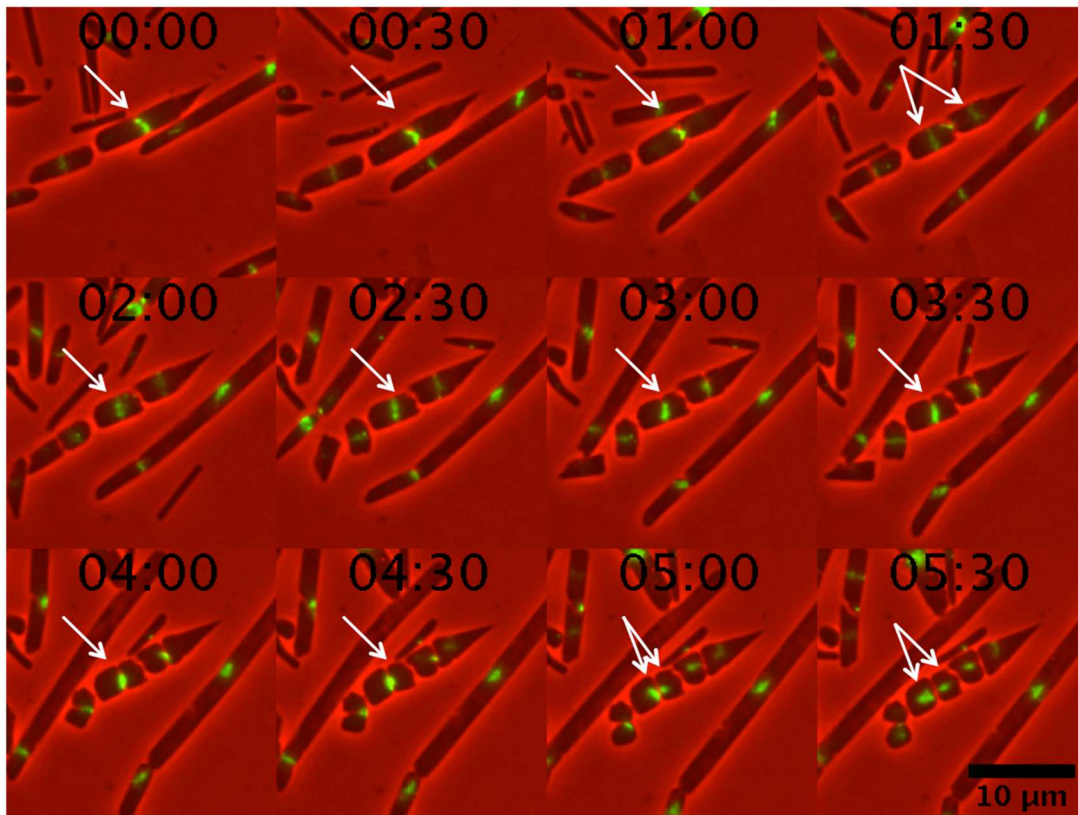


Figure 17: FtsZ1 localizes to the septum prior to constriction of the cell (single white arrow). Within 1 hour of cytokinesis, FtsZ1 bundles relocate to the center of each daughter cell (double white arrows). Strain shown above is *ftsZ1-smVenus*.

4.3.3 *H. salinarum* encodes a diverse and expanded MinD/ParA family

In Bacteria, FtsZ placement is determined by the action of the MinCDE system. *H. salinarum* encodes genes for two MinD homologs despite the fact that the genome lacks MinC and MinE homologs (Ng et al., 2000; Makarova et al., 2010). In addition to MinD1 and MinD2, the *H. salinarum* genome encodes genes for 7 other proteins in the MinD/ParA AAA ATPase family (Ng et al., 2000; Table 11). Several of these genes are

found in repeat regions and thus are present in more than one copy. For comparison with bacterial cell division proteins, BLAST searches were performed against the amino acid sequences for *Bacillus cereus* proteins, as bacterial homologs encoded in the genome of *H. salinarum* have greatest homology to the corresponding genes in *Bacillus* species (Ng et al., 2000). As expected, the annotated *H. salinarum* sp. NRC-1 MinD homologs, MinD1 and MinD2, have the highest homology to *B. cereus* MinD as detected by protein BLAST. Mrp protein (Multiple Resistance and pH-adapted) was also identified along with the ParA and MinD proteins and shares significant amino acid identity.

Table 11: *H. salinarum* AAA ATPase protein family homology to MinD from *Bacillus cereus*

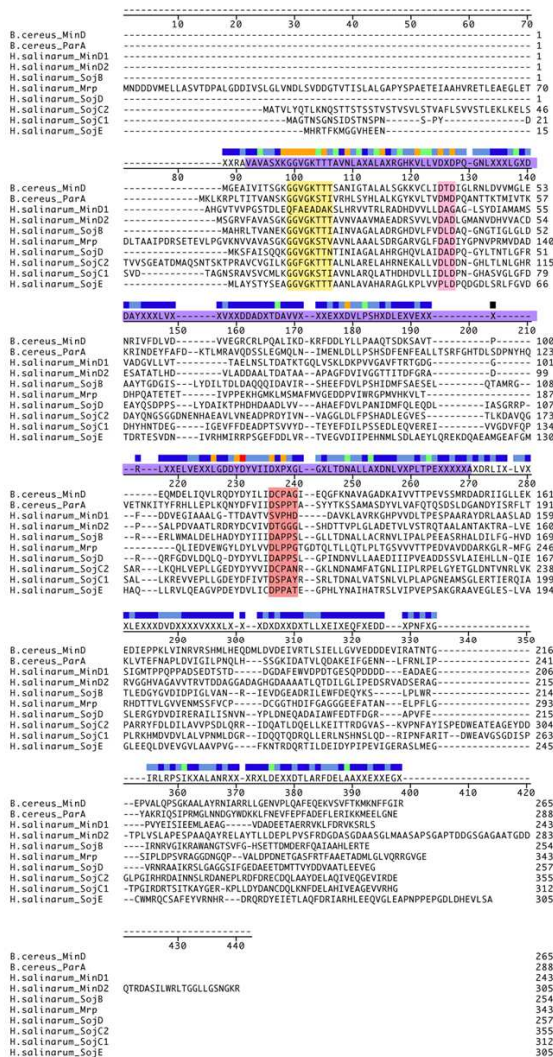
Gene Name	Gene Alias	% Identity	E-value	Gene copy #
VNG0946G	<i>minD1</i>	33	9e-23	1
VNG2130G	<i>minD2</i>	28	1e-21	1
VNG6034G	<i>sojB</i>	27	2e-12	2
VNG1131G	<i>mrp</i>	24	4e-09	1
VNG6123G	<i>sojD</i>	30	4e-07	2
VNG6153G	<i>sojC2</i>	26	1e-06	1
VNG6086G	<i>sojC1</i>	25	7e-05	3
VNG6337G	<i>sojE</i>	26	1e-04	1
VNG6010G	<i>sojA</i>	18	6e-04	2

Amino acid sequences from *H. salinarum* and bacterial AAA ATPases were aligned to examine a) conservation of essential residues, b) relatedness between proteins, and c) whether the annotated MinD homologs are distinct from the ParA family. Immediately apparent from the alignment is that halobacterial proteins in the ParA/MinD family are diverse and there is very little conservation outside of the AAA

ATPase domain (Fig. 18A, consensus sequence highlighted in purple). The most highly conserved region is the P-loop domain (Fig. 18A, highlighted in yellow). This is required for ATP binding and thus is the hallmark of a functional AAA ATPase. Surprisingly, MinD1 does not show conservation of these residues, suggesting that MinD1 cannot bind ATP. MinD1 is also divergent in the Switch 2 region which, in combination with the Switch 1 region, represents the portions of the protein that change conformation in response to the binding of ATP. These results call into question the functionality of MinD1 as an ATPase and suggest that the other AAA ATPase family members may be functional despite high sequence variability in their other protein domains.

The phylogram generated from this alignment confirms the annotations of *H. salinarum* AAA ATPase proteins as two distinct classes: one containing the two MinD homologs and another containing 13 ParA homologs, also known as Soj proteins (Fig. 18B). SojB, -C, -D, and -E share a branch with ParA from *Bacillus cereus*. Conservation of SojA was too poor to generate an alignment (see Table 11). As expected, the two MinD homologs cluster with *B. cereus* MinD, rather than ParA, but they appear to be highly divergent from the bacterial protein, sharing more similarity with *H. salinarum* Mrp. Overall, there is little conservation in this protein family outside the AAA ATPase domain, and MinD2 is the strongest candidate for a functional MinD homolog in this organism.

A



B

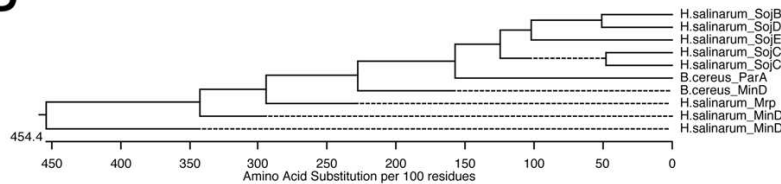


Figure 18: Alignment of MinD/ParA superfamily proteins from *H. salinarum* and *B. cereus*. A) Protein alignment. Consensus sequence highlighted in purple represents the AAA ATPase domain. Yellow: P-loop domain involved in ATP binding; pink: switch 1 domain; orange: switch 2 domain. B) Phenogram of MinD and Para amino acid conservation.

4.3.4 MinD2 localizes to the plane of division prior to membrane constriction and remains at the new poles

Based on the protein alignment and amino acid conservation, we reasoned that MinD2 is the most likely functional homolog to the bacterial MinD cell division protein. Therefore, we generated a strain in which the chromosomal copy of MinD2 was tagged at its N-terminus with smVenus. Fluorescence imaging of smVenus-MinD2 shows that this protein forms discrete puncta at the poles as well as at the plane of division (Fig. 19A). This pattern of localization is retained over the entirety of the cell cycle and is observed in stationary phase cells (Fig. 19B). By plotting the lengthwise distribution of fluorescence over many cells, we observed that the localization of MinD is tightly restricted to the poles (Fig. 19D). Even when cells have more than 3 puncta, localization still tends to be either in the center of the cell or at the poles. By correlating the number of puncta per cell with various features of cell shape (Fig. 19C), we find that the primary determinant of the number of distinct MinD2 puncta per cell is length ($R = 0.73$).

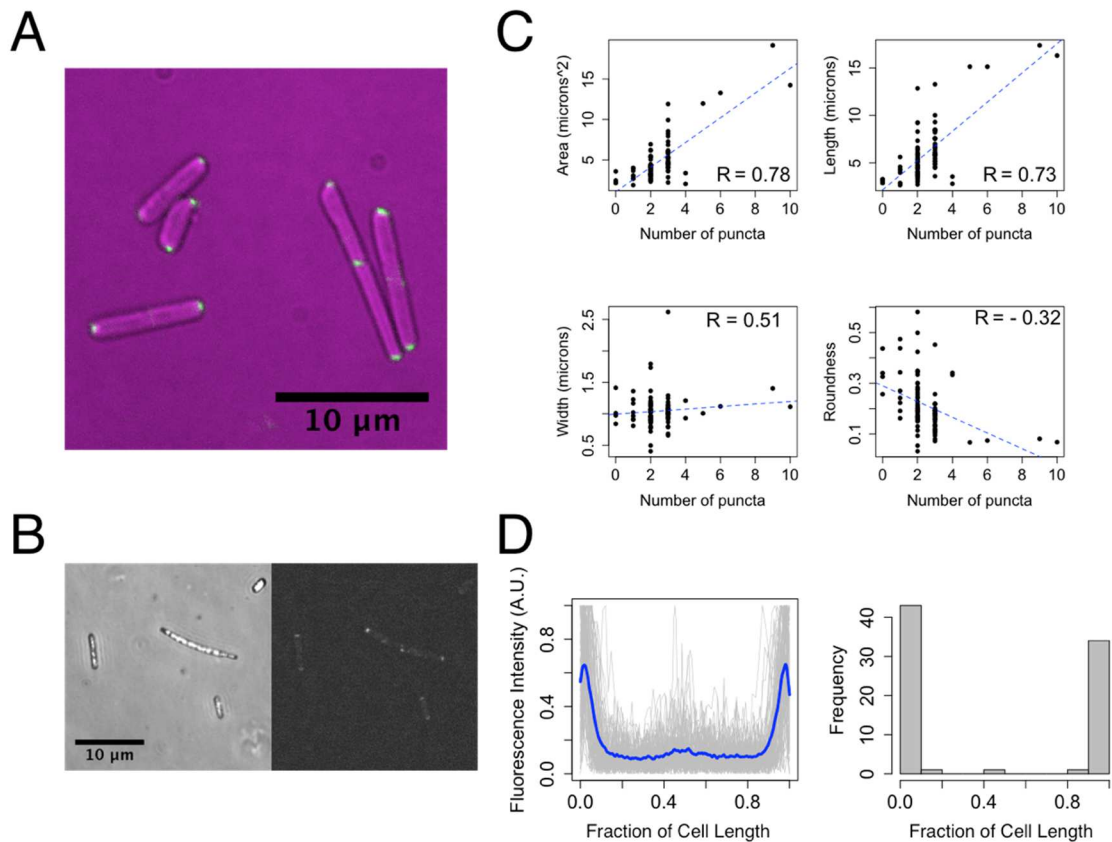


Figure 19: *H. salinarum* MinD2 localizes to the poles and mid-cell. A) Fluorescence microscopy of smVenus-MinD2 in actively dividing cells. B) MinD2 remains at the poles of dormant, stationary phase cells. C) Number of MinD2 puncta per cell is proportional to cell length and area (N = 91). D) MinD localizes to poles and midcell in actively dividing cultures. Shown here are fluorescent traces for individual cells, normalized to cell length (left) and the intracellular location of the highest intensity spots (right).

Comparison of averaged fluorescence localization in *smVenus-minD2* versus *NRC-1/smVenus-minD2* shows that the two strains exhibit similar localization patterns (Fig. 20A). While smVenus-MinD2 puncta are much brighter in the overexpression strain, the number of puncta per cell is also positively correlated with length (data not

shown). This suggests that, although MinD2 does not form a lengthwise gradient as observed in bacteria, localization of MinD2 is closely related to length.

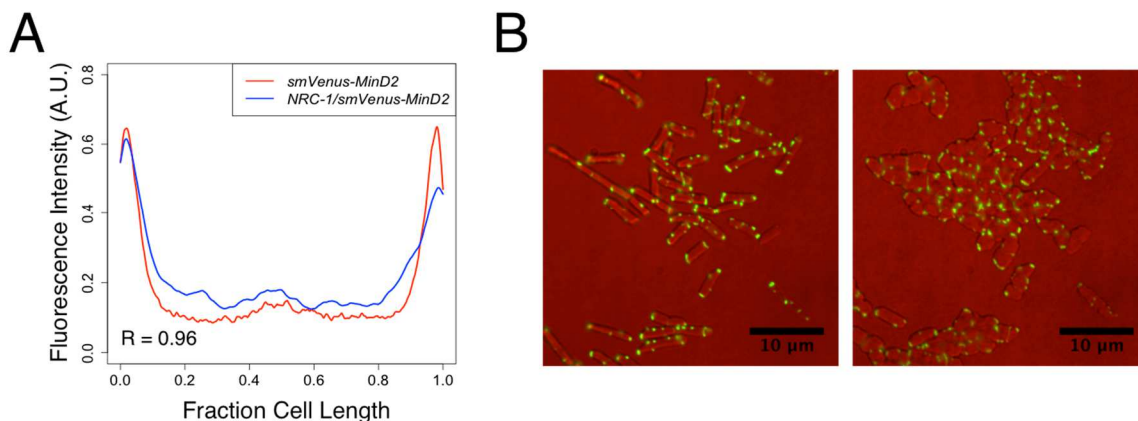


Figure 20: *smVenus-MinD2* localizes to the cell periphery A) Cells from the integrant (*smVenus-minD2*) and overexpression (*NRC-1/smVenus-minD2*) strains show nearly identical patterns of localization as seen in an overlay of average normalized fluorescence intensity (Pearson correlation; $R = 0.96$). $N = 80$. B) Fluorescent imaging of *NRC-1/smVenus-minD2* before (left) and after (right) flattened growth shows that localization of the puncta is restricted to the periphery.

In bacteria, MinD is recruited to the periphery of the cell. The localization of MinD2 to the polar regions of the cell as well as the site of septum formation in *H. salinarum* suggests that this protein could sense or be recruited to regions of concave local curvature. While this is difficult to determine in normal cells with oval or circular cross-sections, cells from *H. salinarum* can be induced to grow into flattened irregular shapes. This leads to a decrease in curvature at the center of the cell to and an increase in curvature at the edges. Log-phase cultures of the strain *NRC-1/smVenus-minD2*, which produces brighter puncta than *smVenus-minD2*, were incubated on an agarose

pad to induce growth as flattened cells. After several hours of incubation in these conditions, cells became devoid of any visible internal puncta (Fig. 20B). This suggests that MinD2 is indeed associated to the periphery of the cell and that this localization may be induced by regions of high local curvature, just as in Bacteria (Ramamurthi & Losick, 2009).

Time-lapse imaging of *smVenus-MinD2* shows that MinD2 localization is dependent on the cell cycle. MinD2 localizes to the plane of division and will remain at this site in the daughter cells as a polar body. In contrast to Bacteria, we find that MinD2 moves to the site of division prior to membrane constriction and septum formation, suggesting that MinD2 itself may be involved in the initiation of division (Fig. 21A). Visualization of asymmetrically dividing cells shows that MinD2 puncta are still observable throughout the cell, including at the midline and asymmetrically-placed septa (Fig. 21B). Therefore there must be additional MinD2-independent criteria for choosing the plane of division. Cells were also imaged at several different intervals to determine whether MinD2 oscillates from pole-to-pole as seen in *E. coli*. Overexpression cells were imaged at 3 second, 30 second, and 30 minute intervals. Cells with the integrated *smVenus* tag were imaged at 5 second, 30 second, 4 minute, and 20 minute intervals (data not shown). Oscillation of MinD2 puncta was not observed at any time interval in either strain. Therefore, although MinD2 has cell-division dependent localization, this protein is not positioned by any known model of oscillation or

diffusion. Taken together, these results suggest that MinD2 functions in cell division, marking the poles, mid-line, and areas of local curvature of each cell.

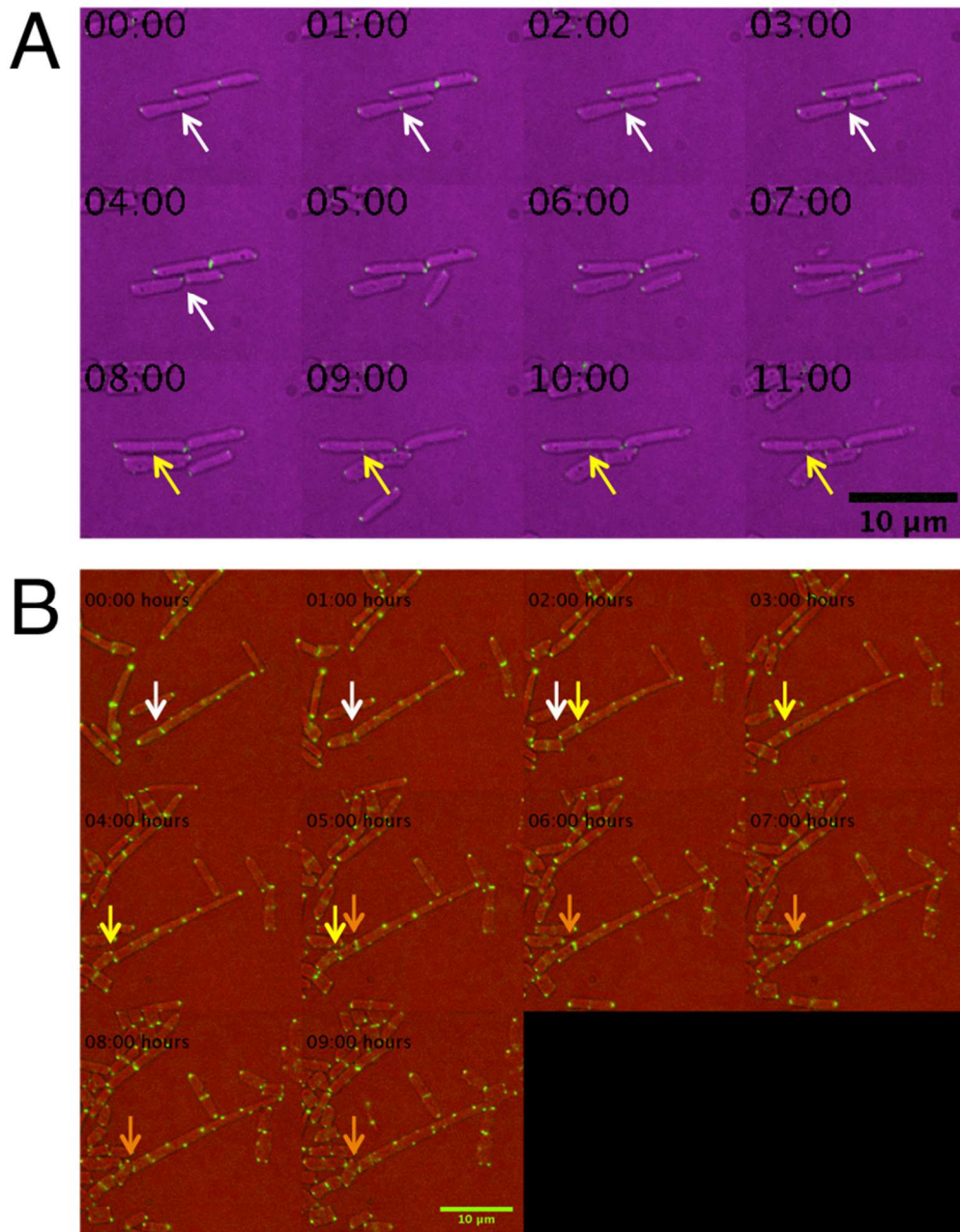


Figure 21: MinD2 localizes to cell division plane and remains at new poles after division. A) Representative cells of the *smVenus-minD2* integrant strain undergoing two successive rounds of symmetric cell division. B) Representative cells of the *smVenus-minD2* overexpression strain undergoing three successive rounds of asymmetric cell division. Color-coding of arrows groups independent cell division events. Time is shown in hours and minutes.

4.5 Discussion

4.5.1 Unique features of growth and cell division in a haloarchaeon

The ability to observe single-cell dynamics is essential to the study of cell division.

Therefore, we have developed a time-lapse protocol that allows us to observe features of wild-type cell division and to image fluorescently labeled molecules in live cells. We find that both symmetric and asymmetric cell division are common in isogenic populations. In addition, *H. salinarum* appears to undergo bimodal growth, either in the stereotypical rod-shaped form or as flattened microcolonies.

Upon first glance, *H. salinarum* cell division appears to be quite similar to that of Bacteria: rod-shaped cells grow and undergo fission at a centrally located septum. It is intuitive to compare *H. salinarum* to *Bacillus* species, as many *H. salinarum* homologs to bacterial genes are closely related to those of *B. subtilis* (Ng et al., 2000). However, the appearance of cell wall constriction at the site of division prior to development of a septum is a feature of *E. coli* rather than *Bacillus* species (Adams & Errington, 2009). Unlike either species, however, asymmetric division is frequently observed in cells of *H. salinarum*, suggesting that the placement of the septum is controlled by a different mechanism. Symmetric cell division is considered to be an important element in the equal partitioning of DNA for the daughter cells. In Bacteria, inhibition of the shape-determining proteins MreB and PBP2 leads to a higher rate of anucleate cells (Young,

2006). While we have not yet developed a nontoxic DNA staining system for live, rapidly-dividing cells, preliminary experiments in fixed log-phase cells suggest that *H. salinarum* does not segregate its DNA during fission (data not shown). We suggest that the aberrant placement of the septum, along with the lack of DNA segregation, is acceptable due to polyploidy of the genome, with 15–30 copies of the chromosome per cell (Breuert et al., 2006). If chromosome copies are randomly distributed, then daughter cells are likely to receive at least one copy of the genome regardless of the placement of the septum. In keeping with this, asymmetric cell division was also observed in stationary phase cultures of *Methanococcus jannaschii* (Malandrin et al., 1999). *M. jannaschii* is also a euryarchaeon and thus is polyploid, whereas the monoploid crenarchaeal species *Sulfolobus sulfataricus* and *Sulfolobus acidocaldarius* primarily exhibit symmetric cell division (Poplawski & Bernander, 1997; Soppa, 2011). Further experiments should seek to confirm this hypothesis by tracking DNA during asymmetric and symmetric division of *H. salinarum*.

Many halophilic archaea are pleiomorphic, and *H. salinarum* is one of the few species that is described in the literature as a pure rod-shaped population (Oren et al., 2009). However, we show that this rod shape appears to be contingent on a free-swimming state and rapidly disappears upon vertical restriction of the cell. Interestingly, the cell can rapidly regain the rod shape once mechanical pressure is relieved. This immediate transition from rod to flattened cell and back again suggests

that the S-layer, despite its crystalline arrangement, is not a rigid scaffold, and is consistent with the lack of turgor pressure in this organism (Margolin et al., 1996; Perez-Fillol & Rodriguez-Valera, 1986). Indeed, turgor pressure and cell walls are described as the major determinants of rod shape in the Bacteria, and the absence of one of these elements could lead to instability of the rod form (Chang & Huang, 2014). A rod–flat cell transition under mechanical pressure, similar to what was observed here, has also been observed in *E. coli*, despite the internal turgor pressure of this organism (Si et al., 2015). However, constriction past 0.5 microns, half the height of a wild-type cell, completely halts growth. Whether *H. salinarum* has greater tolerance to mechanical pressure due to null turgor pressure should be examined.

This bi-modal growth has never before been described in the archaeal literature and thus presents a new perspective on the flexibility of the S-layer and the possibly unique biology of the Haloarchaea. Flat cells have a greater surface-area-to-volume ratio; this can improve the absorption of nutrients; therefore, this form of growth may be an adaptation to the plight of stationary cells that are unable to undergo chemotaxis (Young, 2006). In the haloarchaeon *Haloquadratum walsbyi*, which grows as a thin (0.1–0.2 micron) square-shaped cell, the high surface-area-to-volume ratio is hypothesized to maximize light absorption for the generation of energy by the proton-pump bacteriorhodopsin (Walsby, 2005). Therefore, flattened cell morphology may be an

alternative form of growth for the Haloarchaea in nutrient- or light-limited environments.

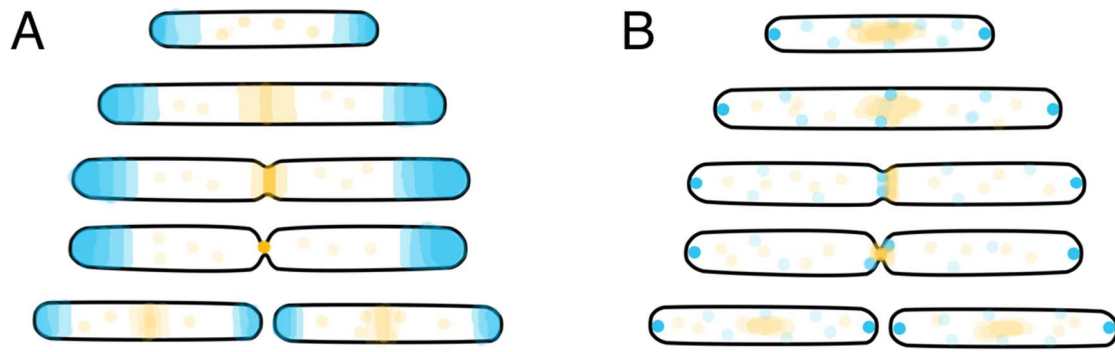


Figure 22: Comparison of FtsZ and MinD dynamics between *Bacillus subtilis* and *Halobacterium salinarum*. A) Model of cell division in *B. subtilis*, where a diffusion gradient set by MinD leads to positioning of FtsZ at the midcell. B) Model of cell division in *H. salinarum*, where MinD1 localizes to the site of constriction prior to formation of the Z ring. Blue color represents MinD; orange color represents FtsZ.

4.5.2 FtsZ1 and MinD2 participate in cell division in *H. salinarum*; observed dynamics are distinct from those of Bacteria

Due to the presence of FtsZ homologs in all sequenced euryarchaeal species, the predominant hypothesis in the field prior to this study was that cell division in the Euryarchaea resembled that of the Bacteria (Makarova 2010; Lindas & Bernander, 2013). Our time-lapse data indicates that although the *H. salinarum* proteins FtsZ1 and MinD2 do appear to be involved in cell division, this process is distinct from that of Bacteria (Fig. 22). Time-lapse confirmed previously known features of cell division protein dynamics while revealing several novel characteristics. Both FtsZ1 and FtsZ2 localized

to the site of constriction, and while the coordination of this localization with cell division dynamics was confirmed only for FtsZ1, we hypothesize that FtsZ2 copolymerizes with FtsZ1 and may be functionally redundant. Surprisingly, we find that FtsZ1 bundles or puncta may localize anywhere in the cell, including the poles. While this seemingly unregulated localization pattern is not strictly counter-indicated in an organism that undergoes asymmetric cell division, polar localization is not compatible with any known mechanisms of FtsZ-mediated cell division (Romberg & Levin; 2003). Indeed, we find that MinD2, although it forms puncta rather than diffusion gradients, is almost exclusively localized to the poles and does not oscillate; this is compatible with the behavior of the *B. subtilis* MinD protein. However, if MinD localization leads to the negative regulation of FtsZ polymerization, as it does in Bacteria, then polar FtsZ localization would never be observed. Furthermore, we observe that MinD2 localizes to the plane of cell division prior to cell constriction, arguing that MinD2 is present at the septum both prior to and during the polymerization of FtsZ1. This implies that MinD2 is not involved in negative regulation of FtsZ1 polymerization and may even be involved in positive regulation.

These potential interactions between FtsZ1 and MinD2 would be in direct contrast to the bacterial system of cell division. However, the inhibitory function of MinD is conveyed by the interactive partner MinC, a protein without any orthologs in the Euryarchaea (Gregory et al., 2008; Makarova et al., 2010). The dynamics observed

here imply that functional homologs of MinC are also absent. Other essential proteins required for cell division in bacterial species have not been detected in the genomes of any Euryarchaea; this includes the FtsZ-positioning proteins ZipA, ZapA, and FtsZA, and the MinD system proteins MinJ, MinE, and DivIVA, among many others (Bach et al., 2014; Rowlett & Margolin, 2015). While a structural homolog of SepF, which mediates tethering FtsZ to the membrane, has been computationally identified, studies have not yet confirmed the functionality of this protein (Makarova et al., 2010; Duman et al., 2013). Our novel model of FtsZ and MinD dynamics in the Haloarchaea, coupled with the absence of cell division partner proteins, suggests that euryarchaeal cell division is distinct from that of the Bacteria and likely requires its own independent set of proteins to regulate and form the cell division apparatus.

4.6 Acknowledgements

I would like to thank Callan Corcoran for cloning assistance and Eli Wilber for the generous donation of plasmids pMTF1044GFP-CHA, PMTF-smVenus, pEPW01, pEPW02, and pEPW03. Portions of the data presented here were generated using instrumentation from the Duke University Light Microscope Core Facility. Thanks to Horia Todor for suggestions on data processing. I am especially grateful to members of the Garner lab at Harvard University for the use of equipment, helpful discussions, and the tireless efforts of Ye-jin Eun. This research was supported by grants to AKS from the

National Science Foundation (NSF-MCB-10-52290 and 1417750) and the Duke Arts and Sciences Council Committee on Faculty Research and a Duke Biology Grant-in-aid to KAD.

5. Conclusions and Future Directions

5.1 Histones in the Archaea and evolution of histone function

Prior to this study, electron microscopy and micrococcal nuclease digestion of proteinaceous DNA suggested that Haloarchaea employ nucleosome particles composed of histone protein to compact their genomes (Shioda et al., 1989; Takayanagi et al., 1991; Ammar et al., 2013). However, these studies did not provide a link between histone protein and the structures or patterns observed. We have shown here that the sole histone protein of *Halobacterium salinarum* is unlikely to function as a canonical eukaryotic histone. Regulation of only 37 genes in log phase (Fig. 4, Section 2.3.3), the relationship of histone dosage to cell morphology (Fig. 3, Section 2.3.2), and the viability of the histone deletion strain (Fig. 2, Section 2.3.2), all suggest that HpyA serves as a transcription factor but not as the major compactor of DNA. We have also expanded the field of archaeal chromatin in that we compare histone mutant phenotypes and chromatin composition in two phases of growth without assuming that histone function is static. Proteomics of rapidly-sedimenting DNA from these cells yielded very little protein from histone or putative nucleoid-associated proteins (NAPs), suggesting that DNA packaging is not primarily mediated by protein, in direct contrast to evidence in other Euryarchaea.

How then do the Haloarchaea package their DNA? First, we must acknowledge that haloarchaeal DNA exists in an unusual biochemical environment: *H. salinarum*

cytoplasm contains 5 M potassium (Perez-Fillol & Rodriguez-Valera, 1986). Electron microscopy of naked DNA has shown that sodium concentrations as low as 150 mM can condense DNA into nucleosome-like particles (Eickbush & Moudrianakis, 1976). The addition of ionic salts (e.g., NaOAc, LiCl, NaCl) to lysate is an essential step in DNA purification and facilitates precipitation. However, 5 M potassium is well beyond the upper limit of most DNA structure and crystallization studies, so it is unclear what levels of condensation could be expected under this condition. We hypothesize that the presence of acidic, salt-adapted protein solvates DNA and may be required for preventing the precipitation of DNA, rather than packaging. However, even if DNA may be condensed *in vivo*, rapidly-sedimenting DNA in our study was isolated from sucrose gradients containing only 150 mM NaCl. The most parsimonious explanation for the continued rapid sedimentation of what should be denatured DNA is that chromosomal DNA from *H. salinarum* is supercoiled, as are all circular chromosomes in bacteria. Supercoiling does not necessitate the binding of a great number of proteins to result in compaction. Topoisomerases are required to generate supercoiling and a subset of NAPs, such as H-NS or SMC, may be involved in constraining individual supercoiled domains (Luijsterburg et al., 2008; Dame et al., 2011). This theory is compatible with the small fraction of NAPs detected in our proteomics experiment, and chromosomal super-structure could be experimentally verified by several assays including mapping of biotin-labeled psoralen by chromatin immunoprecipitation (ChIP)

to identify underwound regions and chromosome conformation capture (3C) to capture associations between proximal regions of DNA (Bermudez et al., 2010; Dekker et al., 2002).

Given these alternative options for DNA compaction and our evidence that histone is unlikely to compact DNA (Fig. 7 & Table 7, Section 2.3.6 & 2.3.7), what then is the role of histone in the Haloarchaea? Our data suggest that histone may act as a transcription factor. While the primary function of eukaryotic histone protein is to wrap DNA, the histone fold motif is not restricted to proteins involved in genome packaging (Arents & Moudrianakis, 1995). Histone variants may be involved in DNA repair, recruitment of RNA polymerase II, or may mark the centromeres (Talbert & Henikoff, 2010). Eukaryotic TBP-associated factors (TAFs) contain structural homologs of histone fold domains; these regions are required for formation of heteromeric complexes and thus the assembly of TFIID (Shao et al., 2005). Additionally, the B and C subunits of the human NF-Y complex also contain histone fold domains; together with subunit A this complex regulates transcription at the highly conserved CCAAT box (Ceribelli et al., 2008). Therefore, there is precedent for non-canonical roles of this structure. Currently, chromatin immunoprecipitation experiments are underway to determine the extent and location of histone–DNA binding throughout the genome.

If histone is indeed shown to act as a transcription factor, several facets of binding activity should be explored. Histone has been shown to be capable of binding

RNA (Singh et al., 2010), and the archaeal NAP Alba binds both DNA and RNA (Marsh et al., 2005); therefore, immunoprecipitated samples should be examined for the presence of both nucleotide species to rule out the possibility that HpyA mediates gene expression changes through direct binding to mRNAs. The ability of HpyA to compact DNA should also be measured, as it is unclear whether the polar distribution of positive charges on the surface of this protein are sufficient to wrap DNA as in eukaryotic nucleosomes. Electrophoretic mobility shift assay (EMSA) can be used to distinguish the mode of binding, since the proteins that simply bind to DNA will slow the electrophoretic mobility, while those proteins that wrap DNA will accelerate the electrophoretic mobility (Pavlov et al., 2002). EMSA is usually performed in low ionic strength buffer; however, there is precedent for effectively mimicking the high salt intracellular environment of *M. kandleri* using 0.1 M potassium glutamate (Pavlov et al., 2002). Should all of the above-mentioned assays fail to find an association of HpyA with nucleic acids, co-immunoprecipitation should be performed in order to detect histone-associated proteins that may facilitate DNA binding activity *in vivo*.

If indeed HpyA functions as a DNA-binding protein that directly regulates transcription, HpyA provides an intriguing model for the evolution of histone proteins. The conservation of a histone-deacetylase in the histone-encoding operons of halophiles is unique among the Archaea and strongly suggests that the haloarchaeal histone is post-translationally modified by this protein. If post-translational modification (PTM) of

histone can be proven to exist in a class of archaeal species, this then suggests an archaeal origin for the epigenetic code. While post-translational modifications (PTMs) of histones have not been detected in other archaea (Zhang et al., 2012; Nalabothula et al., 2013), detection of histone PTMs typically requires different sample preparation than other proteins due to the higher density of arginines and lysines. These amino acids are targets of trypsinization and their high density in histone protein leads to an abundance of small peptides that can be lost during mass spectrometry detection (Tan et al., 2011). Indeed, HpyA, a 150 amino acid protein, encodes 17 arginines and one lysine residue. Therefore, in any further investigation of HpyA PTM, samples should be treated with non-trypsin proteases such as Glu-C.

The evidence presented here is consistent with a novel hypothesis in which the primary function of histone in the Haloarchaea is likely to be transcriptional regulation, rather than genome compaction. This has ramifications for the proposed evolutionary history of histone protein and suggests that the original function was that of a transcription factor, not a genome compactor, as the regulation of gene expression appears to be the only conserved function. However, this hypothesis requires additional biochemical evidence and many questions about haloarchaeal genome compaction remain.

5.2 Genomic instability and fitness

Results presented in Chapter 3 suggest that IS elements are associated with copy number variants (CNVs) in *H. salinarum*, an observation consistent with the hypothesis that genomic instability abounds in the chromosome in addition to the megaplasmid. Megaplasmid instability is strongly correlated with gene expression changes as determined by qPCR and a novel implementation of a CGH analysis algorithm. Given these results, further experiments should examine the relationship between gene expression and copy number as well as genomic instability and fitness.

These results are consistent with observations in other archaeal species. For example, the chromosome of *Sulfolobus sulfataricus* and the megaplasmid of halophilic archaeal species are frequently rearranged due to the abundance of insertion sequence (IS) elements (Ng et al., 1998, Pfeiffer et al., 2008, Martusewitsch et al., 2000; Hawkins et al., 2013). IS elements are small (0.5–2 kb) mobile DNA elements that encode a transposase gene between direct repeat sequences (Brugger, 2002). Mobility of IS elements can lead to frequent insertion into genes or regulatory elements, and illegitimate recombination between homologous IS elements can lead to genomic rearrangements such as inversions, duplications, and deletions (Darmon & Leach, 2014). Our data suggest that IS elements are strongly associated with DNA amplifications and deletions in certain regions of the major chromosome (Fig. 13, Section 3.3.2). This phenomenon was originally observed in the megaplasmid (Sapienza et al., 1982, Ng et

al., 1998) of *H. salinarum* NRC-1. While CNVs were not predicted to be common in the chromosome, IS-mediated integration of a megaplasmid into the chromosome of *Haloferax volcanii*, a halophilic archaeal species related to *H. salinarum*, had been noted in the literature (Hawkins et al., 2013). It was previously unknown whether large-scale rearrangements had direct effects on gene expression in the Archaea. We have shown that a 90 kb duplication in the megaplasmid pNRC100 leads to increased gene expression across the entirety of the region, which is inconsistent with dosage compensation or silencing of extraneous DNA in this organism.

To investigate these phenomena, we implemented a novel application of a DNA array-segmenting algorithm developed for comparative genome hybridization (CGH; Venkatraman & Olshen, 2007). Using randomly sheared genomic DNA controls from ChIP-chip data as input into the segmentation function, we generated a map of deletions and amplifications. Aside from the scientific investigation of genetic instability, these methods are general and can be applied to any gene expression or ChIP experiment in order to test for gross genetic amplifications or deletions. In the future, this will allow researchers to ensure that strains and clonal isolates are stable and have not accumulated unexpected DNA deletions or amplifications, all without performing additional experiments.

Here we have discovered that the amplification of plasmid DNA is strongly associated with increases in gene expression of the amplified region, suggesting that *H.*

salinarum has no mechanism for dosage compensation. This has ramifications for the variable genetic content between the chromosome and megaplastids as previously observed by Breuert and colleagues (2006). Because the chromosome appears to exist in higher copy number than either of the two megaplastids, this raises the question of whether genes are more highly expressed on the chromosomes. Preliminary computational analysis comparing raw gene expression levels from several gene expression microarrays in different regions of the chromosome yielded mixed results. However, extending this analysis to include the entirety of all 1154 gene expression microarrays in our dataset (see Appendix E for references) in the future would likely clarify this relationship. Further validation should also be performed by quantitative RT-PCR in order to compare absolute quantities of mRNA in each genomic region. Genomic instability and the presence of extrachromosomal genes is not unique to the Haloarchaea, but is observed in other archaeal systems including *Sulfolobus*, *Thermococcus*, and *Pyrococcus* species (Brugger et al., 2002). Therefore, such experiments could indicate that polyploidy and segmental duplications affect gene expression in other classes of the Archaea.

These observations of genomic variability raise the question of the evolutionary advantage of genomic instability in the chromosome, which encodes most of the essential genes in halophiles (Ng et al., 2000). Many Archaea, and the Haloarchaea in particular, encode homologs of bacterial genes, many acquired by horizontal gene

transfer (HGT; Makarova et al, 2007; Ng et al., 2000; Wolf et al., 2012). In order for HGT to be productive, organisms must have a method for integration of the new DNA into the chromosome (Darmon & Leach, 2014). In pathogenic bacteria, genomic instability is commonly used as a way to avoid host response by rapid generation of mutated membrane or cell wall components (phase variation; Oliverio & Katz, 2014; Wisniewski-Dyé & Vial, 2008). The glycosylation clusters of *H. salinarum* cell wall S-layer protein showed frequent CGH events in our analysis, suggesting that a process related to phase variation may generate variability in the cell wall S-layer of this organism, possibly to avoid viral predation, since archaeophages are common and diverse in hypersaline environments (Porter et al., 2007). In addition, Haloarchaea possess highly efficient DNA-damage repair systems, which use homologous recombination to maintain a low rate of variation at the individual nucleotide level (Mackwan et al., 2007). Therefore, gross rearrangements of the genome may be an efficient mechanism to generate diversity and alter regulatory programs. Undoubtedly, other mechanisms by which *H. salinarum* tolerates genomic instability in the chromosome exist; however, the examples from the literature discussed here all share the common theme of generating genetic variation and maximizing fitness.

While we have offered evidence of genetic rearrangements that are associated with transcriptional changes, further experiments would clarify the connections between gene dosage, gene expression, and organism fitness. Gene expression and comparative

genomic hybridization (CGH) arrays performed simultaneously with samples derived from the same culture would be a strategy for investigating these connections. In order to differentiate whether genome copy number variation is due to genomic rearrangements or natural increase in genomic content during replication, the growth phases of cultures used for such an experiment could be synchronized by chemical (e.g. aphidicholin; Herrmann & Soppa, 2002) or culturing (e.g. pre-growth to stationary phase) techniques. In addition, time course experiments in which cultures are sampled for genomic variation during growth curve from lag to stationary phase would control for differential replication dynamics of the chromosomal elements (Breuert et al., 2006). Competition experiments conducted on a collection of strains that harbor stable deletions, inversions, duplications, or transpositions would also clarify the relationship of rearrangements to microbial fitness. Future experiments to generate such a collection could yield significant information about the causes and consequences of CNVs. For example, cells could be serially passaged with increasing levels of non-lethal stress such as peroxide, mevinolin, or novobiocin, thus selecting for mutations conferring improved fitness under stress. DNA sequencing, CGH, or pulsed-field gel electrophoresis (PFGE) of digested chromosomal DNA could then be used to detect CNVs or rearrangements that confer fitness advantages.

5.3 Euryarchaeal shape and cell division

Little is known about the single-cell dynamics of division in the Archaea, with only four papers describing broad characteristics of cell division in six species at the time of writing (Horn et al., 1999; Herrmann & Soppa, 2002; Poplawski & Bernander, 1997; Duggin et al., 2015). In addition, although homologs of the proteins comprising the cell division apparatus of the Bacteria are detectable across the Euryarchaea, until now, protein localization studies have been restricted to investigations into the function of the Z ring-forming protein, FtsZ (Margolin et al., 1996; Poplawski et al., 2000; Wang et al., 1996; Duggin et al., 2015). This experimental focus on FtsZ ignores a handful of identified homologs related to bacterial cell division, including MinD, SepF, and the chromosome-segregation proteins in the ParA family (Makarova et al., 2010). Currently, no investigation into other euryarchaeal cell division proteins has been reported in the literature.

Protein localization studies in the Archaea lag behind those in the Bacteria, primarily due to a paucity of methods for tracking single cells in real time. To address this challenge, in Chapter 4 we present a method for time-lapse microscopy of live *H. salinarum* that is suitable for use with any halophilic archaeon. We found evidence that *H. salinarum* readily grows as flattened, amorphous cells in microcolonies in response to minimal physical pressure, suggesting that in absence of turgor pressure (Oren et al., 1999), the rod form may not be actively maintained by *H. salinarum*. To our knowledge,

our evidence also represents the first observation of the dynamic localization of FtsZ and MinD homologs in actively dividing cells. Although our evidence suggests that both proteins are involved in cell division, observed dynamics are distinct from the Bacteria, in which MinD positioning anticorrelates with that of FtsZ. Therefore, further research is necessary to determine the underlying mechanisms of cell division in this organism.

Experiments are currently underway to develop a more comprehensive model of cell division in this organism. Refinement of the time-lapse protocol presented here will enable consistent retention of rod-shaped morphology, allowing quantitation of basic properties of cell division, including the area of mother and daughter cells, the rate of cell division on the single-cell level, and the heterogeneity of cell size throughout the population. Although asymmetric division was observed, the frequency and underlying mechanisms regulating this phenomenon remain unclear. These observations also raise questions regarding whether asymmetry is heritable or whether asymmetric daughter cells default to symmetric division. Once a baseline for these cell division parameters is established, we can proceed to fully describe division in flattened cells. The fluorophore-labeled proteins we have established in this work will provide invaluable markers for tracing individual steps in the process of cell division.

Experiments are also underway to examine the function of additional FtsZ and MinD homologs. The genome of *H. salinarum* encodes five FtsZ homologs and two MinD homologs (Ng et al., 2000). While our preliminary studies demonstrate that FtsZ1 forms

a Z ring at the site of division (Fig. 16 & 17; Section 4.3.2), positioning of this Z ring is incompatible with any bacterial models to date, as our evidence suggests that MinD2 localizes to the site of division prior to septation. Therefore, if the Z ring is indeed the major driver of constriction, then MinD2 localization cannot lead to inhibition of FtsZ1 polymerization as in Bacteria (Gregory et al., 2008). These differences suggest that several follow-up experiments are imperative to determine: a) the relative positioning of FtsZ1 and MinD2, perhaps using two-color fluorescence microscopy to simultaneously image these proteins, b) the dynamics of how the remaining FtsZ and MinD homologs are positioned during cell division, and c) which other proteins are involved in cell division. Strains with which to address a) and b) are already under construction. To identify other cell division proteins, co-immunoprecipitation on FtsZ1 or MinD2 would identify binding partners. Alternatively, a transposon screen (Kiljunen et al., 2014) for mutants with disturbed morphology would also yield additional players in cell division. Follow-up experiments would address many questions regarding the dynamics of MinD2 and FtsZ1 subcellular localization during cell division in *H. salinarum*.

While much remains to be discovered regarding the mechanisms governing cell division in the Euryarchaea, the current work has made significant progress in the elucidation of FtsZ and MinD function. Much of our current knowledge regarding cell division in Archaea is based on studies in the Haloarchaea, and it is likely that many breakthroughs will continue to come from these species. Their mesophilic nature,

flexible metabolisms, variety of morphologies, and developed genetic tools makes these attractive models to characterize euryarchaeal cell division (Soppa, 2006).

5.4 Summary

Our research using the haloarchaeon *H. salinarum* has made contributions to the study of histones, genomic instability, and cell division in the Euryarchaea. We have developed new scientific tools, including the application of a CGH segmenting algorithm to detect CNVs in both ChIP data and gene expression microarrays, and a time-lapse protocol for imaging cell division and fluorescently-labeled molecule dynamics. This also includes the first reported use of a yellow fluorescent protein in the Archaea. We have generated answers to persistent questions, finding that the beads-on-a-string morphology observed in *H. salinarum* DNA is not due to histone protein as previously assumed, observing that IS elements generate CNVs regardless of the locus and that these events may lead to gene expression changes, and showing that bacterial cell division protein homologs FtsZ1 and MinD2 are indeed involved in Euryarchaeal cell division, albeit with distinct dynamics. Consequently, our research has raised several new questions about the relationship of HpyA with gene expression and the organization and architecture of the *H. salinarum* genome. It is also unclear how genome instability is related to diversity and fitness in the Haloarchaea. Finally, the mechanisms that govern the transition to flat cell growth from rod shape, the interactions between FtsZ and MinD homologs, and the

other proteins involved in Euryarchaeal cell division still remain to be discovered. In conclusion, we have presented here a series of steps towards answering the above questions, which will be instrumental to the understanding of basic archaeal physiology.

Appendix A

Complete Defined Medium (CDM) for *Halobacterium salinarum* (protocol as described by Linda Cao)

For 4L:

Fill bucket with 2 L of milliQ water (from chemistry lab next door).

Turn stir plate on.

Weigh out salts and powders:

- 500 g NaCl (sodium chloride) - *You need 1000 g total but doing this in two separate additions will help the salts dissolve. Wait until the NaCl is completely dissolved and continue adding the other salts and powders*
- 39.4 g MgSO₄ 7H₂O (magnesium sulfate heptahydrate)
- 33.48 g MOPS
- 8.052 g KCl (potassium chloride)
- 0.08 g NaH₂PO₄ (sodium phosphate monobasic)
- 1 g L-asp (aspartic acid)
- 6.26 g L-glu (glutamine acid)
- 0.448 g L-tyr (tyrosine)
- 20 mg Folic acid
- 20 mg Biotin (powder)

Start adding the liquid reagents (Grab the amino acid box from the cold room)

- 40 mL of 100X L-ala (alanine)
- 40 mL of 100X L-arg (arginine)
- 40 mL of 100X L-asn (asparagine)
- 160 mL of 25X L-gln (glutamine)
- 40 mL of 100X L-his (histidine)
- 80 mL of 50X L-ile (isoleucine)
- 160 mL of 25X L-leu (leucine)

80 mL of 50X L-gly (glycine) (**this needs to be 2X Schmid 2009 amount per HT**)

Add another 500 g NaCl (sodium chloride)

Finish adding liquid reagents

- 35.6 mL 100X thiamine
- 40 mL of 100X L-lys (lysine)
- 40 mL of 100X L-met (methionine)
- 40 mL of 100X L-phe (phenylalanine)
- 40 mL of 100X L-pro (proline)
- 40 mL of 100X L-ser (serine)
- 40 mL of 100X L-thre (threonine)
- 40 mL of 100X L-trp (tryptophan)
- 40 mL of 100X L-val (valine)
- Add metals (*found in the -20 degree freezer*)
- 400 µL of 10000X (100mM) MnSO₄ – final conc. 0.01mM (10uM)
- 400 µL of 10000X (100mM) FeSO₄ – final conc. 0.01mM (10uM)
- 400 µL of 10000X (100uM) CuSO₄ – final conc. 0.00001mM (0.01uM)
- 400 µL of 10000X (100uM) ZnSO₄ – final conc. 0.00001mM (0.01uM)

Adjust pH of medium to 6.8 with KOH

Filter sterilize the medium using 1L filter sterilizer tops in sterile glass reagent bottles.

Store at 4 degrees Celsius.

Appendix B

Primers used in this study

Primer Name	Direction	Purpose	Sequence (5'-3')	Source
K51	R	Sequencing, Southern probe	GCAACTGGGAGTGCAGGGTC	This study
K52	F	Sequencing, Southern probe	CTGATGGCGTCCGACTTCGAG	This study
K53	F	Sequencing	CGTGACTCGCGTCGAGGTAAC	This study
K54	R	KO screen, sequencing	CGTCGACCCGAAGGCCTGATG	This study
HpyA_gen_seq_F	F	KO screen	GACAGCCTCGAACAGGAC	This study
pNBK05	F	Plasmid screen, sequencing	GACAGATCGAGCAGACGCATC	This study
pNBK07	R	Plasmid screen, sequencing	TGTGCTGCAAGGCGATTAAG	This study
seq_reporter	R	Sequence OE plasmid	AGAACGAGTAGCACACCAAAG	This study
1044_GFP_CHA_ Kpn1_fwd91	F	Plasmid screening, sequencing	CGCGGAAACGATGAAATGCG	This study
K71	F	HpyA insert for overexpression plasmid	ACGTACGT CAT ATGAGCGTCGAACTCCCGTTC	This study
K72	R	HpyA insert for overexpression plasmid	ATCGTGAC AAGCTT TCATTCCACGAGCGTGACGTAC G	This study

K13	F	HpyA, Gateway cloning primer A	GGGG CAAGTTTGTACAAAAAAGCA GGCT CGACATTCCGGTTGCGTTGTG	This study
K68	R	HpyA, Gateway cloning primer D	GGGG CCACTTTGTACAAGAAAGCT GGGT GGCGTTGTTACGAAGCA	This study
HpyA-B3	R	HpyA, deletion primer B	GAACGGGAGTTCGACGCTCAT	This study
HpyA-C3	F	HpyA, deletion primer C	ATGAGCGTCGAACTCCCGTTC GAATGAGATTCCGATTCAGCG	This study
k_vng1673g_d_R	R	<i>Δura3</i> genotype confirmation	GGTCCCGGGACCGCCCCCG	Wilbanks et al., 2012
k_vng1673g_e_F	F	<i>Δura3</i> genotype confirmation	GGGGACGTGGCCGCGACGTT	Wilbanks et al., 2012
pRPA200_iso	F	200bp <i>rpa</i> promoter fragment for complementation plasmid	CCTCCCATGCCACTCTTCACA CGCGGTAC TGGCTCCGCAAGCCA	This study
pRPA_hpyA_iso	R	200bp <i>rpa</i> promoter fragment for rescue plasmid	GAGTTCGACGCTCAT TGTCGGTTCAGGCCA	This study
HpyA_tail_pMTF_iso	R	<i>hpyA</i> insert for complementation plasmid	CGTACGGGTACGCCGAAAGCT TGGATCCG TCATTCCACGAGCGTG	This study

HpyA_ATG_F	F	<i>hpyA</i> insert for complementation plasmid	TGGCCTGAACCGACA ATGAGCGTCGAACTC	This study
0134_qpcr_F2	F	qRT-PCR validation of microarray data	GGAGATCGACGACCGGTTCC	This study
0134_qpcr_R2	R	qRT-PCR validation of microarray data	CATTCCACGAGCGTGACGTA	This study
1123_qpcr_F	F	qRT-PCR validation of microarray data	GAGGGGACCCGCACACTC	This study
1123_qpcr_R	R	qRT-PCR validation of microarray data	CTCGTCGCTGATCACGTCACC	This study
1317_qpcr_F	F	qRT-PCR validation of microarray data	GGACGTGGACGCCGATTTC	This study
1317_qpcr_R	R	qRT-PCR validation of microarray data	GTCCGGATGCTCTCGAACG	This study
1641_qpcr_F	F	qRT-PCR validation of microarray data	TGTTGGTCGCGCTGTATGC	This study
1641_qpcr_R	R	qRT-PCR validation of microarray data	ACGACCAGTCCGTATTTGAACC	This study

1815_qpcr_F3	F	qRT-PCR validation of microarray data	GCATCTGCCTGGGTCAACA	This study
1815_qpcr_R3	R	qRT-PCR validation of microarray data	GACGGTATAGCCGTGGTTCTG	This study
5019_qpcr_F	F	qRT-PCR validation of microarray data	GAGCCAACAAAAGACGAGAC	This study
5019_qpcr_R	R	qRT-PCR validation of microarray data	GTGGTCATGCCAGCAATC	This study
5192_qpcr_F	F	qRT-PCR validation of microarray data	CCGTTATTCAGCGGTCCTC	This study
5192_qcpr_R2	R	qRT-PCR validation of microarray data	CATGCTGCGGTTCAATTCTG	This study
VNG1756_Fwd	F	qRT-PCR validation of microarray data	ACGAAGTGTTTGCGGTCG	This study
VNG1756_Rev	R	qRT-PCR validation of microarray data	TCCACGTCCGGTATTTTCATG	This study
A206K_F2	F	Site-directed mutagenesis -> monomeric venus	CCTACCAATCTAAGCTTTTCGAA AGATCCCAAC	This study

A206K_R2	R	Site-directed mutagenesis -> monomeric venus	TTCGAAAGCTTAGATTGGTAG GACAGGTAATGG	This study
HCY-A3	F	C-terminal tagging of HpyA with smVenus	GGTGAGCTCTGGTACCCTCTAG TCAAGGG TGACTCTCGGCCAGGAGG	This study
HCY-B2	R	C-terminal tagging of HpyA with smVenus	AGAACCTCCGCCACC TTCCACGAGCGTGACGTAC	This study
HCY-C	F	C-terminal tagging of HpyA with smVenus	GTCACGCTCGTGGAA GGTGCGGAGGTTCT	This study
HCY-D	R	C-terminal tagging of HpyA with smVenus	CGCTGAATCCGAATC TTATTTGTATAGTTCATCCATG CC	This study
HCY-E2	F	C-terminal tagging of HpyA with smVenus	GAACTATACAAATAA ATGAGATTCGGATTCAGCG	This study
HCY-F	R	C-terminal tagging of HpyA with smVenus	<i>CGGCCAGTCCGTAATACGACTCA</i> <i>CTTAAGG</i> GTCCTGGGTGCCGTT	This study
F1CY-A2	F	C-terminal tagging of FtsZ1 with smVenus	GGTGAGCTCTGGTACCCTCTAG TCAAGG GGCGTTCAAGGTGTCC	This study
F1CY-B	R	C-terminal tagging of FtsZ1 with smVenus	AGAACCTCCGCCACC CTCCACGTAATCGATGTCC	This study

F1CY-C	F	C-terminal tagging of FtsZ1 with smVenus	ACATCGATTACGTGGAG GGTGGCGGAGGTTCT	This study
F1CY-D	R	C-terminal tagging of FtsZ1 with smVenus	GGCACAGACCCCGCG TTATTTGTATAGTTCATCCATGCC	This study
F1CY-E	F	C-terminal tagging of FtsZ1 with smVenus	GGATGAACTATACAAATAA CGCGGGGTCTGTGCC	This study
F1CY-F	R	C-terminal tagging of FtsZ1 with smVenus	<i>CGGCCAGTCCGTAATACGACTCA</i> <i>CTTAAGG</i> CGATGCCCTCGACGTCC	This study
MDNY-A	F	N-terminal tagging of MinD2 with smVenus	<i>TAGGTGAGCTCTGGTACCCTCTA</i> <i>GTCAAGG</i> ACCGCGCCCGAGAAC	This study
MDNY-B	R	N-terminal tagging of MinD2 with smVenus	TCCTTTACTCACCAT ATGGTTCATCTTCGCACACC	This study
MDNY-C	F	N-terminal tagging of MinD2 with smVenus	CGAAGATGGAACCAT ATGGTGAGTAAAGGAGAAGAA	This study
MDNY-D	R	N-terminal tagging of MinD2 with smVenus	CACACGGCCACTCAT TGAGCCACCACCTCC	This study
MDNY-E	F	N-terminal tagging of MinD2 with smVenus	GGAGGTGGTGGCTCA ATGAGTGGCCGTGTG	This study

MNDY-F2	R	N-terminal tagging of MinD2 with smVenus	CGGCCAGTCCGTAATACGACTC ACTTAAGG GACTCGCTCGACGAG	This study
---------	---	---	---	---------------

Appendix C

Construction of smVenus fluorophore for use in *Halobacterium salinarum*

Acknowledgements:

The smVenus fluorophore was designed and constructed by Eli Wilber (Kiehart Lab, Duke Biology). Plasmids bearing smVenus(S99F), eVenus, and eVenus(F99S) were constructed by Janice Crawford (Keihart Lab, Duke Biology). I constructed the mVenus(A206K) plasmid and all *H. salinarum* strains used in this section. The following tests and analyses of these strains are my own.

Introduction:

Although a green fluorescent protein (GFP) had already been adapted for use in the halophilic archaea (Reuter & Maupin-Furlow, 2004), whose hypersaline cytosol may present difficulties for the folding of heterologous proteins (Kennedy et al. 2000), we sought to develop a yellow Venus variant of this protein to improve brightness. This Venus variant, referred to as smVenus, was shown to be fluorescent upon expression and excitation in *Halobacterium salinarum*. However, upon comparison with standard eukaryotic Venus and a “revertant” smVenus, we suggest that eukaryotic Venus or monomeric smVenus are the better choices for further fluorescent labeling.

Methods:

Protein alignment:

Amino acids were aligned using the CLUSTAL W algorithm as implemented by MegAlign (DNASTar Lasergene suite, v11.0).

Growth conditions:

Cultures were routinely incubated at 42°C with agitation in complex medium (CM; 250 g L⁻¹ NaCl, 20 g L⁻¹ MgSO₄•7H₂O, 3 g L⁻¹ sodium citrate, 2 g L⁻¹ KCl, 10 g L⁻¹ peptone). Mevinolin was added to a final concentration of 1 µg mL⁻¹ to maintain the fluorophore-expressing plasmids. Agarose pads for time-lapse were composed of 80% Basal Salts buffer.

Single cell fluorescence:

Cells were grown to mid-log phase and spun to concentrate. 5 µl of culture was placed on a slide with an agarose pad and immediately imaged. All images were taken with a Zeiss Axio Observer A1 inverted microscope connected to a Hamamatsu digital camera model ORCA-ER with HBO arc lamp excitation. All cells were imaged under 100X magnification (1000X total) with a phase contrast objective. Image analysis was performed using ImageJ: cell outlines were generated by threshold of phase contrast images, cell outlines were set as ROIs, background was subtracted from fluorescence images, and total cellular fluorescence was calculated for each ROI.

Measurement of whole-culture fluorescence:

Cultures were spun 5 min at 10,000 r.p.m and the pellet was resuspended in 100ul basal salts. Fluorescence was measured using an FLx800 BioTek microplate reader with a 485/20 filter. Reported values are normalized for background fluorescence of basal salts buffer and autofluorescence of the empty-vector control strain *ura3/pMTFCHA*.

Results and Discussion:

Development of smVenus and Halobacterium salinarum fluorescent strains:

A “salt-adapted” yellow fluorescent protein or smVenus was adapted from smGFP1044, a variant of green fluorescent protein previously optimized for stability in *Halobacterium salinarum* (Reuter & Maupin-Furlow, 2004). The smVenus fluorophore was developed through successive rounds of site-directed mutagenesis, resulting in 12 amino acid changes, most of them concerning folding and solubility. In addition, smVenus lacks the *ssr1* tag (N-terminal alanine) and is likely to have a decreased turnover rate relative to smGFP1044 (Fig. 1; Reuter & Maupin-Furlow, 2004). As standard Venus is already optimized for folding and solubility, mutagenesis generated a protein nearly identical to Venus fluorophores already in use with Eukaryotic and Bacterial tagging systems.



Figure 1: Amino acid comparison between fluorescent proteins. In blue is the site of the solubility mutation S99F, in purple is the site of the monomeric mutation A206K.

The sole amino acid difference between standard Venus and smVenus is a serine at amino acid 99, rather than a phenylalanine (Fig. 1). Because F99S has been shown to increase protein solubility, we hypothesized that this might be important for protein expression in *H. salinarum*, as these organisms contain extremely high (up to 5M) intracellular potassium (Pérez-Fillol & Rodríguez-Valera, 1986). To test this hypothesis, a revertant smVenus (S99F), a Venus optimized for eukaryotic systems (hereafter referred to eVenus), and an eVenus with the solubility mutation (F99S) were ligated into the same vector backbone as smVenus and each was transformed into the NRC-1 wild-type host. Concerns over fluorescent protein aggregation led us to develop a monomeric smVenus (hereafter referred to as mVenus), using site directed mutagenesis

to produce the A206K amino acid change. mVenus was also transformed into NRC-1 and was compared along with the solubility test strains.

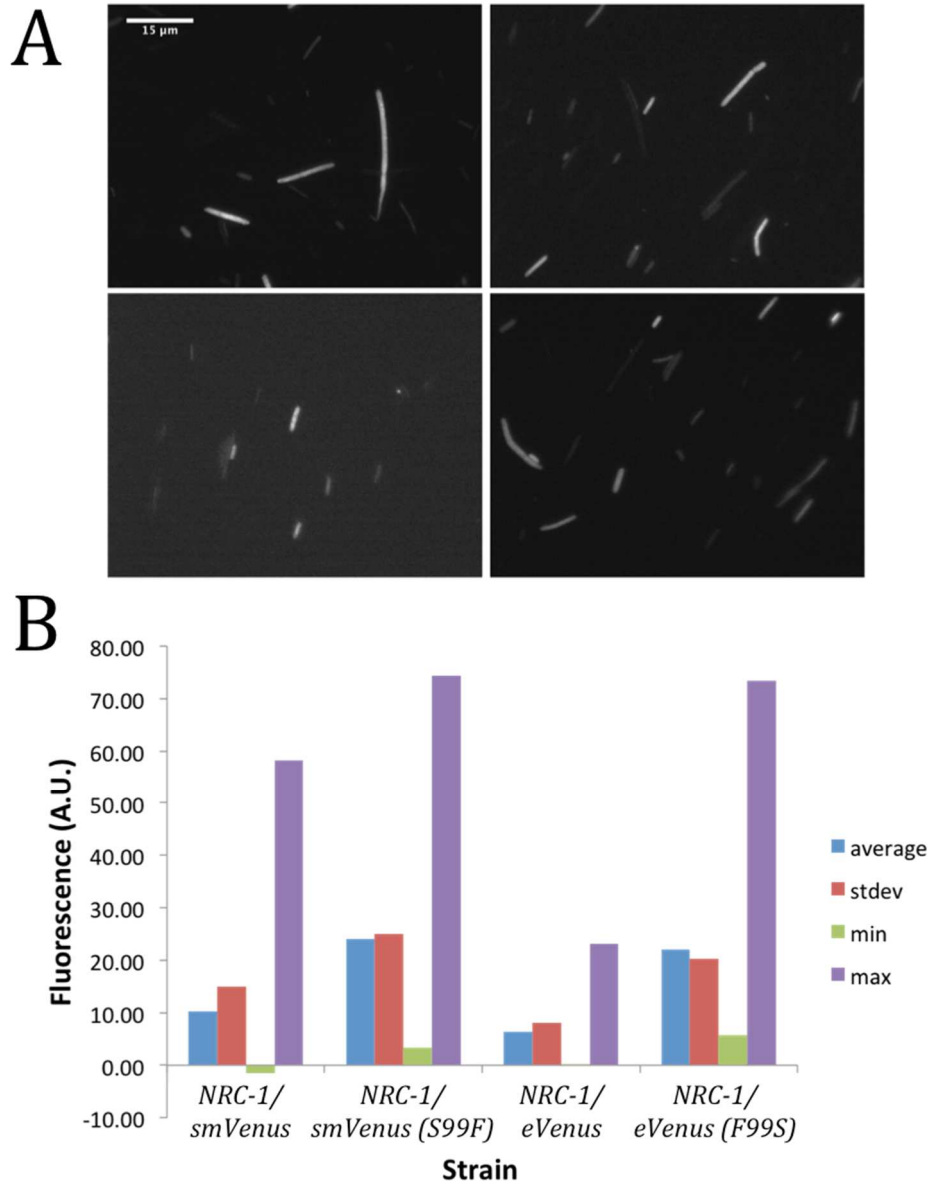


Figure 2. Fluorescence imaging of Venus constructs. Clockwise from upper left: smVenus, smVenus (S99F), eVenus, eVenus (F99S). B) Quantitation of fluorescence micrographs. OD600 of strain cultures (from left): 0.4, 0.5, 0.2, 0.5.

Fluorescence comparison of Venus variants:

Fluorescence of smVenus was compared to eVenus, the smVenus revertant smVenus(S99F), and eVenus with the solubility-promoting mutation F99S. Initial comparison was by fluorescence microscopy (Fig. 2). All strains fluoresced in the yellow spectrum, including the eVenus strain, which was not optimized for expression in Haloarchaea. This was not unexpected, as standard Venus contains several mutations to increase solubility and folding relative to GFP. On average, the revertant smVenus(S99F) was brighter than smVenus, and eVenus(F99S) was brighter than eVenus. Variability in fluorescence was very high for all four strains tested. Therefore, whole-culture fluorescence was also measured by microplate reader (Fig. 3).

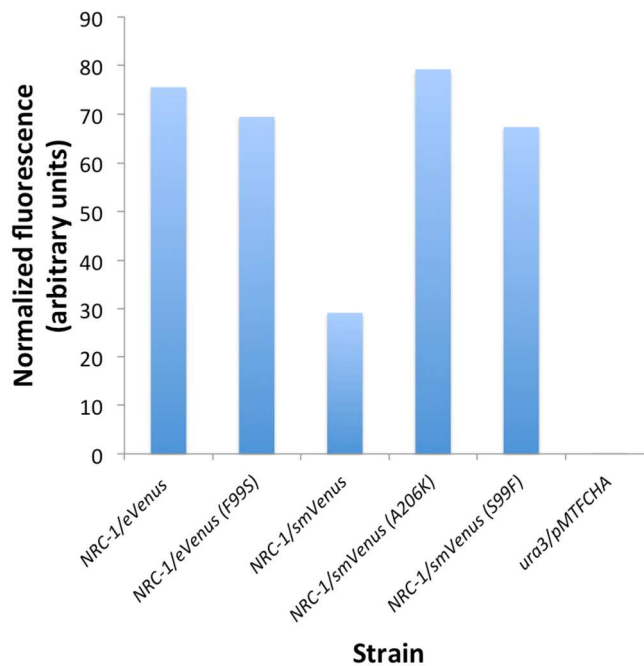


Figure 3: Quantitation of whole culture fluorescence by strain. Average values shown represent data from a total of at least 21 replicates from 4 biological replicates where samples were taken over the entire growth curve (OD600 = 0.1-1.7).

We did not observe dependence between total fluorescence and growth phase, thus all time-points were combined when determining average fluorescence of each strain. The previous four strains were also compared with the monomeric (A206K) smVenus. All strains were approximately equally fluorescent, with the exception of smVenus, which was the least bright. Anecdotally, *NRC-1/smVenus* grew much slower than the other strains. Whether slow growth has led to diminished fluorescence or whether the smVenus construct is toxic to the cell is unknown at this point. We did not see significant difference in fluorescence for eVenus with or without the F99S mutation. The S99F smVenus revertant was brighter than smVenus, but this difference may be due to differential growth rates. Regardless, it is clear from these results that the solubility mutation F99S does not improve fluorescence of Venus protein in *H. salinarum*. We also find that the standard eVenus, optimized for use in *Drosophila melanogaster*, is an acceptable fluorophore for the Haloarchaea. This suggests that many other standard fluorophores may not need adaptation for use in the archaea, provided they are already optimized for folding. Finally, it is the recommendation of the authors that all further experiments be performed with monomeric Venus, as the A206K strain was the brightest tested in this study.

Conclusions:

The yellow fluorescent protein variant Venus is suitable for use in *H. salinarum*, regardless of the genetic origin. There are no functional advantages to using the “salt-

adapted" smVenus, as the F99S mutation does not improve fluorescence. Indeed, this variant should only be used in the monomeric form, smVenus(A206K). The ease of use of the eukaryotic Venus suggests that other standard fluorophores may also function in the archaea without any further optimization.

References:

Pérez-Fillol M., Rodríguez-Valera F. Potassium ion accumulation in cells of different Halobacteria. Microbiologia. 1986 Oct;2(2):73-80.

Reuter C.J., Maupin-Furlow J.A. Analysis of proteasome-dependent proteolysis in Haloferax volcanii cells, using short-lived green fluorescent proteins. Appl Environ Microbiol. 2004 Dec;70(12):7530-8.

Appendix D

Supplementary data for Chapter 3: FtsZ sequence conservation and FtsZ and MinD expression in *Halobacterium salinarum*.

Methods:

Amino acid alignment was performed as described in Chapter 4, quantitation of cellular fluorescence was performed as described in Appendix C, and gene expression analysis was performed as described in Chapter 2. Pearson correlation of gene expression profiles was implemented using the R base package.

H. salinarum encodes five FtsZ homologs of variable conservation:

Of the full contingent of essential bacterial cell division proteins, homologs of only three of these proteins are found in the Euryarchaeota: FtsZ, MinD, and SepF.

Both the FtsZ and MinD protein families are expanded: 5 FtsZ homologs can be found encoded in the genome of *H. salinarum*. Two of these homologs, FtsZ1 and FtsZ2, have much higher amino acid homology to *Bacillus cereus* FtsZ than the other three (Table 1). In particular, FtsZ5 appears to be highly degenerate, with a protein BLAST E-value of 1.6.

Table 1: *H. salinarum* protein homology to FtsZ from *Bacillus cereus*

Gene Name	Gene Alias	% Identity	E-value
VNG0376G	<i>ftsZ1</i>	46	1e-82
VNG0192G	<i>ftsZ2</i>	43	3e-72
VNG1933G	<i>ftsZ3</i>	35	6e-05
VNG0265G	<i>ftsZ4</i>	30	0.061
VNG6260G	<i>ftsZ5</i>	30	1.6

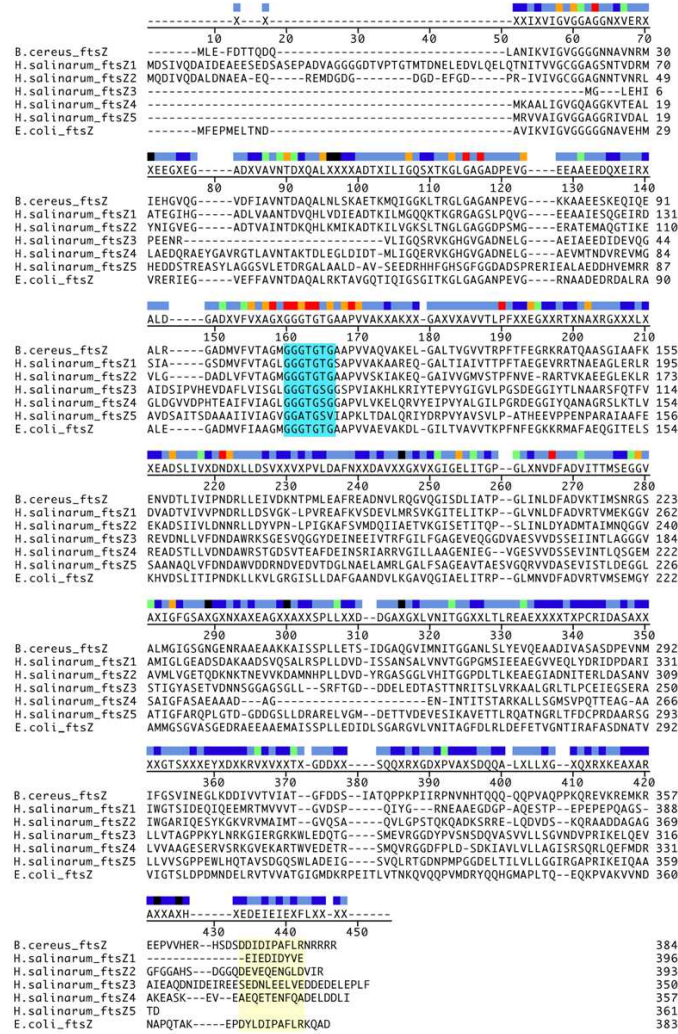
We performed an alignment of *H. salinarum* FtsZ amino acid sequences with full length FtsZ sequences from *B. cereus* and *E. coli*. Distinct from the bacterial proteins, FtsZ1 and FtsZ2 appear to have an extended N-terminal domain, although this may be due to spurious automated annotation. Indeed, much of the N-terminal amino acid sequence of FtsZ1 was not detected by proteomics experiments (Dulmage et al., 2015); however, the complete amino acid sequence of the N-terminal portion of FtsZ2 was detected. Because the initial amino acids of FtsZ1 and FtsZ2 are nearly identical, this suggests that the extended N-terminal domain of FtsZ1 is conserved and might indeed encode protein, despite our inability to detect it in untargeted proteomics experiments (Dulmage et al., 2015). Among the proteins represented here, there is very little conservation of individual residues. Regions of high homology are confined to the tubulin signature motif, which is required for binding of GTP and confirms the membership of these proteins as part of the tubulin family. Interestingly, although the

tubulin signature motif is conserved between the FtsZ amino acid sequences, a threonine residue present in the bacterial FtsZs as well as *H. salinarum* FtsZ1 and FtsZ2 is instead a serine residue at this locus in FtsZ3-5. This serine is a common feature of eukaryotic tubulin proteins. We also examined conservation of C-terminal residues required for recruitment of FtsZ to the septum through binding of ZipA (Fig. 1A, highlighted in yellow); these residues are not conserved between Bacteria and *H. salinarum*.

The loss of conservation at the C-terminus seen in all *H. salinarum* FtsZ proteins is unsurprising, as the function of this region in bacterial proteins is to bind to FtsZA followed by ZipA, which are not found in *H. salinarum*. Despite the loss of conservation at the C-terminus, a SepF structural homolog has been detected in the genome of *H. salinarum* (Makarova et al. 2010). This protein is a functional homolog of FtsA in *B. subtilis* and both targets FtsZ to the membrane of the cell and serves to bundle filaments (Duman et al., 2013, Makarova et al. 2010); it remains to be seen whether it serves similar functions in *H. salinarum*.

Generation of a phylogram from the alignment of Halobacterial FtsZs with those of *B. cereus* and *Escherichia coli* shows that these proteins are highly divergent from their bacterial counterparts (Fig. 1B). Significantly, only FtsZ1 and FtsZ2 group with the bacterial FtsZ genes, while FtsZ3-5 form their own outgroup. Together, while alignment of these proteins confirms that the *H. salinarum* FtsZs are likely to be members of the FtsZ/tubulin family, FtsZ3-5 are diverged and may serve some role outside of Z ring formation and cell division.

A



B

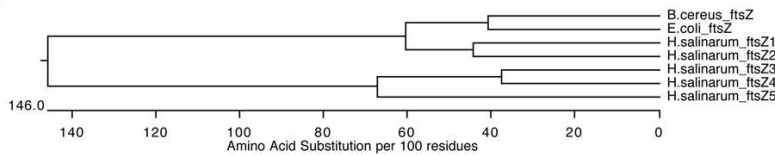


Figure 1: A) Alignment of FtsZ protein sequences from *Halobacterium salinarum* and bacterial species *Bacillus cereus* and *Escherichia coli*. Bars of color above the sequence represent the strength of conservation at a residue, with blue being the lowest degree of conservation and red being the highest. Shown above the alignment is the consensus sequence, where dashes indicate no consensus or gaps in the alignment, and X's represent non-conserved residues. The blue highlighted region indicates the tubulin signature motif, while the yellow highlighted region indicates the bacterial conserved region required for binding with ZipA. B) Phenogram of sequence conservation between FtsZ proteins, showing that while FtsZ1 and FtsZ2 group with their bacterial counterparts, FtsZ3, FtsZ4, and FtsZ5 are divergent.

Dynamics of FtsZ and MinD homolog expression in H. salinarum:

Comparison of FtsZ1-smVenus and smVenus-MinD2 fluorescent signals showed clear differences not only in localization but also in the overall level of fluorescence intensity. Total cellular fluorescence over background was calculated for populations of log phase cells with smVenus-tagged proteins. We found that the fluorescence signal from MinD2 strains is much lower than that of FtsZ1 strains (approximately 100 fold), suggesting that there is a higher abundance of FtsZ1 protein in the cell (Fig. 2). Observed intensity differences hold true even when comparing overexpression strains; this is significant because the smVenus fusion proteins are all under the control of the same promoter; suggesting that these differences are due to differential protein stability. The variability of fluorescence intensity between cells is very low in MinD2 and FtsZ1 strains, indicative of a constant level of protein under tight regulation. FtsZ2 shows the greatest variability – in an asynchronous population this could be attributed to the behavior of a protein that is transiently upregulated. Importantly, intensity is consistent between strains in which the fluorophore is integrated in the genome or expressed from a plasmid, indicating that although differences in localization due to disruption of protein function are observed, overall protein abundance is largely unchanged.

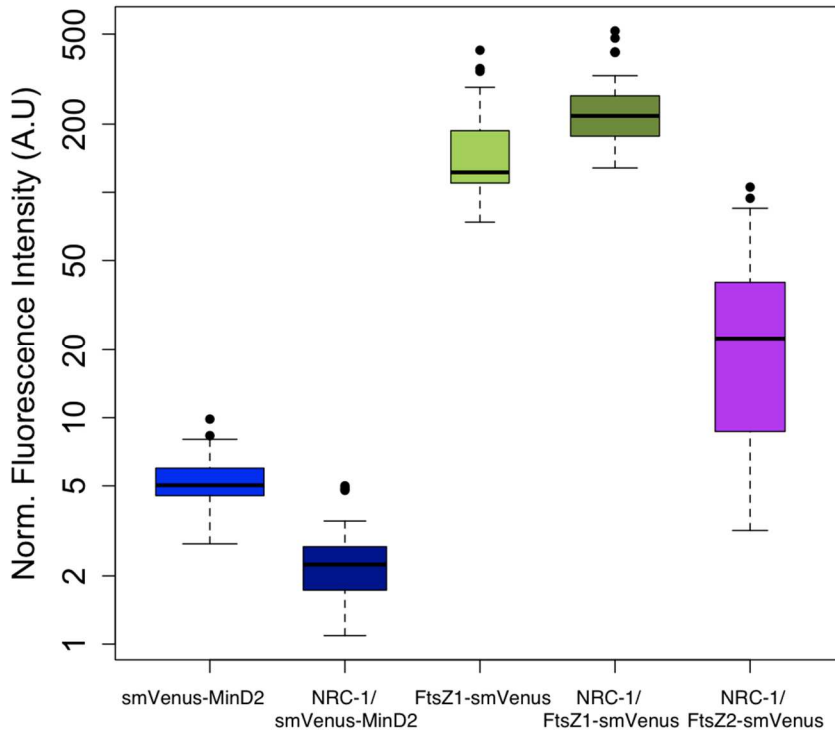


Figure 2: Fluorescence intensity of Venus-tagged cell division proteins MinD2, FtsZ1, and FtsZ2 in logarithmic phase cells. Boxplot width is proportional to the number of cells quantitated.

While we do not have long-term dynamic fluorescence intensity data for the FtsZ1 and MinD2 strains, we can infer general patterns of activation and repression from previously published gene expression data, looking for growth phase-specific activity. We hypothesize that those genes required for cell division must be more strongly expressed just before or during mid-log phase, when most cells are undergoing division. Of the 5 putative FtsZ homologs in the genome of *H. salinarum*, only two genes, *ftsZ1* and *ftsZ3*, show a marked preference for activation during log phase (Fig. 3A). *FtsZ2* is upregulated during transition from logarithmic to stationary growth. In contrast, *ftsZ4*

and *ftsZ5* have a very narrow dynamic range and show only minor upregulation during stationary phase (Fig. 3A). *MinD1* and *minD2* are also upregulated in stationary phase and, as shown in Figure 14B, have correlated expression profiles with *ftsZ4* and *ftsZ5*. The upregulation of these genes in stationary phase suggests that they are not required at high levels for cell division to occur, as the number of actively dividing cells decreases significantly during this stage of growth. Strikingly, expression of these 4 genes is anticorrelated with that of *ftsZ1* and *ftsZ3* (Fig. 3B). While these data suggest that only FtsZ1 and FtsZ3 are required for the initiation of cell division, gene expression does not necessarily correspond to protein abundance and will need to be confirmed at the level of protein activity.

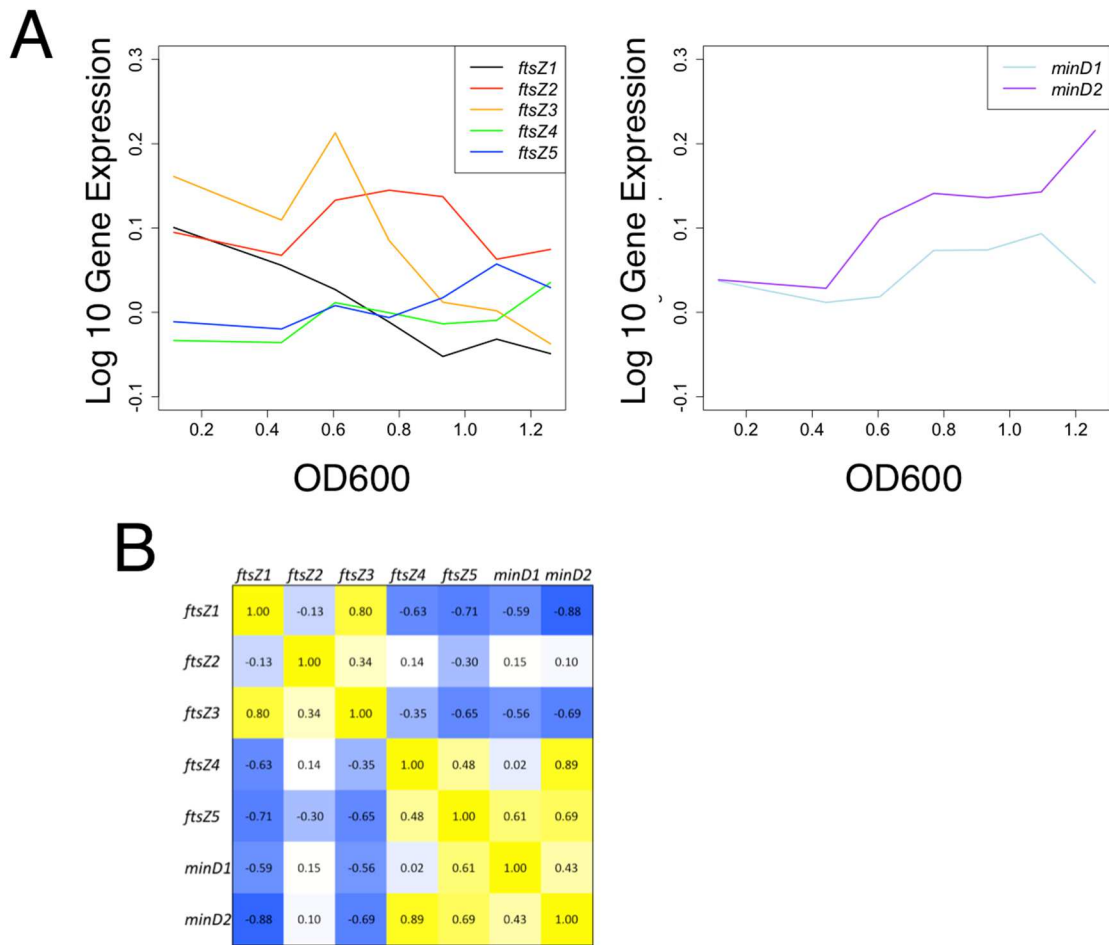


Figure 3: A) Gene expression dynamics of *ftsZ* and *minD* homologs over growth. B) Pearson correlation of gene expression profiles for *ftsZ* and *minD* homologs. Yellow boxes represent positive correlations; blue boxes represent negative correlations.

Appendix E

Supplementary data for Chapter 2: Summary of 1154 microarray experiments

A) Stress conditions and environmental perturbations:

- Metal stress including: cobalt, copper, iron, manganese, nickel, and zinc
- Oxidative stress including: gamma radiation, hydrogen peroxide, Paraquat, and ultraviolet radiation
- Light vs. dark
- Additional environmental factors: glucose, uracil, EDTA chelation, water-bath vs. air-incubation

B) Oxygen conditions:

Oxygen levels	# of arrays
Standard	973
High	67
High to low	8
Low	52
Low to high	8
Stress	46

C) Phases of Growth

Growth Phase	# of arrays
Early log	83
Mid log	960
Late Log	58
Stationary	53

D) Overexpression mutant strains:

Strain	# of arrays
No plasmid	1028
Empty vector	24
<i>hpyA</i>	6
<i>tbpA-cmyc</i>	6
<i>tbpB-cmyc</i>	6
<i>tbpC-cmyc</i>	6
<i>tbpD-cmyc</i>	6
<i>tbpE-cmyc</i>	6
<i>tbpF-cmyc</i>	6
<i>tfbA-cmyc</i>	7
<i>tfbB-cmyc</i>	8
<i>tfbC-cmyc</i>	6
<i>tfbD-cmyc</i>	12
<i>tfbE-cmyc</i>	12
<i>tfbF-cmyc</i>	8
<i>tfbG-cmyc</i>	7

E) Knock-out mutant strains:

VNG0750C 4
zim 3

Strain	# of arrays
<i>NRC-1 (wild-type)</i>	696
<i>afsQ2</i>	4
<i>ark</i>	4
<i>bat</i>	6
<i>boa1</i>	4
<i>boa4</i>	4
<i>bop</i>	3
<i>cspD1</i>	3
<i>gap</i>	4
<i>hpyA</i>	12
<i>htlD</i>	4
<i>htr1</i>	4
<i>htr2</i>	4
<i>htr8</i>	4
<i>idr1</i>	14
<i>idr1idr2</i>	21
<i>idr2</i>	30
<i>kinA2</i>	6
<i>phr1</i>	13
<i>phr1 phr2</i>	3
<i>phr2</i>	3
<i>sirR</i>	27
<i>sop2</i>	8
<i>tbpA</i>	7
<i>tbpC</i>	7
<i>tbpD</i>	7
<i>tfbA</i>	7
<i>tfbB</i>	7
<i>tfbC</i>	7
<i>tfbD</i>	6
<i>tfbE</i>	6
<i>troR</i>	8
<i>ura3</i>	134
VNG0258H (<i>RosR</i>)	45
VNG1179C	6
VNG1296C	3
VNG1451C (<i>trmB</i>)	11
VNG0019H	4
VNG2334C	1

F) References

Publication	# of Arrays
Baliga et al., 2004, Genome Research 14(6):1025-34	7
Bonneau et al., 2007, Cell 131(7):1354-64	160
Facciotti et al., 2007, PNAS 104(11):4630-5	158
Facciotti et al., 2010, BMC Systems Biology 4:64	46
Kaur et al., 2006, Genome Research 16(7):841-53	100
Kaur et al., 2010, Molecular Systems Biology 6:393	120
Schmid et al., 2007, Genome Research 17(10):1399-412	58
Schmid et al., 2009, Molecular Systems Biology 5:281	23
Schmid et al., 2011, Nucleic Acids Research 39(7):2519-32	41
Sharma et al., 2012, BMC Genomics 13:351	79
Unpublished	190
Whitehead et al., 2006, Molecular Systems Biology 2:47	29
Whitehead et al., 2009, PLoS One 4(5):e5484	139

Appendix F

Supplementary data for Chapter 2: ChIP-chip arrays

A) Microarray metadata

Strain	Biological Replicate	Growth Phase	Condition	Publication
<i>NRC-1/cmyc empty vector</i>	A	Stationary	Standard (Complex medium)	Ref 1
<i>NRC-1/cmyc empty vector</i>	B	Stationary	Standard	Ref 1
<i>NRC-1/cmyc empty vector</i>	C	Stationary	Standard	Ref 1
<i>NRC-1/hrg-cmyc</i>	B	Stationary	Standard	Unpublished
<i>NRC-1/hrg-cmyc</i>	A	Stationary	Standard	Unpublished
<i>NRC-1/idr2-cmyc</i>	C	Logarithmic	100uM iron	Ref 2
<i>NRC-1/idr2-cmyc</i>	D	Logarithmic	100uM iron	Ref 2
<i>NRC-1/idr2-cmyc</i>	B	Logarithmic	No iron added	Ref 2
<i>NRC-1/idr2-cmyc</i>	B	Logarithmic	100uM iron	Ref 2
<i>NRC-1/idr2-cmyc</i>	A	Logarithmic	100uM iron	Ref 2
<i>NRC-1/idr2-cmyc</i>	E	Logarithmic	100uM iron, Defined medium	Unpublished
<i>NRC-1/idr2-cmyc</i>	A	Logarithmic	No iron added	Ref 2
<i>NRC-1/idr2-cmyc</i>	F	Logarithmic	100uM iron	Ref 2
<i>NRC-1/idr2-cmyc</i>	C	Logarithmic	No iron added	Ref 2
<i>NRC-1/idr2-cmyc</i>	A	Logarithmic	500uM iron	Unpublished
<i>NRC-1/sirR-cmyc</i>	A	Stationary	0.8mM Manganese	Unpublished
<i>NRC-1/sirR-cmyc</i>	B	Stationary	0.4mM Manganese	Unpublished
<i>NRC-1/sirR-cmyc</i>	B	Stationary	0.8mM Manganese	Unpublished
<i>NRC-1/sirR-cmyc</i>	A	Stationary	Standard	Unpublished
<i>NRC-1/sirR-cmyc</i>	B	Stationary	Standard	Unpublished
<i>NRC-1/sirR-cmyc</i>	A	Stationary	Standard	Unpublished
<i>NRC-1/tbp-cmyc</i>	A	Stationary	Standard	Ref 1
<i>NRC-1/tbp-cmyc</i>	B	Stationary	Standard	Ref 1
<i>NRC-1/tbp-cmyc</i>	C	Stationary	Standard	Ref 1
<i>NRC-1/tbp-cmyc</i>	D	Stationary	Standard	Ref 1
<i>NRC1/tfbA-cmyc</i>	A	Stationary	Standard	Ref 1
<i>NRC1/tfbA-cmyc</i>	B	Stationary	Standard	Ref 1

<i>NRC-1/tfbD-cmyc</i>	A	Stationary	Standard	Unpublished
<i>NRC-1/tfbD-cmyc</i>	B	Stationary	Standard	Unpublished
<i>NRC-1/tfbD-cmyc</i>	A	Stationary	Standard	Unpublished
<i>NRC-1/trh4-cmyc</i>	B	Stationary	Standard	Unpublished
<i>NRC-1/trh4-cmyc</i>	A	Stationary	Standard	Unpublished
<i>NRC-1/trmB-cmyc</i>	C	Stationary	Standard	Ref 3
<i>NRC-1/trmB-cmyc</i>	A	Stationary	0.1% glucose	Ref 3
<i>NRC-1/trmB-cmyc</i>	B	Logarithmic	0.1% glucose	Ref 3
<i>NRC-1/trmB-cmyc</i>	A	Logarithmic	0.1% glucose	Ref 3
<i>NRC-1/trmB-cmyc</i>	D	Stationary	Standard	Ref 3
<i>NRC-1/trmB-cmyc</i>	E	Stationary	Standard	Ref 3
<i>NRC-1/trmB-cmyc</i>	B	Stationary	Standard	Ref 3
<i>NRC-1/trmB-cmyc</i>	A	Stationary	Standard	Ref 3
<i>NRC-1/trmB-cmyc</i>	F	Stationary	Standard	Ref 3
<i>NRC-1/trmB-cmyc</i>	B	Stationary	0.1% glucose	Ref 3
<i>NRC-1/trmB-cmyc</i>	A	Stationary	1ml/L glycerol	Ref 3
<i>NRC-1/trmB-cmyc</i>	C	Stationary	Standard	Ref 3
<i>NRC-1/trmB-cmyc</i>	B	Stationary	10% glucose	Ref 3
<i>NRC-1/trmB-cmyc</i>	A	Stationary	10% glucose	Ref 3
<i>NRC-1/trmB-cmyc</i>	B	Stationary	Standard	Ref 3
<i>NRC-1/trmB-cmyc</i>	A	Stationary	Standard	Ref 3

B) References

1. Facciotti et al., 2007, PNAS 104(11):4630-5
2. Schmid et al., 2011, Nucleic Acids Research 39(7):2519-32
3. Schmid et al., 2009, Molecular Systems Biology 5:281

References

1. **Adams DW, Errington J.** 2009. Bacterial cell division: assembly, maintenance and disassembly of the Z ring. *Nat Rev Microbiol* **7**:642-653.
2. **Adler HI, Fisher WD, Cohen A, Hardigree AA.** 1967. MINIATURE *Escherichia coli* CELLS DEFICIENT IN DNA. *Proc Natl Acad Sci U S A* **57**:321-326.
3. **Alam M, Oesterhelt D.** 1984. Morphology, function and isolation of halobacterial flagella. *J Mol Biol* **176**:459-475.
4. **Allers T, Mevarech M.** 2005. Archaeal genetics - the third way. *Nat Rev Genet* **6**:58-73.
5. **Altman-Price N, Mevarech M.** 2009. Genetic evidence for the importance of protein acetylation and protein deacetylation in the halophilic archaeon *Haloferax volcanii*. *J Bacteriol* **191**:1610-1617.
6. **Ammar R, Torti D, Tsui K, Gebbia M, Durbic T, Bader GD, Giaever G, Nislow C.** 2012. Chromatin is an ancient innovation conserved between Archaea and Eukarya. *Elife* **1**:e00078.
7. **Annunziato A.** 2008. DNA Packaging: Nucleosomes and Chromatin. *Nature Education* **1**:26.
8. **Arber W.** 2000. Genetic variation: molecular mechanisms and impact on microbial evolution. *FEMS Microbiol Rev* **24**:1-7.
9. **Arents G, Moudrianakis EN.** 1995. The histone fold: a ubiquitous architectural motif utilized in DNA compaction and protein dimerization. *Proc Natl Acad Sci U S A* **92**:11170-11174.
10. **Atomi H, Imanaka T, Fukui T.** 2012. Overview of the genetic tools in the Archaea. *Front Microbiol* **3**:337.
11. **Azam TA, Ishihama A.** 1999. Twelve species of the nucleoid-associated protein from *Escherichia coli*. Sequence recognition specificity and DNA binding affinity. *J Biol Chem* **274**:33105-33113.
12. **Bach JN, Albrecht N, Bramkamp M.** 2014. Imaging DivIVA dynamics using

photo-convertible and activatable fluorophores in *Bacillus subtilis*. *Front Microbiol* **5**:59.

13. **Baliga NS, Bjork SJ, Bonneau R, Pan M, Iloanusi C, Kottemann MC, Hood L, DiRuggiero J.** 2004. Systems level insights into the stress response to UV radiation in the halophilic archaeon *Halobacterium NRC-1*. *Genome Res* **14**:1025-1035.
14. **Banfalvi G.** 2011. Overview of cell synchronization. *Methods Mol Biol* **761**:1-23.
15. **Becker EA, Seitzer PM, Tritt A, Larsen D, Krusor M, Yao AI, Wu D, Madern D, Eisen JA, Darling AE, Facciotti MT.** 2014. Phylogenetically driven sequencing of extremely halophilic archaea reveals strategies for static and dynamic osmo-response. *PLoS Genet* **10**:e1004784.
16. **Bell SD.** 2005. Archaeal transcriptional regulation--variation on a bacterial theme? *Trends Microbiol* **13**:262-265.
17. **Bermudez I, Garcia-Martinez J, Perez-Ortin JE, Roca J.** 2010. A method for genome-wide analysis of DNA helical tension by means of psoralen-DNA photobinding. *Nucleic Acids Res* **38**:e182.
18. **Betzig E, Patterson GH, Sougrat R, Lindwasser OW, Olenych S, Bonifacino JS, Davidson MW, Lippincott-Schwartz J, Hess HF.** 2006. Imaging intracellular fluorescent proteins at nanometer resolution. *Science* **313**:1642-1645.
19. **Bhutta MS, McNerny CJ, Gould GW.** 2014. ESCRT function in cytokinesis: location, dynamics and regulation by mitotic kinases. *Int J Mol Sci* **15**:21723-21739.
20. **Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, Kiefer F, Cassarino TG, Bertoni M, Bordoli L, Schwede T.** 2014. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res* **42**:W252-258.
21. **Bisler B, Schmidt A, Scheibe B, Klein C, Tebbe A, Kellermann J, Siedler F, Pfeiffer F, Lottspeich F, Oesterhelt D.** 2006. Quantitative profiling of the membrane proteome in a halophilic archaeon. *Mol Cell Proteomics* **5**:1543-1558.
22. **Blonder J, Conrads TP, Yu LR, Terunuma A, Janini GM, Issaq HJ, Vogel JC,**

- Veenstra TD.** 2004. A detergent- and cyanogen bromide-free method for integral membrane proteomics: application to *Halobacterium* purple membranes and the human epidermal membrane proteome. *Proteomics* **4**:31-45.
23. **Breuert S, Allers T, Spohn G, Soppa J.** 2006. Regulated polyploidy in halophilic archaea. *PLoS One* **1**:e92.
24. **Brugger K, Redder P, She Q, Confalonieri F, Zivanovic Y, Garrett RA.** 2002. Mobile elements in archaeal genomes. *FEMS Microbiol Lett* **206**:131-141.
25. **Burns DG, Janssen PH, Itoh T, Kamekura M, Li Z, Jensen G, Rodriguez-Valera F, Bolhuis H, Dyall-Smith ML.** 2007. *Haloquadratum walsbyi* gen. nov., sp. nov., the square haloarchaeon of Walsby, isolated from saltern crystallizers in Australia and Spain. *Int J Syst Evol Microbiol* **57**:387-392.
26. **Cantone I, Fisher AG.** 2013. Epigenetic programming and reprogramming during development. *Nat Struct Mol Biol* **20**:282-289.
27. **Ceribelli M, Dolfini D, Merico D, Gatta R, Vigano AM, Pavesi G, Mantovani R.** 2008. The histone-like NF-Y is a bifunctional transcription factor. *Mol Cell Biol* **28**:2047-2058.
28. **Chang F, Huang KC.** 2014. How and why cells grow as rods. *BMC Biol* **12**:54.
29. **Chartier F, Laine B, Belaiche D, Touzel JP, Sautiere P.** 1989. Primary structure of the chromosomal protein MC1 from the archaeobacterium *Methanosarcina* sp. *CHTI 55*. *Biochim Biophys Acta* **1008**:309-314.
30. **Clark DJ.** 2010. Nucleosome positioning, nucleosome spacing and the nucleosome code. *J Biomol Struct Dyn* **27**:781-793.
31. **Cohen-Rosenzweig C, Yurist-Doutsch S, Eichler J.** 2012. AglS, a novel component of the *Haloferax volcanii* N-glycosylation pathway, is a dolichol phosphate-mannose mannosyltransferase. *J Bacteriol* **194**:6909-6916.
32. **Coker JA, DasSarma P, Capes M, Wallace T, McGarrity K, Gessler R, Liu J, Xiang H, Tatusov R, Berquist BR, DasSarma S.** 2009. Multiple replication origins of *Halobacterium* sp. strain *NRC-1*: properties of the conserved *orc7*-dependent *oriC1*. *J Bacteriol* **191**:5253-5261.

33. **Cosgrove MS, Boeke JD, Wolberger C.** 2004. Regulated nucleosome mobility and the histone code. *Nat Struct Mol Biol* **11**:1037-1043.
34. **Crooks GE, Hon G, Chandonia JM, Brenner SE.** 2004. WebLogo: a sequence logo generator. *Genome Res* **14**:1188-1190.
35. **Cubonova L, Katano M, Kanai T, Atomi H, Reeve JN, Santangelo TJ.** 2012. An archaeal histone is required for transformation of *Thermococcus kodakarensis*. *J Bacteriol* **194**:6864-6874.
36. **Dame RT, Kalmykova OJ, Grainger DC.** 2011. Chromosomal macrodomains and associated proteins: implications for DNA organization and replication in gram negative bacteria. *PLoS Genet* **7**:e1002123.
37. **Darmon E, Leach DR.** 2014. Bacterial genome instability. *Microbiol Mol Biol Rev* **78**:1-39.
38. **DasSarma S.** 1993. Identification and analysis of the gas vesicle gene cluster on an unstable plasmid of *Halobacterium halobium*. *Experientia* **49**:482-486.
39. **DasSarma S, Berquist BR, Coker JA, DasSarma P, Muller JA.** 2006. Post-genomics of the model haloarchaeon *Halobacterium sp. NRC-1*. *Saline Systems* **2**:3.
40. **de Boer P, Crossley R, Rothfield L.** 1992. The essential bacterial cell-division protein FtsZ is a GTPase. *Nature* **359**:254-256.
41. **Decanniere K, Babu AM, Sandman K, Reeve JN, Heinemann U.** 2000. Crystal structures of recombinant histones HMfA and HMfB from the hyperthermophilic archaeon *Methanothermus fervidus*. *J Mol Biol* **303**:35-47.
42. **Dekker J, Rippe K, Dekker M, Kleckner N.** 2002. Capturing chromosome conformation. *Science* **295**:1306-1311.
43. **DeLange RJ, Green GR, Searcy DG.** 1981. A histone-like protein (HTa) from *Thermoplasma acidophilum*. I. Purification and properties. *J Biol Chem* **256**:900-904.
44. **Delius H, Worcel A.** 1974. Letter: Electron microscopic visualization of the folded chromosome of *Escherichia coli*. *J Mol Biol* **82**:107-109.
45. **Duggin IG, Aylett CH, Walsh JC, Michie KA, Wang Q, Turnbull L, Dawson**

- EM, Harry EJ, Whitchurch CB, Amos LA, Lowe J.** 2015. CetZ tubulin-like proteins control archaeal cell shape. *Nature* **519**:362-365.
46. **Dulmage KA, Todor H, Schmid AK.** 2015. Growth-Phase-Specific Modulation of Cell Morphology and Gene Expression by an Archaeal Histone Protein. *MBio* **6**.
47. **Duman R, Ishikawa S, Celik I, Strahl H, Ogasawara N, Troc P, Lowe J, Hamoen LW.** 2013. Structural and genetic analyses reveal the protein SepF as a new membrane anchor for the Z ring. *Proc Natl Acad Sci U S A* **110**:E4601-4610.
48. **Edgar R, Domrachev M, Lash AE.** 2002. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* **30**:207-210.
49. **Eickbush TH, Moudrianakis EN.** 1978. The compaction of DNA helices into either continuous supercoils or folded-fiber rods and toroids. *Cell* **13**:295-306.
50. **Erickson HP, Anderson DE, Osawa M.** 2010. FtsZ in bacterial cytokinesis: cytoskeleton and force generator all in one. *Microbiol Mol Biol Rev* **74**:504-528.
51. **Ettema TJ, Bernander R.** 2009. Cell division and the ESCRT complex: A surprise from the archaea. *Commun Integr Biol* **2**:86-88.
52. **Eun YJ, Kapoor M, Hussain S, Garner EC.** 2015. Bacterial Filament Systems: Toward Understanding Their Emergent Behavior and Cellular Functions. *J Biol Chem* **290**:17181-17189.
53. **Facciotti MT, Pang WL, Lo FY, Whitehead K, Koide T, Masumura K, Pan M, Kaur A, Larsen DJ, Reiss DJ, Hoang L, Kalisiak E, Northen T, Trauger SA, Siuzdak G, Baliga NS.** 2010. Large scale physiological readjustment during growth enables rapid, comprehensive and inexpensive systems analysis. *BMC Syst Biol* **4**:64.
54. **Fahrner RL, Cascio D, Lake JA, Slesarev A.** 2001. An ancestral nuclear protein assembly: crystal structure of the *Methanopyrus kandleri* histone. *Protein Sci* **10**:2002-2007.
55. **Fox GE, Magrum LJ, Balch WE, Wolfe RS, Woese CR.** 1977. Classification of methanogenic bacteria by 16S ribosomal RNA characterization. *Proc Natl Acad Sci U S A* **74**:4537-4541.

56. **Gibson DG.** 2011. Enzymatic assembly of overlapping DNA fragments. *Methods Enzymol* **498**:349-361.
57. **Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA, 3rd, Smith HO.** 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* **6**:343-345.
58. **Greer EL, Shi Y.** 2012. Histone methylation: a dynamic mark in health, disease and inheritance. *Nat Rev Genet* **13**:343-357.
59. **Gregory JA, Becker EC, Pogliano K.** 2008. *Bacillus subtilis* MinC destabilizes FtsZ-rings at new cell poles and contributes to the timing of cell division. *Genes Dev* **22**:3475-3488.
60. **Grogono-Thomas R, Blaser MJ, Ahmadi M, Newell DG.** 2003. Role of S-layer protein antigenic diversity in the immune responses of sheep experimentally challenged with *Campylobacter fetus subsp. fetus*. *Infect Immun* **71**:147-154.
61. **Grohmann D, Werner F.** 2011. Recent advances in the understanding of archaeal transcription. *Curr Opin Microbiol* **14**:328-334.
62. **Guan Z, Naparstek S, Kaminski L, Konrad Z, Eichler J.** 2010. Distinct glycan-charged phosphodolichol carriers are required for the assembly of the pentasaccharide N-linked to the *Haloferax volcanii* S-layer glycoprotein. *Mol Microbiol* **78**:1294-1303.
63. **Guex N, Peitsch MC, Schwede T.** 2009. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: a historical perspective. *Electrophoresis* **30 Suppl 1**:S162-173.
64. **Gustafsson MG.** 2000. Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy. *J Microsc* **198**:82-87.
65. **Guy L, Ettema TJ.** 2011. The archaeal 'TACK' superphylum and the origin of eukaryotes. *Trends Microbiol* **19**:580-587.
66. **Harrison FC, Kennedy ME.** 1922. The red discolouration of cured codfish. F. A. Acland, printer, Ottawa.

67. **Hedges DJ, Deininger PL.** 2007. Inviting instability: Transposable elements, double-strand breaks, and the maintenance of genome integrity. *Mutat Res* **616**:46-59.
68. **Herrmann U, Soppa J.** 2002. Cell cycle-dependent expression of an essential SMC-like protein and dynamic chromosome localization in the archaeon *Halobacterium salinarum*. *Mol Microbiol* **46**:395-409.
69. **Heyer WD.** 2015. Regulation of Recombination and Genomic Maintenance. *Cold Spring Harb Perspect Biol* **7**.
70. **Horn C, Paulmann B, Kerlen G, Junker N, Huber H.** 1999. In vivo observation of cell division of anaerobic hyperthermophiles by using a high-intensity dark-field microscope. *J Bacteriol* **181**:5114-5118.
71. **Jafri S, Evoy S, Cho K, Craighead HG, Winans SC.** 1999. An Lrp-type transcriptional regulator from *Agrobacterium tumefaciens* condenses more than 100 nucleotides of DNA into globular nucleoprotein complexes. *J Mol Biol* **288**:811-824.
72. **Jarrell KF, Sprott GD.** 1984. Formation and Regeneration of *Halobacterium* Spheroplasts. *Current Microbiology* **10**:147-152.
73. **Kahramanoglou C, Seshasayee AS, Prieto AI, Ibberson D, Schmidt S, Zimmermann J, Benes V, Fraser GM, Luscombe NM.** 2011. Direct and indirect effects of H-NS and Fis on global gene expression control in *Escherichia coli*. *Nucleic Acids Res* **39**:2073-2091.
74. **Keller A, Nesvizhskii AI, Kolker E, Aebersold R.** 2002. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal Chem* **74**:5383-5392.
75. **Kennedy SP, Ng WV, Salzberg SL, Hood L, DasSarma S.** 2001. Understanding the adaptation of *Halobacterium species NRC-1* to its extreme environment through computational analysis of its genome sequence. *Genome Res* **11**:1641-1650.
76. **Kiljunen S, Pajunen MI, Dilks K, Storf S, Pohlschroder M, Savilahti H.** 2014. Generation of comprehensive transposon insertion mutant library for the model archaeon, *Haloferax volcanii*, and its use for gene discovery. *BMC Biol* **12**:103.

77. **Klein C, Aivaliotis M, Olsen JV, Falb M, Besir H, Scheffer B, Bisle B, Tebbe A, Konstantinidis K, Siedler F, Pfeiffer F, Mann M, Oesterhelt D.** 2007. The low molecular weight proteome of *Halobacterium salinarum*. *J Proteome Res* **6**:1510-1518.
78. **Klein C, Garcia-Rizo C, Bisle B, Scheffer B, Zischka H, Pfeiffer F, Siedler F, Oesterhelt D.** 2005. The membrane proteome of *Halobacterium salinarum*. *Proteomics* **5**:180-197.
79. **Koide T, Reiss DJ, Bare JC, Pang WL, Facciotti MT, Schmid AK, Pan M, Marzolf B, Van PT, Lo FY, Pratap A, Deutsch EW, Peterson A, Martin D, Baliga NS.** 2009. Prevalence of transcription promoters within archaeal operons and coding sequences. *Mol Syst Biol* **5**:285.
80. **Kolodner RD, Putnam CD, Myung K.** 2002. Maintenance of genome stability in *Saccharomyces cerevisiae*. *Science* **297**:552-557.
81. **Kornberg RD.** 2007. The molecular basis of eukaryotic transcription. *Proc Natl Acad Sci U S A* **104**:12955-12961.
82. **Kottemann M, Kish A, Iloanusi C, Bjork S, DiRuggiero J.** 2005. Physiological responses of the halophilic archaeon *Halobacterium sp. strain NRC1* to desiccation and gamma irradiation. *Extremophiles* **9**:219-227.
83. **Laine B, Chartier F, Imbert M, Lewis R, Sautiere P.** 1986. Primary structure of the chromosomal protein HMB from the archaeobacteria *Methanosarcina barkeri*. *Eur J Biochem* **161**:681-687.
84. **Lange C, Zerulla K, Breuert S, Soppa J.** 2011. Gene conversion results in the equalization of genome copies in the polyploid haloarchaeon *Haloferax volcanii*. *Mol Microbiol* **80**:666-677.
85. **Lindas AC, Bernander R.** 2013. The cell cycle of archaea. *Nat Rev Microbiol* **11**:627-638.
86. **Livak KJ, Schmittgen TD.** 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2⁻(Delta Delta C(T)) Method. *Methods* **25**:402-408.
87. **Luijsterburg MS, Noom MC, Wuite GJ, Dame RT.** 2006. The architectural role

of nucleoid-associated proteins in the organization of bacterial chromatin: a molecular perspective. *J Struct Biol* **156**:262-272.

88. **Luijsterburg MS, White MF, van Driel R, Dame RT.** 2008. The major architects of chromatin: architectural proteins in bacteria, archaea and eukaryotes. *Crit Rev Biochem Mol Biol* **43**:393-418.
89. **Mackwan RR, Carver GT, Drake JW, Grogan DW.** 2007. An unusual pattern of spontaneous mutations recovered in the halophilic archaeon *Haloferax volcanii*. *Genetics* **176**:697-702.
90. **Madren SM, Hoffman MD, Brown PJ, Kysela DT, Brun YV, Jacobson SC.** 2012. Microfluidic device for automated synchronization of bacterial cells. *Anal Chem* **84**:8571-8578.
91. **Makarova KS, Yutin N, Bell SD, Koonin EV.** 2010. Evolution of diverse cell division and vesicle formation systems in Archaea. *Nat Rev Microbiol* **8**:731-741.
92. **Malandrin L, Huber H, Bernander R.** 1999. Nucleoid structure and partition in *Methanococcus jannaschii*: an archaeon with multiple copies of the chromosome. *Genetics* **152**:1315-1323.
93. **Malik HS, Henikoff S.** 2003. Phylogenomics of the nucleosome. *Nat Struct Biol* **10**:882-891.
94. **Mancera E, Bourgon R, Brozzi A, Huber W, Steinmetz LM.** 2008. High-resolution mapping of meiotic crossovers and non-crossovers in yeast. *Nature* **454**:479-485.
95. **Marc F, Sandman K, Lurz R, Reeve JN.** 2002. Archaeal histone tetramerization determines DNA affinity and the direction of DNA supercoiling. *J Biol Chem* **277**:30879-30886.
96. **Margolin W, Wang R, Kumar M.** 1996. Isolation of an *ftsZ* homolog from the archaeobacterium *Halobacterium salinarium*: implications for the evolution of FtsZ and tubulin. *J Bacteriol* **178**:1320-1327.
97. **Marsh VL, Peak-Chew SY, Bell SD.** 2005. Sir2 and the acetyltransferase, Pat, regulate the archaeal chromatin protein, Alba. *J Biol Chem* **280**:21122-21128.

98. **Martusewitsch E, Sensen CW, Schleper C.** 2000. High spontaneous mutation rate in the hyperthermophilic archaeon *Sulfolobus solfataricus* is mediated by transposable elements. *J Bacteriol* **182**:2574-2581.
99. **Maruyama H, Harwood JC, Moore KM, Paszkiewicz K, Durley SC, Fukushima H, Atomi H, Takeyasu K, Kent NA.** 2013. An alternative beads-on-a-string chromatin architecture in *Thermococcus kodakarensis*. *EMBO Rep* **14**:711-717.
100. **Maruyama H, Shin M, Oda T, Matsumi R, Ohniwa RL, Itoh T, Shirahige K, Imanaka T, Atomi H, Yoshimura SH, Takeyasu K.** 2011. Histone and TK0471/TrmBL2 form a novel heterogeneous genome architecture in the hyperthermophilic archaeon *Thermococcus kodakarensis*. *Mol Biol Cell* **22**:386-398.
101. **Matsubara K, Sano N, Umehara T, Horikoshi M.** 2007. Global analysis of functional surfaces of core histones with comprehensive point mutants. *Genes Cells* **12**:13-33.
102. **McCready S, Muller JA, Boubriak I, Berquist BR, Ng WL, DasSarma S.** 2005. UV irradiation induces homologous recombination genes in the model archaeon, *Halobacterium sp. NRC-1*. *Saline Systems* **1**:3.
103. **Menzel R, Gellert M.** 1983. Regulation of the genes for *E. coli* DNA gyrase: homeostatic control of DNA supercoiling. *Cell* **34**:105-113.
104. **Mescher MF, Hansen U, Strominger JL.** 1976. Formation of lipid-linked sugar compounds in *Halobacterium salinarium*. Presumed intermediates in glycoprotein synthesis. *J Biol Chem* **251**:7289-7294.
105. **Mescher MF, Strominger JL.** 1978. Glycosylation of the surface glycoprotein of *Halobacterium salinarium* via a cyclic pathway of lipid-linked intermediates. *FEBS Lett* **89**:37-41.
106. **Mishra M, Huang J, Balasubramanian MK.** 2014. The yeast actin cytoskeleton. *FEMS Microbiol Rev* **38**:213-227.
107. **Nakanishi S, Sanderson BW, Delventhal KM, Bradford WD, Staehling-Hampton K, Shilatifard A.** 2008. A comprehensive library of histone mutants identifies nucleosomal residues required for H3K4 methylation. *Nat Struct Mol Biol* **15**:881-888.

108. **Naor A, Lapierre P, Mevarech M, Papke RT, Gophna U.** 2012. Low species barriers in halophilic archaea and the formation of recombinant hybrids. *Curr Biol* **22**:1444-1448.
109. **Newton IL, Bordenstein SR.** 2011. Correlations between bacterial ecology and mobile DNA. *Curr Microbiol* **62**:198-208.
110. **Ng WV, Ciufu SA, Smith TM, Bumgarner RE, Baskin D, Faust J, Hall B, Loretz C, Seto J, Slagel J, Hood L, DasSarma S.** 1998. Snapshot of a large dynamic replicon in a halophilic archaeon: megaplasmid or minichromosome? *Genome Res* **8**:1131-1141.
111. **Ng WV, Kennedy SP, Mahairas GG, Berquist B, Pan M, Shukla HD, Lasky SR, Baliga NS, Thorsson V, Sbrogna J, Swartzell S, Weir D, Hall J, Dahl TA, Welti R, Goo YA, Leithauser B, Keller K, Cruz R, Danson MJ, Hough DW, Maddocks DG, Jablonski PE, Krebs MP, Angevine CM, Dale H, Isenbarger TA, Peck RF, Pohlschroder M, Spudich JL, Jung KW, Alam M, Freitas T, Hou S, Daniels CJ, Dennis PP, Omer AD, Ebhardt H, Lowe TM, Liang P, Riley M, Hood L, DasSarma S.** 2000. Genome sequence of *Halobacterium species NRC-1*. *Proc Natl Acad Sci U S A* **97**:12176-12181.
112. **Oliverio AM, Katz LA.** 2014. The dynamic nature of genomes across the tree of life. *Genome Biol Evol* **6**:482-488.
113. **Oren A, Arahall DR, Ventosa A.** 2009. Emended descriptions of genera of the family Halobacteriaceae. *Int J Syst Evol Microbiol* **59**:637-642.
114. **Papke RT, Zhaxybayeva O, Feil EJ, Sommerfeld K, Muise D, Doolittle WF.** 2007. Searching for species in haloarchaea. *Proc Natl Acad Sci U S A* **104**:14092-14097.
115. **Pavlov NA, Cherny DI, Jovin TM, Slesarev AI.** 2002. Nucleosome-like complex of the histone from the hyperthermophile *Methanopyrus kandleri* (MkaH) with linear DNA. *J Biomol Struct Dyn* **20**:207-214.
116. **Peck RF, DasSarma S, Krebs MP.** 2000. Homologous gene knockout in the archaeon *Halobacterium salinarum* with *ura3* as a counterselectable marker. *Mol Microbiol* **35**:667-676.
117. **Peeters E, Driessen RP, Werner F, Dame RT.** 2015. The interplay between

- nucleoid organization and transcription in archaeal genomes. *Nat Rev Microbiol* **13**:333-341.
118. **Pereira SL, Grayling RA, Lurz R, Reeve JN.** 1997. Archaeal nucleosomes. *Proc Natl Acad Sci U S A* **94**:12633-12637.
 119. **Perez-Fillol M, Rodriguez-Valera F.** 1986. Potassium ion accumulation in cells of different halobacteria. *Microbiologia* **2**:73-80.
 120. **Pfeifer F, Blaseio U.** 1989. Insertion elements and deletion formation in a halophilic archaeobacterium. *J Bacteriol* **171**:5135-5140.
 121. **Pfeifer F, Blaseio U, Ghahraman P.** 1988. Dynamic plasmid populations in *Halobacterium halobium*. *J Bacteriol* **170**:3718-3724.
 122. **Pfeiffer F, Schuster SC, Broicher A, Falb M, Palm P, Rodewald K, Ruepp A, Soppa J, Tittor J, Oesterhelt D.** 2008. Evolution in the laboratory: the genome of *Halobacterium salinarum* strain *R1* compared to that of strain *NRC-1*. *Genomics* **91**:335-346.
 123. **Phillips R, Kondev J, Theriot J.** 2009. *Physical biology of the cell*. Garland Science, New York.
 124. **Plaisier CL, Lo FY, Ashworth J, Brooks AN, Beer KD, Kaur A, Pan M, Reiss DJ, Facciotti MT, Baliga NS.** 2014. Evolution of context dependent regulation by expansion of feast/famine regulatory proteins. *BMC Syst Biol* **8**:122.
 125. **Poplawski A, Gullbrand B, Bernander R.** 2000. The *ftsZ* gene of *Haloferax mediterranei*: sequence, conserved gene order, and visualization of the FtsZ ring. *Gene* **242**:357-367.
 126. **Porter K, Russ BE, Dyall-Smith ML.** 2007. Virus-host interactions in salt lakes. *Curr Opin Microbiol* **10**:418-424.
 127. **Postow L, Hardy CD, Arsuaga J, Cozzarelli NR.** 2004. Topological domain structure of the *Escherichia coli* chromosome. *Genes Dev* **18**:1766-1779.
 128. **Prieto AI, Kahramanoglou C, Ali RM, Fraser GM, Seshasayee AS, Luscombe NM.** 2012. Genomic analysis of DNA binding and gene regulation by homologous nucleoid-associated proteins IHF and HU in *Escherichia coli* K12.

Nucleic Acids Res **40**:3524-3537.

129. **Ramakrishnan V.** 1997. Histone structure and the organization of the nucleosome. *Annu Rev Biophys Biomol Struct* **26**:83-112.
130. **Ramamurthi KS, Losick R.** 2009. Negative membrane curvature as a cue for subcellular localization of a bacterial protein. *Proc Natl Acad Sci U S A* **106**:13541-13545.
131. **Raskin DM, de Boer PA.** 1999. [1] Rapid pole-to-pole oscillation of a protein required for directing division to the middle of *Escherichia coli*. *Proc Natl Acad Sci U S A* **96**:4971-4976.
132. **Raskin DM, de Boer PA.** 1999. [2] MinDE-dependent pole-to-pole oscillation of division inhibitor MinC in *Escherichia coli*. *J Bacteriol* **181**:6419-6424.
133. **Redder P, Garrett RA.** 2006. Mutations and rearrangements in the genome of *Sulfolobus solfataricus* P2. *J Bacteriol* **188**:4198-4206.
134. **Reindel S, Schmidt CL, Anemuller S, Matzanke BF.** 2002. Characterization of a non-haem ferritin of the Archaeon *Halobacterium salinarum*, homologous to Dps (starvation-induced DNA-binding protein). *Biochem Soc Trans* **30**:713-715.
135. **Reuter CJ, Maupin-Furlow JA.** 2004. Analysis of proteasome-dependent proteolysis in *Haloferax volcanii* cells, using short-lived green fluorescent proteins. *Appl Environ Microbiol* **70**:7530-7538.
136. **Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK.** 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**:e47.
137. **Rothfuss H, Lara JC, Schmid AK, Lidstrom ME.** 2006. Involvement of the S-layer proteins Hpi and SlpA in the maintenance of cell envelope integrity in *Deinococcus radiodurans* R1. *Microbiology* **152**:2779-2787.
138. **Rowlett VW, Margolin W.** 2015. The Min system and other nucleoid-independent regulators of Z ring positioning. *Front Microbiol* **6**:478.
139. **Rust MJ, Bates M, Zhuang X.** 2006. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nat Methods* **3**:793-795.

140. **Saeed AI, Sharov V, White J, Li J, Liang W, Bhagabati N, Braisted J, Klapa M, Currier T, Thiagarajan M, Sturn A, Snuffin M, Rezantsev A, Popov D, Ryltsov A, Kostukovich E, Borisovsky I, Liu Z, Vinsavich A, Trush V, Quackenbush J.** 2003. TM4: a free, open-source system for microarray data management and analysis. *Biotechniques* **34**:374-378.
141. **Sandman K, Reeve JN.** 2006. Archaeal histones and the origin of the histone fold. *Curr Opin Microbiol* **9**:520-525.
142. **Sapienza C, Rose MR, Doolittle WF.** 1982. High-frequency genomic rearrangements involving archaeobacterial repeat sequence elements. *Nature* **299**:182-185.
143. **Schmid AK, Pan M, Sharma K, Baliga NS.** 2011. Two transcription factors are necessary for iron homeostasis in a salt-dwelling archaeon. *Nucleic Acids Res* **39**:2519-2533.
144. **Schmid AK, Reiss DJ, Kaur A, Pan M, King N, Van PT, Hohmann L, Martin DB, Baliga NS.** 2007. The anatomy of microbial cell state transitions in response to oxygen. *Genome Res* **17**:1399-1413.
145. **Schmid AK, Reiss DJ, Pan M, Koide T, Baliga NS.** 2009. A single transcription factor regulates evolutionarily diverse but functionally linked metabolic pathways in response to nutrient availability. *Mol Syst Biol* **5**:282.
146. **Schweizer N, Weiss M, Maiato H.** 2014. The dynamic spindle matrix. *Curr Opin Cell Biol* **28**:1-7.
147. **Selmecki AM, Dulmage K, Cowen LE, Anderson JB, Berman J.** 2009. Acquisition of aneuploidy provides increased fitness during the evolution of antifungal drug resistance. *PLoS Genet* **5**:e1000705.
148. **Shao H, Revach M, Moshonov S, Tzuman Y, Gazit K, Albeck S, Unger T, Dikstein R.** 2005. Core promoter binding by histone-like TAF complexes. *Mol Cell Biol* **25**:206-219.
149. **Sharma K, Gillum N, Boyd JL, Schmid A.** 2012. The RosR transcription factor is required for gene expression dynamics in response to extreme oxidative stress in a hypersaline-adapted archaeon. *BMC Genomics* **13**:351.

150. **Shih YL, Zheng M.** 2013. Spatial control of the cell division site by the Min system in *Escherichia coli*. *Environ Microbiol* **15**:3229-3239.
151. **Shioda M, Sugimori K, Shiroya T, Takayanagi S.** 1989. Nucleosomelike structures associated with chromosomes of the archaebacterium *Halobacterium salinarium*. *J Bacteriol* **171**:4514-4517.
152. **Si F, Li B, Margolin W, Sun SX.** 2015. Bacterial growth and form under mechanical compression. *Sci Rep* **5**:11367.
153. **Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Soding J, Thompson JD, Higgins DG.** 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**:539.
154. **Singh RK, Liang D, Gajjalaiahvari UR, Kabbaj MH, Paik J, Gunjan A.** 2010. Excess histone levels mediate cytotoxicity via multiple mechanisms. *Cell Cycle* **9**:4236-4244.
155. **Smyth GK, Michaud J, Scott HS.** 2005. Use of within-array replicate spots for assessing differential expression in microarray experiments. *Bioinformatics* **21**:2067-2075.
156. **Soares D, Dahlke I, Li WT, Sandman K, Hethke C, Thomm M, Reeve JN.** 1998. Archaeal histone stability, DNA binding, and transcription inhibition above 90 degrees C. *Extremophiles* **2**:75-81.
157. **Soares DJ, Sandman K, Reeve JN.** 2000. Mutational analysis of archaeal histone-DNA interactions. *J Mol Biol* **297**:39-47.
158. **Sopko R, Huang D, Preston N, Chua G, Papp B, Kafadar K, Snyder M, Oliver SG, Cyert M, Hughes TR, Boone C, Andrews B.** 2006. Mapping pathways and phenotypes by systematic gene overexpression. *Mol Cell* **21**:319-330.
159. **Soppa J.** 2011. Ploidy and gene conversion in Archaea. *Biochem Soc Trans* **39**:150-154.
160. **Spang A, Saw JH, Jorgensen SL, Zaremba-Niedzwiedzka K, Martijn J, Lind AE, van Eijk R, Schleper C, Guy L, Ettema TJ.** 2015. Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature* **521**:173-179.

161. **Steensland H, Larsen H.** 1969. A study of the cell envelope of the halobacteria. *J Gen Microbiol* **55**:325-336.
162. **Stothard P.** 2000. The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniques* **28**:1102, 1104.
163. **Sumper M, Berg E, Mengele R, Strobel I.** 1990. Primary structure and glycosylation of the S-layer protein of *Haloferax volcanii*. *J Bacteriol* **172**:7111-7118.
164. **Takayanagi S, Morimura S, Kusaoke H, Yokoyama Y, Kano K, Shioda M.** 1992. Chromosomal structure of the halophilic archaeobacterium *Halobacterium salinarium*. *J Bacteriol* **174**:7207-7216.
165. **Tan M, Luo H, Lee S, Jin F, Yang JS, Montellier E, Buchou T, Cheng Z, Rousseaux S, Rajagopal N, Lu Z, Ye Z, Zhu Q, Wysocka J, Ye Y, Khochbin S, Ren B, Zhao Y.** 2011. Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. *Cell* **146**:1016-1028.
166. **Thomm M, Sandman K, Frey G, Koller G, Reeve JN.** 1992. Transcription in vivo and in vitro of the histone-encoding gene *hmfB* from the hyperthermophilic archaeon *Methanothermus fervidus*. *J Bacteriol* **174**:3508-3513.
167. **Todor H, Dulmage K, Gillum N, Bain JR, Muehlbauer MJ, Schmid AK.** 2014. A transcription factor links growth rate and metabolism in the hypersaline adapted archaeon *Halobacterium salinarum*. *Mol Microbiol* **93**:1172-1182.
168. **Tomschik M, Karymov MA, Zlatanova J, Leuba SH.** 2001. The archaeal histone-fold protein HMf organizes DNA into bona fide chromatin fibers. *Structure* **9**:1201-1211.
169. **Tonner PD, Pittman AM, Gulli JG, Sharma K, Schmid AK.** 2015. A regulatory hierarchy controls the dynamic transcriptional response to extreme oxidative stress in archaea. *PLoS Genet* **11**:e1004912.
170. **Treuner-Lange A, Sogaard-Andersen L.** 2014. Regulation of cell polarity in bacteria. *J Cell Biol* **206**:7-17.

171. **Tse-Dinh YC.** 1985. Regulation of the *Escherichia coli* DNA topoisomerase I gene by DNA supercoiling. *Nucleic Acids Res* **13**:4751-4763.
172. **Uchiyama S, Kobayashi S, Takata H, Ishihara T, Hori N, Higashi T, Hayashihara K, Sone T, Higo D, Nirasawa T, Takao T, Matsunaga S, Fukui K.** 2005. Proteome analysis of human metaphase chromosomes. *J Biol Chem* **280**:16994-17004.
173. **van Noort J, Verbrugge S, Goosen N, Dekker C, Dame RT.** 2004. Dual architectural roles of HU: formation of flexible hinges and rigid filaments. *Proc Natl Acad Sci U S A* **101**:6969-6974.
174. **Van PT, Schmid AK, King NL, Kaur A, Pan M, Whitehead K, Koide T, Facciotti MT, Goo YA, Deutsch EW, Reiss DJ, Mallick P, Baliga NS.** 2008. Halobacterium salinarum NRC-1 PeptideAtlas: toward strategies for targeted proteomics and improved proteome coverage. *J Proteome Res* **7**:3755-3764.
175. **Vega DE, Young KD.** 2014. Accumulation of periplasmic enterobactin impairs the growth and morphology of *Escherichia coli tolC* mutants. *Mol Microbiol* **91**:508-521.
176. **Venkatraman ES, Olshen AB.** 2007. A faster circular binary segmentation algorithm for the analysis of array CGH data. *Bioinformatics* **23**:657-663.
177. **Vitha S, McAndrew RS, Osteryoung KW.** 2001. FtsZ ring formation at the chloroplast division site in plants. *J Cell Biol* **153**:111-120.
178. **Walsby AE.** 2005. Archaea with square cells. *Trends Microbiol* **13**:193-195.
179. **Wang X, Lutkenhaus J.** 1996. FtsZ ring: the eubacterial division apparatus conserved in archaeobacteria. *Mol Microbiol* **21**:313-319.
180. **Weidenbach K, Gloer J, Ehlers C, Sandman K, Reeve JN, Schmitz RA.** 2008. Deletion of the archaeal histone in *Methanosarcina mazei Go1* results in reduced growth and genomic transcription. *Mol Microbiol* **67**:662-671.
181. **Whitehead K, Kish A, Pan M, Kaur A, Reiss DJ, King N, Hohmann L, DiRuggiero J, Baliga NS.** 2006. An integrated systems approach for understanding cellular responses to gamma radiation. *Mol Syst Biol* **2**:47.
182. **Wilbanks EG, Larsen DJ, Neches RY, Yao AI, Wu CY, Kjolby RA, Facciotti MT.**

2012. A workflow for genome-wide mapping of archaeal transcription factors with ChIP-seq. *Nucleic Acids Res* **40**:e74.
183. **Wilkinson SP, Ouhammouch M, Geiduschek EP.** 2010. Transcriptional activation in the context of repression mediated by archaeal histones. *Proc Natl Acad Sci U S A* **107**:6777-6781.
184. **Winzler EA, Shoemaker DD, Astromoff A, Liang H, Anderson K, Andre B, Bangham R, Benito R, Boeke JD, Bussey H, Chu AM, Connelly C, Davis K, Dietrich F, Dow SW, El Bakkoury M, Foury F, Friend SH, Gentalen E, Giaever G, Hegemann JH, Jones T, Laub M, Liao H, Liebundguth N, Lockhart DJ, Lucau-Danila A, Lussier M, M'Rabet N, Menard P, Mittmann M, Pai C, Rebischung C, Revuelta JL, Riles L, Roberts CJ, Ross-MacDonald P, Scherens B, Snyder M, Sookhai-Mahadeo S, Storms RK, Veronneau S, Voet M, Volckaert G, Ward TR, Wysocki R, Yen GS, Yu K, Zimmermann K, Philippsen P, et al.** 1999. Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* **285**:901-906.
185. **Wisniewski-Dye F, Vial L.** 2008. Phase and antigenic variation mediated by genome modifications. *Antonie Van Leeuwenhoek* **94**:493-515.
186. **Woese CR, Fox GE.** 1977. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci U S A* **74**:5088-5090.
187. **Woods WG, Dyall-Smith ML.** 1997. Construction and analysis of a recombination-deficient (*radA*) mutant of *Haloferax volcanii*. *Mol Microbiol* **23**:791-797.
188. **Xie Y, Reeve JN.** 2003. In vitro transcription assays using components from *Methanothermobacter thermautotrophicus*. *Methods Enzymol* **370**:66-72.
189. **Yaoi T, Laksanalamai P, Jiemjit A, Kagawa HK, Alton T, Trent JD.** 2000. Cloning and characterization of *ftsZ* and *pyrF* from the archaeon *Thermoplasma acidophilum*. *Biochem Biophys Res Commun* **275**:936-945.
190. **Young KD.** 2006. The selective value of bacterial shape. *Microbiol Mol Biol Rev* **70**:660-703.
191. **Yuan GC, Liu YJ, Dion MF, Slack MD, Wu LF, Altschuler SJ, Rando OJ.** 2005. Genome-scale identification of nucleosome positions in *S. cerevisiae*. *Science*

309:626-630.

192. **Zerulla K, Soppa J.** 2014. Polyploidy in haloarchaea: advantages for growth and survival. *Front Microbiol* **5**:274.
193. **Zhang Z, Guo L, Huang L.** 2012. Archaeal chromatin proteins. *Sci China Life Sci* **55**:377-385.

Biography

Keely Dulmage was born September 7th, 1983 in Victoria, British Columbia, Canada. She received her Bachelors of Arts in Biology from the University of Minnesota (Twin Cities) in 2007. Her dissertation was supported by a Grant-in-aid from the Duke department of Biology in 2013 and Duke Graduate School Travel Grants in 2013 and 2014.

Publications:

- 1) Dulmage K, Todor H, Schmid AK. Growth-phase-specific modulation of cell morphology and gene expression by an archaeal histone protein. *Mbio.* 2015 Sep 8;6(5)
- 2) Todor H, Dulmage K, Gillum N, Bain JR, Muehlbauer MJ, Schmid AK. A transcription factor links growth rate and metabolism in the hypersaline adapted archaeon *Halobacterium salinarum*. *Mol Microbiol.* 2014 Sep;93(6):1172-82.
- 3) Anderson MZ, Baller JA, Dulmage K, Wigen L, Berman J. The three clades of the telomere-associated TLO gene family of *Candida albicans* have different splicing, localization, and expression features. *Eukaryot Cell.* 2012 Oct;11(10):1268-75.
- 4) Selmecki AM, Dulmage K, Cowen LE, Anderson JB, Berman J. Acquisition of aneuploidy provides increased fitness during the evolution of antifungal drug resistance. *PLoS Genet.* 2009 Oct;5(10):e1000705.
- 5) Gerami-Nejad M, Dulmage K, Berman J. Additional cassettes for epitope and fluorescent fusion proteins in *Candida albicans*. *Yeast.* 2009 Jul;26(7):399-406.