POPULATION STRUCTURE AND DIVERSITY OF *HALOMONAS* SPP. FROM THE GREAT SALT PLAINS REVEALED BY MULTI LOCUS SEQUENCE TYPING

By

DEEPAK G RUDRAPPA

Master of Science in Agriculture Microbiology
University of Agricultural Sciences
Bangalore, Karnataka, India
2000

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
December, 2010

POPULATION STRUCTURE AND DIVERSITY OF

HALOMONAS SPP. FROM THE GREAT SALT PLAINS

REVEALED BY

MULTI LOCUS SEQUENCE TYPING

Dissertation Approved:

Dr. Robert V. Miller

Dissertation Adviser

Dr. Robert Burnap

Dr. Elshahed Mostafa

Dr. Carol Bender

Outside Committee Member

Dr. Mark E. Payton

Dean of the Graduate College

.

ACKNOWLEDGMENTS

Writing of this dissertation has been one of the most significant academic challenges I have faced so far. I would like to express my deepest appreciation and gratitude to the many individuals who have supported me to reach this milestone. Without the support, patience and guidance of Dr. Robert V Miller, my major advisor, this study would not have been completed. It is to him that I owe my deepest gratitude. Dr. Miller undertook to act as my major advisor despite his many other academic and professional commitments and provided me with an excellent atmosphere for doing research. His wisdom, knowledge and commitment to the highest standards inspired and motivated me.

I would also like to thank Dr. Robert Burnap, Dr. Carol Bender, and Dr. Elshahed Mostafa for guiding my research for the past several years and helping me to develop my background in microbiology and ecology. A special thanks goes to Dr. Mostafa, who was willing to participate in my defense committee during the final years, whose encouragement, supervision and support enabled me to develop a better understanding of the subject.

I thank Connie Bud, Reema Biswas, Dr.Noha Youssef, Jim Davis, Audra Liggenstoffer, Tammy Austin who as good friends, were always willing to help, motivate and give their best suggestion. It would have been a lonely lab without them. I would also like to thank all my fellow graduate students in the department and undergraduate students in the lab throughout these years.

Finally, I would like to thank the love of my life, my wife Dr. Shubhapriya Bennur, without whom this effort would have been worth nothing. She was always there cheering me up and stood by me through the good times and bad. Your love, support and constant patience have taught me so much about sacrifice, discipline and compromise. A special thanks to Aarav, my son, who came into our lives in the middle of our PhD and got used to our crazy schedule and never complained!!

Heartfelt regards to my parents Dr.Rudrappa and Chandramma, who have always supported, encouraged and believed in me, in all my endeavors.

This dissertation is dedicated to Shubha, Aarav, my parents and to all my immediate family.

TABLE OF CONTENTS

Chapter Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

Figure Page

CHAPTER I

ABSTRACT

The Great Salt Plains (GSP) is an ecologically diverse, hyper-saline environment. Many differences in overall phenotypic characters were observed among *Halomonas* spp. isolates from the GSP that had >97% 16S rDNA sequence similarity. We believe these *Halomonas* spp. isolates have evolved over a long period of time by means of developing certain genetic mechanisms. Multi Locus Sequence Typing (MLST) was used to study the population structure of these isolates. Housekeeping genes *ectB, gyrB, and recA* along with 16S rDNA and the ITS region were assayed in 32 *Halomonas* spp. isolates. MLST analysis identified 28 unique sequence types indicating a high level of genetic diversity. Among the house keeping genes, *recA* demonstrated increased resolution indicating that *Halomonas* spp. isolates contain several genetically distinct phylotypes most likely due to environmental bottlenecks and recombination events. These results support the hypothesis that high level of gene transfer events have occurred among *Halomonas* spp. isolates at the GSP site even though 16S rDNA sequence similarity has been conserved. Phylogenetic relationship of very closely related *Halomonas* spp. isolates was resolved by MLST. These result prompted us to study and compare the influence of recombination in an organism from an environment with more moderate characteristics. I analyzed and compared sequence information from 9 *Shewanella* isolates from both fresh and marine water bodies. The isolates tested showed comparable levels of genetic variations indicating genomic adaptation of the *Shewanella* strains to their environmental conditions, mediated by lateral gene transfer events. Next I studied the survivability of *Halomonas* spp.

isolates exhibited increased level of resistance to DNA damage from UV radiation on par with *E. coli.* This study helped us to understand how environmental factors impact the genetic diversity and stability of microorganisms.

CHAPTER II

REVIEW OF LITERATURE

Halophilic Microorganisms- History:

"An embankment is made and ditches to draw clear sea water. It is left for a long time until the color becomes red. If the south wind blows with force during the summer and autumn the salt may grain overnight. If the south wind does not come all the profits are lost".

This is the translation from a work by Peng-Tzao-Kan-Mu that dates back to about 2700 B.C. that describes the salt production from seawater in the Chinese province of Yai-cheau (7). This is one of the earliest reports of the red brines known to be caused by microbial communities adapted to life at high salt concentrations or near saturated salt conditions. Many such water bodies were reported in ancient times in Egypt. In modern times, blooms of red halophilic archaea have been reported in the Dead Sea (12).

Pierce in 1914 made a study showing that the red growth in brines is from bacteria (88). Klebahn in 1919 was able to isolate "*Bacillus halobius ruber*" from red discolorations on fish and described what is now known as *Halobacterium* (59). Similar details were provided by Browne, and Harrison and Kennedy in 1922 (14, 49). During the late 1920's into early 1930's, a substantial amount of work was reported on halophilic bacteria at the Deft School of Microbiology. Baumgartner (9) made the first isolation of an obligatory anaerobic halophilic bacterium in 1937 and described it as "*Bacteroides halosmophilus*".

Since then many attempts have been made to understand the microbial diversity in different saltern environments. Claude ZoBell and his student Whitney Smith showed the presence of a rich autochthonous bacterial community at the Great Salt Lake (GSL) (109). They estimated that 50% of the bacteria from the GSL require at least 70 g/l of salt to grow and 96% of the total bacterial community could not grow without salt. Wilkansky in 1936 (124) examined the water and sediment samples from the Dead Sea and was able to find several aerobic and anaerobic bacterial species.

Much of the current understanding of biology of halophilic microorganisms is due to the work of Norman Gibbons and his student Donn Kushner in Canada, and Helge Larsen and his student Ian Dundas from Norway during the 1960's (31). They opened a new direction into understanding the diversity and unique adaptation strategies of these salt-tolerant microorganisms from near saturated, hypersaline environments.

Ecology:

Ian Dundas (31) suggested that life might have originated from a hypersaline environment. This was further supported by reports indicating that life originated on Earth from a "primordial soup" of abiotically formed organic molecules that accumulated in the puddles of salty seawater (32**).**

Oceans are by far the largest saline bodies of water. Hypersaline environments can be defined as those with salt concentrations in excess of seawater. Thalassic water bodies are those that are derived from the evaporation of seawater. Athalassic waters are those where the salts are lower than sea water (81). The two most studied hypersaline water bodies are the Great Salt Lake in North America and the Dead Sea in the Middle East. Many small evaporation ponds exist near coastal areas fed by seawater. Hypersaline ponds are also found in Antartica that are stratified with respect to salinity. Alkaline hypersaline soda brines are perfect examples of dynamic environments. They vary significantly in size, salinity and temperature according to seasons.

Subterranean evaporate deposits and deep-sea basins also harbor hypersaline environments. A general characteristic of hypersaline environments is that the seawater is evaporated and a salinity gradient is created. This results in diverse microbial blooms adapted to varied salinity levels (81).

Since the terrestrial environments are considered more heterogeneous, it is safe to assume that a wide range of salinities exists in any soil. Saline soils harbor more halotolerant rather than halophilic microorganisms (120). This is due to periodical changes in salinities according to changes in climate. The microbial species composition varies considerably from aquatic environments. While members belonging to *Salinivibrio* dominate saline lakes and salterns, bacteria belonging to *Bacillus*, *Pseudomonas* and *Halomonadacaea* dominate the saline soils. Gram-negative bacteria are more dominant than the gram-positive ones. Facultative anaerobes are rare. Most of these bacteria require relatively high salt concentration (5 to 10%) for optimum growth but they can also grow at low salinities (0.1 to 2%). Much less attention has been devoted to studying saline soils than the hypersaline aquatic environments. Halotolerant bacteria were isolated from the Salty soils in Death Valley(CA), the Great Salt Plains(OK), and Iraq (16, 120).

Physiology of Halophiles:

Even though salts are essential for survival of living organisms, halophilic microorganisms are dependent on saline to hypersaline conditions for growth. These organisms can be classified according to their salt needs: slight halophiles grow best at 0.2-0.85 M (2-5%) NaCl, moderate halophiles grow best at 0.85-3.4 M (5-20%) NaCl and extreme halophiles grow best at 3.4-5.1 M (20-30%) NaCl. Many of these halotolerant and halophilic microorganisms have a wide range of salt tolerance for growth and sometimes this range depends on environmental and nutritional factors (80).

It is well known that the exclusion of salt and the maintainance of a steep ionic gradient by halotolerant microorganisms is achieved by the constant pumping of ions by an active, energy dependent mechanism or by tightening the membrane permeability barrier. Two mechanisms have been suggested to support this hypothesis: Na^+ / H^+ antiport and the presence of a primary respiration driven Na^+ pump (80)**.** High osmolarity in hypersaline conditions is harmful to most cells as a lot of water is lost to the environment (or the external medium) in order to reach osmotic equilibrium. Halophiles have developed a strategy to avoid this loss of cellular water, and they accumulate high concentrations of solute within the cytoplasm (22). These solutes are called compatible solutes and are usually amino acids, sugars and polyols, such as glycine betaine, ectoine, sucrose, trehalos and glycerol, which do not interfere with the metabolic processes and have no net charge at a physiologically neutral pH. In some haloarchaea and extremely halophilic bacteria, an exception is seen. They accumulate KCl equal to the external concentration of NaCl.

Halophilic Archaea: These organisms grow best at higher salinities (3.4-5 M NaCl), form blooms, and are the cause of red coloration of many brines (22). Haloarchaea species are rod, cocci or disc shaped and many are pleomorphic especially when the salt concentrations are altered in the media. Haloarchaea belong to the family *Halobacteriacaea.* More than 20 genera have been reported. The first haloarchaea genome sequenced was *Halobacterium* spp. NRC-1 (23). Many more have been sequenced recently or are close to completion. Haloarchaea have very high intercellular salt concentration with K^+ ions accumulated internally up to 5M along with Na^+ ions. The K^+ gradient is maintained by a Na/H antiporter and K^+ uniporters. Genomic and proteomic analysis have shown that haloarcheal proteins have an excess proportion of acidic to basic amino acids which is crucial for protein activity at high salt conditions (23). These proteins are resistant to high salt concentrations and need salt for their activity. Bacteriorhodopsin and carotenoids are the two most studied proteins of haloarchaea (63).

The Halotolerant and Halophilic Bacteria:

Halophiles can be defined as salt loving organisms. They are classified as slightly, moderately or extremely halophilic depending on their need for NaCl (81). Halophiles are found in all three kingdoms: Archaea, Bacteria and Eukarya. They are found throught the phyla and orders of the bacterial domain. Bacteria with varying physiological properties such as aerobic and anaerobic, chemoheterotrophs, photoautotrophs, photoheterotrophs as well as chemolithotrophs fall within this group. Recently, 16S rRNA sequence information has been used to get a better idea about the phylogenetic relationship within this halophilic bacterial community. Halophiles are known to exist within the phyla *Cyanobacteria, Proteobacteria, Firmicutes, Actinobacteria, Spirochaetes and Bacteroidetes* (81). Cyanobacterial mats are found in the hypersaline lakes and saltern evaporation ponds. A good description of the cyanobacteria belonging to the genera: *Halospirulina, Cyanothece, Aphanothece, Chroococcidiopsis* and *Myxobactron* can be found in the literature. The halophilic anoxygenic photosynthetic sulfur bacteria of the *Halorhodospira – Ectothiorhodospira* group have been extensively studied (54). Fatty acid composition analysis of the cellular lipids and the 16S rRNA sequence analysis have led to the identification of more extreme halophiles within this group. Based on the analysis of small rRNA subunits, the fermentative obligatory anaerobic halophilic bacteria are classified in the domain *Bacteria* within the phylum *Firmicutes* (84). Many other branches within the domain *Bacteria* contain halophilic microorganisms. Considering the scope of this dissertation, I will focus on the bacteria belonging to the family *Halomonadaceae.* Figure 2.1 shows a universal phylogenetic tree. Branches including halophilic microorganisms have been highlighted. Representative genera are also indicated.

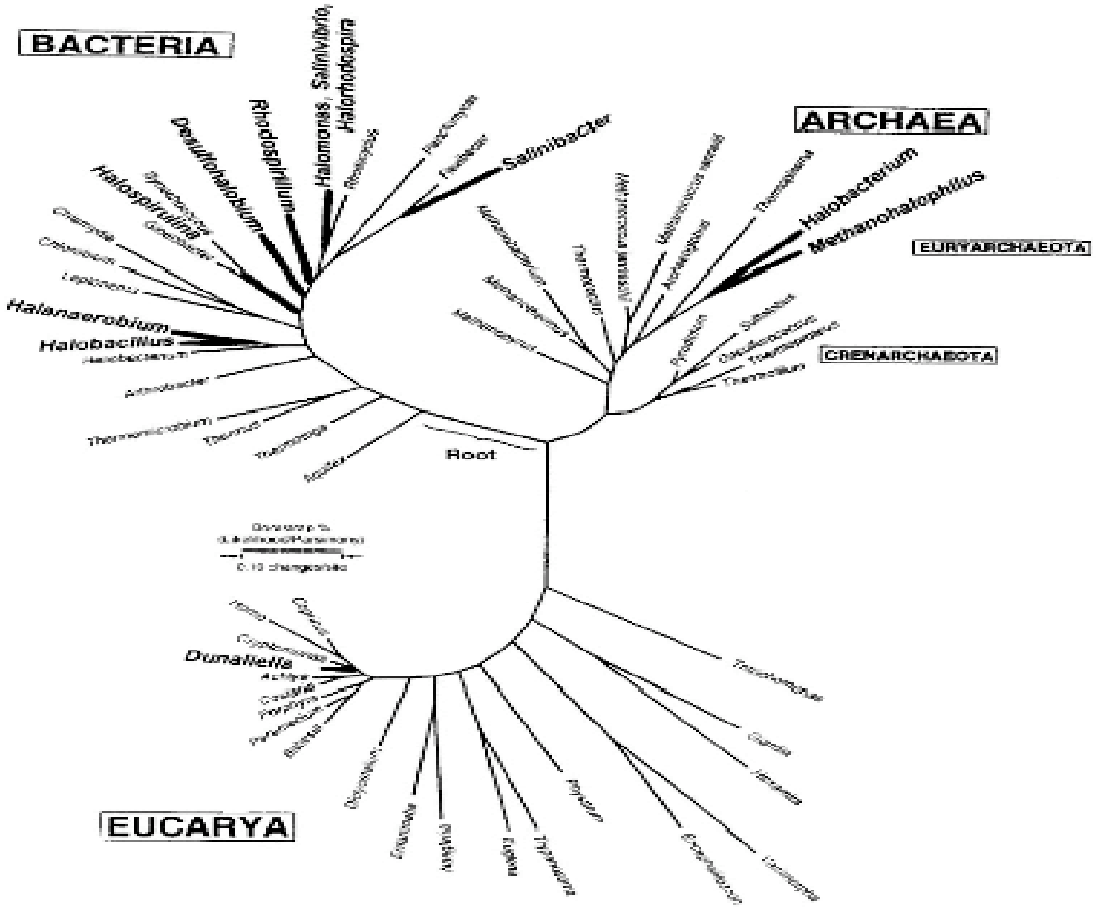


Figure 2.1: Universal phylogenetic tree based on small subunit rRNA sequence comparisons. From Oren A., 2002 (80)**.**

***Halomonadaceae*:**

The family *Halomonadaceae* includes a large group of metabolically versatile, aerobic, moderately halophilic bacteria within the class *Gammaproteobacteria* (Order *Oceanospirillales).* Since its creation in 1988 (40), exhaustive taxanomic rearrangements have been made in this group.

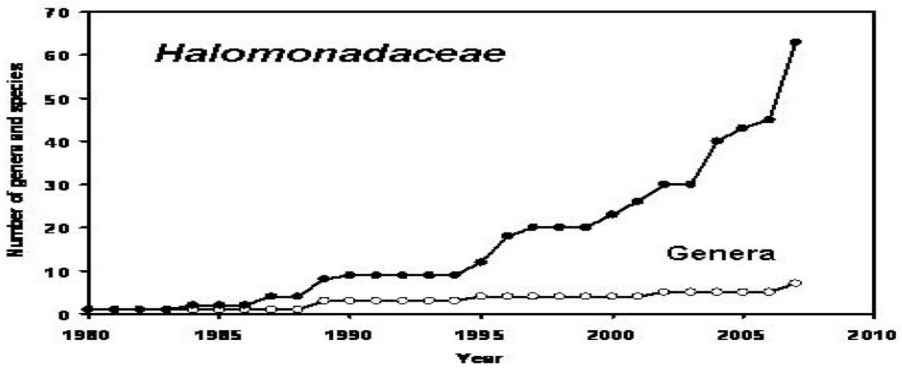


Figure 2.2: Number of different genera and species within the family *Halomonadaceae* whose names have been validly published. From Oren. A.,2008. (81)

Only a few non-halophiles are included in this group. The halophilic members have been discussed extensively in reviews by Ventosa et al. (1998) and Vreeland (1992) (120, 122). Most of these species are gram-positive or gram-negative aerobic or facultatively anaerobic moderately halophilic bacteria. Even though some of the gram-negative species are grouped in several different genera, phenotypic and phylogenetic studies support their close relationship to family *Halomonadaceae*. Arahal et al. and de la Haba et al. (4, 25) reported a detailed review of the phylogeny of the family *Halomonadaceae* based on 16S and 23S rRNA sequence analysis. The family *Halomonadaceae* is comprised of nine genera. *Halomonas* and *Chromohalobacter* are the prominent genera with 55 and 9 species respectively. Six genera are currently comprised of a single species. Very recently a new genus *Kushneria* was described by Sanchez-Porro et al. (99) with four species. Within the genus *Halomonas* the average similarity scores between different 16S rRNA sequences was one unit higher than that of the 23S rRNA sequences suggesting a slower evolutionary rate. This study clearly indicated that genus *Halomonas* is not monopyletic and is comprised of two well separated phylogenetic groups. The mean 16S rRNA and 23S rRNA sequence similarities within each of this two groups were 97.8%, 97.0% and 97.4%, 97.5% respectively. The sequence similarity values between the starains included in each of these two groups ranged 96.1-93.2% for 16S rRNA and 95.3-93.5% for 23S rRNA. This suggests that these

two groups can constitute two different genera. However according to the *Ad Hoc* Committee for the Re-evaluation of Species Definition in Bacteriology (112), the differentiation of taxa should be based on a polyphasic approach including the phenotypic features. Some phenotypic features that differentiate these two groups are: species belonging to group 1 have a higher G+C content of 57-70 mol% than species from the 2nd group, 51.4-62 mol%. Group 1 is comprised of more halophiles than group 2 and all species from group 2 are motile while group 1 is either motile or non-motile (3). The other remaining species that are currently assigned to genus *Halomonas* do not fall into either of these two phylogenetic groups.

There were considerable differences in the topologies of phylogenetic trees from 16S rRNA and 23S rRNA sequences. One possible explanation for this is the fact that recombinational events might have happened in these genes. The authors concluded by saying that both 23S rRNA and 16S rRNA genes still does not resolve the phylogenetic relationship between these very closely related species belonging to the genus *Halomonas*. To overcome this limitation and to further clarify the phylogenetic relationship between the members of this genus, authors suggested the use of housekeeping genes resulting in a Multi Locus Sequence Analysis (MLSA) (25). Until now only one study has been reported that attempted to carry out the MLSA of the species *Halomonas variabilis* (79). The authors used sequences of *gyrB, ectB* and *ectC* along with the 16S rRNA gene from ten strains of *H.variabilis* and found that phylogenetic trees based on *gyrB* and *ectB* genes were very similar to that based on 16S rRNA gene where as the *ectC* based tree was incongruent with other tree topologies. These interesting observations led me to investigate the fine genetic structure (phylogenetic relationships) of bacteria belonging to the genus *Halomonas* obtained from the hyper saline Great Salt Plains soils using MLST.

Great Salt Plains:

The Great Salt Plains (GSP) is located in North Central Oklahoma near the town of Cherokee (Figure 2.3). These plains are part of the Salt Plains National Wildlife Refuge (SPNWR) with an ecologically diverse, extreme, hyper saline environment. Salts are deposited by continuous percolation of underlying brine to the upper surface. Sodium chloride is the main constituent of the brine. The salt concentration varies considerably across the plains. These salt formations dissolve completely during rainy seasons forming many small ponds and streams with varying salt concentrations. The GSP is an extreme environment and microorganisms living there are continuously exposed to harsh environmental conditions such as high temperature, freezing winters, high salinity and direct Ultraviolet (UV) radiation. Survival under these conditions has led to stringent genetic selection among these microbial communities (16).

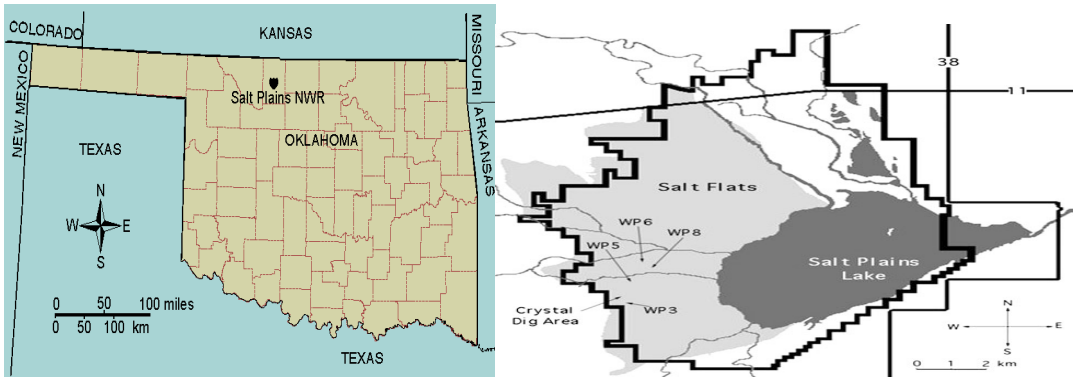


Figure 2.3: Maps showing the location of GSP and common sampling sites within GSP.

The microbial community of these hyper saline terrestrial ecosystems may be completely different from the hyper saline aquatic marine ecosystem. Thus the salt tolerant relatives of common soil bacteria are usually found in these hyper saline soils.

Previous studies at GSP:

Litzner et al. (2006) (69) phenotypically characterized isolates obtained from soil samples at the GSP using carbon substrate and nitrogen utilization patterns in batch cultures. This study extended the initial phenotypic characterization of 105 isolates by Caton et al. (16) to 60 additional isolates by measuring the utilization pattern of 45 carbon substrates. Based on these data the isolates were assigned to seven different phenons along with several singletons. These isolates exhibited different patterns of amino acids, carboxylic acids, carbohydrates and alcohol utilization. Most of these isolates were able to utilize almost all of the nitrogen sources tested. Antibiotic sensitivity was also tested where in little antibiotic resistance was observed. About one third of the isolates were resistant to penicillin and streptomycin. Partha Dutta ,2006 (33) studied the microbial community from the GSP using polymerase chain reaction (PCR) and denaturing gradient gel electrophoresis (DGGE). Salt gradient and core samples were collected in different years. Based on DGGE analysis, low and high salt soil samples showed greater band richness than medium salt samples. Core samples from the same locations exhibited similar microbial communities and samples collected from the same locations in different years were comprised of different microbial communities. This study showed that constantly varying environmental factors select for different microbial communities from samples collected in different years and different locations at the GSP.

Ultraviolet (UV) radiation sensitivity was tested in *Halomonas* spp. isolates by Wilson et al. (125). *Halomonas* spp. isolates were examined for their survivability when exposed to UV radiation in laboratory settings. The UV resistance of *Halomonas vinusta* like isolates was not significantly different than *Escherichia coli* but was significantly more UV resistant than *Pseudomonas aeruginosa*. *Halomonas salina* like isolates showed significant variation in their resistance to UV irradiation (125). Biswas Reema (11) studied the effect of UV radiation on the photosynthetic cyanobacteria *Aphanothece* and examined the role of the *phrA* gene in DNA

damage repair. *Aphanothece* survived higher doses of UV only when incubated in light after UV exposure, indicating the importance of light-requiring DNA repair in this organism when compared to cyanobacteria from regular non-extreme environment.

Five different bacteriophages specific to the genus *Halomonas* have been isolated and characterized in regard to their propagation and adsorption to different *Halomonas* spp. isolates (78). All phages showed relatively broad host range and the phages became highly activated at higher salinities. Beside salinity, the phages showed high levels of tolerance towards variable temperatures and alkaline pH. Restriction digestion using three endonucleases (*Eco*RI, *Hin*d III, and *Bam*HI) suggested a small (46 kb) phage size.

Initially halotolerant culturable heterotrophic bacteria were isolated from GSP soils on the basis of morphological differences, size and shape of colonies, pigmentation and time required for growth (16). Colonies were isolated on different types of media. One hundred five isolates belonging to different phylotypes were identified using heterotrophic enrichments and 16S rDNA sequence (phylogenetic) analysis. *Bacillus* and *Halomonas spp*. were the dominant Gram +ve and Gram -ve bacteria phylotypes respectively. *Halomonas* spp. isolates were closely related to bacteria isolated from aquatic systems. Many differences in overall phenotypic characteristics were observed when compared with isolates obtained from abandoned saltern soils (95). Very few GSP isolates were motile, oxidase positive, did not produce H_2S, grow at temperature >45°C and were less halophilic (0.1 to 30%) (16). These observations suggest that *Halomonas* spp. isolates from the GSP have developed certain genetic mechanisms to protect themselves against varying salt concentrations and constant desiccating conditions. Currently not enough information is available to characterize the *Halomonas* spp. population structure in hyper saline environments. In order to understand and resolve this inconsistency, I performed MLST analysis on the *Halomonas* spp. bacteria from the GSP.

Molecular microbial ecology:

Steady advances in recent years have improved our understanding of the biogeochemical cycles and ecosystems. Since many crucial processes are driven by microorganisms in nature, it is essential for us to understand and appreciate how microbial communities function. Theories developed to understand microbial ecological patterns focus on the distribution and the structure of diversity within a microbial population (94).

Speciation in bacterial communities:

Most of these theories depend on the understanding of the concept of species. Population ecology counts individuals within a species and community ecology counts the number of species. A common definition of species using a genetic concept is, group of interbreeding individuals that is isolated from other such groups by barriers of recombination (94). However, prokaryotes are asexual and this definition doesn't hold true for them. An alternative ecological species concept defines a species as set of individuals that can be considered identical in all relevant ecological traits. Cohan (18) postulated that "bacteria occupy discrete niches and that periodic selection will purge genetic variation within each niche without preventing divergence between the inhabitants of different niches". This assumes that molecular diversity relates directly to ecological diversity. When the environmental variables are continuous this might not be relevant for bacteria and might not lead to simple mapping between molecular markers and an ecological niche. This implies that speciation and ecological species definitions should consider gene transfer events in bacteria, which are capable of transferring only a small part of the genome and serve as a mechanism for maintaining the biological species.

The importance of gene exchange mechanisms in bacteria arises from the fact that bacterial genomes are divided into two distinct parts, the core genome and the accessory genome (18). The core genome consists of genes that are crucial for the functioning of an organisms and the

accessory genome consists of genes that are capable of adapting to the changing ecosystem through gain and loss of function. Strains that belong to the same species can differ in the composition of accessory genes and therefore their capability to adapt to changing ecosystems. Several studies have been devoted to identify core and accessory genes for certain bacterial species using complete genomic sequence information and genome hybridization using DNA microarray (44, 87, 117). Most of these studies are biased towards clinical isolates, and fewer studies are focused on environmental isolates. Extensive surveys of 16S rRNA sequences have resulted in better understanding of microbial diversity. Additional ecological diversity exists in plasmids, transposons and pathogenisity islands as they can be easily shared between bacteria in a favorable environment but still be absent in the same bacterial species found elsewhere. This poses a major challenge to those studying bacterial population structure and community ecology. We are still a long way from finding a fitting theory that connects the fluid nature of bacterial communities to their ecology (123). Since bacteria are capable of withstanding rapid changes in environmental conditions, high rates of speciation could play a crucial role in influencing temporal diversity patterns within a community (19)

Understanding the nature and contribution of different processes that determine the frequencies of genes in any population is the biggest concern in population and evolutionary genetics. The main focus is to explain these processes upon which selection operates by generating genetic variation. Eukaryotes differ from prokaryotes in how they exchange genetic information among themselves. In eukaryotes, recombination is constantly involved in the shuffling of genes. Bacteria reproduce by binary fission, and it is thought that recombination is a rare event even though Tatum and Lederberg (1947) (116) discovered laterally transferred genes in *E.coli.* However sequencing of the genomes of two isolates of *Helicobacter pylori* revealed that >6% of the genes are unique (2). When recombination is rare and limited to few genes, almost all other genes will be transmitted through vertical inheritance. Mutations will accumulate slowly over the period and will result in

irreversible divergence of lineages. In contrast, when recombinational events are common, genes will release themselves from the rest of the genome and diversity in genes linked to adaptive alleles will be purged through selection (89). This type of panmictic population structure has been reported in *Neisseria gonorrhoeae* and *Rhizobium meliloti* lineages (108). The diversity observed in bacterial sequences of environmental isolates reveals the constraints placed by forces such as population bottlenecks, barriers to gene exchange or simple chance. When these factors strike a balance it will lead to speciation and stable structure within a bacterial community (24**)**. Thus the species can be defined simply as a cluster of strains that are markedly similar to each other, often with reference to the type strain of that cluster (78). But how similar these strains need to be (DNA-DNA similarity) in order for them to be classified as belonging to same species is constantly changing. The ad hoc committee for the reevaluation of the bacterial species definition (112) maintains that the 70% DNA-DNA similarity and 5^0 C or less Δ T_m criteria should continue to define a bacterial species. Thus, both phenotypic characterization and use of phylogenetic marker such as 16S rRNA continue to play a prominent role in describing a species. Looking at the vast amount of diversity, emphasis should be on understanding the genetic potential and different mechanisms that helped bacteria to survive and evolve at the same time (78).

Methods for evaluating population structure of Bacteria:

During the last decade, several methods have been developed which help us to understand inter- and intra-species relationships: DNA typing methods using several house keeping genes, DNA arrays and other sequence based techniques (69). DNA array analysis compares gene content relative to a reference strain and is cost effective compared to whole genome sequencing. However the hybridization technique lacks some of the important evolutionary clues stored in DNA sequences. One major drawback of these techniques is that they have been previously applied to clinical isolates and little attention has been devoted to environmental isolates. Multi Locus Enzyme Electrophoresis (MLEE) is a typing method borrowed from population genetic

studies of higher organisms. MLEE assesses genetic variation by the resolution of electrophoretic variants of metabolic enzymes by means of starch electrophoresis. Variant alleles at each locus infer an electromorph and thus a genotype. Initially MLEE studies involving bacterial populations appeared to be in agreement with the common notion that bacterial populations are in linkage disequilibrium (i.e. accumulation of rare mutational events occurs in nonrandom combinations) (104) and have clonal structures where all loci share a similar phylogenetic history (congruence). The clonal population structure of bacterial species was the dominant model until population studies began to explore nucleotide sequences to study genetic variations in bacterial populations in the late 1980's and early 1990's (69). A major drawback observed from MLEE studies is the relatively low resolution of the genotypes obtained from the electromorph data.

As nucleotide sequencing became more popular, evidence for more frequent genetic exchange events in these asexual bacterial populations of bacteria became stronger. Evidence of so called mosaic genes (sequences which are obtained by recombination events between two divergent organisms) was first observed in genes under strong positive selection that coded for antibiotic resistance and surface-exposed antigens (111). However the same phenomenon was soon observed in house keeping genes (which are under stabilizing selection for conservation of vital metabolic functions). This led to a reassessment of the role of sexual processes in the structuring of bacterial populations. Since then several studies have demonstrated the wide spread occurrence of horizontal gene exchange in bacterial populations mediated by transformation, transduction and conjugation, thus involving only small segments of the chromosome. One interesting observation made from these studies was that many bacteria studied show extensive evidence for horizontal gene exchange but are also comprised of clonally related groups. In these populations, genetic exchange is sufficient to erase clonality over a long period of time but the frequency is not enough to completely erase the clonal signals over evolutionary time points. These data suggest that typing methods employed should be able to accommodate a wide range of population

structures. Any typing scheme employed should be able to generate data that will help us to understand the population structure and resolve the biological relationships among the isolates in question. Advances in sequencing, PCR techniques and information technology have made choosing of targets for typing schemes more rational than conditional.

MLST was first introduced as an improvement over MLEE and other typing methods for studying the epidemiology of bacterial pathogens (70). This method has been tested mainly for pathogenic bacterial populations (28, 70). In recent years this method has been successfully used to understand population structure of environmental isolates (52, 93)**.** MLST databases can be easily used to compare bacterial isolates from different sources without the need for exchange of isolates. This method provides the basis for a universal language for bacterial typing. Seven house keeping genes used in a typical MLST scheme provide sufficient discrimination for bacterial typing without being subjected to diversifying selection that might result in vague relationships among isolates. MLST works on a principle similar to MLEE in that variation at house keeping loci serves as basis for assigning allele numbers and allelic profiles. Point mutation events detected within these house keeping gene sequences indicate variations at a specific locus, and each different sequence is assigned a different allele number. The combination of allele numbers of each of the house keeping loci gives the allele profile of a particular bacterial isolate. This is also known as a sequence type (ST). The resulting phylogenetic relationship among STs (bacterial isolates) can be displayed as a dendrogram or a cluster based analysis called eBURST. Developing a MLST scheme involves three crucial steps; 1. Choice of isolates to be tested; 2. The choice of loci to be characterized; and 3. Design of primers for amplification of those loci. There are a few criteria for the selection of loci as well: the genes should be in a region of the genome devoid of recombination hot spots or genes under diversifying selection, location of the genes on the genome should be uniformly scattered in order to increase the discriminatory power of MLST and present in single copy in the genome (20, 110, 119).

The first MLST scheme was reported for *N. meningitides* in 1998 and was based on a previous MLEE study (70). Since then MLST schemes have been developed for many pathogenic and a few environmental isolates with varying applications (Table 2.1). Some of the applications of MLST include, epidemiological studies, population and evolutionary studies, species identification, isolate characterization and anthropological studies (69). The size of each database is dependent upon the typing and submission of the findings by researchers throughout the world. Isolate collections from wide geographical regions is advantageous in global epidemiology or population biology studies. Table 2.1 summarizes the current list of MLST schemes for different bacterial isolates. More than half of these schemes are hosted at the University of Oxford, UK (http://pubmlst.org) while the rest of the schemes are hosted at Imperial college, UK (http://www.mlst.net), the Max-Planck Institute for infection Biology, Berlin, Germany (http://web.mpiib-berlin.mpg.de/mlst/) and the Pasteur Institute, Paris, France (http://www.pasteur.fr/mlst/) (1, 56).

Table 2.1: Complete list of MLST schemes published on pubMLST website (http://pubmlst.org).

Organism	**Applications to date**
Acinetobacter baumannii#1	Isolate characterisation and population structure analysis
Acinetobacter baumannii#2	-
Arcobacter spp.	Isolate characterisation and population structure
Aspergillus fumigatus	Isolate characterisation
Bacillus cereus	Population and evolutionary analyses.
Batrachochytrium dendrobatidis	Population and evolutionary analyses.
Bordetella spp.	-

Borrelia burgdorferi	Isolate characterisation and epidemiological analyses
Brachyspira hyodysenteriae	Molecular epidemiology and population structure analyses.
Brachyspira intermedia	Molecular epidemiology and population structure analyses.
Brachyspira spp.	Isolate characterisation and population structure analysis
Burkholderia cepacia complex	Isolate characterisation and population structure analyses
Burkholderia pseudomallei	Isolate characterisation and species identification.
Campylobacter fetus	Isolate characterisation and population structure analyses
Campylobacter helveticus	-
Campylobacter insulaenigrae	Isolate characterisation and population structure analysis
Campylobacter jejuni	Epidemiological surveillance of food-borne infections, population structure analyses.
Campylobacter lari	-
Campylobacter upsaliensis	-
Candida albicans	Isolate characterisation and identification of nosocomial outbreaks.
Candida glabrata	Isolate characterisation and population structure analyses
Candida krusei	Isolate characterisation and population structure analyses
Candida tropicalis	Isolate characterisation and epidemiological analyses
Chlamydiales spp.	Isolate characterisation and population structure analysis
Clostridium difficile	Isolate characterisation and population structure analysis
Clostridium septicum	Isolate characterisation and evolutionary analyses
Cronobacter spp.	-
Cryptococcus neoformans	Isolate characterisation and population structure analysis
Enterococcus faecalis	-
Enterococcus faecium	Isolate characterisation and epidemiological analyses.

Escherichia coli#1	Examination of the evolution of virulence.
Escherichia coli#2	-
Haemophilus influenzae	Epidemiological surveillance.
Helicobacter pylori	Tracking the patterns of human migration.
Klebsiella pneumoniae	Isolate characterisation and epidemiological analyses
Lactobacillus casei	Isolate characterisation and population structure analysis
Laribacter hongkongensis	-
Leptospira spp.	Isolate characterisation and epidemiological analyses
Listeria monocytogenes	-
Mannheimia haemolytica	Isolate characterisation
Moraxella catarrhalis	-
Neisseria spp.	Epidemiological surveillance, post-vaccine studies, population and evolutionary analyses.
Pasteurella multocida#1	Isolate characterisation and population structure analysis
Pasteurella multocida#2	Evolutionary analyses between isolates from different host species
Pencillium marneffei	-
Plesiomonas shigelloides	Isolate characterisation and population structure analysis
Porphyromonas gingivalis	Isolate characterisation and population structure analyses
Pseudomonas aeruginosa	Population and evolutionary analyses.
Salmonella enterica	-
Salmonella typhi	Estimation of age and time of emergence of *S. typhi*.
Sinorhizobium spp	-
Staphylococcus aureus	Identification of predominant epidemic MRSA isolates, species evolution, population analyses.

Staphylococcus epidermidis	-
Stenotrophomonas maltophilia	-
Streptococcus agalactiae	Identification of clones associated with neonatal infections.
Streptococcus oralis	Isolate characterisation and population structure analyses
Streptococcus pneumoniae	Identification of antibiotic resistant clones, epidemiological surveillance, population analyses.
Streptococcus pyogenes	Identification of relationships between emm-type and clonal complex.
Streptococcus suis	-
Streptococcus uberis	-
Streptococcus zooepidemicus	Isolate characterisation and population structure analysis
Streptomyces spp	Species identification
Vibrio parahaemolyticus	-
Vibrio vulnificus	Isolate characterisation and epidemiological analyses.
Wolbachia	Isolate characterisation and population structure analyses
Xylella fastidiosa	Isolate characterisation and population structure.
Yersinia pseudotuberculosis	Population and evolutionary analyses.

MLST protocol:

MLST indexes the neutral variations in sequences of house keeping gene fragments. Since this involves use of multiple loci, MLST provides a robust platform for typing organisms that are subjected to frequent recombination instead of relying on phylogenetic analysis using a single locus. A more general protocol for MLST is shown in figure 2.4.

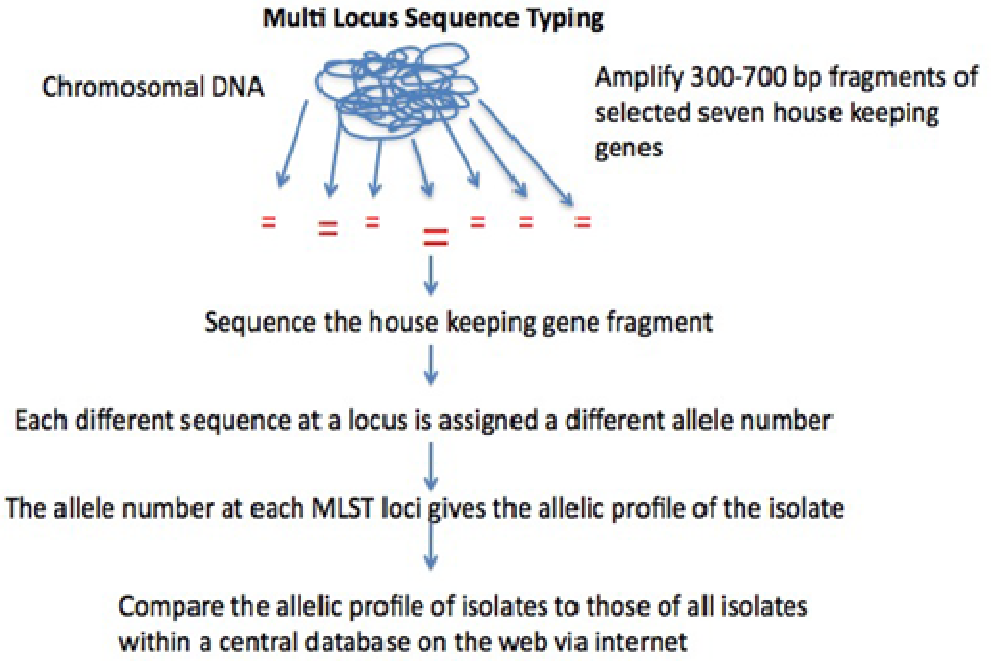


Figure 2.4: Schematic diagram of steps involved in MLST analysis.

MLST requires purification of chromosomal DNA from pure cultures and amplification of ~ 300-700 bp internal fragments of selected house keeping loci. This step calls for use of specific primers to ensure that a single PCR product is achieved for each locus. The resulting PCR fragments are purified and sequenced using the dideoxy-chain-termination (Sanger) method in both forward and reverse directions. Using the sequence alignment software ClustalX or MEGA (64, 62) sequences are assembled and aligned with a reference allele. Sequences are trimmed to

the desired length for comparison and examined for miscalled bases, insertions, gaps and any other sequence differences. Careful attention needs to be paid to as a single base pair change from a previously reported allele sequence will result in new allele number designation for the locus. Once allele numbers have been assigned to the sequenced house keeping loci for an isolate, combination of all the allele numbers will result in a ST for that isolate. Each ST represents a unique sequence of nucleotides and summarizes a large amount of nucleotide sequence data (69). Once the allele numbers and the corresponding STs have been determined for a collection of isolates, the resulting data can be analyzed using various software programs.

Data Analysis:

A large amount of data is generated by MLST and therefore making careful selection of a suitable method for data organization and interpretation is necessary. First relationships based on STs and nucleotide sequences are determined. An appropriate method is then determined by the structure of bacterial population under investigation. In populations where the predominant mechanism for intrastrain variation is recombination, analysis based on the allele assignment data (ST) can be employed by using various clustering algorithms like UPGMA (unweighted pair group method with arithmetic mean) (90), split decomposition (53) and eBURST (based upon related sequence types) (39). UPGMA algorithms are useful for examining a small number of isolates since they identify isolates that are closely related. Often it is difficult to resolve the deeper genetic relationship within a bacterial population with these methods. eBURST was developed to improve upon UPGMA for interpreting larger data sets by Feil et al (37). It is freely available as a java application on the MLST website (http://eburst.mlst.net). The eBURST algorithm provides interpretation of the origin and pattern of descent among bacterial isolates in a population without relying on constructing phylogenetic trees. This algorithm was specifically designed for MLST analysis. eBURST works on a model of recent bacterial evolution where it assumes that due to selection pressure some genotypes will occasionally increase in frequency within a population

and diversify by accumulating mutations or by horizontal gene transfer resulting in variants of the ancestor. It allows us to select two or more possible founder STs to view the evolution of that bacterial species. A clonal complex within a bacterial population resembles a phylogenetic network of linked STs with continued genetic extension resulting in a population snapshot of that bacterial population (39).

In clonal organisms where recombination is rare, an allele profile based analysis will likely be misleading. For these organisms it is advisable to use nucleotide sequence analysis either based on individual loci or a concatenated data set. Thus an important question to be addressed with any MLST data set is whether it is in agreement with a clonal model of population structure. If the organism is clonal, then the analysis becomes simple, and a conventional phylogenetic tree approach can be used. Clonality is tested by the congruence test (85) where the phylogenetic signals observed at different loci are similar or congruent. However, most bacterial data sets analyzed by MLST are nonclonal according to the congruence test. For such bacterial populations, the clonal complex is an ideal concept that groups genetically related organisms. A clonal complex is a group of related but not identical isolates that are likely to have been derived from a common ancestor. The MLST website has access to analysis software within the START2 package (http://pubmlst.org/software/analysis/) (55). Recent rapid developments in mathematical genetics have resulted in development of various tools and algorithms for analysis of MLST data to accommodate variable amounts of recombination events (MEGA, DnaSP, ClonalFrame, Ldhat, SplitTree) (62, 68, 27, 71, 53). Within these algorithms, various methods are incorporated to identify recombination events and identify the rate of recombination: Sawyer's Runs Test examines nucleotide sequence data to determine whether more consecutive identical polymorphic sites occur than would be expected by chance (30, 100). The maximum Chi-Squared Test is used to identify potential recombination events between sequences by comparing the distribution of polymorphic sites with those expected to occur by chance (73). The Index of Association ($\mathbf{I_A}$)

method measures the extent of linkage equilibrium within a population by quantifying the amount of recombination between a set of sequences and detecting association between alleles at different loci (53).

Influence of recombination on adaptive evolution of bacterial population:

Genetic recombination between different strains and species of bacteria plays a crucial role as a source of variation. Increasing rates of adaptive evolution with the help of recombinational events can provide a selective advantage over identical and fit populations that are unable to engage in horizontal gene transfer. Recombination estimation using MLST analysis has been done in *E.coli, Streptococcus pneumoniae, Haemophilus influenzae, Helicobacter pylori, Streptomyces* spp. (29, 35, 36, 39, 121). These studies suggest that rates of homologous recombination in some bacteria (e.g.: *S. pneumoniae, H. pylori*) exceed that of mutation by a factor of 10 or more.

Bacterial population structure has a strong influence on processes like environmental bottlenecks as well as periodic selection of adaptive clones. Population bottlenecks and selection reduce diversity and in the presence of recombination, increase linkage disequilibrium. This makes estimation of the rate of recombination difficult, as the estimated rates appear to be lower than the real recombination rates. As natural populations of many bacteria are composed of multiple lineages and are genetically variable at many loci, high rates of recombination will certainly influence the pace of adaptive evolution. Table 2.2 lists the inferred mean values of recombination over mutation (**r/m**) estimated for different bacteria and archaea (121). These authors noticed that a high variation in recombination rates exists in pathogenic bacteria. Among the truly free-living, non-animal associated bacteria marine and aquatic species have very high recombination rates in contrast to terrestrial bacteria where low to intermediate rates were observed. Currently most available data is for pathogenic or agriculturally important bacteria. As MLST is gaining popularity among microbial ecologists, more data can be expected on the

population structure of free-living, non-pathogenic bacteria. Scientists will be able to test how some ecological variables correlate with high rates of recombination. The current study was performed in order to answer these questions relating to the population structure of *Halomonas* spp. from an extremeand variable environment.

Table 2.2**:** The ratio of nucleotide changes as the result of recombination relative to point mutation **(*r/m*)** for different bacteria and archaea estimated from MLST data using ClonalFrame. Adapted from Vos and Didelot, 2009 (121).

Species	**Ecology**	**STs**	**loci**	**r/m**
Flavobacterium psychrophilum	Obligate pathogen	33	7	63.6
Pelagibacter ubique (SAR 11)	Free-living, marine	9	8	63.1
Vibrio parahaemolyticus	Free-living, marine (OP)	20	7	39.8
Salmonella enterica	Commensal (OP)	50	7	30.2
Vibrio vulnificus	Free-living, marine (OP)	41	5	26.7
Streptococcus pneumoniae	Commensal (OP)	52	6	23.1
Microcystis aeruginosa	Free-living, aquatic	79	7	18.3
Streptococcus pyogenes	Commensal (OP)	50	7	17.2
Helicobacter pylori	Commensal (OP)	117	8	13.6
Moraxella catarrhalis	Commensal (OP)	50	8	10.1
Neisseria meningitidis	Commensal (OP)	83	7	7.1
Plesiomonas shigelloides	Free-living, aquatic	58	5	7.1
Neisseria lactamica	Commensal	180	7	6.2
Myxococcus xanthus	Free-living, terrestrial	57	5	5.5
Haemophilus influenzae	Commensal (OP)	50	7	3.7
Wolbachia b complex	Endosymbiont	16	5	3.5

Campylobacter insulaenigrae	Commensal (OP)	59	7	3.2
Mycoplasma hyopneumoniae	Commensal (OP)	33	7	3
Haemophilus parasuis	Commensal (OP)	79	7	2.7
Campylobacter jejuni	Commensal (OP)	110	7	2.2
Halorubrum sp.	Halophile	28	4	2.1
Pseudomonas viridiflava	Free-living, plant pathogen	92	3	2
Bacillus weihenstephanensis	Free-living, terrestrial	36	6	2
Pseudomonas syringae	Free-living, plant pathogen	95	4	1.5
Sulfolobus islandicus	Thermoacidophile	17	5	1.2
Ralstonia solanacearum	Plant pathogen	58	7	1.1
Enterococcus faecium	Commensal (OP)	15	7	1.1
Mastigocladus laminosus	Thermophile	34	4	0.9
Legionella pneumophila	Protozoa pathogen	30	2	0.9
Microcoleus chthonoplastes	Free-living, marine	22	2	0.8
Bacillus thuringiensis	Insect pathogen	22	6	0.8
Bacillus cereus	Free-living, terrestrial (OP)	13	6	0.7
Oenococcus oeni	Free-living, terrestrial	17	5	0.7
Escherichia coli ET-1 group	Commensal (free-living?)	44	7	0.7
Listeria monocytogenes	Free-living, terrestrial (OP)	34	7	0.7
Enterococcus faecalis	Commensal (OP)	37	7	0.6
Porphyromonas gingivalis	Obligate pathogen	99	7	0.4
Yersinia pseudotuberculosis	Obligate pathogen	43	7	0.3
Chlamydia trachomatis	Obligate pathogen	14	7	0.3
Klebsiella pneumoniae	Free-living, terrestrial (OP)	45	7	0.3
Bordetella pertussis	Obligate pathogen	32	7	0.2

Brachyspira sp.	Commensal (OP)	36	7	0.2
Clostridium difficile	Commensal (OP)	34	6	0.2
Bartonella henselae	Obligate pathogen	14	7	0.1
Lactobacillus casei	Commensal	32	7	0.1
Staphylococcus aureus	Commensal (OP)	53	7	0.1
Rhizobium gallicum	Free-living, terrestrial	33	3	0.1
Leptospira interrogans	Commensal (OP)	61	7	0.02

Abbreviations: STs, sequence types; OP, opportunistic pathogen.

DNA Repair Potential of the *Halomonas* spp. from the GSP:

Microorganisms living in the GSP are continuously exposed to desiccating conditions, high surface temperatures, freezing winters, high salinity and direct UV radiation (16). Survival under these conditions may lead to selection for unique survival strategies. DNA repair mechanisms associated with UV light or chemically induced DNA damage have been shown to be important in protecting the microorganisms from desiccation (8). Because very little is known about the DNA repair capacity of the microorganisms from hyper saline terrestrial environments, I decided to study the DNA repair potential of the *Halomonas sp*. isolates, which are the dominant gram negative bacterial group from the GSP. Survival after exposure to UV light was used as a tool to assess the DNA repair capacity.

SPECIFIC AIMS

The Great Salt Plains (GSP) located in North Central Oklahoma is an ecologically diverse, extreme, hyper saline environment. Thus microorganisms living in these salt plains are continuously exposed to harsh environmental conditions such as high temperature, freezing winters, high salinity, high desiccating conditions and direct UV rays. Initially 16 isolates of culturable heterotrophic bacteria belonging to *Halomonas* spp. were identified from GSP soils using phenotypic characterization and phylogenetic analysis using 16S rRNA sequences (16). Later 16 additional isolates were added to this study. *Halomonas* spp. isolates were the dominant group among the gram-negative bacteria identified from GSP. Many differences in overall phenotypic characteristics were observed when compared with *Halomonas* spp. isolates obtained from abandoned saltern soils (16, this study). 16S rRNA sequence analysis identified these isolates as closely related to several different species of bacteria within the genus *Halomonas* even though the isolates had around 97% 16S rRNA sequence similarity. In an earlier evaluation of phylogenetic status of family *Halomonadaceae*, the authors found that most of the species belonging to the genus *Halomonas* had around 97% 16S/23S rDNA sequence similarity and very few of them had lesser percentage of sequence similarity. Furthermore the bacteria belonging to *Halomonas* have a range of G+C content (52-68 mol%) which is considered too wide as it is accepted that members of same genus should not have G+C content range >10 mol% (4). These observations led us to hypothesize that since the *Halomonas* spp. at GSP are constantly subjected to extreme environmental conditions, this has led to the development of adaptive mechanisms over a long period of time that elevate genetic diversity within the phylotype and hence increased speciation**.** In order to study the fine genetic structure of this bacterial population from GSP soils,

I decided to use Multi Locus Sequence Typing (MLST) approach. MLST schemes have been extensively used to study many bacterial species, especially pathogenic bacteria. Only a handful of MLST studies have been reported for environmental isolates of bacteria. In order to test our hypothesis, the following aims were developed:

1. To develop a MLST scheme for *Halomonas* spp. isolates from GSP.
2. To provide insight into the genetic separation of isolates belonging to the genus *Halomonas* from GSP soils
3. To test whether this population structure is unique to the GSP environment.

In addition we believe that the *Halomonas* spp. may have developed certain mechanisms to protect themselves from constantly exposure to harsh, rapidly changing environment at GSP. It has been reported that desiccation produces the same kind of cellular damage on microorganisms as exposure to high UV irradiation (8). Therefore UV radiation can be used as an effective tool to study the effect of desiccation on these microorganisms. Most of the *Halomonas* spp. isolates from GSP showed greater levels of resistance to the DNA damage caused by the UV irradiation (125). This was greater than the UV resistance observed for the isolates from hyper saline aquatic environments. Also, these isolates reported a higher spontaneous mutation rates following UV-C exposure. Thus, we hypothesize that *Halomonas* spp. bacteria surviving under extreme desiccating conditions at GSP might have developed an efficient constitutively expressed DNA repair mechanism**.** The gene product RecA positively controls the regulation of this type of DNA repair system. In order to understand the molecular mechanism of the DNA repair system, the following aim was developed:

To evaluate the *Halomonas* spp. isolates from GSP for their survival abilities against UV radiation and phylogenetic evaluation of the *recA* gene sequences from *Halomonas* spp. isolates

Dissertation Organization:

This dissertation consists of six chapters: Review of literature, specific aims and abstract, three different chapters focusing on the research findings of this work and an overall conclusion. Each of the three research chapters are written in a format consisting of introduction, methods, results and discussion.

CHAPTER III

DEVELOPMENT OF MLST SCHEME AND RECOMBINATION ANALYSIS

ABSTARCT

The Great Salt Plains (GSP) is an ecologically diverse, hyper-saline environment. Many differences in overall phenotypic characters were observed among *Halomonas* spp. isolates from the GSP that had >97% 16S rDNA sequence similarity. We believe these *Halomonas* spp. isolates have evolved over a long period of time by means of developing certain genetic mechanisms. Multi Locus Sequence Typing (MLST) was used to study the population structure of these isolates. Housekeeping genes *ectB, gyrB,* and *recA* along with 16S rDNA and the ITS region were assayed in 32 Halomonas spp. isolates.MLST analysis identified 28 unique sequence types indicating a high level of genetic diversity. Among the house keeping genes, *recA* demonstrated increased resolution indicating that *Halomonas* spp. isolates contain several genetically distinct phylotypes most likely due to environmental bottlenecks and recombination events. Evidence for frequent LGT and recombination among these isolates was detected. Recombination rates were on par with other sexual bacteria. The rates of recombination observed are high enough to overcome the effects of population bottlenecks and periodic selection that would otherwise purge the genetic variation generated. These results support the hypothesis that a high level of speciation has occurred among *Halomonas* spp. isolates at the GSP site even though 16S rDNA sequence similarity has been conserved. Phylogenetic relationship of very closely related *Halomonas* spp. isolates was resolved by MLST.

INTRODUCTION

Lateral gene transfer (LGT) plays a crucial role in evolution and creating genetic diversity (5). Theories developed to understand the ecological patterns of microorganisms focus on the distribution and structure of diversity within microbial populations (94). Although it is believed that LGT is rare in bacteria, it is frequent enough to allow rapid acquisition of novel functions needed to promote adaptation and speciation.

Establishment of genetic barriers between closely related bacteria is a necessary condition for speciation to occur. Bacteria are promiscuous and homologous recombination occurs even between bacteria that have as much as 25% sequence divergence (92). Recombinational events in bacteria are capable of relaxing the diversity purging effects of periodic selection and preserve the genetic diversity (41). By providing entirely new functions, heterologous gene transfer places the emerging ecotype out of reach of any periodic selection to which the parental ecotype may be susceptible (65). Since bacteria are capable of withstanding rapid changes in environmental conditions, high rates of speciation would play a crucial role in influencing temporal diversity patterns within a community (19). The diversity observed in bacterial sequences of environmental isolates reveals the constraints placed by forces such as population bottlenecks and barriers to gene exchange. When these two factors maintain equilibrium, it leads to speciation and structure within a bacterial community (24). It is likely that the number of loci actively involved in differentiating ecological adaptation is small and thus, recombination barriers depend on the accumulation of variations at neutral loci rather than adaptive loci. It has been suggested that strains within an ecotype have just enough time between selective sweeps to accumulate differences at one or two loci due to mutation or recombination (18).

An important criterion for identifying species clusters of evolutionary significance is to find characteristic features (ecological or genetic) that distinguish one phylotype from its close

relatives. One way to do this is to place them onto an environmental resources map to show divergent but closely related groups of bacteria (41). For example, fine scale resource partitioning has been observed between coastal *Vibrio* isolates obtained from ecologically distinct samples. These data suggest a recent divergence from the sympatric ancestral population (41). However it is still unclear whether the differentiation of bacterial populations based on genetic-ecological patterns is maintained by LGT-recombination events and selection. The rates of LGT can vary greatly among populations of bacteria. The majority of the data available is for pathogens (121).

Sequence based population structure analysis methods are gaining popularity among microbial ecologists. This has resulted in large datasets that will help us to better understand evolutionary adaptations of free-living, non-pathogenic bacterial populations. Molecular typing is one such approach that includes techniques such as pulsed field gel electrophoresis (PFGE), Restriction Fragment Length Polymorphism (RFLP), Randomly Amplified Polymorphic DNA (RAPD) and Repetitive Element PCR (69). However, these techniques have less discriminatory power when it comes to defining underlying phylogenetic relationships.

Multi Locus Sequence Typing (MLST) (70) that involves the use of several housekeeping genes was developed to study the population and evolutionary genetic structure of bacteria and to explore the contribution of recombination in creating variations in housekeeping genes under study (36, 38). This helps to eliminate the focus on using a single gene in such studies.

Halomonas belongs to the family *Halomonadaceae* (class: *Gammaproteobacteria*) and is currently comprised of more than fifty species (**25**). Members of this genus are Gram-negative, rod-shaped, non-sporulating chemoorganotrophs with mostly respiratory metabolism. Most *Halomonas* species are moderate halophiles with a few halotolerants. Species macro restriction fingerprinting (PFGE) has been successfully used to clarify the taxonomic position of isolates

within the genus *Halomonas.* The genome size of eleven *Halomonas* strains tested ranged from 1.45 to 2.83 Mbp (33).

Very little work has been done on the genetic aspects of *Halomonas* species isolates due to lack of suitable genetic tools (80). Natural transformation has not been reported and transduction methods have not yet been developed even though bacteriophages specific for *Halomonas* spp. that are capable of interspecies gene transfer have been reported (102, 120). Conjugation is the only mechanisms available for genetic manipulation of *Halomonas* bacteria (80). Although members of the genus share a high 16S rRNA sequence similarity (~97%), the genus *Halomonas* is not monophyletic and is comprised of two clearly separated phlogenetic groups**.** The sequence similarity values between the strains included in group 1 and those in group 2 ranged from 96.1-93.2% for 16S rRNA and 95.3-93.5% for 23S rRNA (25). The remaining species assigned to genus *Halomonas* do not fall into either of these two phylogenetic groups. There are considerable differences in the topologies of phylogenetic trees for *Halomonas* spp. based on 16SrRNA and 23SrRNA sequences. The difference in topologies suggests that recombinational events have occured in the rRNA genes.

Analysis using both 16S rRNA and 23S rRNA does not help to resolve the phylogenetic relationship between very closely related *Halomonas* species (25). In order to overcome this limitation and to provide novel insights into the evolutionary relationship between the members of genus *Halomonas*, recommendations have been made to include studies involving sequencing of house keeping genes or other protein-coding genes (MLST) (25). Okomoto et al. (79) reported an attempt at phylogenetic analysis of *Halomonas variabilis and* related organisms based on 16S rRNA, *ectBC*, and *gyrB* sequences. Isolates tested shared similar physiological and phylogenetic characteristics among all the genes except *ectB.*

The Great Salt Plains (GSP) located in north-central Oklahoma is an ecologically diverse, extreme, hypersaline environment (16). The salt concentration varies considerably across the plains and during different seasons. Microorganisms living in these plains are continuously exposed to harsh environmental conditions such as high temperature, freezing winters, high salinity, desiccation and direct UV radiation (125). Partha Dutta (33) studied the microbial population at GSP using polymerase chain reaction (PCR) and denaturing gradient gel electrophoresis (DGGE). Results from his study suggest that the constantly varying environmental factors at the GSP select for different microbial communities in different years and at different salt gradient sites. Several *Halomonas* spp. isolates have been identified from the GSP using heterotrophic enrichments and 16S rRNA sequence analysis (16). Many differences in overall phenotypic characteristics were observed when compared with isolates obtained from abandoned saltern soils (95). Very few GSP isolates are motile, oxidase positive, do not produce H_2S, and grow at temperature >45°C, few are capable of anaerobic growth and are less halophilic (0.1 to 30%). *Halomonas* spp. isolates group into several phylotypes on the basis of 16S rRNA sequence analysis indicating high speciation within this group of bacteria at GSP (16).

These observations led to the hypothesis that *Halomonas* spp. in the GSP have developed certain mechanisms allowing them to adapt to the extreme conditions at the GSP. We proposed that widespread gene loss and acquisitions has occurred during the evolution of the *Halomonas* spp. most likely through recombination following LGT mediated transduction or conjugation. Seaman and Day(102) studied two novel bacteriophages isolated from GSP and reported that one of these phages has an unusually large genome (~340 kb) with the capability of inter-species gene transfer among *Halomonas* spp. Conjugation has been successfully used to transfer plasmid derived vectors and broad host range plasmids in *Halomonas* spp.

In order to test our hypothesis and to estimate the rate of recombination among closely related *Halomonas* spp. at the GSP, we employed MLST. The main objective of our study was to better

understand the mechanism of evolutionary adaptation of *Halomonas* spp. to the extreme environmental conditions at GSP. A collection of 32 *Halomonas* isolates from GSP were used to (i) develop a MLST scheme for *Halomonas* isolates; (ii) provide insight into the genetic separation of these isolates; and (iii) identify niche specific phenotypic and genotypic characters. To our knowledge, this is the first MLST based approach to examine interspecies relationships among bacteria belonging to the genus *Halomonas*.

METHODS

Site description and Isolate selection:

Bacterial strains used in this study are listed in Table 3.1. Isolates in Table 3.1(GSP 1- GSP 58) were obtained in 2001-2002 from two different dry, salt-crusted sites WP3 (N 36° 42.856' and W 98° 15.725') and WP8 (N 36° 42.750' and W 98° 15.584) (Figure 3.1).

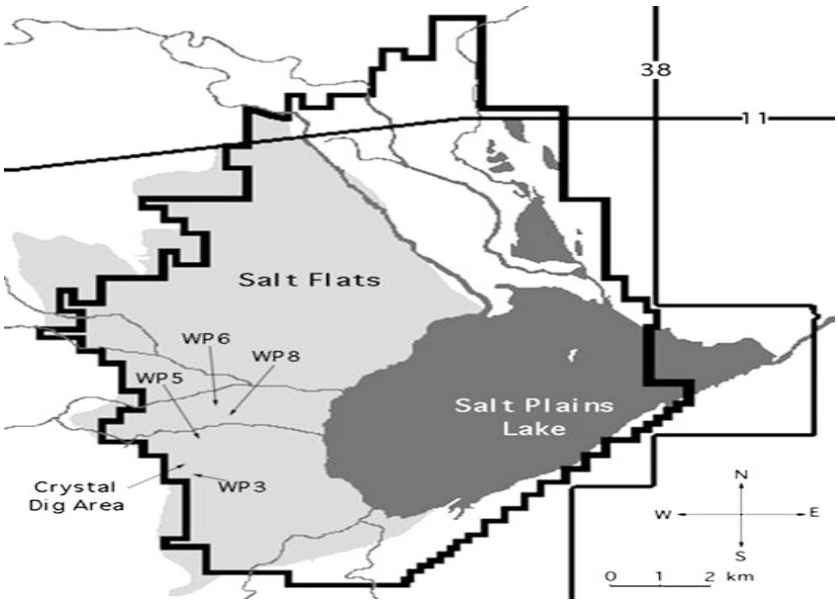


Figure 3.1: Map showing the GSP area and sampling sites at the GSP.

Strains from Table 3.1(GSP 1001 – GSP 1018) were obtained in 2005 from a salt gradient site WP68 (N 36° 42.856' and W 98° 15.725'). *Chromohalobacter salexigenes* DSM 3043 was also included in the study since this is the only close relative of *Halomonas* spp. for which a complete genome sequence is available. Isolates were obtained using enrichment cultures and were grown and maintained on SP-1 medium (16). Isolates were streaked several times on fresh SP-1 agar plates to ensure purity. They were identified as *Halomonas* based on phenotypic (microscopic, gram characteristics and biochemical tests) and phylogenetic (16S rRNA gene sequence analysis) characterization (16 and this study). Based on this information, 32 *Halomonas* spp. isolates from GSP were included in this study (Table 3.1).

MLST of *Halomonas* species:

Genomic DNA was extracted following the method of in Caton et al.(16). Genomic DNA was the target for PCR amplification of selected house keeping genes: ectoine synthase (*ectB*), DNA gyrase subunit B (*gyrB*) and recombinase A (*recA*) along with 16S rRNA gene and the 16S-23S internal transcriber spacer (ITS) region. Primer sequences used for PCR amplification are listed under Table 3.2. Primers for the *gyrB* and *recA* genes were designed based on the genome sequence information available for *C. salexigenes* using OligoPerfect™ Designer (Invitrogen). Primers for 16SrRNA, the ITS region and *ectB* were selected from previous publications (Table 3.2).

Amplification and Sequencing:

Each primer set had different reaction conditions. Briefly the PCR parameters were as follows- For 16S rRNA: DNA was denatured at 95° for 2 min, followed by 40 cycles of 95°C for 1 min, 50°C for 1 min, and 72°C for 1 min, with a final extension period of 5 min (16). For *ectB*, initial denaturation of 94°C for 4 min, followed by 35 cycles of 94°C for 20 s, 58°C for 40 s, 72°C for 40s, and final extension of 8 min. For *gyrB* and *recA*: initial denaturation

Table 3.2: Primers used for MLST of the *Halomonas* species

Locus	Gene product	Pimer (5'→3')	Primer Source
16S rRNA	16S ribosomal RNA	AGAGTTTGATCCTGGCTCAG	Caton et al(2004)
ectB	Ectoine synthesis	AAGGAGGTGATCCAGCCGCA ACGGCACGTTCCGTGGTTT	Okamoto et al(2009)
gyrB	DNA binding protein	GTCTTCACGGTGTACTTCGTTACC GAC GGA AGA AGA ACG TCA GC CGA AGT TTT CTT CGC AGA CC	This study
ITS	Internal transcribed spacer region	TTGTACACACCGCCCGTC	Su et al (2009)
recA	DNA repair and recombination protein	TTCGCCTTTCCCTCACGGTA AGG CGT ACG CTG GAG TAG AA GGT GAA CAG GCA CTG GAA AT	This study

of 5 min at 94°C followed by 35 cycles of 94°C for 2 min, 57±3°C for 1 min, 72°C for 2 min, and final extension of 10 min at 72 °C (this study). For ITS: Initial denaturation of 94°C for 5 min, 8 cycles of 94°C for 2 min, 55°C for 1 min, 72 °C for 1 min followed by 30 additional cycles of 94°C for 1 min, 60°C for 1 min, and 72°C for 1 min. PCR was carried out in a 50 µl reaction mixture containing: 2 µl of genomic DNA, 1X PCR buffer (Promega), 2.5 mM $MgSO_4$, 0.2 mM deoxynucleside triphosphate mixture, 2.5 U of GoTaq Flexi DNA polymerase (Promega, Madison,WI) and 10 µM of each of the forward and reverse primers. PCR products of the appropriate sizes were purified using PureLink™ PCR Purification Kit (Invitrogen) and DNA sequences of the PCR amplified gene fragments were generated in both directions on an ABI Model 3730 DNA Analyzer (Applied Biosystems) using BigDye™"-terminated reactions at the Oklahoma State University DNA sequencing core facility.

Sequence data analysis:

Both forward and reverse raw sequence files were manually inspected and assembled using MEGA 4.0 (62). All the sequences were confirmed through BLAST searches to determine

whether the resultant data matched with the correct gene in the database. Multiple sequence alignments were obtained using CLUSTALX (64) for all sequences. A database for *Halomonas* spp. sequences (http://pubmlst.org/halomonas/) was created by Keith Jolley (University of Oxford, UK). Alleles and sequence types (STs) generated in this study are deposited in the database. Each different allele within a locus was assigned a different number and the combination of all the numbers resulted in a ST for a strain (Table 3.1). These data were used for subsequent analysis.

Pairwise distances between individual and concatenated nucleotide sequences were determined using PAUP* 4.0 (113). Estimations for parameters of the number of polymorphic sites, the average pair-wise nucleotide differences per site (π), GC content and Tajima's D, a test for alleles departing from neutral evolution were calculated using DnaSP 5 (68). Standardized index of association (I^S_A), a measure of linkage disequilibrium and ratios of non-synonymous to synonymous substitutions (d_N/d_S), a measure of purifying selection were estimated using START2 (55). The Index of Association (IA) is calculated as follows: IA =VO/VE -1. If V0 is the observed variance of K and VE is the expected variance of K, where K is the number of loci at which two individuals differ.

Measure of recombination rate and detection of recombination:

Rate of recombination relative to point mutation (r/m) was calculated using ClonalFrame (27). ClonalFrame was run with 100,000 burn-ins followed by 100000 more iterations. Population scale rates of recombination (ρ/θ_W) were estimated with LDhat 2.1 (71). θ_W is the average mutation rate while ρ is the rate of gene conversion, $\rho = 2N_e r/2$. Calculations were performed for individual isolates and for the concatenated data set. Individual recombination events within the concatenated data set were detected using the recombination detection program (RDP) 3.0 (51) with all the methods at default settings. *P*-values were calculated using the Bonferroni correction

implemented within the program. Pair-wise homeoplasy index ($\mathbf{\Phi_W}$) (phi test for recombination) values and Neighbornet phylogenetic network trees for the concatenated data set were estimated using Splitstree v4.10 (53).

Phylogenetic analysis:

Using the hierarchical ratio test within the MODELTEST version 3.7 (91), the most appropriate model of DNA substitution for each locus was determined. Neighbor –Joining (NJ) trees were constructed using the program PAUP* 4.0 (113) with Jukes-Cantor correction. Maximum likelihood (ML) trees were inferred for each individual loci and the concatenated data set using PhyML v3.0 (46) with DNA substitution model as suggested by MODELTEST**.**

Population structure analysis:

Allelic profile data was used to generate groups, clonal complex (CC) of each group and to understand the allelic profile diversity between the isolates using eBURST (39). eBURST classifies the sequence types as groups of single-locus variants (SLVs), double locus variants (DLVs) or singletons (STs differing at three or more loci). The external/internal branch length ratio test within the Clonal Frame GUI (27) was used to understand the population snapshot of the *Halomonas* spp. isolates. This test calculates the ratio of the sum of the external branches (that connect a leaf of the tree) to the sum of the internal branches (that connects two internal nodes of a tree). This program also computes the expected distribution of the external to internal branch length ratio under a coalescent model. If the observed value is significantly smaller than expected, it suggests that the genealogy under test is “star like”, which is in agreement with either a recent expansion of population size due to a population bottleneck or a selective sweep.

RESULTS

Identification of *Halomonas* spp. isolates:

Isolates were confirmed as *Halomonas* spp. based on physiological characterization and 16S rRNA sequence analysis. Caton et al. (16), previously identified 16 isolates listed in Table 3.1 (GSP-1 to GSP-58). Among those isolates, GSP-58 belongs to genus *Chromohalobacter*. Sixteen additional isolates (tables 3.1 and 3.3) were characterized in this study (GSP 1001- GSP 1018). After this study was initiated, *H.indialinina, H. marisflavi and H. avicenniae* were reclassified into genus *Kushneria* by Sanchez-Porro et al (99). Thus, GSP 1015 and GSP 1016 belong to the genus *Kushneria.* Ribosomal RNA sequence data obtained in this study were compared to that of several type strains within the genera *Halomonas, Chromohalobacter and Kushneria* obtained from Genbank. A neighbor-joining dendrogram was generated using *Bacillus haloalkaliphilus* and *Pseudomonas halophila* as outgroups (figure 3.2). Within the genus *Halomonas,* isolates group into two clearly separated phylogenetic clusters as suggested by de la Haba et al.(25). GSP isolates are distributed in several clusters throughout the tree indicating higher speciation.

Table 3.1: *Halomonas* species isolates used in the study and the distribution of STs and allele profiles

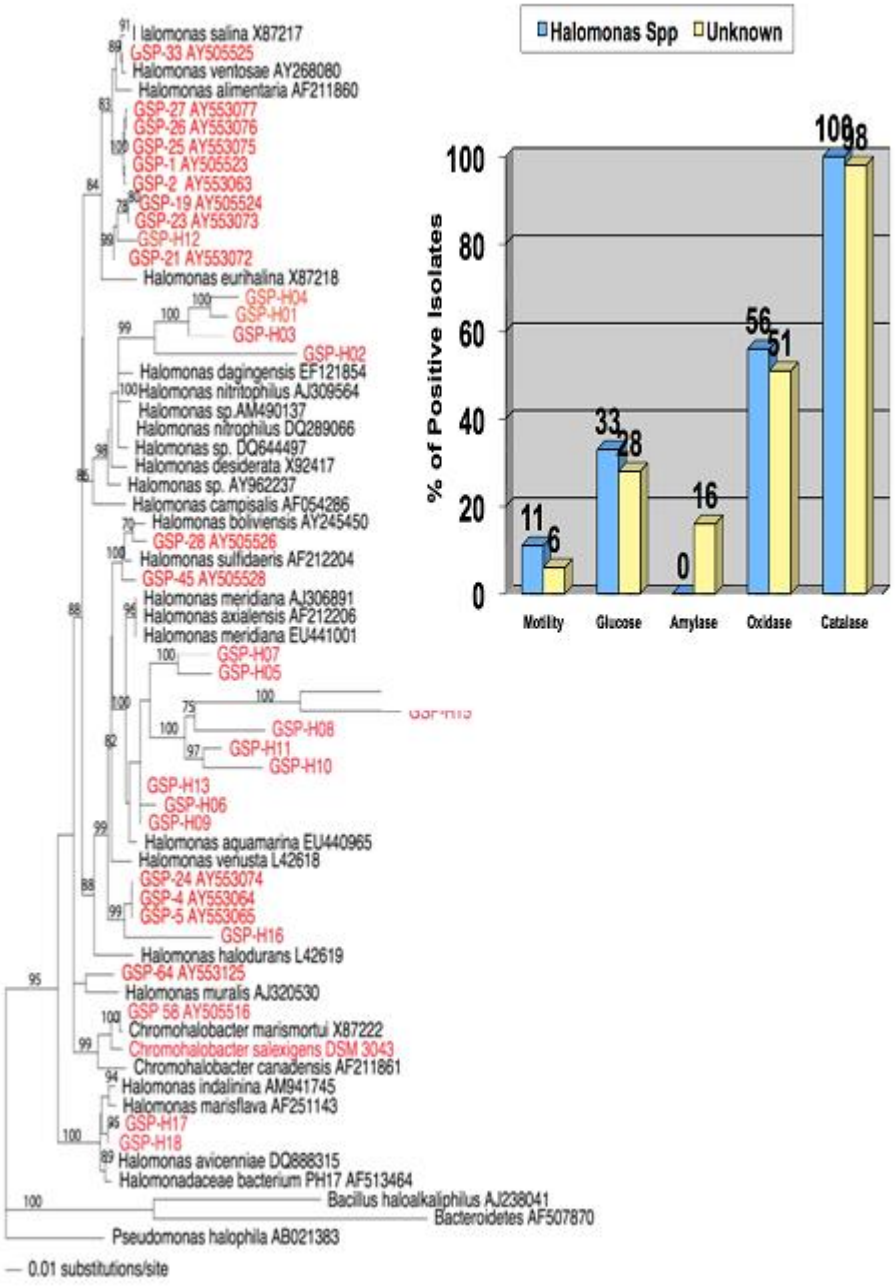
Isolate	Country	Region	Year	Source	Species	STs and Allele profiles					
						STs	16S rRNA	*ectB*	*gyrB*	ITS	*recA*
GSP1001	USA	Oklahoma	2005	terrestrial	*Halomonas spp.*	1	1	1	1	1	1
GSP1002	USA	Oklahoma	2005	terrestrial	*Halomonas spp.*	2	2	1	1	2	2
GSP1003	USA	Oklahoma	2005	terrestrial	*Halomonas spp.*	3	1	1	2	1	3
GSP1004	USA	Oklahoma	2005	terrestrial	*Halomonas spp.*	4	1	1	3	1	4
GSP1005	USA	Oklahoma	2005	terrestrial	*Halomonas spp.*	5	3	1	4	3	5
GSP1006	USA	Oklahoma	2005	terrestrial	*Halomonas spp.*	6	3	1	4	3	6
GSP1007	USA	Oklahoma	2005	terrestrial	*Halomonas spp.*	7	3	1	4	3	7
GSP1008	USA	Oklahoma	2005	terrestrial	*Halomonas spp.*	8	4	1	4	3	8
GSP1009	USA	Oklahoma	2005	terrestrial	*Halomonas spp.*	5	3	1	4	3	5
GSP1010	USA	Oklahoma	2005	terrestrial	*Halomonas spp.*	9	3	1	4	3	9
GSP1011	USA	Oklahoma	2005	terrestrial	*Halomonas spp.*	10	5	1	4	4	10
GSP1012	USA	Oklahoma	2005	terrestrial	*Halomonas spp.*	11	6	1	5	5	11
GSP1013	USA	Oklahoma	2005	terrestrial	*Halomonas spp.*	5	3	1	4	3	5
GSP1016	USA	Oklahoma	2005	terrestrial	*Halomonas spp.*	12	7	1	5	6	12
GSP1017	USA	Oklahoma	2005	terrestrial	*Halomonas spp.*	13	8	1	5	6	13
GSP1018	USA	Oklahoma	2005	terrestrial	*Halomonas spp.*	14	8	1	5	7	14
GSP1	USA	Oklahoma	2002	terrestrial	*Halomonas spp.*	16	10	3	5	9	16
GSP2	USA	Oklahoma	2002	terrestrial	*Halomonas spp.*	17	10	3	5	9	17
GSP3	USA	Oklahoma	2002	terrestrial	*Halomonas spp.*	18	11	1	5	10	18
GSP4	USA	Oklahoma	2002	terrestrial	*Halomonas spp.*	18	11	1	5	10	18
GSP5	USA	Oklahoma	2002	terrestrial	*Halomonas spp.*	18	11	1	5	10	18
GSP19	USA	Oklahoma	2002	terrestrial	*Halomonas spp.*	19	12	4	6	10	19
GSP21	USA	Oklahoma	2002	terrestrial	*Halomonas spp.*	20	12	1	5	10	20
GSP23	USA	Oklahoma	2002	terrestrial	*Halomonas spp.*	21	12	4	7	11	19
GSP24	USA	Oklahoma	2002	terrestrial	*Halomonas spp.*	22	11	1	8	10	21
GSP25	USA	Oklahoma	2002	terrestrial	*Halomonas spp.*	23	10	1	5	9	22
GSP26	USA	Oklahoma	2002	terrestrial	*Halomonas spp.*	24	10	1	9	9	23
GSP27	USA	Oklahoma	2002	terrestrial	*Halomonas spp.*	25	10	1	5	9	24
GSP28	USA	Oklahoma	2002	terrestrial	*Halomonas spp.*	26	11	1	9	12	25
GSP33	USA	Oklahoma	2002	terrestrial	*Halomonas spp.*	27	13	1	2	13	26
GSP45	USA	Oklahoma	2002	terrestrial	*Halomonas spp.*	28	11	1	2	14	27
GSP58	USA	Oklahoma	2002	terrestrial	*Halomonas spp.*	15	9	2	5	8	15
DSM3043	Netherlands	Bonaire Island	1974	terrestrial	*Chromohalobacter salexigenes*	15	9	2	5	8	15

Table 3.3: Results of phenotypic assays performed on the *Halomonas* species isolates from GSP

Isolate	Gram reaction	Sulfide reduction	Indole production	Test for motility	Glucose Fermentation	Starch Hydrolysis	Oxidase test	Catalase test
GSP1001	-	-	-	-	+	-	+	+
GSP1002	-	-	-	-	+	-	+	+
GSP1003	-	-	-	-	+	-	+	+
GSP1004	-	-	-	-	+	-	+	+
GSP1005	-	-	-	-	+	-	+	+
GSP1006	-	-	-	-	-	-	+	+
GSP1007	-	-	-	-	-	-	+	+
GSP1008	-	-	-	+	+	-	-	+
GSP1009	-	-	-	-	-	-	-	+
GSP1010	-	-	-	+	+	-	+	+
GSP1011	-	-	-	-	-	-	+	+
GSP1012	-	-	-	-	-	-	+	-
GSP1013	-	-	-	-	-	-	-	+
GSP1014	-	-	-	-	-	-	+	+
GSP1015	-	-	-	-	-	-	+	+
GSP1016	-	-	-	-	-	-	-	+
GSP1017	-	-	-	-	-	-	+	+
GSP1018	-	-	-	-	-	-	-	+
GSP 1	-	-	-	-	-	-	+	+
GSP 2	-	-	-	-	-	-	+	+
GSP 3	-	-	-	-	-	-	-	+
GSP 4	-	-	-	+	+	-	-	+
GSP 5	-	-	-	-	-	-	-	+
GSP19	-	-	-	-	-	-	+	-
GSP21	-	-	-	-	+	-	+	+
GSP23	-	-	-	-	-	-	+	-
GSP24	-	-	-	-	+	-	-	+
GSP25	-	-	-	-	-	-	+	+
GSP26	-	-	-	-	-	-	+	+
GSP27	-	-	-	-	-	-	+	+
GSP28	-	-	-	-	-	-	+	+
GSP33	-	-	-	-	-	-	+	+
GSP45	-	-	-	-	-	-	+	+
GSP58	-	-	-	-	-	-	-	+

+ indicates a positive reaction and – indicates a negative reaction.

Figure 3.2: NJ phylogenetic tree of 32 *Halomonas* species from GSP (shown in red) based on the 16S rRNA sequence analysis. The chart shows the results of physiological tests comparing previously identified isolates (16) to the ones identified in this study.



MLST fragments and allelic profiles:

The fragment lengths (obtained from sequencing PCR amplified product) for 16S rRNA, ITS and three housekeeping genes are listed in Table 3.4**.** Allele numbers (each sequence variant of a given gene) and the corresponding profiles (STs) were assigned upon depositing the sequences at the pubMLST website (http://pubmlst.org/halomonas/). Since this is the first MLST database for *Halomonas* species bacteria, all the profiles obtained are new and do not correspond to any other database. A total of 28 STs were detected for the 33 isolates in the collection. Thirteen, 4, 9, 14 and 27 alleles were identified for 16S rRNA, *ectB, gurB,* ITS and *recA* genes, respectively (Table 3.1). Both ITS and *recA* genes were highly polymorphic (64.44% and 57.33% respectively) i.e., they had increased levels of allelic variations among all the loci tested. Identical alleles for each of the genes were recognized in some isolates (GSP 1005, 1009, 1013; GSP 3, 4, 5; and GSP58, DSM 3043) that are grouped close to each other on the 16S rRNA based NJ tree (figure 3.2)**.** Except for GSP-58 and DSM 3043, isolates that share identical allelic profiles were obtained from a same sampling site.

Table 3.4: Sequence characteristics of the loci used in MLST of *Halomonas* spp.

Loci	Length	Number of alleles	Average distance(%)	Polymorphic sites(%)	Nucleotide diversity (π)	d_N/d_S	Tajima's D	r/m	ρ/θ_W
16SrRNA	711	13	0.041(95.9)	217(30.5)	0.08013	NA	-1.59241 NS	1.59	0.173
ectB	399	4	0.065(93.5)	55(13.78)	0.0919	1.023	-1.72648 NS	1.61	0.178
gyrB	495	9	0.046(95.4)	138(27.88)	0.11616	0.3424	-1.27434 NS	4.27	0.249
ITS	270	14	0.134(86.6)	174(64.44)	0.21474	NA	-1.14077 NS	8.39	1.106
recA	300	27	0.149(85.1)	172(57.33)	0.14025	0.193	-1.27476 NS	14.41	2.37
Concat	2175	28	0.109(89.1)	757(34.80)	0.08437	NA	-1.04198 NS	18.5	1.53

NA- not applicable, NS- values for Tajima's D were non-significant, π- Nucleotide diversity/site, dN/dS- Ratio of Non-synonymous substitutions Vs Synonymous substitutions, r/m – Rate of recombination relative to point mutation, ρ/θ_w – Ratio of rates at which recombination and mutation occur.

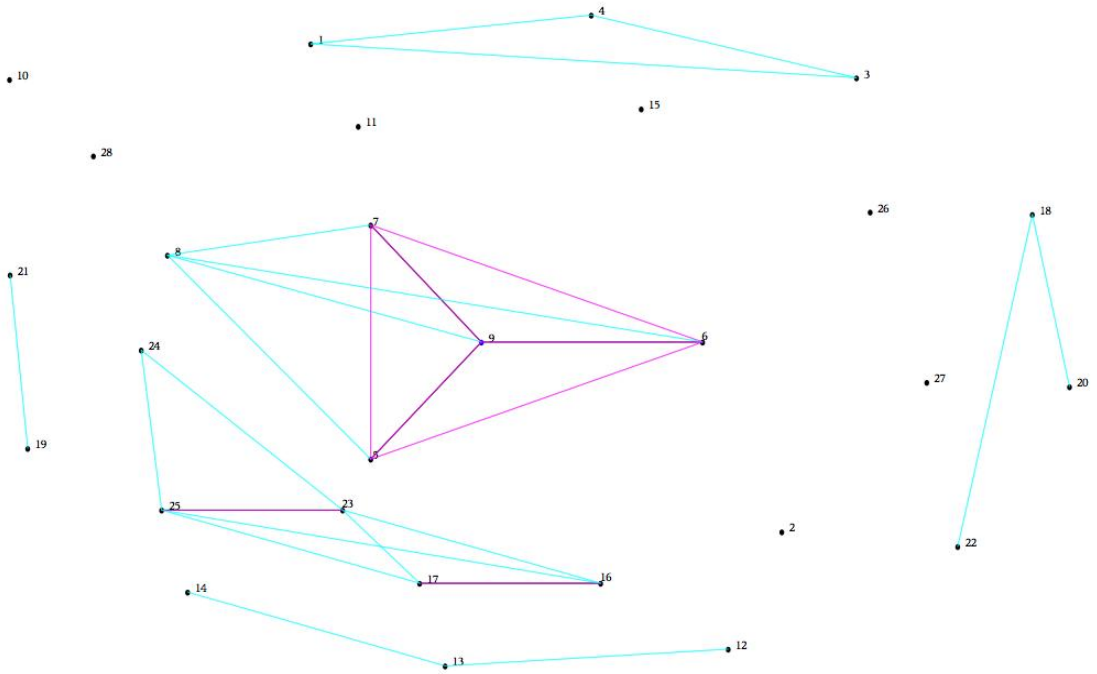
Sequence characteristics:

Detailed characteristics of nucleotide sequences for each of the loci tested are presented in Table 3.4. The average interspecies (GSP isolates) sequence similarity values for each of the gene were: 16S rRNA (95.9%), *ectB* (93.5%), *gyrB* (95.4%), ITS (86.6%), *recA* (85.1%). For the concatenated gene sequences (2175 bp), the average similarity value was 89.1%. ITS region sequences are known to be highly variable and hence the low sequence similarity value between isolates is not surprising. However, the value obtained for the *recA* gene is unusual since this gene is thought to be highly conserved among different genera. All the loci tested varied greatly in their nucleotide sequence variations. More polymorphic sites were detected in ITS and *recA* genes (64.4% and 57.3% of total length) than in the other genes tested. The nucleotide diversity (**π**), the average number of nucleotide differences per site between two randomly selected sequences values were high (0.08013 -0.21474) indicating the diverse nature of the ITS and *recA* loci. The *recA* gene had the lowest d_N/d_S ratio (<0.5) of 0.1932 indicating the presence of a selective pressure restricting amino acid changes. Tajima's D values were not significant for any of the loci tested (-1.27476 to –1.72648) suggesting neutral selection of these loci.

Population structure of *Halomonas* spp.:

A population snapshot and pattern of evolutionary relationships between *Halomonas* spp. isolates from the GSP were prepared and analyzed using the eBURST algorithm (39). This program generates clonal complexes (CC) by linking single/double locus variants based on a recent ancestor model. It assumes that, due to selection pressure, certain genotypes will increase in frequency in a population and diversify by accumulating mutations or by gene exchange events (37). This test revealed the presence of seven unlinked individual sequence types (ST) and six CC's (linked ST's) with varying numbers of single (8) and double (18) locus variants (figure3.3) suggesting that the *Halomonas* spp. have evolved as result of genetic recombination rather than

Figure 3.3: eBURST generated figure showing association between STs. Purple lines connect STs which are similar in 4 or more alleles. Blue lines connect STs with double (DLV) or triple (TLV) locus variants. All the STs which are not grouped are singletons (no matching alleles).

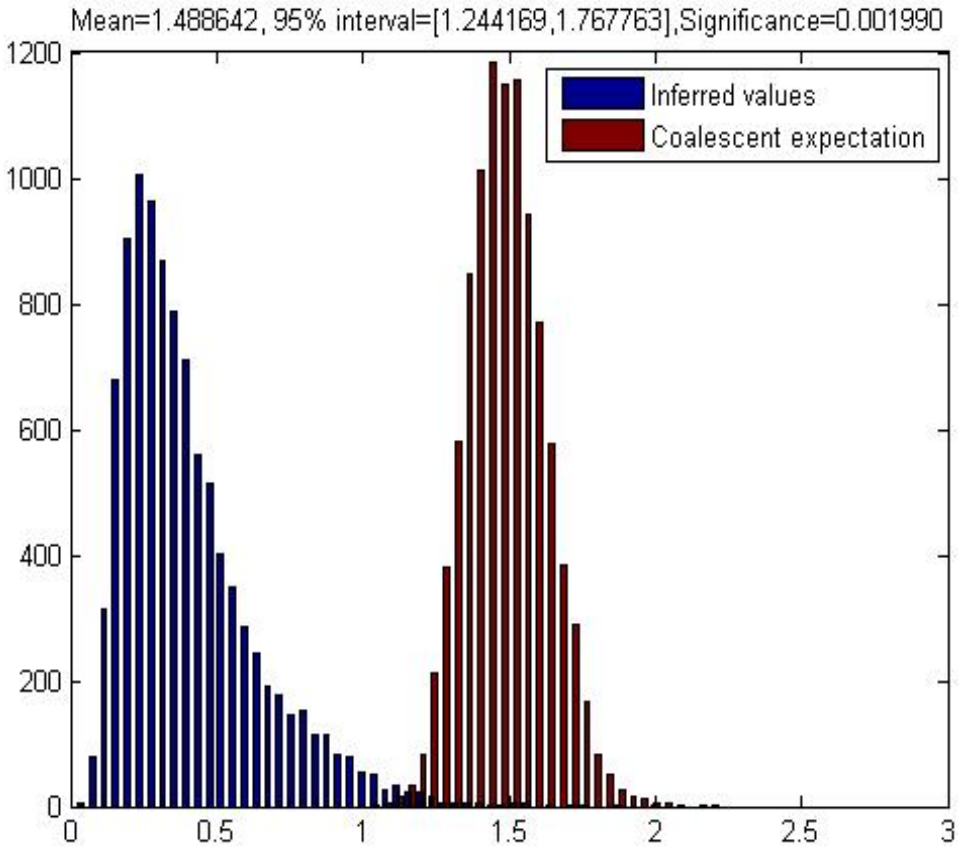


by accumulating point mutations as they are distantly related and the population does not appear to be structured into highly dominant clonal complexes. The internal-to- external-branch-length-ratio test showed that *Halomonas* spp. exhibit a significantly lower ratio (0.3, $P = 0.00199$) than expected under the coalescent model (figure 3.4). This suggests that this population has gone through a recent expansion in size that is consistant with a recent population bottleneck. The high level of genetic diversity observed within the gene sequences, even in the absence of selection could be explained by LGT that results in rapid diversification of the alleles (26).

Recombination among *Halomonas* spp. isolates:

In order to estimate the comparative impacts of recombination over mutation on the diversification of the alleles, we estimated the frequency of recombination by looking at the

Figure 3.4: Distribution of the Interior/exterior branch length ratio of trees resulting from ClonalFrame analysis of concatenated data as compared to trees simulated under the coalescent model. *Halomonas* species exhibit a significantly lower ratio (0.3, P = 0.00199) than expected under the coalescent model indicating that this population has gone through a recent expansion of population size.



n

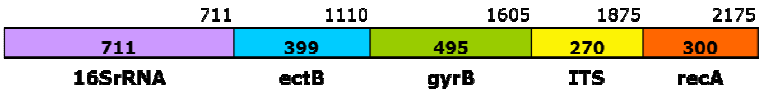
umber of nucleotide polymorphisms among all the alleles in highly similar isolates or single/double locus variants. All 44 locus variants had an excessive number of single-nucleotide polymorphisms (SNPs). Since the chance of more than 3 mutations occurring in a single locus is extremely small, these variant loci could only result from recombination. An LGT event followed by homologous recombination leads to multiple-nucleotide changes in an allele whereas mutation leads to a single nucleotide change.

To quantify the rate of recombination among Halomonas isolates from the GSP, ClonalFrame analysis was carried out on individual gene sequences and the concatenated data set. The rate of recombination relative to point mutation (r/m), which is a measure of how important the effect of recombination is in the diversification of the sequence relative to point mutation, was high (18.5 for concatenated data and 1.59-14.41 for each of the loci tested)(Table 3.6).These numbers exceed the values determined for other groups of bacteria and archaea (29). Vos and Didelot (121) found that r/m values reflected the greater importance of recombination over mutation when the value exceeds 10 for a given species. Doroghazi and Buckley (2010) (29) reported an r/m value of 19.5 for *Streptomyces* isolates obtained from different geographical locations.

An estimate of the relative contribution of mutation and recombination in generating diversity within the *Halomonas* population was calculated using the LDhat 2.1 (71) program for the concatenated data set. The ratio of recombination to mutation rates, ρ/θ_W (a measure of how often recombination events occur relative to mutation) was estimated to be 1.53:1 (Table 3.6). To examine the effect of recombination on *Halomonas* spp. isolates from the GSP, we analyzed DNA sequences from all loci tested for mosaic patterns suggestive of recombination using RDP 3.0 software package. Sixteen recombination events were detected in 58% of the *Halomonas* STs. These events were detected as having a portion of the gene sequence similar to parental genotype genotype and the remaining sequence similar to a different parental genotype. For each of the detected events, the number of recombinant ST with the evidence of the event, breakpoint positions corresponding to the ends of the recombined fragments, putative parent STs (donor and recipient), and *p* value for all the seven detection methods used (Table 3.5**)** were recorded. All events were detected by at least four of the seven methods indicating confidence in the detection methods rather than statistical overcalling. These recombination events were distributed in all the five loci tested (Table 3.5**)** with 50% of the events occurring within *recA* gene**.** All the events

detected were predicted to have both breakpoints beyond the ends of the gene fragment. This data indicates that LGT events occur often enough at GSP to transfer entire gene.

Table 3.5: Recombination events detected using RDP 3.0. Gene boundaries are shown below.



Daughter	Donor	Region	RDP	GENECONV	Bootscan	Maxchi	Chimaera	SiScan	3Seq
ST2	ST9	1-1109	4.24E-16	3.31E-13	1.68E-16	2.22E-16	3.17E-14	1.75E-25	6.66E-16
ST14	Unknown (ST12)	1-1163	1.18E-29	1.31E-21	1.39E-30	1.61E-21	5.05E-21	NS	7.47E-40
ST13	S 18	300-940	NS	1.01E-03	5.51E-04	1.01E-02	NS	5.30E-07	NS
ST15	S 11	643-1103	1.68E-32	4.11E-33	6.67E-40	3.80E-19	1.04E-06	7.87E-26	1.15E-06
ST29	S 12	680-1616	4.56E-53	7.78E-42	2.57E-38	2.48E-27	3.59E-20	4.73E-25	2.08E-40
ST17	S 26	686-1139	5.39E-15	3.58E-12	7.30E-14	3.72E-12	NS	6.07E-11	7.42E-23
ST21	S 12	741-1507	7.59E-31	1.49E-29	1.90E-32	2.44E-14	1.47E-14	4.33E-08	2.60E-20
ST15	Unknown (ST21)	1110-1584	1.33E-34	1.56E-33	3.74E-06	1.23E-12	1.44E-12	1.91E-07	2.45E-23
ST11	S 12	1680-1888	7.67E-14	1.41E-12	1.92E-10	7.04E-13	2.30E-15	NS	1.64E-23
ST12	S 20	1873-2167	2.65E-02	NS	NS	1.54E-04	NS	5.28E-05	5.07E-03
ST1	ST3	1887-2079	1.58E-05	1.87E-05	9.35E-04	5.06E-07	1.97E-06	NS	9.90E-09
ST10	Unknown (ST28)	1888-2131	4.41E-12	4.04E-10	3.90E-10	2.63E-09	9.89E-10	1.53E-11	1.92E-15
ST29	ST5	1888-2147	4.59E-02	NS	0.01921549?	2.00E-02	NS	NS	1.59E-02
ST18	S 26	1896 2085	0 015737903	1.37E 02	NS	8.37E 05	NS	NS	4.52E 04
ST8	Unknown (ST24)	1900-2156	1.25E-11	7.61E-11	1.52E-12	1.26E-13	9.86E-12	1.86E-20	2.02E-21
ST9	ST5	1900-2167	6.57E-05	NS	NS	4.97E-06	3.06E-06	NS	7.79E-08

It is difficult to precisely determine the evolutionary history of the *Halomonas* species isolates using a vertical model of inheritance. Even though recombination events were detected in most of the isolates, significant linkage disequilibrium was also detected indicating the conservation of clonal population structure among these isolates. The standardized index of (I_A^S = 0.1006) significantly differed from zero in all isolates (Table 3.6) indicating significant linkage disequilibrium among the alleles. The pairwise homoplasy index test (Φ_W test) rejected the hypothesis of no recombination (Φ_W = 0.0). That is no case of expected homoplasies higher than the true homoplasies were detected out of 1000 trials (Table 3.6).

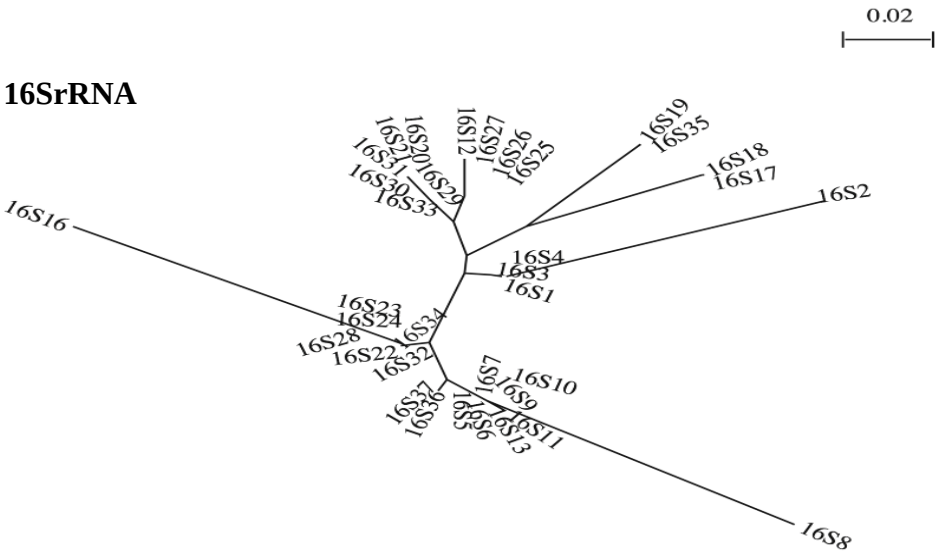
Table 3.6: Population parameters for the concatenated data set

	Length	π	Sites	Φ_W	I^S_a	r/m	ρ/θ_W
Concatenated data	2175	0.09522	757	0.00(P-sig)	0.1006	18.5	1.53

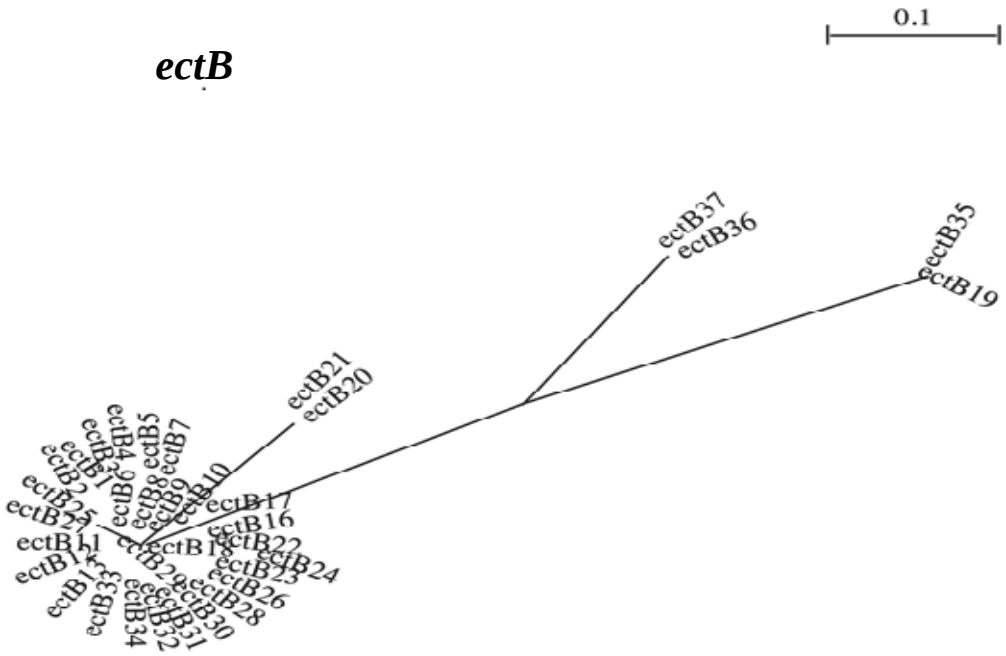
π – Nucleotide diversity/site, Sites – Number of polymorphic sites, Φ_w – Pairwise homoplasy index, r/m – Rate of recombination relative to point mutation, ρ/θ_w – Ratio of rates at which recombination and mutation occur.

The phylogenies for all loci tested were incongruent as inferred from the topology of the ML trees generated using the PhyML program (Figure 3.5**)**. The *recA* gene again showed higher resolution within the isolates grouped in different clusters from those observed in 16S rRNA tree.

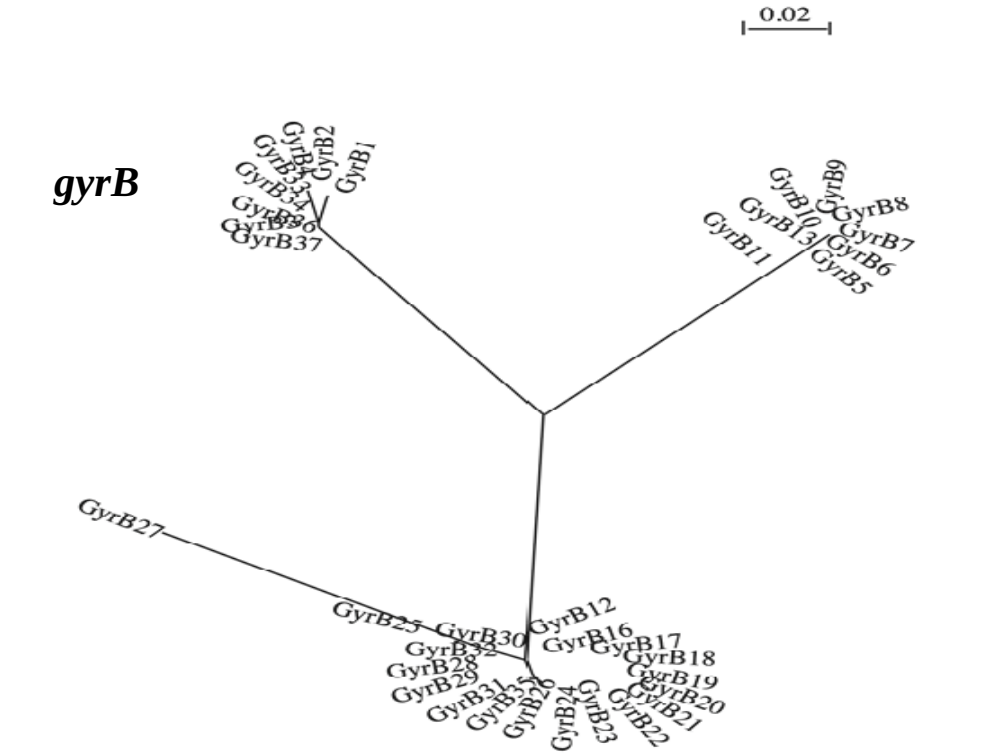
Figure 3.5: Maximum Likelihood (ML) trees for each loci tested for all 33 isolates. ML trees were created using the substitution model as suggested by MODELTEST. The scales of branch length vary for different loci.

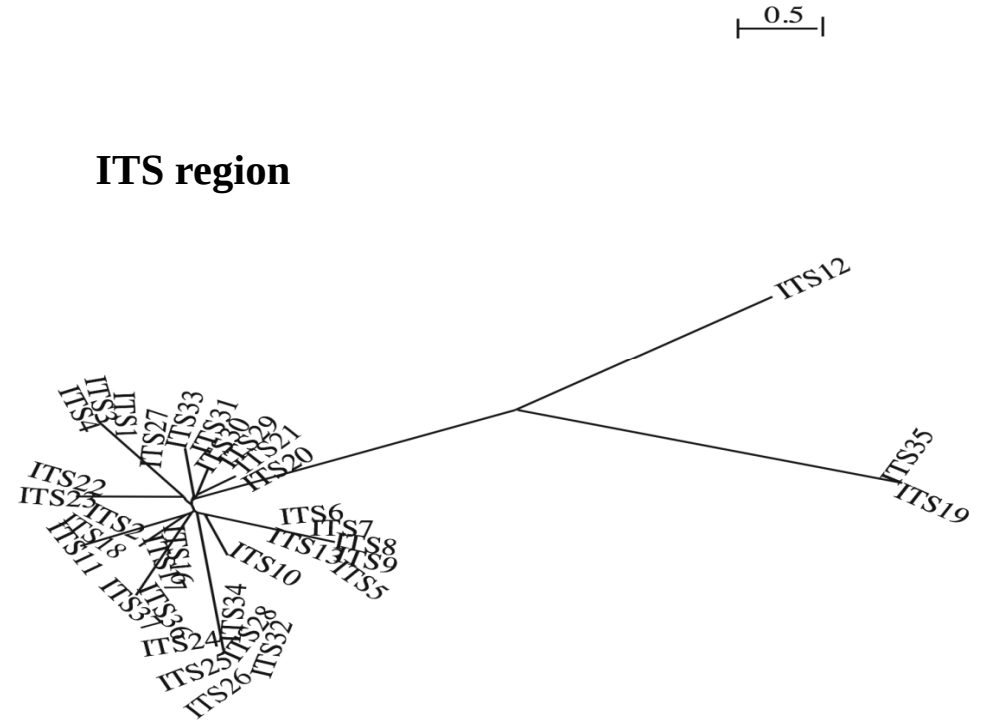


ectB

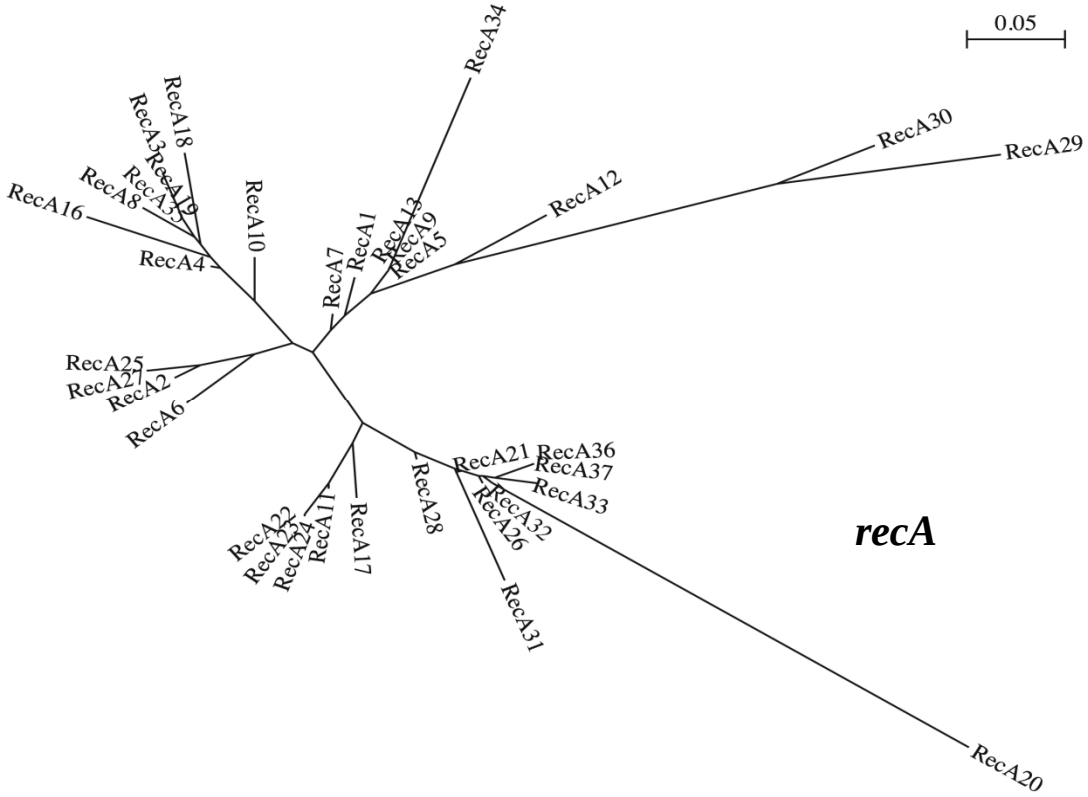

0.1
ectB37
ectB36
ectB35
ectB19
ectB21
ectB20
ectB5
ectB7
ectB10
ectB17
ectB16
ectB22
ectB24
ectB23
ectB26
ectB28
ectB30
ectB31
ectB32
ectB34

gyrB


0.02
GyrB2
GyrB1
GyrB37
GyrB9
GyrB8
GyrB7
GyrB6
GyrB5
GyrB11
GyrB27
GyrB25
GyrB30
GyrB12
GyrB16
GyrB17
GyrB18
GyrB19
GyrB20
GyrB21
GyrB22
GyrB23
GyrB24
GyrB26
GyrB28
GyrB29
GyrB31



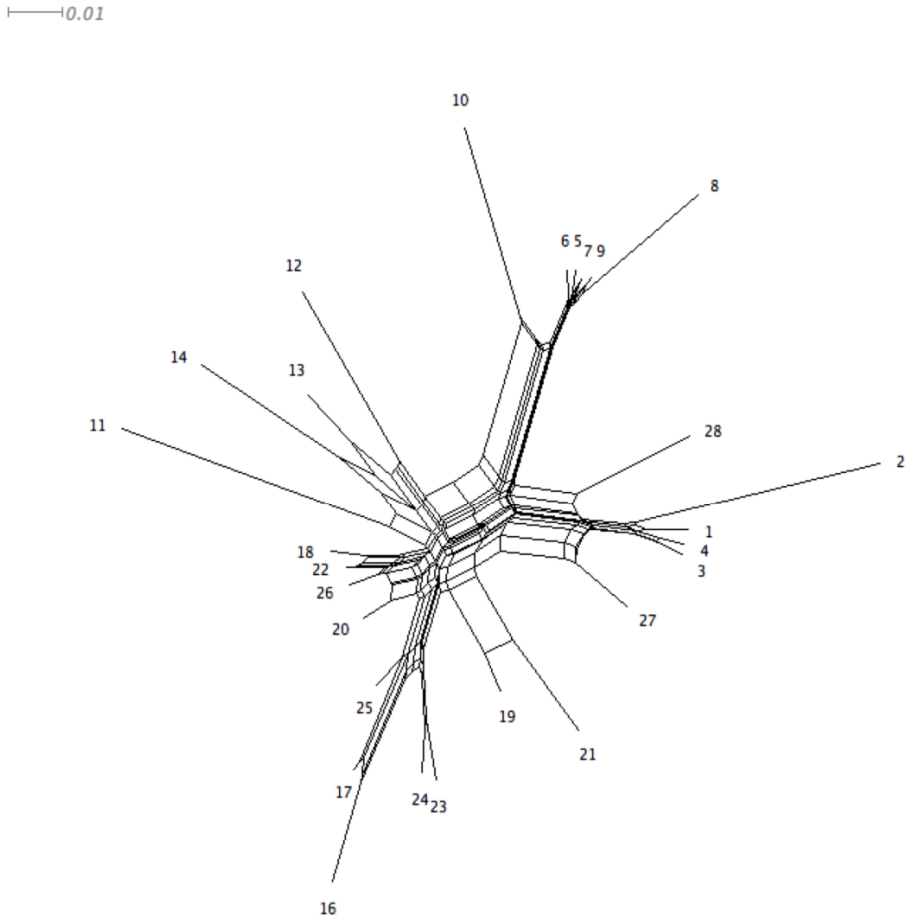
ITS region



recA

To better define the evolutionary history of isolates, NeighborNet analysis implemented within Splitstree v4.10 was used (53). This program scans the entire nucleotide sequence for inconsistencies and connects the STs through a reticulate creating a network diagram (Figure 3.6). Reticulation is prevalent in *Halomonas* spp. in the GSP resulting in the rise of new lineages.

Figure 3.6: NeighborNet analysis of the MLST data supporting the reticulate evolution within *Halomonas* species at the GSP.



DISCUSSION

It is known that bacterial populations consist of distinct genotypic clusters. Homologous recombination among these clusters disturbs the linkage association creating greater genotypic diversity. High genetic diversity is unusual in clonal organisms such as bacteria because diversity is purged across loci by recurring selective sweeps (periodic selection) (115). To reconcile this apparent contradiction, Cohan (18) suggested that each ecotype creates its own sequence-based cluster through time as a result of high genetic diversity. With a better knowledge of niche-adaptive genes and phenotypes, we will be able to connect the observed genetic/phenotypic diversity to the above mentioned model (18). Our data produce discrete clusters with diverse, incongruent topologies. This lack of congruence between each of the MLST loci trees suggests the presence of high rates of recombination at the GSP. This conclusion is supported by the statistical tests and analyses used. In spite of significant linkage disequilibrium between alleles, the rate of recombination is on par with some of the bacteria and archaea that exhibit high rates of recombination (121, 83) suggesting that recombination has not completely erased the patterns of vertical inheritance thus maintaining clonality within the *Halomonas* isolates. However, we have to be careful in drawing conclusions here as the constant population size used to calculate ρ in LDhat is yet to be validated for microbial data (29).

Our data support a biological species model, where frequent recombinational events maintain a single unified species cluster within an ecosystem. The existence of high genotypic diversity could be due to weak selection or to LGT events that detach genes from selection at the genome level (89). Since our data show that diversity is maintained at all the loci tested including core house keeping genes used commonly for phylogenetic analyses, we can attribute the presence of high nucleotide diversity to frequent recombinational events even in the presence of selection. Recombination results in shuffling of variations between lineages, causing an unexpectedly high level of homoplasy (convergent evolution) (39) as evident from our data where the pairwise

homoplasy index test rejected the hypothesis of no recombination. Our data also supports the presence of widespread reticulation within the *Halomonas* phylogeny, a characteristic of recombination.

Fuzzy species formed due to reticulate evolutionary processes have been shown in other microbial groups (29, 83). Sometimes fuzzy species are identified as a result of sequencing of amplified fragments from mixed non-viable cultures. In our study all the sequences were obtained from single colonies of viable, pure cultures. Under these conditions only a small number of isolates from the population are required for meaningful phylogenetic analysis (39).

Fuzzy species boundaries created by LGT events could occur via transformation, conjugation or transduction (74). At present very little information is available on the relative importance of these LGT mechanisms in *Halomonas.* This group of bacteria has not been shown to be naturally competent for transformation (120). Conjugation has been used effectively for genetic manipulation of *Halomonas* (120). Although transduction has not been very well characterized in *Halomonas,* bacteriophages specific for *Halomonas* spp. have been isolated from the GSP (102, 78). Seaman and Day (102) demonstrated that one of the phages with an unusually large genome (~340 kb) is capable of mediating transduction. Thus it is safe to predict that transduction is a factor in producing this high level of genetic diversity and recombination detected within the *Halomonas* population at the GSP.

Our data suggest that *Halomonas* spp. at the GSP have been subjected to a recent expansion in population size. This could have resulted from two separate events, a population bottleneck or a selective sweep. It is very difficult to differentiate between these two as they leave a similar genetic hint within a population. It is clear from our study that the *Halomonas* spp. isolates from the GSPare not structured into major clonal complexes and several unrelated individual isolates were identified from eBURST analysis. This may be due to frequent recombinational events

among these closely related isolates. Using ClonalFrame, external/internal branch length ratios were calculated that showed significant support for a population bottleneck due to environmental changes. It is speculated that the GSP area was covered with seawater and the briny remains of this ancient sea rise to the surface and leave crusts of white salt on the surface. During rainy season, rainfall dissolves the salt creating temporary streams and ponds across the salt flat with varying salt concentrations (16). We speculate that these conditions led to the recent evolutionary bottleneck. We speculate that the population that survived this bottleneck has developed genetic mechanisms to adapt and grow in this constantly changing, hypersaline environment.

Interestingly the *Halomonas* spp. *recA* gene showed high levels of nucleotide diversity and recombination rates than did the other loci tested. Close to 50% of the recombination events identified by RDP 3.0 had breakpoint positions within the *recA* gene. The *recA* gene may correspond to a hotspot of recombination or it may be under a strong selective pressure that provides a selective benefit to new variants emerging due to recombination. Wilson et al.(16) showed that *Halomonas* spp. isolates from the GSP may have developed constitutively expressed SOS DNA repair system. The RecA protein regulates the SOS regulon as well as being actively involved in homologous recombination (76). Since the closely related *Halomonas* isolates from the GSP live in a rapidly changing saline environment, the *recA* gene and its protein product may act as a valuable repository and resource for reconstructing and visualizing processes of natural selection and adaptive evolution within the GSP. The constant fluctuation in salinity and other environmental conditions at the GSP provide a driving force for evolutionary changes. Thus our data show that gene transfer among *Halomonas* isolates not only involves assimilation of novel sequences, but also gene replacement by LGT events as these isolates are constantly exposed to donors of genetic material in this plastic environment. This mechanism is expected to occur if the acquired genes/sequences have adaptive significance in that ecotype. Further characterization of

the *recA* gene and its protein product will shed more light on the molecular mechanisms leading to heightened levels of genetic diversity among *Halomonas* spp. isolates at the GSP.

CONCLUSION

An MLST scheme for *Halomonas* species at the GSP was developed. Application of this scheme demonstrated high rates of genetic diversity through recombination although clonality was preserved. Our data provide clear insights into the phylogenetic discreteness and evolutionary origins of *Halomonas* spp. at the GSP. These results may provide clues towards clarifying the phylogenetic inconsistencies observed within the genus *Halomonas.* Our findings support the crucial role played by recombination in evolution and building structure of bacterial populations from diverse environments. However, some may argue that this study provides only a limited snapshot of overall genetic diversity and population structure. It is important to note that the isolates in this study belong to different species. We do not have access to complete genome sequence information for any isolate within the genus *Halomonas* limiting our understanding of arrangements of these loci on the chromosome. This study significantly contributes to a more detailed understanding of importance of gene flow, selection and recombination in the adaptation and survival strategies of microorganisms in extreme environments. Understanding of these concepts has important consequences in studying the evolution of these group of bacteria under extreme conditions along with their usefulness in several biotechnological applications (25, 16).

GLOSSARY

Pulsed-field gel electrophoresis (PFGE): Variation between strains is detected using rare-cutting restriction enzymes and the resulting genomic fragments are separated on an agarose gel.

Multilocus enzyme electrophoresis (MLEE): Enzyme polymorphisms between strains are detected on the basis of differing electrophoretic mobility of the encoded proteins on a starch gel.

Multilocus Sequence Typing (MLST): Allelic variation at multiple housekeeping loci is indexed directly by nucleotide sequencing of internal fragments of around 450 base pairs. The resulting data is stored on a central online database.

PubMLST: An international network of publicly available MLST databases.

Allelic profile: The alleles at each of the housekeeping genes used for MLST that define a strain.

Sequence Type (ST): Defined by the allelic profile. It is a unique combination of allele designations used in an MLST scheme.

Clonal complex: A group of related sequence types.

Clonal population structure: A feature of asexual populations in which a all members are related by a consistent tree-like phylogeny.

Housekeeping genes: These genes are ubiquitous within a population, encoding proteins essential for central metabolism. These genes evolve at a moderate rate compared with informational genes.

Synonymous substitution: A mutation which does not result in a change in the amino acid sequence of the protein.

Positive selection: Genes under strong selective pressure for rapid evolution, such that non-synonymous changes are common than synonymous changes.

Homologous recombination: Replacement of alleles by DNA sequences with localized regions of high homology and involving RecA protein.

Population bottleneck: An evolutionary event resulting in a loss of a large proportion of the population. The surviving members of the population will only contain a fraction of the original genetic diversity.

Congruence: Agreement between the topologies (shape) of phylogenetic trees derived from different gene loci.

Terms used in statistical analysis:

S- Polymorphic sites

π- Nucleotide diversity/site: Average number of nucleotide differences per site between two sequences .

d_N/d_S- Ratio of Non-synonymous substitutions Vs Synonymous substitutions

Tajima's D- Indicator of population growth or selection pressure. It is based on the differences between the no. of segregating sites and the avg. no. of nucleotide differences. This is based on the hypothesis that all mutations are selectively neutral.

Φ_W- Pair wise homoplasy index: It is the phi test for recombination. P-value 0.00 indicates statistically significant evidence for recombination. This test examines incompatibility in phylogenetic signals. Incompatible sites could have two possible histories, one with the mutation and one where recombination was present in different lineages.

I_a – Standardized index of association (test for recombination).

r/m- Rate of recombination relative to point mutation: Ratio of probabilities that a given site is altered through recombination and mutation. A measure of how important the effect of recombination is in the diversification of the sequence relative to mutation.

ρ/θ_W – Ratio of rates at which recombination and mutation occur: It is a measure of how often recombination events occur relative to mutation.

CHAPTER IV

PHYLOGENETIC ANALYSIS OF *SHEWANELLA* SPECIES

ABSTRACT

Multi Locus Sequence Typing (MLST), a typing tool developed for characterization of bacterial isolates based on the sequence information from internal fragments of several house keeping genes has provided a valuable platform for estimation of the role of recombination and point mutation in shaping bacterial communities. The Great Salt Plains (GSP) located in north-central Oklahoma is an ecologically diverse, extreme, hyper-saline environment. Results from MLST analysis of *Halomonas* species (Chapter 3) adds to the reports supporting the crucial role played by recombination in evolution of bacterial populations from diverse environments like GSP. In order to obtain an idea of the relative influence of recombination to the evolution of bacteria in extreme environments such as the GSP, genetic diversity observed for *Halomonas* spp. was compared to an organism from an environment with more moderate characteristics. Nine species belonging to the genus *Shewanella* that are commonly found in marine and fresh water environments from different geographic locations were tested for sequence level comparison. Comparable levels of genetic diversity and recombination rates were observed among the loci tested. The data from this study support our hypothesis that, higher rates of recombination observed within *Halomonas* spp. isolates acts as a genetic mechanism that has led to increased genetic diversity among these bacteria from the GSP.

INTRODUCTION

The influence of recombination in understanding the structure of bacterial species is very unclear at this time. Multi Locus Sequence Typing (MLST), a typing tool developed for characterization of bacterial isolates based on the sequence information from internal fragments of several house keeping genes has provided a valuable platform for estimation of the role of recombination and point mutation in clonal diversity (69). This approach uses the sequences to gain an estimate of the recombination rates and point mutations that have occurred during the initial stages of diversification of bacterial clones thus avoiding the problem of identifying the ancient recombinational events within the sequences of distantly related isolates.

The Great Salt Plains (GSP) located in north-central Oklahoma is an ecologically diverse, extreme, hyper-saline environment. The microorganisms living in the GSP are exposed to harsh environmental conditions including high temperature, freezing winters, high salinity, desiccation and direct ultra-violet radiation. *Bacillus* and *Halomonas* are the dominant Gram-positive and Gram-negative culturable heterotrophic bacteria identified from GSP soils (16). MLST analysis (Chapter 3) has shown that *Halomonas* spp. isolates from GSP soils contain several genetically distinct groups most likely due to the effects of environmental bottlenecks and recombination events. These results support the presence of high speciation among *Halomonas* spp. isolates at the GSP site even though there is high 16S rDNA sequence similarity (Chapter 3).

Results from MLST analysis of *Halomonas* species (Chapter 3) adds to the reports supporting the crucial role played by recombination in evolution of bacterial populations from diverse environments. Results from our study demonstrated the importance of population parameters such as lateral gene transfer and recombination in shaping bacterial communities. The adaptation and survival strategies of microorganisms in extreme environments have important consequences in understanding their evolution and their usefulness in several biotechnological applications.

In order to obtain an idea of the relative influence of recombination to the evolution of bacteria in extreme environments such as the GSP, we wished to compare our results for *Halomonas* to an organism from an environment with more moderate characteristics. *Shewanella* is a genus that contains gram-negative, facultatively anaerobic, metal-reducing bacteria commonly found in marine and fresh water environments (96). These organisms often predominate in chemically stratified (with high inputs of organic matter and fluctuating redox conditions) water bodies (96). The ability of these bacteria to utilize a diverse range of final electron acceptors (nitrite, nitrate, thiosulfate, iron, manganese, uranium) in the absence of oxygen has contributed to their survival in diverse environments and makes *Shewanella* potentially important in bioremediation of contaminated areas. They are easy to culture in the laboratory with commonly used growth media (50). Since *Shewanella* is closely related to *E.coli*, genetic tools designed for *E.coli* can also be used with *Shewanella* spp. isolates. Currently there are approximately 48 recognized species within the genus *Shewanella* based primarily on DNA-DNA hybridization and 16SrRNA sequences. The distribution of *Shewanella* in many parts of the world is primarily based on two important physiological parameters: (96) their capacity to metabolize diverse group of compounds found in the environment and their ability to survive at low temperatures (50).

A total of 23 *Shewanella* isolates have been completely sequenced (42). The sequence information has provided us with the tool to investigate their genomic diversity from an evolutionary viewpoint. Caro-Quintero et al.(15)reported that several *Shewanella baltica* isolates obtained from the Baltic Sea had exchanged as much as 20% of their genome based on complete genome sequence and transcriptome analysis (15). Therefore our findings and the results of Quintero et al (15) suggests that genetic exchange in response to environmental settings may be enhanced. This observation has implications in better understanding of bacterial speciation and evolution.

To identify to what extent the similarities in sequences of house keeping genes among a group of isolates is determined by the genetic relatedness and ecological adaptation of the microorganism in question, we have analyzed and compared the partial sequence information of several house keeping genes of 9 *Shewanella* isolates (Table 4.1) from diverse geographic locations and habitats including both fresh and marine water bodies. Since *Halomonas* isolates used in our study were isolated from the terrestrial environment which was once believed to be submerged in sea water, we believed it would be ideal to compare them with isolates obtained from aquatic environment. Results from this study will signify the sequence-level comparison of environmental representatives from *Proteobacteria,* an important and diverse group of organisms. This study identified several trends from the comparison of available sequence information that may apply to other environmentally adaptable bacteria besides *Halomonas* and *Shewanella* for which genome sequences of closely related species are available.

METHODS

Bacterial strains:

Table 4.1 List of *Shewanella* isolates used in this study

Genome Name	Source	Location
Shewanella amazonensis SB2B	Marine deposits	River,Brazil
Shewanella baltica OS155	Marine	Baltic sea
Shewanella baltica OS185	Marine	Baltic sea
Shewanella baltica OS195	Marine	Baltic sea
Shewanella baltica OS223	Marine	Baltic sea
Shewanella denitrificans OS217	Marine	Baltic sea
Shewanella frigidimarina NCIMB400	Marine	North sea,UK
Shewanella sp. MR-4	Sea-water, oxic zone	Black sea
Shewanella sp. W3-18-1	Marine sediment	Wash,USA

The list of *Halomonas* isolates used is provided in Chapter 3. The nine *Shewanella* isolates used in this study along with the location and source of isolation are provided in Table 4.1. The

sequence information for the selected house keeping genes from the *Shewanella* isolates was obtained from GeneBank.

House-Keeping gene selection and sequence data analysis:

In the *Halomonas* MLST study (Chapter 3), we used 16S rRNA, *gyrB*, *ectB*, Internal Transcribed Spacer (ITS) region and *recA* genes to understand the population structure of the isolates. We decided to use the same genes in this comparison study using *Shewanella* isolates. The *ectB* gene which codes for ectoine synthase was not detected in any of the *Shewanella* isolates within the GeneBank database. Instead, the *rpoB* gene that codes for beta subunit of RNA polymerase was used. All the sequences were confirmed through BLAST searches to verify the resultant data matches with the right gene in the database. Multiple sequence alignments were obtained using CLUSTALX (64). Each different allele within a locus was assigned a different number and the combination of all the numbers resulted in a ST for a strain (Table 4.2). These data were used for subsequent analysis.

Pair-wise distances between individual and concatenated nucleotide sequences were determined using PAUP* 4.0 (113). Estimations for parameters of the number of polymorphic sites, the average pair-wise nucleotide differences per site (π), GC content and Tajima's D, a test for alleles departing from neutral evolution were calculated using DnaSP 5 (68). Standardized index of association (I^S_A), a measure of linkage disequilibrium and ratios of non-synonymous to synonymous substitutions (d_N/d_S), a measure of purifying selection were estimated using START2 (55). The Index of Association (IA) is calculated as follows: IA =VO/VE -1. If V0 is the observed variance of K and VE is the expected variance of K where K is the number of loci at which two individuals differ.

Measure of Recombination rate and detection of recombination:

Rate of recombination relative to point mutation (r/m) values were calculated using ClonalFrame (27). ClonalFrame was run with 100000 burn-ins followed by 100000 more iterations. Population scale rates of recombination (ρ/θ_W) were estimated with LDhat 2.1 (71). θ_W is the average mutation rate while the ρ is the rate of gene conversion, $\rho = 2N_e r/2$. All the calculations were performed for the concatenated data set only. Pair wise homeoplasy index ($\mathbf{\Phi_W}$) (phi test for recombination) values and Neighbornet phylogenetic network trees for the concatenated data set were estimated using Splitstree v4.10 (53).

Phylogenetic analysis:

Using the hierarchical ratio test within the MODELTEST version 3.7 (91), the most appropriate model of DNA substitution for each locus was determined. Neighbor –Joining (NJ) trees were constructed using the program PAUP* 4.0 (113) with Jukes-Cantor correction. Maximum likelihood (ML) trees were inferred for each individual loci and the concatenated data set using PhyML v3.0 (46) with DNA substitution model as suggested by MODELTEST**.**

RESULTS

Sequences of *Shewanella* isolates:

Currently 23 isolates belonging to genus *Shewanella* have been completely sequenced and the sequence data is publicly available at GeneBank (http://www.ncbi.nlm.nih.gov/genbank). Sequence information from 9 of the isolates for all the loci were obtained from GeneBank. These sequences were made into equal length by eliminating poorly aligned positions and highly divergent regions of an alignment using a computer program Gblocks (114). Highly divergent regions may not be homologous or may have been saturated by multiple substitutions and it is convenient to eliminate them prior to analysis. All the isolates were unrelated by the criterion that

they resulted in 9 ST's and 8, 9, 6, 9, 9 alleles were detected for 16S rRNA, *gyrB*, ITS, *recA* and *rpoB* genes respectively (Table 4.2). This level of divergence is unusual for clonal organisms like bacteria as the diversity is purged across all loci by recurrent selective sweeps (periodic selection) (115).

Table 4.2 Sequence types (ST) and the allele profiles generated using the sequences obtained from GeneBank

ST	*16SrRNA*	*gyrB*	*ITS*	*recA*	*rpoB*
1	1	1	1	1	1
2	2	2	2	2	2
3	1	3	2	3	3
4	3	4	2	4	4
5	4	5	2	5	5
6	5	6	3	6	6
7	6	7	4	7	7
8	7	8	5	8	8
9	8	9	6	9	9

Sequence characteristics:

Detailed characteristics of nucleotide sequences for each of the loci tested are presented in Table 4.3**.** Among all the loci tested, the *recA* gene had the highest number of polymorphic sites (72.1% of the total length) indicating that the *recA* gene has an increased level of allelic variation. None of the isolates had similar allelic profiles.

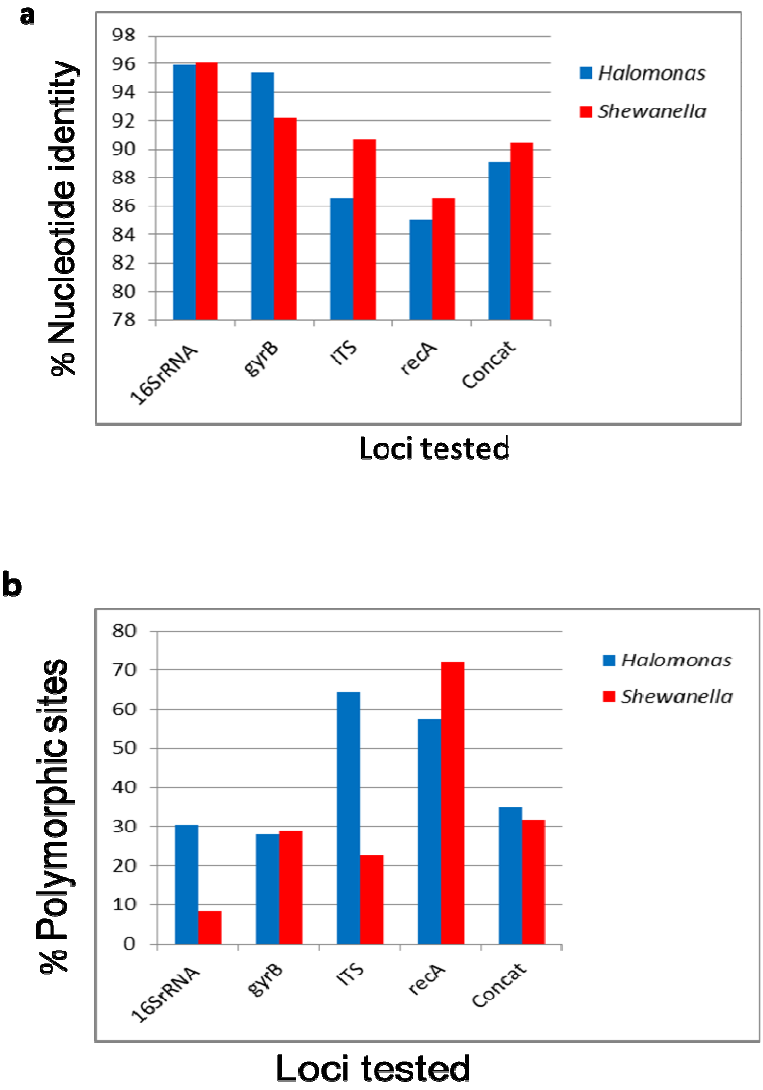
Table 4.3 Sequence characteristics of the loci used in this analysis

loci	Length	Number of alleles	Average Distance (%)	Polymorphic sites (%)	Nucleotide diversity	dN/dS	Tajima's D
16SrRNA	1521	8	0.0394 (96.06)	129 (8.48)	0.02836	NA	NS-1.10497
gyrB	2417	9	0.078 (92.2)	701 (29.00)	0.14503	0.107	NS-0.97764
ITS	230	6	0.093 (90.7)	52 (22.61)	0.10383	NA	NS0.33285
recA	956	9	0.134 (86.6)	689 (72.1)	0.37878	0.801	NS0.42810
rpoB	3964	9	0.0404 (95.96)	1124 (28.35)	0.10839	0.12	NS-0.84514
concat	9012	9	0.0955 (90.45)	2854 (31.67)	0.13161	NA	NS-0.58574

NA- not applicable, NS- values for Tajima's D were non-significant, dN/dS- Ratio of Non-synonymous substitutions Vs Synonymous substitutions.

The average interspecies sequence similarity values for each of the gene were: 16S rRNA (96.06%), *gyrB* (92.2%), ITS (90.7%), *recA* (86.6%) and *rpoB* (95.96%). For concatenated gene sequences (9012 bp), the average similarity value was 90.45%. We observed a similar sequence divergence value (85.1%) for the *recA* gene from our *Halomonas* MLST study (Chapter 3). The *recA* gene is considered to be highly conserved among different genera and individual species of bacteria. The average number of nucleotide differences per site between two randomly selected sequences measured as nucleotide diversity (**π)** was high in the *recA* gene (0.379) indicating the diverse nature of this locus. All the housekeeping genes showed smaller (>1, Table 4.2) ratio of non-synonymous to synonymous substitutions (dN/dS) indicating the presence of a strong purifying selection limiting amino acid changes. Tajima's D values were non-significant for all the loci tested (-1.10497 to 0.42810) verifying the presence of neutral selection on the alleles of these loci. The sequence characteristics observed for all the loci tested in *Shewanella* isolates were similar to *Halomonas* species isolates suggesting that genetic diversity observed can be related to the ecological adaptation of microorganisms living in diverse, extreme environments. Figure 4.2a and 4.2b compare the sequence similarity values and the number of polymorphic sites (% of total) for all the loci tested except for *ectB* and *rpoB* in *Halomonas* and *Shewanella* respectively.

Figure 4.2 All the loci tested in *Halomonas* isolates are compared with their orthologs from *Shewanella* isolates. (a) Chart comparing the sequence similarity values for different loci tested (b) Chart comparing the polymorphic sites observed (% of total length).



Recombination among *Shewanella* isolates:

To further examine the rate of recombination among *Shewanella* strains used in this study, ClonalFrame (27) analysis was performed on the concatenated data set. The rate of recombination relative to point mutation (r/m) value of 9.6 (Table 4.4) was similar to the values observed for several other groups of bacteria and archaea (121). An estimate of the relative contribution of mutation and recombination in generating diversity within the *Shewanella* isolates was obtained using LDhat 2.1 (71) program for the concatenated data set. The ratio of recombination to mutation rates (ρ/θ_W) was estimated to be 0.32:1 (Table 4.4), suggesting that recombination loses strength as sequence divergence increases between species of bacteria. The pairwise homoplasy index test (Φ_W test) rejected the hypothesis of no recombination ($\Phi_W = 1.316x\ 10^{-10}$) i.e., assumption that expected homoplasy is higher than the true homoplasy was non-significant out of 1000 trials (Table4.4).

Table 4.4 Population parameters for the concatenated data set

	Length	π	Sites	Φ_W	I_a	r/m	ρ/θ_W
Concatenation	9012	0.13161	2854	$1.316x\ 10^{-10}$(P-sig)	0.1445(NS)	9.6	0.32

π – Nucleotide diversity/site, Sites – Number of polymorphic sites, Φ_w – Pairwise homoplasy index, r/m – Rate of recombination relative to point mutation, ρ/θ_w – Ratio of rates at which recombination and mutation occur.

The phylogenies for all the loci tested were incongruent as inferred from the topology of ML trees generated using PhyML program (Figure 4.3). Again, the *recA* gene showed higher resolution i.e., the isolates were clustered in different groups than what was observed for 16SrRNA tree. It is challenging to precisely show the evolutionary history of *Shewanella* using a vertical mode of inheritance following observations of the wide spread occurrences of horizontal gene exchange.

NeighborNet analysis (53) implemented within Splitstree v4.10 scans the entire nucleotide sequence length for inconsistencies and connects the STs through a reticulate thus creating a network graph (Figure 4.4). The concatenated sequence data for all the *shewanella* strains tested in this study suggests that reticulation is prevalent among these isolates resulting in rise of new lineages as a result of recombination between species clusters.

Figure 4.3 Maximum Likelihood (ML) trees of each of the loci tested from all the *Shewanella* isolates used in this study. ML trees were created using the substitution model as suggested by MODELTEST. The scales of branch length vary for different loci.

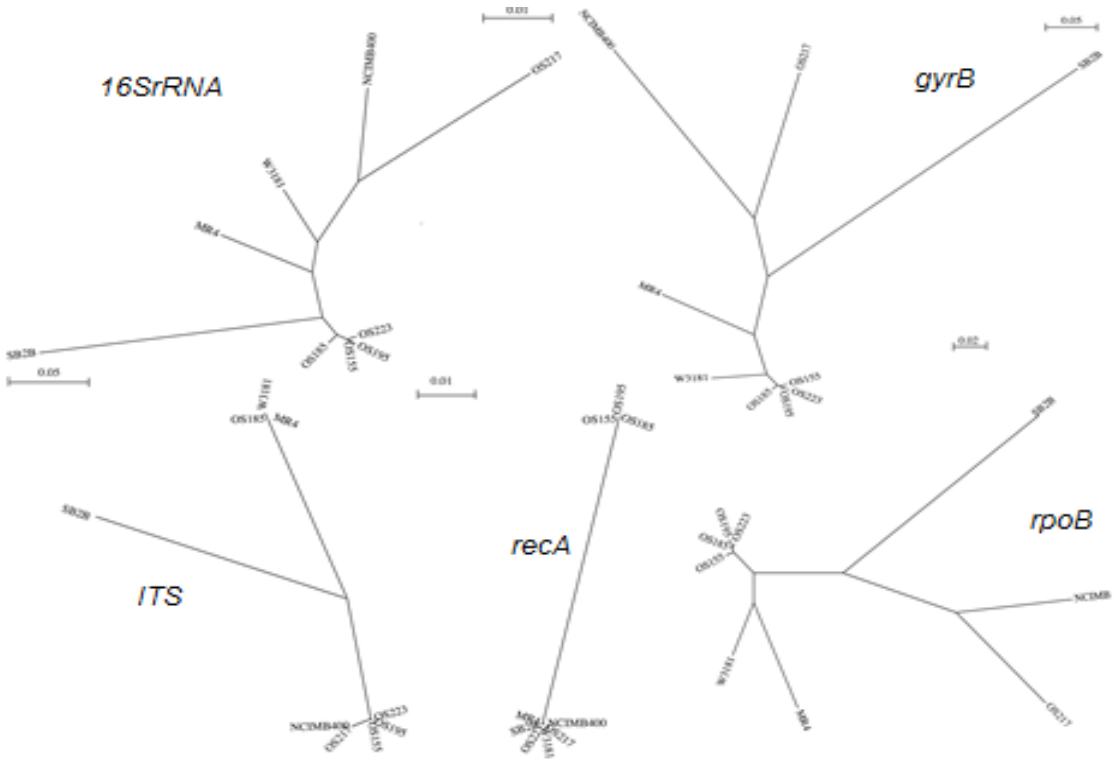
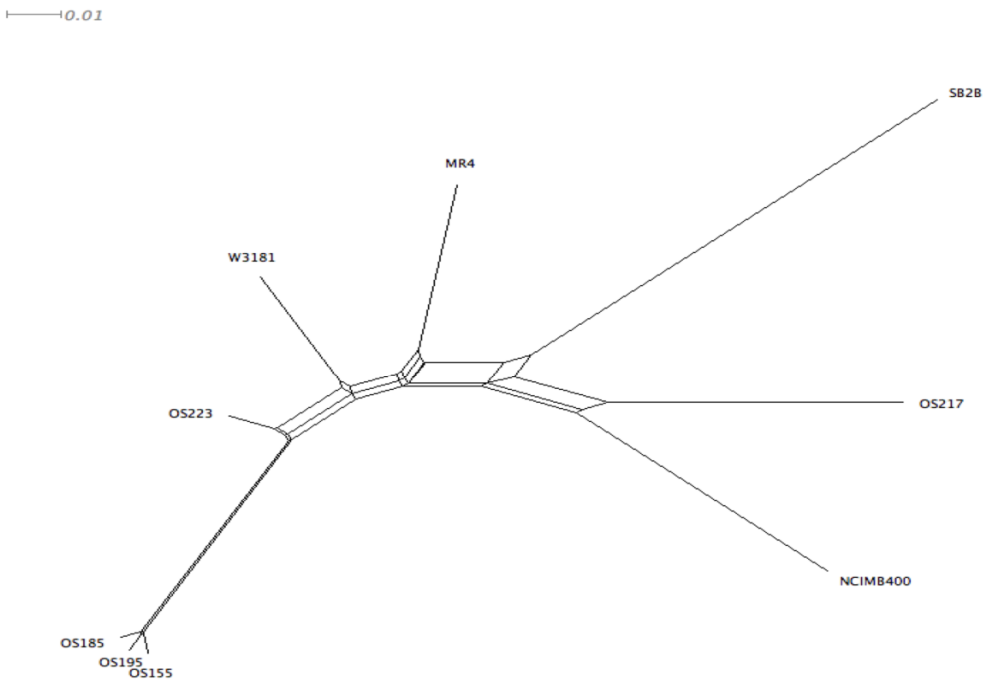


Figure 4.4 NeighborNet analysis of the concatenated data supporting the reticulate evolution between *Shewanella* strains.



DISCUSSION

Caro-Quintero et al (15) tested several *Shewanella* isolates obtained from Baltic Sea for the evidence of horizontal gene transfer. Although these isolates showed comparable evolutionary divergence, strains obtained from more similar depth shared more genes compared with strains from different depths. This cannot be attributed just to the higher evolutionary relatedness between these genomes. Instead, these findings are most likely attributed to recent events of horizontal exchange between these isolates or their immediate ancestors. Further, based on the examination of the nucleotide identity patterns of the recently exchanged core genes (95-98% range) suggested that these genes were brought into the genomes through homologous recombination mechanisms (15). These patterns could be best explained through the mechanisms of transformation, transduction and conjugation for genetic exchange and homologous recombination as the process through which the exchanged DNA could be inserted into the genome. The genomes of *Shewanella* isolates encode several genes with amino acid similarities to conjugative transfer genes and a complete *recA*-dependent homologous recombination protein complex.

One interesting observation made during this study supports the findings of our report on *Halomonas* species divergence at GSP (Chapter 3). Caro-Quintero et al (2010) (15) report that more anaerobic metabolism genes were exchanged between strains from more similar water depths indicating genomic adaptation of the *Shewanella* strains to their environmental conditions, mediated by horizontal gene exchange leading to species divergence (rapid speciation). To the best of our knowledge, such rapid, genome wide adaptation mediated by genetic exchange between species of bacteria in response to environmental settings has rarely been reported. Widespread genetic exchange between co-existing strains has been shown by metagenomic studies of natural bacterial populations (96).

The sequence information obtained from selected *Shewanella* genomes during this study suggests that genetic exchange through homologous recombination events could constitute an important apparatus for population structure among spatially co existing bacteria, similar to the role of sexual reproduction in higher organisms. Thus, this study of comparing the population structure and adaptation of two different bacterial genera helps in understanding of the rapidity and mode of adaptation and highlighting the relationship between ecological setting and genetic mechanisms that form and endure microbial population structure.

In conclusion, it appears from our study that bacterial genomes adapt through continuous genome wide genetic exchange events in a highly dynamic, nutrient rich environment. This differs from what was observed previously in other marine bacteria (45, 19) living in a rather constant, nutrient poor environments, with a streamlined genome. The patterns observed in our study may be applicable to other bacteria that experience frequent environmental fluctuations.

CHAPTER V

DNA REPAIR POTENTIAL OF *HALOMONAS* SPP. FROM THE GREAT SALT PLAINS

ABSTRACT

Great Salt Plains (GSP) National wildlife reserve located in North Central Oklahoma is the ecologically diverse, extreme, hypersaline environment. Evaporated mineral salt crusts form the surface layer of these plains. Microorganisms living in the GSP are continuously exposed to desiccating conditions, high surface temperatures, freezing winters, high salinity and direct Ultra Violet (UV) radiation. Survival under these conditions has led to the development of unique survival strategies. DNA repair mechanisms associated with UV light or chemically induced DNA damage have been shown to be important in protecting the microorganisms from desiccation. DNA repair potential of the *Halomonas* spp. isolates, which are the dominant Gram Negative bacterial group from the GSP was studied. Survival after exposure to UV light was used as a tool to assess the DNA repair capacity. Several *Halomonas* spp. isolated from the GSP were tested and compared with *E.coli* and *P. aeruginosa* for DNA repair capacity. Tolerance to UV induced DNA damage of few of the *Halomonas* spp. isolates was on par with *E.coli* but was greater than the *P. aeruginos*. Wilson et al (125) have suggested that these GSP isolates may have developed constitutively expressed inducible DNA repair mechanisms positively regulated by *recA* gene. Partial sequence information for the *recA* gene was obtained from all the isolates tested for phylogenetic comparison with 16S rRNA sequences. *recA* gene based phylogenetic tree provided a a greater degree of resolution of the relationship between closely related *Halomonas* spp. isolates obtained from a diverse and extreme environment like GSP.

INTRODUCTION

Extreme environment at GSP and DNA repair:

Great Salt Plains (GSP) is part of the Salt Plains National Wildlife Refuge (SPNWR) with ecologically diverse, extreme, hyper saline environment. Salts are deposited by continuous percolation of underlying brine to the upper surface, sodium chloride being the main constituent of the brine. The salt concentration varies considerably across the plains. These salt formations dissolve completely during rainy seasons forming many small ponds and streams with varying salt concentrations across them. GSP is an extreme environment and microorganisms living in these plains are continuously exposed to harsh environmental conditions such as high temperature, freezing winters, high salinity, high desiccating conditions and direct exposure to Ultraviolet (UV) rays. Survival under these extreme conditions calls for selection among the microbial communities (16). The microbial community of these hyper saline terrestrial ecosystems is different from the hyper saline aquatic ecosystem, thus the salt tolerant relatives of the common soil bacteria are usually found in these hyper saline soils.

Culturable heterotrophic bacteria were identified from GSP soils using enrichment cultures and phylogenetic analysis. Different clades of GSP isolates were enriched for Gram-positive and Gram-negative bacteria. Bacillus and *Halomonas* spp. were the dominant ones among these bacteria respectively. These isolates were closely related to the bacteria isolated from aquatic systems. Many differences in overall phenotypic characteristics were observed when compared with isolates obtained from the abandoned saltern soils (98). Very few GSP isolates were motile, oxidase positive, did not produce H_2S, grow at temperature >45°C and were less halophilic (0.1 to 30%)(16). Survival under extreme environmental conditions has led to development of certain genetic mechanisms to protect these bacteria against varying salt concentrations and constant desiccating conditions (Chapter 3). Studies have shown that desiccation produces the same kind

of cellular damage on microorganisms as high UV irradiation (6). UV radiation imparts a significant amount of DNA damage in majority of organisms and this is dependent on the source and intensity of the radiation and duration of exposure. Microorganisms have developed unique survival strategies against UV radiation by producing chemicals or pigments to shield themselves against the UV damage and in addition to this they may have developed efficient repair mechanisms to restore the damage caused by the UV radiation. Therefore we believe that UV radiation can be used as an effective tool to study the effect of desiccation on these microorganisms.

Ultraviolet Radiation:

UV radiation is a short wavelength electromagnetic radiation that falls between the visible region and the x-ray region. UV radiation is known to cause deleterious effects in almost all living organisms ranging from prokaryotes, eukaryotes including higher plants and animals. UV radiation can be grouped into three major classes, UV-A (315-400nm), UV-B (280-315nm) and UV-C (<280nm). Each of these radiations causes different types of damages to DNA. Because of its short wavelength, UV-C is the most detrimental to the living cells as it is absorbed directly by the DNA and, cyclobutane dimers and single stranded breaks are formed. But most of the UV-C is screened out by the stratospheric ozone layer before it reaches the earth's surface. UV-B is absorbed in a small amount and it causes both direct and indirect damage (by forming reactive oxygen species) to the DNA. UV-A wavelengths impart least amount of damage and are responsible for indirect damage to the DNA by producing secondary photoreactions of existing DNA photoproducts (86).

DNA damage:

Mutagenic DNA lesions induced by UV radiation are usually of two kinds: 1) Cyclobutane-pyrimidine dimers (CPDs)- A cyclobutane dimer is a four-membered ring structure formed

between adjacent pyrimidines (thymine or cytosine) resulting from the saturation of their 5 and 6 carbon double bonds. 2) 6-4 photoproducts (6-4 PPs)- These photoproducts are produced specifically at 313 nm at a dose between 100-500 J/m^2. These 6-4 PPs are linkages between the 6 carbon of 3' cytosine or thymine and the 4 carbon of a 5' cytosine. During this process, in addition to the C4-C6 bond formation between the adjacent pyrimidines, C4 of one migrates to the C6 of other pyrimidine.

CPDs are the most abundant and most cytotoxic lesions but the 6-4 PPs have more lethal and mutagenic effects. The major portion of solar UV radiation borders between UV-A and UV-B and produces a higher proportion of photoproducts since the photoisomerization is efficient around 320nm (107). In both microorganisms and mammals, the effects of CDPs on key biological functions have been extensively studied. It has been reported that CPDs stall the progress of DNA polymerases and if unprepaired, it can completely stop the expression of a transcriptional unit. Therefore, if these CPDs are left unrepaired, they can block both transcription and replication (107).

DNA repair:

For all organisms the key to survival is the efficient transfer of genetic material from one daughter cell to another. This requires efficient and accurate DNA replication machinery and the ability to withstand continuous DNA damage with less heritable mutations. In order to achieve this goal, organisms have developed efficient DNA repair mechanisms so that they can survive the lethal effects of DNA damage. These repair mechanisms continuously scan the genome of the organism for the DNA damage and when necessary, trigger DNA repair and restore genetic information (107). UV radiation imparts significant amount of DNA damage in organisms and this is dependent on the source and intensity of the radiation and duration of exposure. Microorganisms have developed unique survival strategies against UV radiation. Organisms

produce chemicals or pigments to shield themselves from UV radiation and have developed efficient repair mechanisms to restore the damage caused by the UV radiation. In bacteria there are two major types of DNA repair system: 1) light repair and 2) dark repair (13).

Light Repair: This mechanism involves light and an enzyme known as photolyase. This enzyme binds to CPD's or 6-4 PP's and reverses the damage using light as energy source. This process is known as photo reactivation. This was first observed in Streptomyces griseus (13). Photolyases have been reported in bacteria, fungi, plants, invertebrates and some vertebrates. But this enzyme seems to be absent or silent in humans. Photolyases are monomeric flavin containing enzymes with molecular weight of around 50-65 K Da. These enzymes work very well when there is light present in the range of 300-450 nm. Photo reactivation is extremely efficient when the photolyase is bound to CPD's. In bacteria,the phr gene encodes photolyase (98). It has been reported that photolyase enzymes also carry out an alternate function. This suggestion resulted from the observation that in the absence of light the enzyme binds to Pyrimidine dimers and stimulates the nucleotide excision repair system (48).

Dark Repair**:** The DNA damage left unrepaired by the photo reactivation is taken up by the dark repair system. There are three major types of dark repair mechanisms:

Excision repair**:** In contrast to photo reactivation, dark repair system replaces the damaged DNA with new undamaged nucleotides. There are two different systems in excision repair: Base Excision Repair (BER): BER protects the cell from DNA damages caused by hydrolysis, reactive oxygen species and ionizing radiations (118). The key enzyme involved in BER is DNA glycosylases. There are different types of DNA glycosylases and the specificity of the repair pathway depends on the kind of glycosylase present. Glycosylase function by cleaving the N-glycosidic bonds between the bases and 2-deoxyribose moieties of the nucleotide residues (103). Once the bases are removed, AP endonucleases and phosphodiesterases act to remove the

remaining residues. DNA polymerase fills the gap and DNA ligase seals the strand. Enzymatic base excision was first identified for Uracil. The lesions produced at uracil are highly mutagenic and hence uracil glycosylases are well conserved among all the organisms. In addition to a number of other DNA glycosylases, many organisms have UVendonuclease enzyme. This enzyme acts at the site of the pyrimidine dimers. UV endonucleases are usually found in organisms that are resistant to UV radiation (77). Nucleotide Excision Repair (NER): NER is involved in repairing different kinds of DNA damages including CPD's and 6-4 PP's. It is present in most of the organisms and highly conserved in eukaryotes. NER employs a product of around 30 genes to repair a lesion from the DNA. In humans where there is a defective NER, it may lead to cancer prone genetic disorders such as xeroderma pigmentosum, cockayne's syndrome and trichothiodystrophy (107). NER can be divided in two sub categories: a) Transcription Coupled Repair (TC-NER): Repair of transcribed strands in active genes. b) Global Genome Repair (GG-NER): Repair of non-transcribed parts of the genome. In prokaryotes, protein products UvrA, Uvrb and UvrC perform the damage recognition and repair. The uvrABC nuclease recognizes the dimer and makes a break on either side of the lesion. Helicase releases the complex, DNA polymerase I fills the gap and ligase seals the ends.

Recombinational Repair**:** Recombination is an important process in DNA repair. The *recA* gene plays an important role in this process through DNA replication and strand exchange. Double strand breaks and single strand gaps are repaired by this process (105). During the post replication recombinational repair daughter strand gaps are filled by transferring a complimentary strand from the homologous region of DNA to the site opposite the damage. The lesion is unrepaired and the damaged base is still available for the excision repair. The complimentary strand obtained during this process is error free. The RecA protein performs the homology search function followed by the strand exchange (60).

SOS Repair: RecA is a key regulatory protein involved in SOS mutagenic repair (105). RecA activates the SOS network by cleaving repressor LexA and stimulating the activity of series of gene products. Inducing the SOS leads to expression of UmuC and UmuD. These proteins inhibit the cell cycle so that DNA repair can occur before cell division (75, 13). Mutaions occurring as a result of damage by UV radiation and various chemicals usually involves polymerases and when it comes across a noncoding or miscoding lesion, inserts an incorrect nucleotide opposite the lesion (57). In bacteria, the umuCD gene products binds to DNA polymerase and alters its proof reading fidelity thus allowing it to move past the DNA lesion.

***recA* gene product:**

The *recA* gene and its protein product RecA are key elements in the process of homologous recombination in bacteria. Bacterial recA mutants are completely devoid of recombinational activity. The first evidence to this process came from experiments in *E.coli* where RecA protein promotes the annealing of single-stranded DNA and double-stranded DNA molecules with complementary sequences is an ATP dependent reaction. Thus the RecA protein favors pairing and strand exchange between homologous DNA molecules. RecA plays the role of synaptase in this strand exchange reaction (66). The *recA* gene product alo plays an important role in the bacterial cell's response to several environmental stress factors that cause damages to the DNA. The DNA repair process is part of a regulon known a SOS regulon that is induced by various stress factors. This regulon is comprised of more than 20 genes including *recA, lexA, umuCD, recN, uvrB and uvrD.* These genes are under the control of *lexA* repressor that binds to sites called as SOS boxes in the promoter region. When a bacterial DNA is hit by stress factors such as UV radiation, chemicals etc, a signal is generated which includes the RecA protein and this acts as a coprotease i.e., it speeds up the auto catalysis of LexA repressor protein. This results in decrease in the concentration of LexA protein and induction of SOS regulon. This induction takes place when the replication fork passes the lesion on the DNA. Thus, this regulon acts as a last

ditch effort to save the DNA from damaging lesions (76, 13, 43). Apart from inducing the SOS regulon, RecA protein is directly involved in repair of bacterial DNA damages after exposure to UV radiation. This process is called post-replication recombinational repair. Here two types of lesions are processed: a) Daughter strand gaps formed due to skipping of a lesion in the template during the DNA replication. b) Double strand breaks formed due to cutting of a DNA strand opposite a gap.

RecA protein regulates the mutagenesis in bacteria by two different mechanisms: a) By cleaving LexA and allowing the derepression of umuDC operon which is responsible for SOS mutagenesis. b) By post-translational activation of UmuD protein using proteolytic cleaving. In E.coli it has been shown that double mutants of lexA and recA that can efficiently express UmuC and UmuD were unable to produce any UV-induced mutation. This indicates that RecA is more directly involved in the SOS mutagenesis than just being a coprotease (76). Several studies have showed that recA mutants are highly sensitive solar UV radiation. It is reported that recA mutants of marine isolates of p.aeruginosa are significantly more sensitive to UV-A and UV-B than the wild type isolates. These isolates were able to restore their UV resistance capacity when a functional recA gene was introduced into their cells. RecA has also been shown to be important for survival of many terrestrial bacterial and archeal isolates (10, 101, 58, 6, 61). RecA is highly conserved among many prokaryotes. These findings suggest that RecA originated very early during the evolution of prokaryotes. Thus recA gene can be used as a good indicator for bacterial response to UV radiation induced DNA damage.

Why *Halomonas*?

Several studies of the DNA repair capacity have been carried out on bacterial isolates from saline aquatic environment. Similar studies not been carried out on microorganisms from hyper saline terrestrial environment. Few studies have been reported on the effect of UV radiation in

Halomonas spp. isolates from the hyper saline aquatic environments (17). Wilson et al. analyzed the UV survivability of a group of Halomonas spp. isolates from the GSP. These bacteria were chosen because they were the dominant Gram –ve isolates at GSP. In an earlier evaluation of phylogenetic status of family Halomonadaceae, the authors found that Halomonas spp. isolates shared around 97% 16S/23S rDNA sequence similarity. These isolates have a wide range of G+C content (52-68 mol%). It is accepted that members of same genus should not have G+C content range >10 mol%(3, 4). Since *Halomonas* exceed this range, it would appear that these groups of organisms are evolutionarily plastic with a potential of high levels of horizontal gene transfer. Most of the Halomonas spp. isolates from the GSP showed greater levels of resistance to the DNA damage caused by the UV irradiation. This was greater than the resistance observed for the hyper saline aquatic isolates. The UV survivability for these GSP isolates was also compared to those of *E. coli* and *P. aeruginosa*. The UV resistance of Halomonas vinusta like isolates was not significantly different than *E.coli* but was significantly more resistant than *P.aeruginosa*. Halomonas salina like isolates showed significant variation in their resistance to UV irradiation (125). Thus we hypothesize these bacteria surviving under extreme desiccating conditions at GSP might have developed an efficient constitutively expressed DNA repair mechanisms that increases genetic variations. This is supported by the observations made by (125) where the GSP isolates reported a higher spontaneous mutation rates following UV-C exposure. The regulation of DNA repair system is positively controlled by the gene product RecA, which is thought to be highly conserved among all the bacterial species (76). Evaluation of these Halomonas isolates from GSP for their survival abilities against UV radiation will help us to understand the molecular mechanism of DNA repair in these hyper saline, terrestrial organisms.

METHODS

Bacterial Strains used in this study: Bacterial strains used in this study are listed in Table 5.1. *Halomonas* spp. isolates listed in Table 5.1(GSP 1- GSP 58) were obtained in 2001-2002 from two different dry, salt-crusted sites WP3 (N 36° 42.856' and W 98° 15.725') and WP8 (N 36° 42.750' and W 98° 15.584) at the GSP (Figure 1). Strains GSP 1001 – GSP 1018 were obtained in 2005 from a salt gradient site WP68 (N 36° 42.856' and W 98° 15.725'). Isolates were obtained using enrichment cultures and were grown and maintained on SP-1 medium (16). Colonies were streaked several times on fresh SP-1 agar plates to ensure purity. These isolates were identified as belonging to *Halomonas* based on phenotypic (microscopic, gram characteristics and biochemical tests) and phylogenetic (16S rRNA gene sequence analysis) characterization (16 and this study). Based on this information, 32 *Halomonas* spp. isolates from GSP were included in this study (Table 5.1). *Chromohalobacter salexigenes* DSM 3043 was also included in the study since this is the only close relative of *Halomonas* spp. for which a complete genome sequence is available. *E. coli* and *P. aeruginosa* strains were used as controls and grown on LB medium. Subgroups of isolates listed in Table 5.1 were used in this study for UV sensitivity testing and phylogenetic analysis.

Table 5.1 Bacterial strains used in this study

Isolate	Species
GSP1	*Halomonas* spp.
GSP2	*Halomonas* spp.
GSP3	*Halomonas* spp.
GSP4	*Halomonas* spp.
GSP5	*Halomonas* spp.
GSP19	*Halomonas* spp.
GSP21	*Halomonas* spp.
GSP23	*Halomonas* spp.
GSP24	*Halomonas* spp.
GSP25	*Halomonas* spp.
GSP26	*Halomonas* spp.
GSP27	*Halomonas* spp.
GSP28	*Halomonas* spp.
GSP33	*Halomonas* spp.
GSP45	*Halomonas* spp.
GSP58	*Halomonas* spp.
GSP1001	*Halomonas* spp.
GSP1002	*Halomonas* spp.
GSP1003	*Halomonas* spp.
GSP1004	*Halomonas* spp.
GSP1005	*Halomonas* spp.
GSP1006	*Halomonas* spp.
GSP1007	*Halomonas* spp.
GSP1008	*Halomonas* spp.
GSP1009	*Halomonas* spp.
GSP1010	*Halomonas* spp.
GSP1011	*Halomonas* spp.
GSP1012	*Halomonas* spp.
GSP1013	*Halomonas* spp.
GSP1016	*Halomonas* spp.
GSP1017	*Halomonas* spp.
GSP1018	*Halomonas* spp.
DSM3043	*Chromohalobacter salexigenes*
JC158	*Escherichia coli*
PA01	*Pseudomonas aeruginosa*

UV Sensitivity: GSP Isolates were grown in 10 ml of SP-1 medium. SP-1 is a low magnesium medium containing (per liter): NaCl (98 g), KCl (2.0 g), $MgSO_4$.7H2O (1.0 g), $CaCl_2$.2H2O (360 mg), $NaHCO_3$ (60 mg), NaBr (230 mg), $FeCl_3$.6H2O (1 mg), Bacto tryptone (5.0 g), yeast extract (10.0 g), and glucose (1.0 g), final pH 6.8. *C. salexigenes*, *E.coli* and *P. aeruginosa* strains were grown in LB medium. After 16-18 hours, 1 ml of this overnight culture was transferred to a 125 ml side arm flask containing 20 ml of medium (SP-1/LB). All cultures were grown to mid exponential phase. Ten milliliters of this culture was centrifuged at 4960g for 10 minutes. The pellet was suspended in an equal volume of sterile 10% saline and transferred to a sterile empty petri dish for exposure to various doses of UV-C radiation. Broadband UVR was generated by ultraviolet bench lamps (Spectronics corp., Westbury, NY). Model XX-15F that produces a peak at 254 nm of approximately 1.1 J/m^2 at 15 cm distance was used to generate UV-C. Measurement of irradiation was performed with a UVX Digital Radiometer with a probe suitable for UV-C (Ultra-Violet Inc., San Gabriel, CA).

All experiments were conducted in the dark under amber light to minimize photoreactivation. For each exposure time, appropriate dilutions were made and plated onto duplicate plates (SP-1/LB). Plates were incubated 24-48 hours and the colonies were counted to determine CFU. Each experiment was conducted three times to quantify surviving CFU. Percent survival was calculated using the following equation,

Percent Survival = CFU of post treatment growth / CFU of control growth.

Survival curves for each isolate were generated separately. Means and standard errors of the mean were calculated (Table 5.2). UV sensitivity coefficients (S_{UV}) were determined for selected isolates at LD_{90} using the formula,

$$S_{UV} = \ln \{ (CFU)_D / (CFU)_0 \} / D$$

Where, $(CFU)_0$ is the concentration of CFU in the unexposed sample and $(CFU)_D$ is the recoverable concentration after exposure to dose D of UV radiation (106).

Primer design, PCR amplification and sequencing: DNA extraction and PCR amplifications were carried out as previously mentioned (Chapter 3). Genomic DNA was the target for PCR amplification of partial fragments of 16S rRNA and the *recA* gene. For 16S rRNA: DNA was denatured at 95° for 2 min, followed by 40 cycles of 95°C for 1 min, 50°C for 1 min, and 72°C for 1 min, with a final extension period of 5 min (16). Since, prior to this study, no *recA* sequence has been determined for any *Halomonas* species, primers for PCR amplification of interior fragment of *recA gene* of GSP-5 were designed using a highly conserved region from the alignment of previously characterized *recA* gene sequences including *C.salexigenes*. ~200 bp product was successfully amplified, gel purified, cloned into the pDrive cloning vector (QIAGEN PCR Cloning Kit) and sequenced. The sequence information obtained was used to design primers for amplification of a larger region of *recA* for phylogenetic studies. Using these primers, a ~500 bp product was successfully amplified from several *Halomonas* spp. isolates from GSP. PCR parameters used for amplification of the *recA* gene fragment were as follows, initial denaturation of 5M in at 94°C followed by 35 cycles of 94°C for 2 min, 57±3°C for 1 min, 72°C for 2 min, and final extension of 10 min at 72 °C.

Phylogenetic analysis: Sequences obtained for 16S rRNA and *recA* gene fragments were aligned using CLUSTALX (64), trimmed to equal lengths and 711 and 300 bases were used respectively for phylogenetic analysis. Neighbor –Joining (NJ) trees were constructed using the program PAUP* 4.0 (113) with Jukes-Cantor correction. *P.aeruginosa* 16S rRNA and *recA* gene sequences were used as out group.

RESULTS AND DISCUSSION

Caton et al. (16) have described in detail the isolation procedure and physiological characterization of isolates GSP 1- GSP 58. In this study, we subjected the isolates GSP 1001-GSP1018 to a series of biochemical and physiological tests in order to identify them. This analysis confirmed then as *Halomonas* spp. (Figure 5.1). Only few isolates were motile and able to metabolize sugar and amylase. More than 50% of the isolates tested positive for oxidase and catalse tests. Most of these isolates were thermotolerant and were able to grow at a wide range of temperatures from 4°C to 45°C with an optimum at 37°C. Most of the isolates were able to grow at salt concentration from 1%-15%. They did not require hypersaline conditions for growth. Optimum growth rates were observed at 10% NaCl. These results are not surprising considering the rapid fluctuations in temperature and salt concentrations at the GSP site during different times of the year.

A

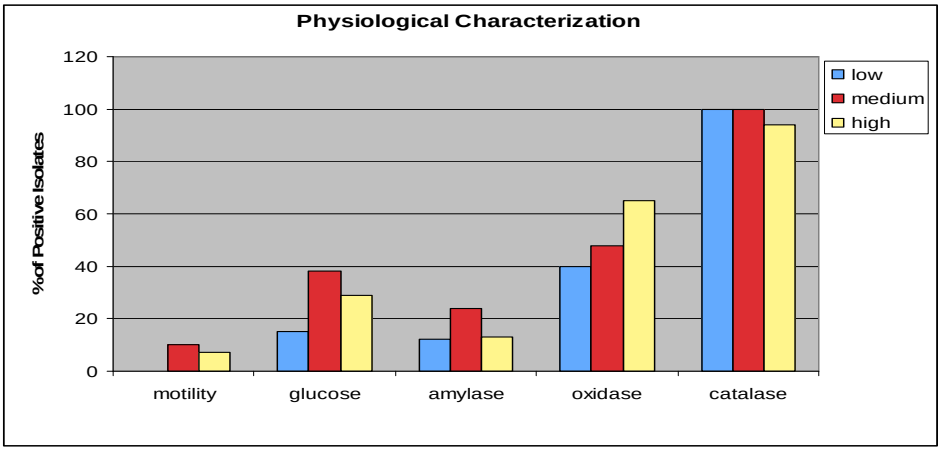
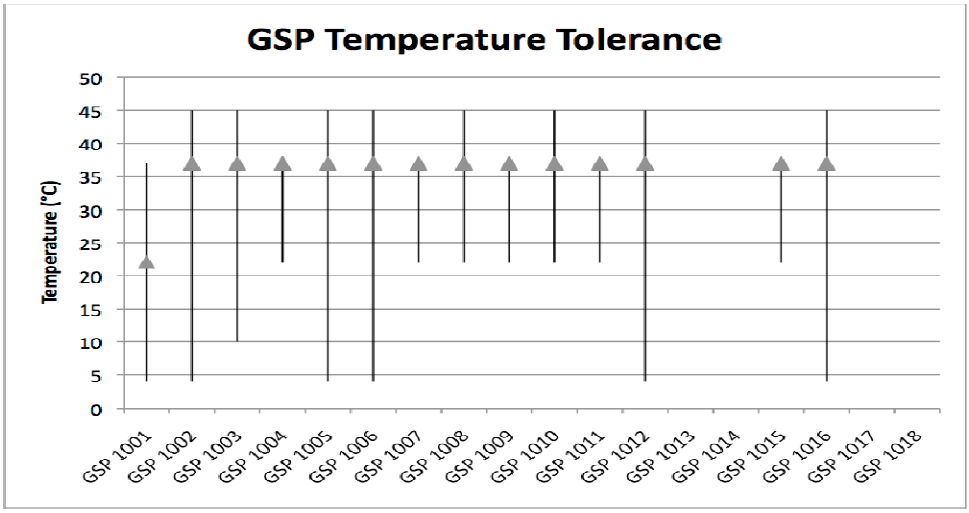
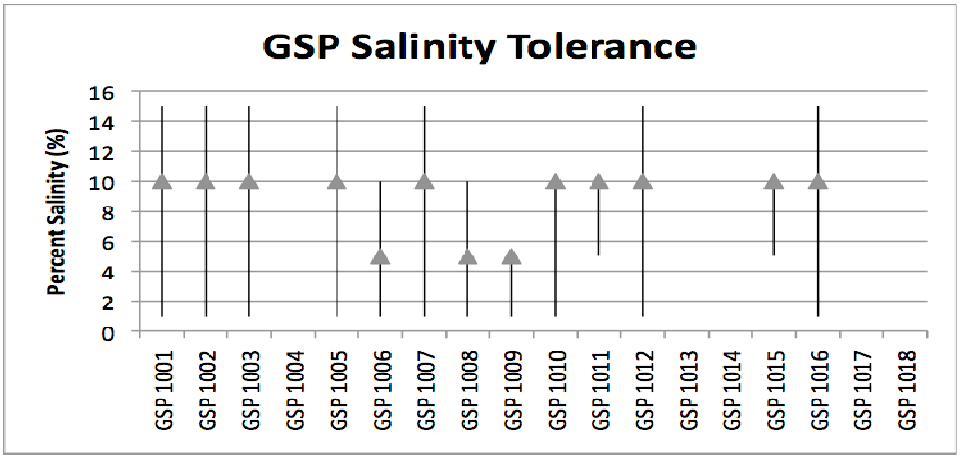


Figure 5.1: Biochemical and phenotypic characterization of GSP isolates. (**A**) Chart comparing biochemical test results of isolates obtained at different salt gradient sites (**B**) Chart showing temperature tolerance of isolates GSP 1001-GSP1018 (**C**) Chart showing Salt tolerance of isolates GSP 1001-GSP1018.

B



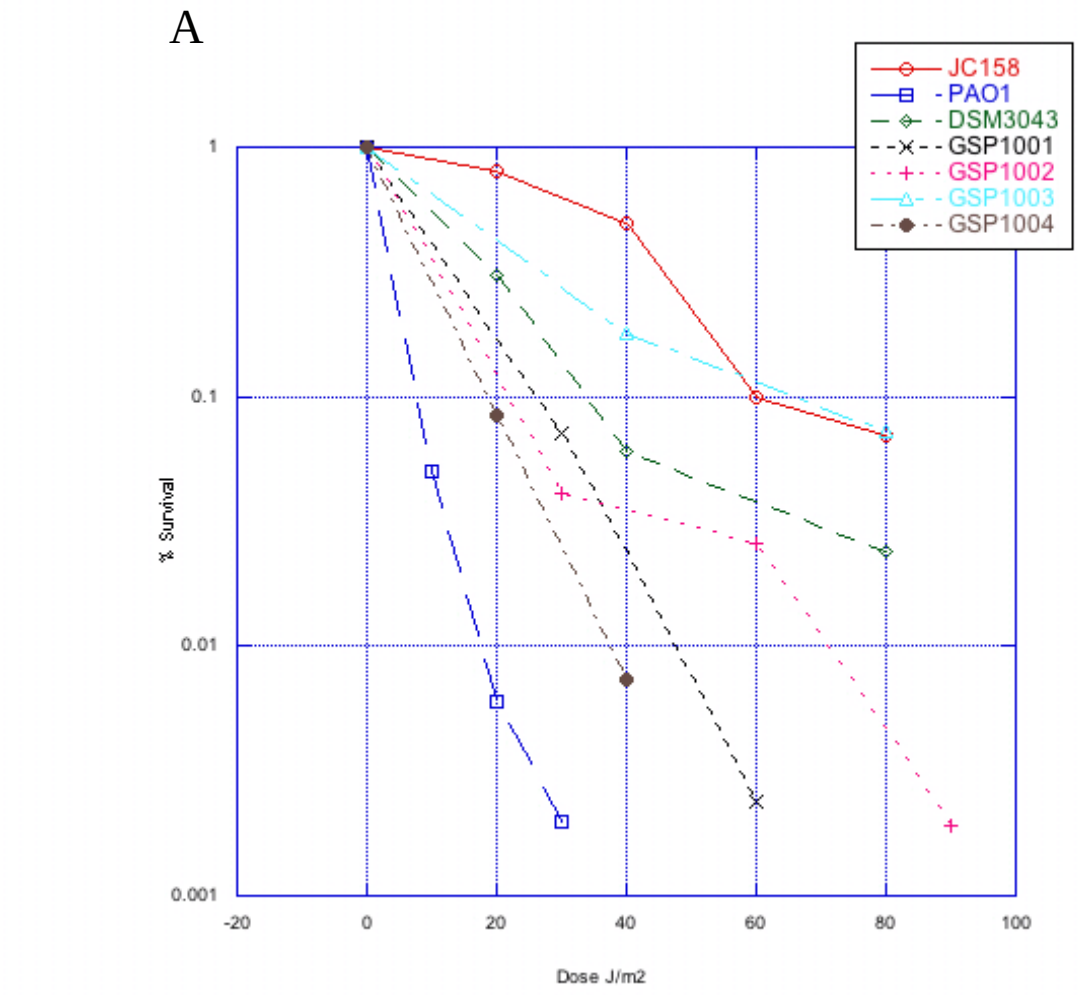
C



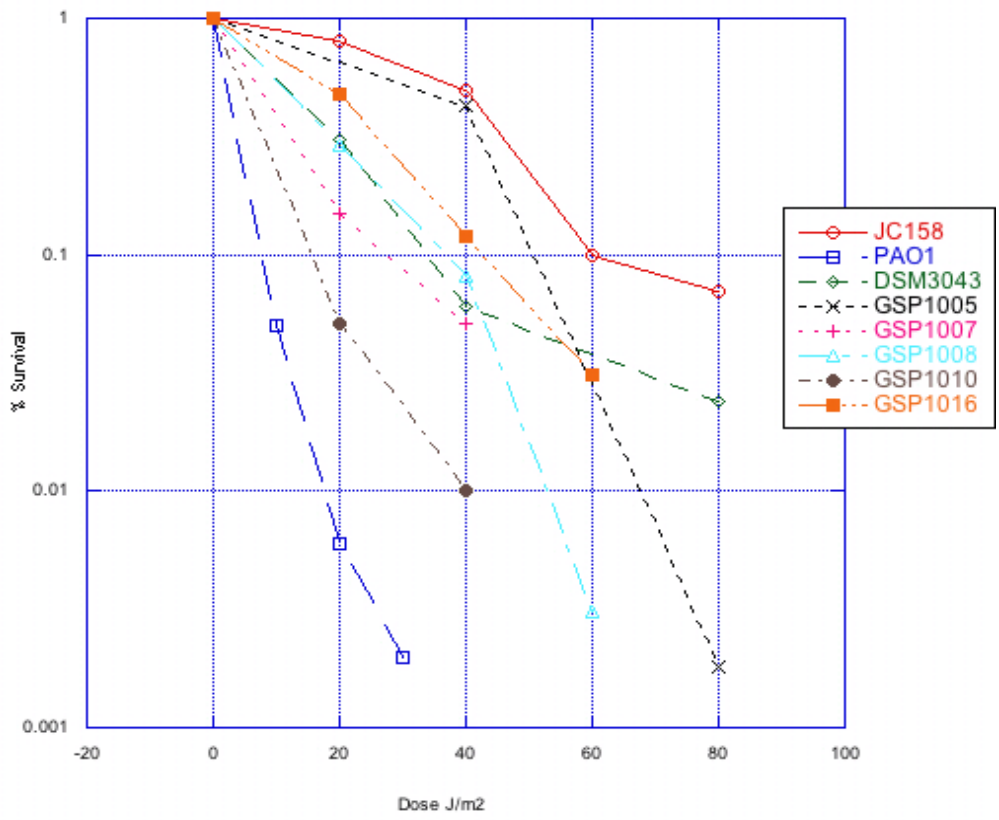
UV Sensitivity:

Wilson et al. (2004) (125) reported the DNA repair potential of *Halomonas* spp. Isolates exposed to different wavelengths of UV radiation. In this report, I studied the DNA repair potential of nine *Halomonas* spp. isolates (GSP- 1001, 1002, 1003, 1004, 1005, 1007, 1008, 1010, 1016), as well as *C. salixegens* and well characterized strains of *E.coli* and *P.aeruginosa* when exposed to UV-C radiation (Figure 5.2). UV sensitivity coefficients (S_{UV}) were calculated and compared with previously studied *Halomonas* spp. Isolates (Table 5.2).

Figure 5.2 UV sensitivity curves for *Halomonas* spp. Isolates. (A) GSP-1001, 1002, 1003, 1004. (B) GSP-1005, 1007, 1008, 1010, 1016. JC158(*E.coli*), PAO1(*P.aeruginosa*), DSM3043(*C.salexigens*).



B



All the GSP isolates tested in this study were quite resistant to UV-C. GSP-1003, and GSP-1005showed increased level of resistance at levels that were not significantly different from *E.coli.* GSP-1001, 1004 and 1010 were less resistant at higher doses of UV-C, but were significantly more resistant than *P.aeruginosa* (Figure 5.2 and Table 5.2).

Martin. et al.(95) studied UV resistance of several isolates of *Halomonas elongata* obtained from hypersaline aquatic environments. *Halomonas* isolates from GSP appear to be more resistant to UV radiation than their counterparts from aquatic environments (S_{UV} of -92 $x10^{-2}$ in *H.elongata*). This suggests that *Halomonas* isolates from terrestrial environment have

acquired/retained DNA repair mechanism with elevated capacities than the isolates from aquatic environments.

Table 5.2 UV Sensitivity coefficients of *Halomonas* spp. isolates

Strains	UVC LD90[a] (J/m2) B	CFUD/CFU0	ln(CFUD/CFU0) D	SUVC (D/B)	SUVC(x10-2)[b]
PAO1	8	0.1	-2.302585093	-0.287823	-28.8
GSP1010	16	0.1	-2.302585093	-0.143912	-14.4
GSP1004	18	0.1	-2.302585093	-0.127921	-12.8
GSP1002	22	0.1	-2.302585093	-0.104663	-10.5
GSP1001	25	0.1	-2.302585093	-0.092103	-9.2
GSP1007	28	0.1	-2.302585093	-0.082235	-8.2
DSM3043	34	0.1	-2.302585093	-0.067723	-6.8
GSP1008	38	0.1	-2.302585093	-0.060594	-6.1
GSP1016	42	0.1	-2.302585093	-0.054824	-5.5
GSP1005	50	0.1	-2.302585093	-0.046052	-4.6
JC158	60	0.1	-2.302585093	-0.038376	-3.8
GSP1003	62	0.1	-2.302585093	-0.037139	-3.7

[a]UV dose producing a 90% reduction in viability, [b]The more negative the value, the more sensitive the organism.

Wilson et al. (125) reported a high frequency of spontaneous mutation in the *Halomonas* spp. Isolates from GSP. These observations suggest that *Halomonas* spp. Isolates from GSP have adapted to the constantly fluctuating, extreme desiccating conditions at GSP and this has led to the development of constitutively expressed error prone DNA repair mechanisms. Regulation of this mechanism is positively controlled by the gene product RecA (76). Detailed analysis of the *recA* gene from these *Halomonas* isolates will answer questions about the molecular mechanisms of DNA repair in this and other hypersaline, terrestrial environments.

Sequencing of 16SrRNA and *recA* genes from *Halomonas* spp. isolates:

Since the *Halomonas* spp. isolates from the GSP site demonstrated elevated levels of constitutively expressed error prone DNA repair strategies positively controlled by *recA* (76), we decided to study the sequences of *recA* gene and compare it with the sequences from 16SrRNA

gene to understand the role of the *recA* gene in shaping the population structure of these isolates at this hypersaline, extreme environment. Since the *recA* gene sequences showed increased level of variations in our study (Chapter 3), they can provide better resolution to differentiate the *Halomonas* spp. isolates at species level along with the 16SrRNA based identification. The 16SrRNA and *recA* gene sequences were determined for 32 terrestrial *Halomonas* spp. isolates from GSP (Table 5.1). The 16SrRNA and *recA* gene sequences were trimmed to equal lengths and 711 and 300 bases were used respectively for phylogenetic analysis. Parings of the same sequences of the 16S rRNA and *recA* genes were observed in some isolates (GSP 1005, 1009, 1013; GSP 3, 4, 5) that are grouped closely together on the 16S rRNA based NJ tree (figure 5.3)**.** Isolates that shared identical pairs of gene sequences were obtained from a same sampling site. The average interspecies (*Halomonas* isolates) sequence similarity values for each of these genes were 95.9% for the 16S gene and 85.1% for the *recA* gene. Even though some of the isolates shared similar sequences for these genes, they are not identical as determined by the colony morphology and phenotypic characters mentioned earlier. The majority of nucleotide differences in the *recA* sequences among *Halomonas* spp. isolates were synonymous substitutions. The *recA* gene had a low d_N/d_S ratio (<0.5) of 0.1932 (Chapter 3) indicating the presence of a selective pressure restricting amino acid changes.

Phylogenetic analysis:

Phylogenetic trees based on 16SrRNA and *recA* gene sequences were generated by neighbor-joining method implemented within the program PAUP 4.0 (Figure 5.3 and 5.4). Bootstrap values of 50% or greater are shown at the nodes of each tree. Tree topologies were significantly different. The *recA* gene tree was not in general agreement with the tree generated from 16S rRNA analysis. There were more variations in the *recA* gene sequences than in the 16S rRNA gene sequences. It appears that there has been so much change in the *recA* gene sequences among isolates that the phylogenetic relationship between the *Halomonas* spp. isolates has been

obscured by lateral gene transfer events that may have taken place in this plastic environment like GSP. The *recA* gene phylogenetic analysis provides a greater degree of resolution over 16S rRNA gene analysis among closely related isolates in the genus *Halomonas* obtained from a diverse and extreme environment like GSP.

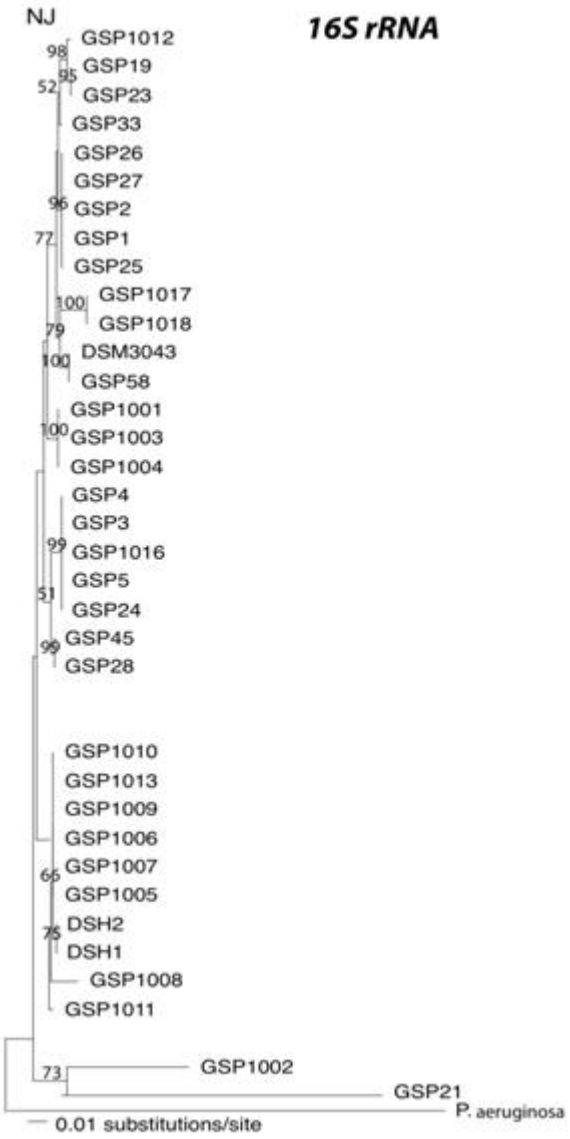


Figure 5.3 Phylogenetic tree based on the neighbor joining analysis of 16SrRNA gene sequences of *Halomonas* spp. Isolates from GSP

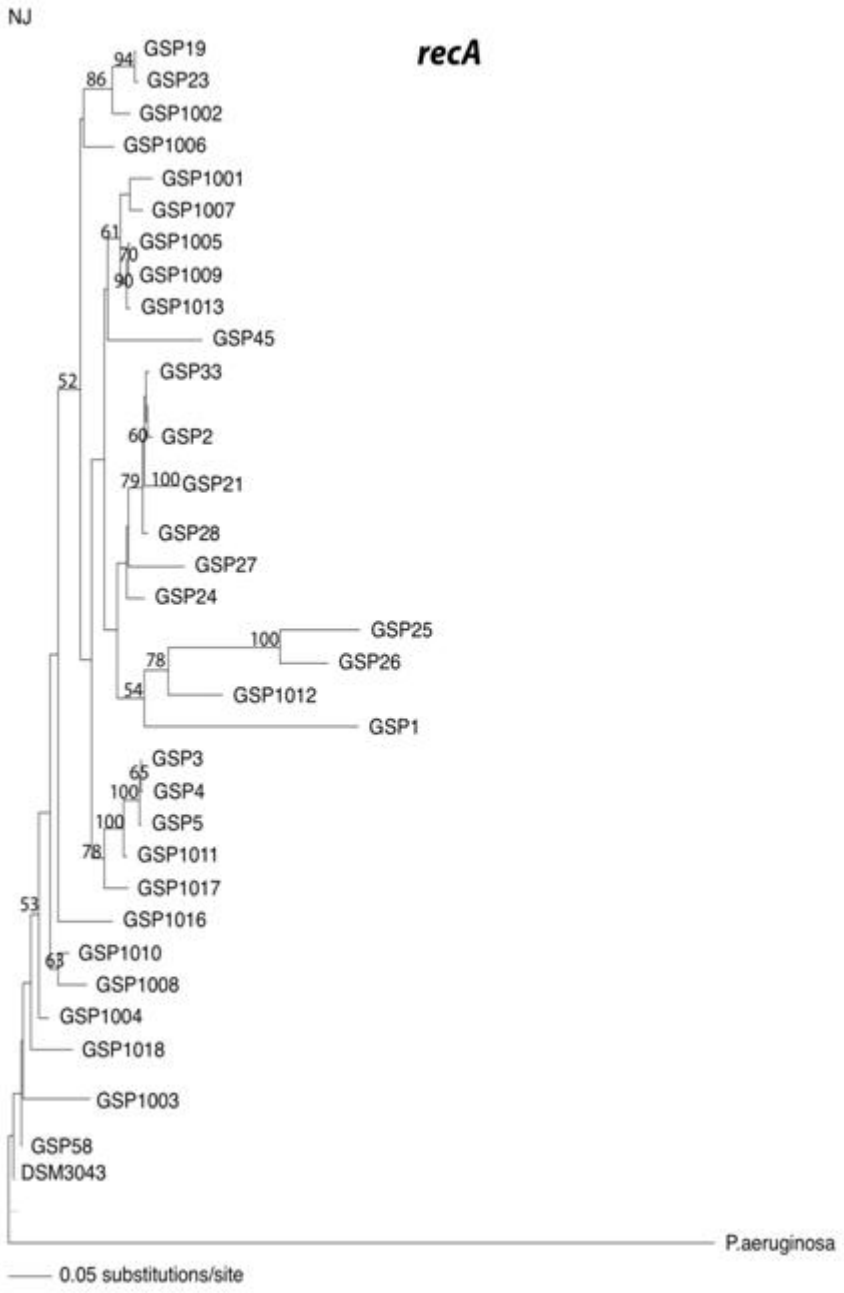


Figure 5.4 Phylogenetic tree based on neighbor joining analysis of *recA* gene sequences of *Halomonas* spp. Isolates from GSP.

In conclusion, our study has shown that *Halomonas* spp. isolates from GSP exhibit an increased level of resistance to damage from UV radiation presumably because expression of the *recA* gene allows for constitutive expression of mechanisms for DNA repair. Based on the sequence information of *recA* gene obtained during this study, we can draw conclusions relating the environment in which these bacteria thrive and the microbial population within that environment. This approach will help develop molecular tools required to characterize microbial populations which are of ecological, pathogenic and biotechnological interest.

CHAPTER VI

CONCLUSION AND FUTURE WORK

Despite extensive study of the behavior and biotechnological applications of *Halomonas* spp., the ecology of this bacterium remains poorly understood. Even though the questions asked and answered in this dissertation are diverse in nature, the common goal has been to understand the population structure of closely related *Halomonas* spp. from the hypersaline environment of the GSP. This knowledge will help us to understand the genetic variability and the adaptive evolution of this group of bacteria. We achieved this by estimating the rate of recombination among this group. This proved to be more complex than originally conceived. This study has opened the door for understanding the genetic diversity and evolutionary strategies of hypersaline terrestrial bacterial populations. To our knowledge this is the first report characterizing *Halomonas* spp. isolates using a sequence-based typing method (MLST). We found evidence for frequent LGT and recombination among these isolates. Recombination rates were on par with other sexual bacteria. The rates of recombination observed are high enough to overcome the effects of population bottlenecks and periodic selection that would otherwise purge the genetic variation generated. However, recombination was not frequent enough to completely disrupt linkage disequilibrium.

According to our findings, the mechanisms of recombination and gene exchange in *Halomonas* spp. are likely to have developed in response to selective pressure for preserving variations allowing for rapid evolution. Under this condition, recombining populations have a slight advantage over non recombining population when they are competing against each other (115). This type of plastic and dispensable genome provides chances for bacterial lineages to adapt to specific native, drastically fluctuating environmental conditions.

The extensive genetic variability observed in chapter 3 encouraged us to ask if the same patterns are observed in a bacterium living in a different ecotype. We analyzed and compared the partial sequence information of several house keeping genes of nine *Shewanella* isolates from diverse geographic locations and habitats including both fresh and marine water bodies. The sequence information analyzed from selected *Shewanella* genomes suggested that genetic exchange through homologous recombination events could constitute an important apparatus for population structure among spatially co-existing bacteria. This study helps us to understand the rapidity and mode of adaptation of bacterial species highlighting the relationship between ecological setting and genetic mechanisms.

While Chapters 3 and 4 focused on understanding the genetic diversity of *Halomonas* and *Shewanella*, in Chapter 5 we enquired whether the mechanisms *Halomonas* spp. isolates might have developed in order to overcome the DNA damages sustained by constant exposure to high desiccating and fluctuating extreme conditions. We used UV-C radiation to test the survivability of these bacteria when exposed to different doses under laboratory settings. *Halomonas* spp. isolates exhibited increased level of resistance to DNA damage from UV radiation on par with *E. coli*, presumably using *recA* gene dependent constitutively expressed error prone repair system as suggested by Wilson et al. (125).

The results of the studies carried out during my PhD have opened new roads for the research topics that range from the molecular evolution of environmental isolates to the significance of recombination in evolution of bacteria. The ability to understand how environmental factors impact the diversity and stability of microorganisms will serve to better understand the population genetics and evolution of bacteria. However, the fitness advantage provided by variations in gene content under environmentally significant conditions has not been tested and is an exciting future research area to put these results in perspective.

Future directions:

The work described in this dissertation used several molecular and bioinformatics tools to examine patterns of genetic diversity among a collection of *Halomonas* spp. isolates from the Great Salt Plains of Oklahoma. However, the population structure described is only based on culturable isolates. We cannot definitely say how well these cultured isolate represent the total population of *Halomonas* spp. at the GSP. So, more work is needed to verify the results obtained in this study using culture independent studies. It is also important to note that targeting core housekeeping genes will allow quantification of most of the isolates, but designing primers for niche-adaptive or variable genes may permit tracking subgroup of isolates within this group. In addition to measuring the degree to which *Halomonas* spp. isolates vary in their genetic content, quantitative measure of their abundance would provide understanding of how these bacteria barriers itself during periods of fluctuations in environmental conditions.

REFERENCES

1. Aanensen, D. M., & Spratt, B. G. (2005). The multilocus sequence typing network: mlst.net. Nucleic Acids Res, 33(Web Server issue), W728-733.

2. Alm, R. A., et al., 1999 Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen Halicobacter pylori Natur, 397 (6715), 176-80

3. Arahal, D. R. & Ventosa, A. (2006). The family Halomonadaceae. In The Prokaryotes. A Handbook on the Biology of Bacteria, 3rd edn, vol. 6, Proteobacteria: Gamma Subclass, pp. 811-835. Edited by M. Dworkin, S. Falkow, E. Rosenberg, K.H. Schleifer & E. Stackebrandt. Springer, New York.

4. Arahal, D. R., Ludwig, W., Schleifer, K. H., & Ventosa, A. (2002). Phylogeny of the family Halomonadaceae based on 23S and 165 rDNA sequence analyses. Int J Syst Evol Microbiol, 52(Pt 1), 241-249.

5. Awadalla P, Eyre-Walker A, Maynard Smith J. 1999. Linkage disequilibrium and recombination in hominid mitochondrial DNA. Science 286:2524–25

6. Baliga, N. S., S. J. Bjork, R. Bonneau, M. Pan, C. Iloanusi, M. C. Kottemann, L. Hood, and J. DiRuggiero. (2004). Systems level insights into the stress response to UV radiation in the halophilic archaeon Halobacterium NRC-1. Genome research 14:1025-35.

7. Baas Becking, L.G.M. 1928. On organisms living in concentrated brine. Tijdschr. Ned. Dierkund. Ver. Ser.III. 1: 6-9.

8. Battista, J. R. 1997. Against all odds: the survival strategies of Deinococcus radiodurans. Annual review of microbiology 51:203-24.

9. Baumgartner, J.G. 1937. The salt limits and thermal stability of a new species of anaerobic halophile. Food Res. 2: 321-329.

10. Better, M., and D. R. Helinski. 1983. Isolation and characterization of the recA gene of Rhizobium meliloti. Journal of bacteriology 155:311-6.

11. Biswas, Reema G., 2008 DNA repair capacity of *Great Salt* Plain cyanobacterium Aphanothece sp. SP24. . ProQuest Dissertations and Theses 73 pages; AAT 1461785

12. Bloch, M.R. 1976. Salt in human history. Interdisc. Sci. Rev. 4: 336-352

13. Booth, M. G. Solar ultraviolet radiation and the role ofrecA in marine bacteria. ProQuest Dissertations and Theses 0410:107 p.

14. Browne, W.W. 1922. Halophilic bacteria. Proc. Soc. Exp. Biol. Med. 19: 321-322.

15. Caro-Quintero, A., Deng, J., Auchtung, J., Brettar, I., Hofle, M. G., Klappenbach, J., et al. (2010). Unprecedented levels of horizontal gene transfer among spatially co-occurring Shewanella bacteria from the Baltic Sea. *ISME J*.

16. Caton, T. M., L. R. Witte, H. D. Ngyuen, J. A. Buchheim, M. A. Buchheim, and M. A. Schneegurt. (2004). Halotolerant aerobic heterotrophic bacteria from the Great Salt Plains of Oklahoma. Microbial ecology 48:449-62.

17. Chai, LH, Cui, XL, Wang, T, Xu, LH, Jiang, CL (2002) The microbial diversity of Salt Lake Keke in Western China. GenBank no. AY121439

18. Cohan, F. M. 2001. Bacterial species and speciation. Syst Biol 50:513-524.

19. Coleman ML, Sullivan MB, Martiny AC, Steglich C, Barry K, Delong EF et al. (2006). Genomic islands and the ecology and evolution of Prochlorococcus. Science 311: 1768–1770.

20. Cooper, J. E., & Feil, E. J. (2004).Multilocus sequence typing--what is resolved? *Trends Microbiol, 12*(8), 373-377.

21. Darwin, C. 1839. Journal of researches into the geology and natural history of the various countries visited by H.M.S. Beagle, under the command of Captain Fitzroy, R.N. from 1832 to 1836. Henry Colburn, London

22. DasSarma, S. and P. DasSarma 2006. Halophiles. Encyclopedia of Life Sciences, Wiley, London.

23. DasSarma, S. 2004. Genome sequence of an extremely halophilic archaeon, in Microbial Genomes, pp. 383-399, C.M. Fraser, T. Read, and K.E. Nelson (eds.), Humana Press, Inc, Totowa, NJ.

24. Daublin, V., N.A. Moran, and H. Ochman, 2003 Phylogenetics and the cohesion of bacterial genomes Science, 301 (5634), 829-832.

25. de la Haba, R. R., Arahal, D. R., Marquez, M. C., & Ventosa, A. (2009). Phylogenetic relationships within the family Halomonadaceae based on 23S and 16S rRNA comparative sequence analysis. *Int J Syst Evol Microbiol*.

26. Delétoile, A., V. Passet, J. Aires, I. Chambaud, M.-J. Butel, T. Smokvina, S.Brisse. Species delineation and clonal diversity in four Bifidobacterium species as revealed by multilocus sequencing, Research in Microbiologoy (2010), doi: 10.1016/j.resmic.2009.12.006

27. Didelot, X. and Falush, D. (2007) Inference of Bacterial Microevolution using Multilocus Sequence Data. Genetics 175, 1251-1266.

28. Dingle KE, Colles FM, Falush D, Maiden MCJ. 2005. Sequence typing and comparison of population biology of Campylobacter coli and Campylobacter jejuni. J. Clin. Microbiol. 43:340–47

29. Doroghazi J R and Buckley D H., (2010) Widespread homologous recombination within and between Streptomyces species. The ISME Journal 1-8

30. Drouin, G., Prat, F., Ell, M., and Clarke, G. D. (1999) Detecting and characterizing gene conversions between multigene family members. Mol. Biol. Evol. 16, 1369–1390.

31. Dundas, I. 1974. Halobacteria. BioSystems 6: 66-67.

32. Dundas, I. 1998. Was the environment for primordial life hypersaline? Extremophiles 2: 375-377.

33. Dutta, Partha,2006, Soil community analysis along a *salt* gradient using denaturing gradient gel electrophoresis of PCR-amplified 16S rRNA genes, ProQuest Dissertations and Theses, 91 pages; AAT 1439049

34. Feavers IM, Heath AB, Bygraves JA, Maiden MC. 1992. Role of horizontal genetic exchange in the antigenic variation of the class 1 outer membrane protein of Neisseriameningitidis. Mol. Microbiol. 6:489–95

35. Falush,D., M. Stephens, J. K. Pritchard, (2003) Genetics 164, 1567-1587.

36. Feil, E. J., M. C. J. Maiden, M. Achtman, B. G. Spratt, (1999). Mol. Biol. Evol. 16, 1496-1502.

37. Feil EJ, Enright MC. 2004. Analyses of clonality and the evolution of bacterial pathogens.Curr. Opin. Microbiol. 7:308–13

38. Feil, E. J., Holmes, E. C., Bessen, D. E., Chan, M. S., Day, N. P., Enright, M. C., et al. (2001). Recombination within natural populations of pathogenic bacteria: short-term empirical estimates and long-term phylogenetic consequences. *Proc Natl Acad Sci U S A, 98*(1), 182-187.

39. Feil EJ, Li BC, AanensenDM,HanageWP,Spratt BG. 2004.eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. J. Bacteriol. 186:1518–30

40. Franzmann, P.D., Wehmeyer, U. & Stackerbrandt, E. (1988). Halomonadaceae fam. nov., a new family of the class Proteobacteria to accommodate the genera Halomonas and Deleya. Syst Appl Microbiol 11, 16–19.

41. Fraser, C., Alm, E. J., Polz, M. F., Spratt, B. G., & Hanage, W. P. (2009). The Bacterial Species Challenge: Making Sense of Genetic and Ecological Diversity. *Science, 323*(5915), 741-746.

42. Fredrickson JK, et al. (2008) Towards environmental systems biology of Shewanella. Nat Rev Microbiol 6:592–603.

43. Friedberg., G. C. W., W Siede 1995. DNA repair and mutagenesis. ASM Press Washington D C.

44. Gill, S.R., et al., 2005 Insights on evolution of virulence and resistance from the complete genome analysis of an early methicillin-resistant *Staphylococcus aureus*strain and a biofilm-producing methicillin-resistant *Staphylococcus epidermidis* strain. J Bacteriol, 187 (7), 2426-38

45. Giovannoni SJ, Tripp HJ, Givan S, Podar M, Vergin KL, Baptista D et al. (2005). Genome streamlining in a cosmopolitan oceanic bacterium. Science 309: 1242–1245.

46, Guindon S., Gascuel O. PhyML :"A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood." Systematic Biology. 2003 52(5):696-704.

47. Harder, W., and Schleifer, K.-H. (Eds.), The Prokaryotes. A handbook on the biology of bacteria: ecophysiology, isolation, identification, applications. 2nd ed., Vol. IV. Springer-Verlag, New York.

48. Harm. 1976. Photochemistry and Photobiology of nucleic acids, vol. Academic press, new york.

49. Harrison, F.C., and Kennedy, M.E. 1922. The red discoloration of cured codfish. Trans. Roy. Soc. Canad. Sct.III 16: 101-152.

50. Hau HH, Gralnick JA (2007) Ecology and biotechnology of the genus Shewanella. Annu Rev Microbiol 61:237–258.

51. Heath, L., van der Walt, E., Varsani, A. & Martin D.P. (2006). Recombination patterns in aphthoviruses mirror those found in other picornaviruses. J Virol 80, 11827-11832.

52. Hoffmaster AR, Ravel J, Rasko DA, Chapman GD, Chute MD, et al. 2004. Identification of anthrax toxin genes in a Bacillus cereus associated with an illness resembling inhalation anthrax. Proc. Natl. Acad. Sci. USA 101:8449–54

53. Hudson DH. 1998. SplitsTree: analyzing and visualizing evolutionary data. Bioinformatics 14:68–73

54. Imhoff, J.F. 2001. True marine and halophilic anoxygenic phototrophic bacteria. Arch. Microbiol. 176: 243-254.

55. Jolley KA, Feil EJ, Chan MS, Maiden MCJ. 2001. Sequence type analysis and recombinational tests (START). Bioinformatics 17:1230–31

56. Jolley, K. A. (2009). Internet-based sequence-typing databases for bacterial molecular epidemiology. *Methods Mol Biol, 551,* 305-312.

57. Jun, S. H., T. G. Kim, and C. Ban. 2006. DNA mismatch repair system. Classical and fresh roles. The FEBS journal 273:1609-19.

58. Kim, J. I., A. K. Sharma, S. N. Abbott, E. A. Wood, D. W. Dwyer, A. Jambura, K. W. Minton, R. B. Inman, M. J. Daly, and M. M. Cox. 2002. RecA Protein from the extremely radioresistant bacterium Deinococcus radiodurans: expression, purification, and characterization. Journal of bacteriology 184:1649-60.

59. Klebahn,H. 1919. Die Schädlinge des Klippfisches. Mitt. Inst. Allg. Bot. Hamburg 4: 11-69.

60. Kobayashi, I., and H. Ikeda. 1978. On the role of recA gene product in genetic recombination: an analysis by in vitro packaging of recombinant DNA molecules formed in the absence of protein synthesis. Molecular & general genetics 166:25-9

61. Kottemann, M., A. Kish, C. Iloanusi, S. Bjork, and J. DiRuggiero. 2005. Physiological responses of the halophilic archaeon Halobacterium sp. strain NRC1 to desiccation and gamma irradiation. Extremophiles 9:219-27.

62. Kumar, S., Nei, M., Dudley, J., & Tamura, K. (2008). MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. *Brief Bioinform, 9*(4), 299-306.

63. Lanyi JK (2004) Bacteriorhodopsin. Annual Review of Physiology 66:665–688.

64. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. (2007). Clustal W and Clustal X version 2.0. Bioinformatics, 23, 2947-2948.

65. Lawrence JG, Ochman H. 1998. Molecular archaeology of the Escherichia coli genome. Proc. Natl. Acad. Sci. USA 95:9413–17

66. Lehmann, A. R. 1995. Nucleotide excision repair and the link with transcription. Trends in biochemical sciences 20:402-5.

67. Levin B R. In: The Evolution of Sex. Michod R E, Levin B R, editors. Sunderland, MA: Sinauer; 1988. pp. 194–211.

68. Librado, P. and Rozas, J. 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25: 1451-1452

69. Maiden, M. C. (2006). Multilocus sequence typing of bacteria. *Annu Rev Microbiol, 60*, 561-588.

70. Maiden, M. C., Bygraves, J. A., Feil, E., Morelli, G., Russell, J. E., Urwin, R., et al. (1998). Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci U S A, 95*(6), 3140-3145.

71. McVean G., Awadalla P., Fearnhead P., Awadalla P., Fearnhead P., Fearnhead P. A (2002) coalescent-based method for detecting and estimating recombination from gene sequences. Genetics.;160:1231–1241

72. Maynard Smith J, Smith N H, O'Rourke M, Spratt B G.(1993) How clonal are bacteria? Proc Natl Acad Sci USA.;90:4384–4388

73. Maynard Smith, J. and Smith, N. H. (1998) Detecting recombination from gene trees. Mol. Biol. Evol. 15,590–599

74. Miller, R. V. (1998). Bacterial gene swapping in nature. *Sci Am, 278*(1), 66-71.

75. Miller, R. V., W. Jeffrey, D. Mitchell, and M. Elsari. 1999. Bacterial response to solar ultraviolet light. ASM News 65:535-541.

76. Miller, R. V (2000) *recA*: The gene and its protein product. In: Luria, S (Ed.) Encyclopedia of Microbiology, 2nd ed., Academic Press, San Diego, pp 43-54

77. Morikawa, K., M. Ariyoshi, D. G. Vassylyev, O. Matsumoto, K. Katayanagi, and E. Ohtsuka. 1995. Crystal structure of a pyrimidine dimer-specific excision repair enzyme from bacteriophage T4: refinement at 1.45 A and X-ray analysis of the three active site mutants. Journal of molecular biology 249:360-75.

78. Nassar, Majed S.,2008, Isolation and characterization of Halomonas spp. bacteriophages from the *Great Salt* Plain Wildlife Reserve, Cherokee, OK, ProQuest Dissertations and Theses, 94 pages; AAT 3345079

79. Okamoto, T., Taguchi, H., Nakamura, K., Ikenaga, H., Kuraishi, H. & Yamasato, K. (1993). Zymobacter palmae gen. nov., sp. nov., a new ethanol-fermenting peritrichous bacterium isolated from palm sap. Arch Microbiol 160, 333–337

80. Oren A (2002) "Diversity of halophilic microorganisms: environments, phylogeny, physiology, and applications," Journal of Industrial Microbiology and Biotechnology, vol. 28, no. 1, pp. 56–63.

81. Oren A (2008) Microbial life at high salt concentrations: phylogeneticand metabolic diversity. Saline Systems 4:2

82. Oren, A., Larimer, F., Richardson, P., Lapidus, A., & Csonka, L. N. (2005). How to be moderately halophilic with broad salt tolerance: clues from the genome of Chromohalobacter salexigens. *Extremophiles, 9*(4), 275-279.

83. Papke R. T.,Koenig J. E., Rodríguez-Valera F., Doolittle W. F.2004 Frequent recombination in a saltern population of Halorubrum. Science 306, 1928–1929.

84. Patel, B.K.C., Andrews, K.T., Ollivier, B., Mah, R.A., and Garcia, J.-L. 1995. Reevaluating the classification of Halobacteroides and Haloanaerobacter species based on sequence comparisons of the 16S ribosomal RNA gene. FEMS Microbiol. Lett. 134: 115-119.

85. Pavon A B., and Maiden C J., Multi Locus Sequence Typing. Methods in Molecular Biology. (551), 129-140

86. Peak, M. J., J. G. Peak, M. P. Moehring, and R. B. Webb. 1984. Ultraviolet action spectra for DNA dimer induction, lethality, and mutagenesis in Escherichia coli with emphasis on the UVB region. Photochemistry and photobiology 40:613-20.

87. Pearson, B M., C. Pin, J wright, K. I'Anson, T. Humphrey, and J.M.Wells, 2003 Comparitive genome analysis of *Compylobacter jejuni* using whole genome DNA microarrays. FEBS Lett, 554 (1-2), 224-30

88. Pierce, G.J. 1914. The behavior of certain micro-organisms in brine. Carnegie Institution of Washington Publication no. 193: 49-69.

89. Polz, M F., D.E. Hunt, S.P. Preheim, and D.M. Weinreich, 2006 Patterns and mechanisms of genetic and phenotypic differentiation in marine microbes. Philosophical Transactions of the Royal Society B: Biological Sciences, 361 (1475), 2009-2021

90. Popoff MY, Bockemuhl J, Brenner FW. 1998. Supplement 1997 (no. 41) to the Kauffmann-White scheme. Res. Microbiol. 149:601–4

91. Posada D and Crandall KA 1998. Modeltest: testing the model of DNA substitution. Bioinformatics 14 (9): 817-818.

92. Posada D, Crandall KA, Holmes EC. (2002) Recombination in Evolutionary Genomics. Annu. Rev. Genet, 36: 75-97

93. Priest FG, Barker M, Baillie LWJ, Holmes EC, Maiden MCJ. 2004. Population structure and evolution of the Bacillus cereus group. J. Bacteriol. 186:7959–70

94. Prosser, J.I., et al., 2007 The role of ecological theory in microbial ecology.Nat Rev Micro. 5 (5): 384-392

95. Martin E.L., R.L. Reinhardt, L.L. Baum, M.R. Becker, J.J. Shaffer, and T.A. Kokjohn (2000) The effects of ultraviolet radiation on the moderate halophile *Halomonas elongata* and the extreme halophile *Halobacterium salinarum.* Can. J. Microbiol. **46**(2): 180–187

96. Pinchuk GE, Hill EA, Geydebrekht OV, De Ingeniis J, Zhang X, et al. (2010) Constraint-Based Model of Shewanella oneidensis MR-1 Metabolism: A Tool for Data Analysis and Hypothesis Generation. PLoS Comput Biol 6(6): e1000822. doi:10.1371/journal.pcbi.1000822

97. Quesada, E., V. Bejar, M. J. Valderrama, A. Ventosa, and A. R. Ramos Cormenzana. 1985. Isolation and characterization of moderately halophilic nonmotile rods from different saline habitats. Microbiologia (Madrid, Spain) 1:89-96.

98. Sancar, A. 1996. No "End of History" for photolyases.Science (New York, N.Y272:48-9.

99. Sánchez-Porro, C., de la Haba, R. R., Soto-Ramírez, N., Márquez, M. C.,Montalvo-Rodríguez, R. & Ventosa, A. (2009). Description of Kushneria aurantia gen. nov., sp. nov., a novel member of the family Halomonadaceae, and a proposal for reclassification of Halomonas marisflavi as Kushneria marisflavi comb. nov., of Halomonas indalinina as Kushneria indalinina comb. nov. and of Halomonas avicenniae as Kushneria avicenniae comb. nov. Int J Syst Evol Microbiol 59, 397-405.

100. Sawyer, S. (1989) Statistical tests for detecting gene conversion. Mol. Biol. Evol. 6, 526–538.

101. Schlesinger, D. J. 2007. Role of RecA in DNA damage repair in Deinococcus radiodurans. FEMS microbiology letters 274:342-7.

102. Seaman, P. F., and M. J. Day. 2007. Isolation and characterization of a bacteriophage with an unusually large genome from the Great Salt Plains National Wildlife Refuge, Oklahoma, USA. FEMS Microbiol. Ecol. 60:1-13

103. Seeberg, E., L. Eide, and M. Bjoras. 1995. The base excision repair pathway. Trends in biochemical sciences 20:391-7.

104. Selander RK, Levin BR. 1980. Genetic diversity and structure in Escherichia coli populations. Science 210:545–47

105. Shinohara, A., and T. Ogawa. 1995. Homologous recombination and the roles of double-strand breaks. Trends in biochemical sciences 20:387-91.

106. Simonson, CS, Kokjohn, TA, Miller, RV (1990) Inducible UV repair potential of Pseudomonas aeruginosa PAO. J Gen Microbiol 136: 1241–1249

107. Sinha, R. P., and D. P. Hader. 2002. UV-induced DNA damage and repair: a review. Photochemical & photobiological sciences 1:225-36.

108. Smith, J.M., N.H. Smith, M.O'Rourke, and B.G.Spratt, 1993 How clonal are bacteria? Proc Natl Acad Sci USA, 90 (10), 4384-8

109. Smith, W.W., and ZoBell, C.E. 1937a. An autochthonous bacterial flora in Great Salt Lake. J. Bacteriol. 33:118.

110. Spratt, B. G. 1999. Multilocus sequence typing: molecular typing of bacterial pathogens in an era of rapid DNA sequencing and the internet. Curr Opin Microbiol 2:312-6.

111. Spratt BG, Bowler LD, Zhang QY, Zhou J, Smith JM. 1992. Role of interspecies transfer of chromosomal genes in the evolution of penicillin resistance in pathogenic and commensal Neisseria species. J. Mol. Evol. 34:115–25

112. Stackebrandt, E., Frederiksen, W., Garrity, G. M., Grimont, P. A., Kämpfer, P.,Maiden, M. C., Nesme, X., Rosselló-Mora, R., Swings, J., Trüper, H. G.,Vauterin, L., Ward, A. C. & Whitman, W. B. (2002). Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. Int J Syst Evol Microbiol 52, 1043–1047.

113. Swofford, D. L. 2002. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland,Massachusetts.

114. Talavera, G., and Castresana, J. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Systematic Biology 56, 564-577.

115. Tanabe Y, Kaya K, Watanabe MM. Evidence for recombination in the microcystin synthetase (mcy) genes of toxic cyanobacteria Microcystis spp. J Mol Evol 2004;58:633-641

116. Tatum, E.L. and J. Lederberg, 1947 Gene Recombination in the Bacterium *Escherichia coli*. The Journal of Bacteriology, 53 (6), 673-684.

117. Tettelin, H., et al., 2005 Genome analysis of multiple pathogenic isolates of *Streptococcusagalactiae* implications for the microbial "pan-genome". Proc Natl Acad Sci USA, 102 (39), 13950-5

118. Todo, T., H. Ryo, K. Yamamoto, H. Toh, T. Inui, H. Ayaki, T. Nomura, and M. Ikenaga. 1996. Similarity among the Drosophila (6-4)photolyase, a human photolyase homolog, and the DNA photolyase-blue-light photoreceptor family. Science (New York, N.Y 272:109-12.

119. Urwin, R., and M. C. Maiden. 2003. Multi-locus sequence typing: a tool for global epidemiology. Trends Microbiol 11:479-87.

120. Ventosa, A., Nieto, J. J., & Oren, A. (1998). Biology of moderately halophilic aerobic bacteria. *Microbiol Mol Biol Rev, 62*(2), 504-544.

121. Vos, M., & Didelot, X. (2008). A comparison of homologous recombination rates in bacteria and archaea. *ISME J, 3*(2), 199-208.

122. Vreeland, R.H. 1992. The family Halomonadaceae, pp. 3181-3188 In: Balows, A., Trüper, H.G., Dworkin, M.,

123. Wertz, J. E., Goldstone, C., Gordon, D. M. & Riley, M. A. A molecular phylogeny of enteric bacteria and implications for a bacterial species concept. J. Evol. Biol. 16, 1236–1248 (2003).

124. Wilkansky, B. 1936. Life in the Dead Sea. Nature 138: 467.

125. Wilson, C., Caton, T. M., Buchheim, J. A., Buchheim, M. A., Schneegurt, M. A., & Miller, R. V. (2004). DNA-repair potential of Halomonas spp. from the Salt Plains Microbial Observatory of Oklahoma. *Microb Ecol*, *48*(4), 541-549.

VITA

Deepak G Rudrappa

Candidate for the Degree of

Doctor of Philosophy

Thesis: POPULATION STRUCTURE AND DIVERSITY OF *HALOMONAS* SPP. FROM THE GREAT SALT PLAINS REVEALED BY MULTI LOCUS SEQUENCE TYPING

Major Field: Microbiology and Molecular Genetics

Biographical:

Education: Completed the requirements for the Doctor of Philosophy in Microbiology and Molecular Genetics at Oklahoma State University, Stillwater, Oklahoma in December, 2010. Received the Master of Science in Agricultural Microbiology from University of Agricultural Sciences, Bangalore, India in 2000. Received the Bachelor of Science in Agriculture from University of Agricultural Sciences, Dharwad, India in 1998.

Experience: Employed by Rallies Research center as Research Scientist from 2000 to 2004. Employed by Oklahoma State University, Department of Microbiology and Molecular Genetics, as a graduate teaching/research assistant, during fall 2004 to summer 2010.

Professional Memberships: American Society of Microbiology (ASM), Missouri Valley ASM branch, Oklahoma Academy of Sciences, Microbiology and Molecular Genetics Graduate Student Association and Graduate and Professional Student Government Association, Oklahoma State University.

Name: Deepak G Rudrappa Date of Degree: December, 2010

Institution: Oklahoma State University Location: Stillwater, Oklahoma

Title of Study: POPULATION STRUCTURE AND DIVERSITY OF *HALOMONAS* SPP. FROM THE GREAT SALT PLAINS REVEALED BY MULTI LOCUS SEQUENCE TYPING

Pages in Study: 115 Candidate for the Degree of Doctor of Philosophy

Major Field: Microbiology and Molecular Genetics

Scope and Method of Study: The Great Salt Plains (GSP) is an ecologically diverse, hyper-saline environment. Many differences in overall phenotypic characters were observed among *Halomonas* spp. isolates from the GSP that had >97% 16S rRNA sequence similarity. We believe these *Halomonas* spp. isolates have evolved over a long period of time by means of developing certain genetic mechanisms. Multi Locus Sequence Typing (MLST) was used to study the population structure of these isolates. Housekeeping genes *ectB, gyrB, and recA* along with 16S rRNA and the ITS region were assayed in 32 *Halomonas* spp. isolates.

Findings and Conclusions: MLST analysis identified 28 unique sequence types indicating a high level of genetic diversity. Among the house keeping genes, *recA* demonstrated increased resolution indicating that *Halomonas* spp. isolates contain several genetically distinct phylotypes most likely due to environmental bottlenecks and recombination events. We found evidence for frequent LGT and recombination among these isolates. Recombination rates were on par with other sexual bacteria. The rates of recombination observed are high enough to overcome the effects of population bottlenecks and periodic selection that would otherwise purge the genetic variation generated. These results support the hypothesis that a high level of speciation has occurred among *Halomonas* spp. isolates at the GSP site even though 16S rRNA sequence similarity has been conserved. Phylogenetic relationship of very closely related *Halomonas* spp. isolates was resolved by MLST. The results of this study have opened new roads for the research topics that range from the molecular evolution of environmental isolates to the significance of recombination in evolution of bacteria.

ADVISER'S APPROVAL: Dr. Robert V Miller