

**EVOLUTION AND ECOLOGICAL SPECIALIZATION OF  
A *SHEWANELLA BALTICA* POPULATION**

**By**

**Jie Deng**

**A DISSERTATION**

**Submitted to  
Michigan State University  
in partial fulfillment of the requirements  
for the degree of**

**DOCTOR OF PHILSOPHY**

**Biochemistry and Molecular Biology**

**2011**

## ABSTRACT

### EVOLUTION AND ECOLOGICAL SPECIALIZATION OF A *SHEWANELLA BALTICA* POPULATION

By

Jie Deng

Studying how bacterial strains diverge over time and how divergence leads to specialization to new environmental niches is important for understanding the dynamics of environmental communities. I studied these questions using a collection of 46 *Shewanella baltica* strains isolated over a period of 12 years from the Gotland Deep in the central Baltic Sea, which is characterized by the presence of a stable redox gradient over 140 m in depth. I used several experimental and computational methods to explore the genetic and phenotypic profiles of these strains and determined whether these profiles could be explained by the conditions at their sites of isolation. Specifically, genotyping by both single gene-based and multi-locus sequence typing (MLST) indicated specialization across the redox gradient and over time. Versatility in anaerobic respiration, for example, the ability to respire thiosulfate, was correlated with the redox gradient, while versatility in carbon source utilization was correlated with specialization over time. Comparative Genomic Hybridization (CGH) revealed a heterogeneous distribution of genes across *S. baltica* genomes, where genes potentially important for specialization both over the redox gradient and over time were identified. Further mechanistic investigation of their evolution provided evidence of horizontal gene transfer in the *S. baltica* genomes, whereas predicted recent recombination among strains from two lineages was also supported by this larger dataset. The differential fitness of *S. baltica* strains under particular redox conditions was determined from competition assays, where

competed strains were labeled with *gfp* and *cfp* and the population dynamics during competition was measured by flow cytometry. The two competed genotypes were separated by flow cytometry and their transcriptomes analyzed by RNA-Seq which showed differential expression in response to the nitrate condition as well as to the competitor strain. By integrating these different types of information, I determined that specialization has occurred in this *S. baltica* population both over the redox gradient and over time, and that the genomic plasticity as well as the extensive gene exchange has facilitated the specialization process. Overall, I used *S. baltica* as a model to provide insights into fundamental questions about the interactions among bacterial genomics, physiology, and habitat (ecology).

## ACKNOWLEDGEMENTS

This thesis would not have been possible without the kind support and help from many individuals who, in one way or another, contributed to the preparation and completion of this study.

First and foremost, I owe my sincerest gratitude to my supervisor, Dr. James Tiedje, who supported and mentored me throughout my thesis with his knowledge and patience, while at the same time allowed me the room to work in my own way. His encouraging and optimistic attitude has helped me through the tough and frustrating stages of my work, and has greatly influenced my scientific and personal growth.

I am deeply grateful to a number of colleagues involved with this project. Ingrid Brettar and Manfred Höfle supplied this precious collection of *Shewanella baltica* isolates. Kostas Konstantinidis constantly provided valuable insights and continues to contribute to this project with his specialty in bioinformatics. I also share the credit of my work with Jennifer Auchtung, who guided me for nearly two years and helped me grow to be an independent thinker and researcher.

In my daily work, I have been blessed with a friendly and cheerful group of lab mates, whose critical discussions, support, and encouragement throughout the years have helped propel me forward.

I would also thank the Department of Biochemistry and Molecular Biology, for offering me the opportunity of studying in this university, as well as my committee, for providing valuable advice during my Ph.D.

Lastly, I am indebted to my dearest parents for their love and constant support, and for their understanding of me pursuing advanced degrees abroad. Without the combined input and efforts from all of these friends and colleagues, I am sure this journey would not have been this fantastic - thank you all.

## TABLE OF CONTENTS

LIST OF TABLES .....	x
LIST OF FIGURES .....	xi
LIST OF ABBREVIATIONS .....	xiv
<b>CHAPTER I</b>	
<b>INTRODUCTION</b>	
A perspective on bacterial specialization .....	2
The versatile <i>Shewanella</i> .....	3
The history behind the <i>Shewanella baltica</i> population in the Baltic Sea .....	5
Thesis Outline .....	7
References.....	10
<b>CHAPTER II</b>	
<b>PHYLOGENETIC ANALYSIS OF A <i>SHEWANELLA BALTICA</i> POPULATION USING MULTI LOCUS SEQUENCE TYPING</b>	
<b>Abstract</b> .....	<b>15</b>
<b>Introduction</b> .....	<b>16</b>
<b>Materials and Methods</b> .....	<b>20</b>
Bacterial strains and growth conditions .....	20
MLST Analysis .....	23
<b>Results</b> .....	<b>25</b>
<i>gyrB</i> analysis .....	25
MLST analysis .....	27
Recombination analysis on genes selected for genotyping .....	31
<b>Discussion</b> .....	<b>32</b>
Genotyping indicates niche specialization and genetic continuity over time.....	32
Recombination was detected from sequences of the marker genes .....	33
<b>References</b> .....	<b>35</b>
<b>CHAPTER III</b>	
<b>COMPARATIVE GENOMIC STUDY OF THE <i>SHEWANELLA BALTICA</i> GENOMES USING MICROARRAYS</b>	
<b>Abstract</b> .....	<b>40</b>
<b>Introduction</b> .....	<b>41</b>

<b>Materials and Methods</b> .....	<b>44</b>
Array Design .....	44
DNA microarray hybridization .....	45
Gene expression studies .....	46
Correlation analysis on <i>gyrB</i> , MLST and CGH based distance matrices .....	47
<b>Results and Discussion</b> .....	<b>49</b>
Evaluation of probe binding efficiency and specificity .....	49
Array normalization and setting cutoff value for determining gene presence /absence .....	50
General statistics of CGH .....	51
GC bias in the core and variable genome .....	55
Less conservation of intergenic regions .....	57
Functional bias in the core genome and variable genome .....	59
Similar clustering patterns with MLST phylogeny .....	63
Genes related to specialization over time .....	65
Genes related to specialization across the redox gradient .....	69
Conserved gene core in specific lineages .....	71
Genes specific to strains in Clade A and D .....	72
Genes specific to Clade E .....	74
Genes specific to Clade A .....	74
Genes specific to Clade D .....	75
Core genes in Clade B strains .....	75
High versatility in carbon source utilization was reflected from the OS223 genome .....	76
Horizontal gene transfer and genomic islands .....	77
Recombination among specific lineages was supported .....	79
<b>References</b> .....	<b>83</b>

## CHAPTER IV

### CHARACTERIZATION OF CARBON SOURCE UTILIZATION AND RESPIRATORY CAPABILITIES OF THE *SHEWANELLA BALTICA* STRAINS

<b>Abstract</b> .....	<b>88</b>
<b>Introduction</b> .....	<b>89</b>
<b>Materials and Methods</b> .....	<b>92</b>
Bacterial strains and growth conditions .....	92
Regression analysis on Biolog datasets .....	92
<b>Results</b> .....	<b>94</b>
Examining diversity of carbon source utilization using the Biolog system .....	94

Examining diversity in utilization of different electron acceptors .....	97
<b>Discussion .....</b>	<b>99</b>
Metabolic similarity in carbon source utilization is correlated with evolutionary relatedness .....	99
Temporal distribution has a significant impact on evolution of metabolic capabilities by <i>S. baltica</i> strains .....	102
Respiratory versatility within the <i>S. baltica</i> species is correlated with their genotypic relatedness .....	103
Reduction of nitrate likely via ammonification is a common phenotypic characteristic among all <i>S. baltica</i> strains .....	104
DMSO reduction may be related with specialization to anoxic deep water environment .....	104
Inability to reduce thiosulfate is limited within Clade E .....	105
Inability to reduce TMAO is also limited within Clade E .....	106
<b>References .....</b>	<b>107</b>

## CHAPTER V

### COMPETITION ASSAYS AMONG SELECTED *SHEWANELLA BALTICA* STRAINS WHILE EXPOSED TO DIFFERENT ELECTRON ACCEPTORS

<b>Abstract .....</b>	<b>111</b>
<b>Introduction .....</b>	<b>112</b>
<b>Materials and Methods .....</b>	<b>115</b>
Bacterial strains and growth conditions .....	115
GFP and CFP tagging of <i>Shewanella baltica</i> strains .....	115
Competition assays .....	116
Experimental design for transcriptomic characterization .....	116
RNA extraction and amplification .....	117
Mapping, normalization, and statistical analysis of RNA-Seq data .....	117
<b>Results and Discussion .....</b>	<b>119</b>
Competition under different electron acceptors .....	119
Transcriptional characterization of <i>S. baltica</i> OS185 and OS195 during competition with limited nitrate as electron acceptor .....	121
RNA processing and Illumina RNA-Seq summary .....	123
Overview of transcriptomic profiles .....	126
Differential expression caused by culturing methods (pure culture vs. co-culture) in each strain .....	130
Differential expression of OS195- <i>cfp</i> genes in pure culture and in co-culture (Pc::Cc) .....	130

Differential expression of OS185- <i>gfp</i> genes in pure culture and in co-culture (Pg::Cg) .....	131
Comparison on expression levels of shared gene between OS185 and OS195 in pure culture and in co-culture (Pg::Pc and Cg::Cc) .....	133
Differentially expressed genes specifically in pure culture (Pc::Pg) .....	137
Differentially expressed genes in both pure culture and co-culture (Pc::Pg, Cc::Cg) .....	138
Co-culture specific differentially expressed genes (Cc::Cg).....	141
RNA-Seq technology in transcriptomic analysis .....	142
Mechanisms of differential fitness .....	144
<b>References .....</b>	<b>146</b>

## CHAPTER VI

### THESIS SUMMARY AND OUTLOOK

<b>References .....</b>	<b>159</b>
-------------------------	------------

## LIST OF TABLES

Table 2.1. <i>Shewanella baltica</i> Strains Used in this Study .....	21
Table 2.2. Marker Genes Used for MLST Analysis .....	23
Table 2.3. Primers Used for MLST Analysis .....	24
Table 2.4. Neutrality Test and Estimation of Recombination ( $\rho$ ) and Mutation Rates ( $\theta$ ) .....	28
Table 3.1. Number of Probes Designed from Each Reference Strain .....	45
Table 3.2. Composition of Variable Genes .....	52
Table 3.3. Comparison of GC Content in Core Genome and Variable Genome .....	55
Table 3.4. Description of COG Categories .....	62
Table 3.5. Correlation among Three Distance Matrices for Phylogenetic Inferences .....	63
Table 3.6. Types of Differential Expression .....	68
Table 4.1. Use of Different Electron Acceptors by <i>S. baltica</i> Strains .....	97
Table 4.2. Two Sample T Test Shows that Strains from 1998 Use Significantly More Carbon Sources .....	102
Table 5.1. Strain Description and Growth Conditions .....	115
Table 5.2. Sample Description .....	123
Table 5.3. Total RNA Yield and Amplification Efficiency .....	125
Table 5.4. Summary of Mapped RNA-Seq Reads .....	125
Table 5.5. Number of Genes Classified in Different Cellular Functions .....	136
Table 5.6. Number of Genes Classified in Different COG Categories .....	136

## LIST OF FIGURES

Figure 1.1. Water chemistry profile in the Gotland Deep from summer of 1986 .....	7
Figure 2.1. Neighbor Joining tree of 46 <i>S. baltica</i> strains produced from <i>gyrB</i> sequences .....	26
Figure 2.2. (A) Neighbor Joining tree of 46 <i>S. baltica</i> strains from concatenated sequences of seven genes used for MLST analysis. (B) Phylogenetic network of the MLST loci among the <i>S. baltica</i> strains .....	29
Figure 3.1. Spheres of the diagram represent the gene content of the four sequenced <i>S. baltica</i> strains (OS155, OS185, OS195, and OS223) .....	45
Figure 3.2a. Box-Plot of probe signals from CGH profiles of the four reference strains against percent identity of the probes matched to their genomes .....	50
Figure 3.2b. Sum of false positive and false negative rates under different cut-off values .....	51
Figure 3.3. Heatmap of CGH signals from all genes and intergenic regions of the 46 <i>S. baltica</i> strains .....	53
Figure 3.4. The sizes of core genome calculated from the CGH dataset .....	54
Figure 3.5. GC contents of genes that are represented in the microarray .....	56
Figure 3.6. Histograms illustrating frequency of presence among 46 <i>S. baltica</i> strains of (a) all genes and (b) intergenic regions represented in the microarray .....	58
Figure 3.7. Distribution of genes in different COG functional categories in (a) pangenome (b) genes whose presence varies among the 46 <i>S. baltica</i> strains (c) genes that are present in greater than 90% of the strains .....	60
Figure 3.8. Hierarchical clustering of the CGH profile .....	64
Figure 3.9. Comparison of conserved core genes of (a) strains isolated from the 1980s and 1998 (b) strains recovered from at or below the anoxic-oxic interface and from more oxic water sample .....	66
Figure 3.10. (a) COG functional categories of annotated core genes of strains isolated from the 1980s. (b) Differential expression patterns of core genes of strains from the 1980s in OS185 and OS195 .....	67

Figure 3.11. (a) COG functional categories of annotated core genes of strains isolated from anoxic waters. (b) Differential expression patterns of core genes of strains from anoxic waters in OS185 and OS195 .....	70
Figure 3.12. Core genome comparison of Clade A, D, E and J .....	72
Figure 3.13. Number of strain-specific genes revealed from CGH .....	77
Figure 3.14. The patterns of genetic exchange apply to a larger collection of <i>S. baltica</i> strains .....	81
Figure 4.1. Heatmap of 46 <i>S. baltica</i> strains based on Biolog profiles .....	95
Figure 4.2. Principle Coordinate Analysis (PCoA) of Biolog metabolic profiles of the <i>S. baltica</i> strains .....	96
Figure 4.3. Regression of Biolog similarity and genomic relatedness .....	100
Figure 4.4. Hierarchical clustering of the Biolog profile .....	101
Figure 5.1. Competition patterns observed among five <i>S. baltica</i> isolates in minimal medium under aerobic and anaerobic condition with nitrate as the sole electron acceptor .....	120
Figure 5.2. Growth curves of fluorescence tagged <i>S. baltica</i> OS185 and OS195 strains .....	122
Figure 5.3. Competition pattern between fluorescence tagged <i>S. baltica</i> OS195 and OS185 in minimal medium supplemented with nitrate as sole electron acceptor .....	122
Figure 5.4. BioAnalyzer profile of total RNA extracted from sample Pg-2 (a) before amplification with RNA Integrity Number of 9.3 and (b) after amplification .....	124
Figure 5.5. Hierarchical clustering of normalized mRNA reads from the 12 samples .....	126
Figure 5.6. Canonical Correspondence Analysis of normalized mRNA reads from the 12 samples .....	127
Figure 5.7. Summary of pairwise comparisons performed with RNA-seq profiles of OS185 and OS195 under nitrate condition .....	128
Figure 5.8. Testing for differentially expressed genes in the four pairwise comparisons. Scatter plots of log <sub>2</sub> fold changes versus average numbers of reads were drawn for (a) Cc::Cg (b) Pc::Pg (c) Cg::Pg (d) Cc::Pc .....	129
Figure 5.9. Insertion of phage genes in OS195 chromosome .....	131

Figure 5.10. COG categories of differentially expressed genes by OS185 in co-culture vs. pure culture (Cc::Pc) .....	132
Figure 5.11. Heatmap of differentially expressed genes between OS185 and OS195 in co-culture (Cc::Cg) and in pure culture (Pc::Pg) .....	134
Figure 5.12. Summary of numbers of differentially expressed gene shared by both OS185 and OS195 between the two strains in co-culture and in pure culture .....	135
Figure 5.13. Differentially expressed genes specific to Pc::Pg comparison .....	137
Figure 5.14. Differentially expressed genes from both Pg::Pc and Cg::Cc comparisons .....	140
Figure 5.15. COG categories of differentially expressed genes shared by OS185 and OS195 from Pg::Pc comparison .....	140
Figure 5.16. COG categories of differentially expressed genes shared by OS185 and OS195 from Cg::Cc comparison .....	142
Figure 6.1. COG functional comparison of <i>S. baltica</i> gene core and <i>Burkholderia cenocepacia</i> gene core .....	156

## LIST OF ABBREVIATIONS

**RAPD: Randomly Amplified Polymorphic DNA**

**MLST: Multi Locus Sequence Typing**

**HGT: Horizontal Gene Transfer**

**MLEE: Multi-Locus Enzyme Electrophoresis**

**EM: Electrophoretic Mobilities**

**RFLP: Restriction Fragment Length Polymorphism**

**AFLP: Amplified Fragment Length Polymorphism**

**ANI: Average Nucleotide Identity**

**PCR: Polymerase Chain Reactions**

**RDP: Recombination Detection Program**

**GARD: Genetic Algorithm for Recombination Detection**

**ST: Sequence Type**

**LMW: Low-Molecular-Weight**

**CGH: Comparative Genomic Hybridization**

**Oligoarrays: Oligonucleotide Arrays**

**COG: Clusters of Orthologous Groups**

**LEE: Locus *pf* Enterocyte Effacement**

**EPEC: Enteropathogenic *Escherichia Coli***

**DMSO: Dimethyl Sulfoxide**

**TMAO: Trimethylamine-N-Oxide**

**PM: Phenotype Microarray**

**PCoA: Principle Coordinate Analysis**

**DNRA: Dissimilatory Nitrate Reduction to Ammonium**

**DMS: Dimethylsulfide**

**TMA: Trimethylamine**

**DSN: Duplex-Specific Nuclease**

**PF: Passed Filtering**

**NGS: Next Generation Sequencing**

**ncRNA: Non-Coding RNA**

# **CHAPTER I**

## **Introduction**

**A perspective on bacterial specialization.** Bacteria and Archaea are arguably the most environmentally versatile organisms on earth, and represent the greatest abundance as well as divergence in the tree of life. As the pioneer of life on this planet, originating approximately 3.7 billion years ago, prokaryotic life had evolved for over 3 billion years without eukaryotic competition, having thus occupied most if not all habitable niches on earth (1) (2). Microbial evolution and diversification (speciation) has been a major driver for new genotypic and phenotypic life on Earth. Yet the principles of these processes are not well understood. Since the principle of evolution by means of natural selection was first proposed by Darwin and later accepted into more branches of biology in the 20<sup>th</sup> century (3), numerous efforts have been focused on understanding the connections between natural environment and the biological outcome (4). With the integration of molecular biology and population genetics, the mechanisms underlying adaptation and speciation are also becoming clearer.

Through all the endeavors, a series of theories and concepts were incorporated into the evolutionary framework. While natural selection is still the only known driving force of adaptation, a number of non-adaptive causes of evolution were found including processes such as mutation, genetic drift, genetic hitchhiking (recombination), and horizontal gene transfer, which comprise the known engine for generation of genomic diversity (5) (6) (7) (8) (9) (10) (11). Under this framework, genotypic traits, reflected through phenotypes of organisms, are tested by and subjected to natural selection and random drift, which eventually leads to either fixation or extinction of the genotype.

Prior to the genome sequencing era, the determination of genetic elements attributable to phenotypic traits was a painstaking and time-consuming process. Microbial genetics has

greatly benefited from the automated sequencing technologies. Meanwhile, as more molecular and bioinformatic tools emerged, which greatly facilitated interpretation of the large amount of sequence data, high throughput comparative genomic studies became possible and started to provide valuable insights into what drove the genotypic and phenotypic diversification of life (4) (12).

Microbes are ideal candidates for investigating evolutionary trajectories as well as mechanisms. They occupy the vastest number of niches, providing materials for assessing adaptation in all environmental dimensions; they have relatively small genome sizes, which saves the cost for sequencing or genotyping as well as reduces the amount of analysis efforts; their shorter generation time also makes possible experimental evolution studies that allowed them to evolve tens of thousands of generations in a limited time (7).

Here we report an evolutionary study of a bacterial population *Shewanella baltica*, using isolates taken along an environmental gradient in an estuarine ecosystem, and sampled over nearly two decades. Through studying specialization among these strains, we hope to provide insights into how different the *S. baltica* genotypes are, how they are likely to position themselves in an ecological gradient, and how they interact with each other as a microbial community.

**The versatile *Shewanella*.** *Shewanella* genus belongs to gammaproteobacteria that are well known for their diverse respiratory capabilities. Strains in this genus are capable of utilizing a wide variety of electron acceptors including  $O_2$ ,  $NO_3^-$ ,  $S^0$ , Mn (III), Fe(III), Cr(VI), V(V), dimethylsulfoxid (DMSO), trimethylamine-N-oxide (TMAO), fumarate, and more (13). They are

among the most heavily studied environmental micro-organisms and have become a model system for studying metal bioremediation, biofuel generation, as well as microbial ecology (14). *Shewanella* species are known to inhabit a wide range of niches, including soil, marine, freshwater and sediments, and are often associated with energy-rich environments where anaerobic metabolism of organic matter is carried out by complex microbial communities (15). Many of these environments are subjected to highly variable concentrations of organic and inorganic electron acceptors and hence redox gradients can form, while the remarkable electron accepting ability of *Shewanellae* as well as their capability to sense and taxis towards electron acceptors allow them to be well adapted to inhabiting these organic-rich, redox fluctuating environments (16).

*Shewanella* started to gain much interest after the publication describing a strain *Shewanella oneidensis* MR-1, formerly *Alteromonas putrefaciens*, that was shown to be capable of reducing manganese and iron oxides (17). It was noted for its ability to transfer electrons directly to solid metal as well as its extraordinary respiratory versatility, being able to utilize over ten electron acceptors. The genome sequence of *S. oneidensis* MR-1 was released in 2002 (18), which facilitated establishment of MR-1 as a model organism for genetic and molecular biochemical studies of the *Shewanellae* genus as well as the metal-reducing bacteria in general for the purpose of bioremediation.

The *Shewanella* genus currently includes 56 recognized species (<http://www.bacterio.cict.fr/>), and some have been shown to occur in an even more diverse range of environments. The genomes of 30 strains of *Shewanella* have either been fully sequenced or are in the process of being sequenced or gap closure (according to NCBI Microbial

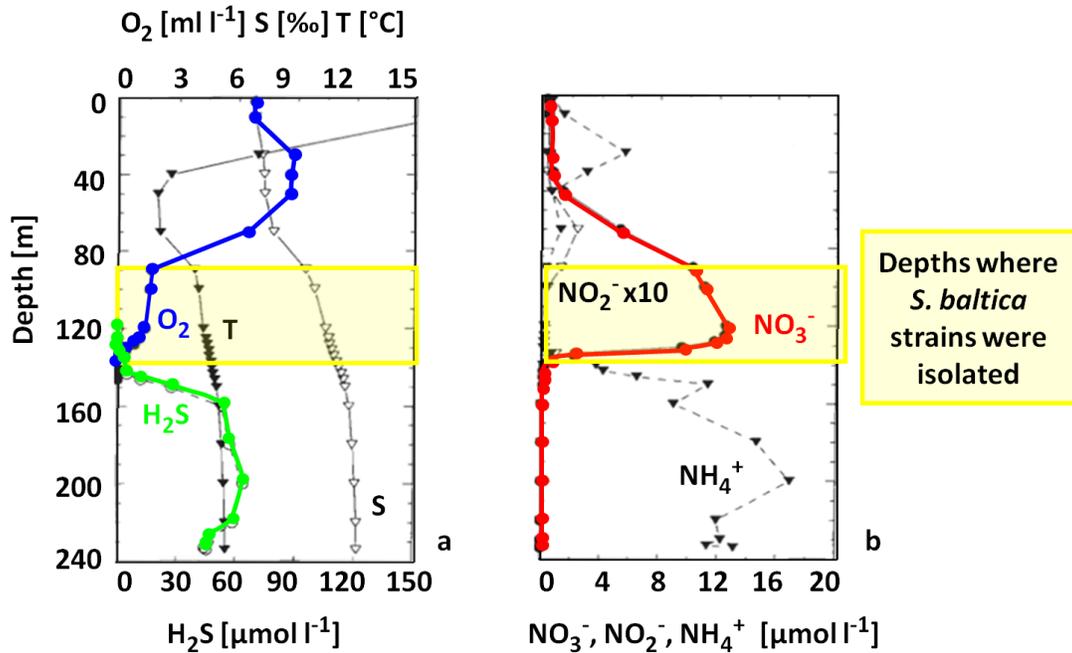
Genome database). Comparative genomic and proteomic analyses have been performed with a number of *Shewanella* strains, and provide valuable insights into genomic and phenotypic evolution among the *Shewanella* species (19). The correlation between genomic variation and phenotypic differences among the *Shewanella* species was examined. Meanwhile, core and variable gene pools were determined through comparative genomics, and it was shown that hypothetical and mobile genetic element-related genes dominate the variable gene pool of the species, indicative of the importance of horizontal gene transfer in evolution of *Shewanella* genomes.

**The history behind the *Shewanella baltica* population in the Baltic Sea.** The Baltic Sea is the world's largest brackish water environment. It is also an estuarine ecosystem with pronounced eutrophication due to anthropogenic activity creating overloads of phosphorus and nitrogen contaminants (20). Gotland Deep is a stratified basin 240 m deep located in the central Baltic Sea (21). A comprehensive set of long-term chemical and hydrographical data has been collected from this site. Vertical mixing of water is inhibited due to a halocline present at 60 to 90 meters depth. As stagnation periods prolong, the deep water displays oxygen deficiency and H<sub>2</sub>S accumulation from the sediment (22). This may have caused the establishment of microbial populations particularly adapted to these conditions. Denitrification was considered the major factor counteracting eutrophication, while organic carbon derived from phytoplankton primary production seemed to be the major electron donor (23).

To study the microbial ecology of this brackish water environment, especially the microbial flora catalyzing this part of the nitrogen cycle in the central Baltic, water samples

were collected from different depths in Gotland Deep in the summers of 1986 and 1987. Different methods were used to isolate denitrifying bacteria (22), yielding 113 *S. baltica* strains representing 77% of the total population of culturable denitrifiers (24). *S. baltica* strains were shown to be specifically enriched in the oxic-anoxic interface from 80 m to 140 m, where vigorous denitrification activity was found (25), as well as strong gradients of oxygen, nitrate and hydrogen sulfide [Figure 1.1]. Thirty-six of these 113 strains were selected for further characterization based each having unique profiles from randomly amplified polymorphic DNA (RAPD) genotyping (25), while we selected four of them, OS155, OS185, OS195 and OS223, isolated from different redox zones, for whole genome sequencing.

The stratification of the redox environment in the Gotland Deep was disrupted by a major water turn over event in 1993 (26) (27), after which a new but similar stratification was established (28). In a cruise in 1998, ten additional *Shewanella baltica* strains were recovered from the same sampling station using a similar procedure. Strikingly, phylogenetic profiles of most of these strains clustered with those already found in the 1980s, indicating the stability over time of the genetic structure of the *S. baltica*, even though the water column had been completely mixed by the turnover event. Therefore, these strains are a valuable resource for investigating both the evolution over space across the redox gradient, and evolution over time across the 12 years.



**Figure 1.1. Water chemistry profile in the Gotland Deep from summer of 1986.** The concentrations of oxygen, hydrogen sulfide and nitrate are highlighted in blue, green and red color. Temperature (T) and salinity (S) are also shown. The yellow box indicates the depth where *S. baltica* strains were isolated. Figure is modified from (25). For interpretation of the references to color in this and all other figures, the reader is referred to the electronic version of this thesis.

**Thesis Outline.** Although systematic studies of comparative physiology, genomics and transcriptomics have been performed on *Shewanella*, this has been done on either a single or a few strains of different species, which limits our ability to understand adaptation at the population level (29) (19). In this thesis, I have chosen the *Shewanella baltica* species as the model to study bacterial specialization because I had access to a large collection (46) of already partially characterized strains that were isolated from defined positions along a stable spatial environmental gradient. This offered the potential to further our understanding of ecology and evolution using the *Shewanella spp.* as the model. I used several methods to characterize

specialization at the level of both genotypic and phenotypic variation, and inferred their evolutionary trajectories and potential underlying mechanisms.

The following chapters are ordered such that the story begins by identifying genetic relatedness at a high level among all strains, followed by comparison of gross differences and commonalities among specific lineages, and ends with a detailed transcriptomic analysis on two competed model strains. Specifically, Chapter 2 introduces a few genotyping systems to characterize the phylogenetic relatedness among all sampled strains, which provides initial evidence for specialization among the *S. baltica* population. Chapter 3 describes my attempt to further exploit variation of genomic contents among these strains through Comparative Genomic Hybridization (CGH), which aims at discovery of genomic signatures potentially important for specialization. Comparative transcriptomic analysis on selected strains also provides additional support for functional importance of differentially distributed genes. Particularly, genomic contents of several lineages representing strains from different niches are contrasted and compared, yielding important insights on specialization and evolution among these lineages. Meanwhile, the *S. baltica* gene core is subjected to comparison with that of *Burkholderia cenocepacia* in order to provide perspective at the species level on how well genetics can reflect the greater ecology of bacterial species.

Chapter 4 focuses on phenotypic examination of *S. baltica* strains, which includes high throughput characterization of carbon source utilization as well as testing respiratory capabilities of several electron acceptors. These phenotypic profiles are further compared with the environmental parameters as well as genomic profiles to reveal how well ecology and genetics can be reflected at the phenotypic level. Lastly, Chapter 5 examines differential fitness

of selected *S. baltica* strains under defined redox conditions through competition assays. Illumina RNAseq technology was used for transcriptomic characterization of two model strains under competition, which also demonstrates the power of Next Generation Sequencing (NGS) platform for generating large amounts of sequence reads starting from very limited RNA material. This chapter's focus is even more specific, and aims at revealing adaptive strategies of *S. baltica* in redox specialization. It is clear that specialization among the *S. baltica* population has occurred at multiple levels, and this is summarized in Chapter 6.

## REFERENCES

## References

1. Staley JT, R. W. Castenholz, R. R. Colwell, J. G. Holt, M. D. Kane, N. R. Pace, A. A. Salyers, and J. M. Tiedje (1997) The Microbial World: Foundation of the Biosphere. *The American Academy of Microbiology*.
2. Brocks JJ, Logan GA, Buick R, & Summons RE (1999) Archean Molecular Fossils and the Early Rise of Eukaryotes. *Science* 285(5430):1033-1036
3. Provine WB (1988) Progress in Evolution and Meaning in Life. *Evolutionary Progress. University of Chicago Press*:49–79.
4. Koonin EV (2008) Darwinian evolution in the light of genomics. *Nucleic Acids Research* 37(4):1011-1034.
5. Brown EW, LeClerc JE, Kotewicz ML, & Cebula TA (2001) Three R's of bacterial evolution: How replication, repair, and recombination frame the origin of species. *Environmental and Molecular Mutagenesis* 38(2-3):248-260.
6. Christophe Fraser EJA, Martin F. Polz, Brian G. Spratt, William P. Hanage (2009) The bacterial species challenge: making sense of genetic and ecological diversity. *Science* 323:6.
7. Cohan FM & Koeppel AF (2008) The Origins of Ecological Diversity in Prokaryotes. *Current Biology* 18(21):R1024-R1034.
8. Fraser C, Hanage WP, & Spratt BG (2007) Recombination and the Nature of Bacterial Speciation. *Science* 315(5811):476-480.
9. Hanage WP, Spratt BG, Turner KME, & Fraser C (2006) Modelling bacterial speciation. *Philosophical Transactions of the Royal Society B: Biological Sciences* 361(1475):2039-2044.
10. Weinbauer MG & Rassoulzadegan F (2003) Are viruses driving microbial diversification and diversity? *Environmental Microbiology* 6(1):1-11.
11. Gogarten JP & Townsend JP (2005) Horizontal gene transfer, genome innovation and evolution. *Nature Reviews Microbiology* 3(9):679-687.
12. Doolittle WF & Papke RT (2006) Genomics and the bacterial species problem. *Genome Biology* 7(9):116.
13. Fredrickson JK, *et al.* (2008) Towards environmental systems biology of *Shewanella*. *Nature Reviews Microbiology* 6(8):592-603.

14. Tiedje JM (2002) *Shewanella* - the environmentally versatile genome. *Nature Biotechnology* 20:2.
15. Hau HH & Gralnick JA (2007) Ecology and Biotechnology of the Genus *Shewanella*. *Annual Review of Microbiology* 61(1):237-258.
16. Bencharit S & Ward MJ (2005) Chemotactic Responses to Metals and Anaerobic Electron Acceptors in *Shewanella oneidensis* MR-1. *Journal of Bacteriology* 187(14):5049-5053.
17. Myers CR & Nealson KH (1988) Bacterial manganese reduction and growth with manganese oxide as the sole electron acceptor. *Science* 240(4857):1319-1321
18. Heidelberg JF, *et al.* (2002) Genome sequence of the dissimilatory metal ion-reducing bacterium *Shewanella oneidensis*. *Nature Biotechnology* 20(11):1118-1123.
19. Konstantinidis KT, *et al.* (2009) Comparative systems biology across an evolutionary gradient within the *Shewanella* genus. *Proceedings of the National Academy of Sciences* 106(37):15909-15914.
20. Elmgren R (1989) Man's Impact on the Ecosystem of the Baltic Sea: Energy Flows Today and at the Turn of the Century. *AMBIO*:326-332.
21. Fonselius S & Valderrama J (2003) One hundred years of hydrographic measurements in the Baltic Sea. *Journal of Sea Research* 49(4):229-241.
22. Brettar I & Rheinheimer G (1991) Denitrification in the Central Baltic: evidence for H<sub>2</sub>S-oxidation as motor of denitrification at the oxic-anoxic interface. *Marine Ecology Progress Series* 77:157-169.
23. Brettar I & Höfle MG (1993) Nitrous oxide producing heterotrophic bacteria from the water column of the central Baltic: abundance and molecular identification. *Marine Ecology Progress Series* 94:253-265.
24. Brettar I (2002) *Shewanella denitrificans* sp. nov., a vigorously denitrifying bacterium isolated from the oxic--anoxic interface of the Gotland Deep in the central Baltic Sea. *International Journal of Systematic and Evolutionary Microbiology* 52(6):2211-2217.
25. Ziemke F, Brettar I, & Höfle MG (1997) Stability and diversity of the genetic structure of a *Shewanella putrefaciens* population in the water column of the central Baltic. *Aquatic Microbial Ecology* 13:63-74.
26. Håkansson B, Broman B, & Dahlin H (The flow of water and salt in the Sound during the Baltic major inflow event in January 1993. *ICES Statutory Meeting Dublin, Paper C.M.1993/C:57*.

27. Matthäus W & Lass H-U (1995) The recent salt water inflow into the Baltic Sea. *J. Phys. Oceanogr.* 25:280-286.
28. Matthäus W, *et al.* (The Baltic Sea in 1996- Continuation of Stagnation and Decreasing Phosphate Concentrations. *Deutsche Hydrographische Zeitschrift* 48(2):161-174.
29. Beliaev AS, *et al.* (2005) Global Transcriptome Analysis of *Shewanella oneidensis* MR-1 Exposed to Different Terminal Electron Acceptors. *Journal of Bacteriology* 187(20):7138-7145.

## CHAPTER II

# Phylogenetic Analysis of a *Shewanella baltica* Population Using Multi Locus Sequence Typing

Thanks to Jorge Rodrigues and Ingrid Brettar, who contributed the *gyrB* gene sequences.

## Abstract

In order to determine the evolutionary relatedness among 46 representative *Shewanella baltica* isolates, a Multi Locus Sequence Typing (MLST) system was developed using seven genes previously identified as phylogenetically informative from the *Shewanella* gene core. Eleven distinct clades were identified. When phylogenetic relatedness was compared with the environmental profiles, niche specialization over space and over time was observed. Eighty-six percent of strains isolated from anoxic water were grouped in one clade, suggesting specialization for the anoxic niche, while strains isolated from other depths were distributed throughout the phylogram. Clustering of strains isolated from different years also indicated specialization over time. Finally, although extensive recombination was observed in four sequenced *S. baltica* strains and thought to dominate their evolutionary paths, this pattern was not strongly supported for the this larger population based upon allelic profiles of the MLST loci.

## Introduction

In the eyes of zoologists, species is one of the basic units of biological organization, i.e. classification. As stated in Wikipedia: “A species is often defined as a group of organisms capable of interbreeding and producing fertile offspring.” When it comes to the microbial world, there is no species concept but there is a species definition which is profoundly different than for higher order biology. Although Bacteria and Archaea have the greatest abundance as well as divergence in the tree of life and are critical to a habitable planet, the principles explaining their ecology and evolution are poorly understood. While the key to the species definition in zoology is the ability of exchanging genetic materials during reproduction, the same principle is not applicable to microbes (1) (2) (3) (4). The boundaries among microorganisms are blurred by the presence of horizontal gene transfer (HGT), which allows one cell to acquire genetic material from cells that are distantly related, and because a considerable portion of any strain’s genome varies within a species, i.e. the pangenome (5) (6). Although homologous recombination is somewhat similar to the sexual behavior in higher organisms, the frequency of recombination can vary greatly among different taxa. Moreover, boundaries of recombination are not yet clear (7) (8) (9).

Early microbiologists used phenotypic characteristics to differentiate bacterial lineages, such as those on the morphological level, or by specific (or combinations of) biochemical assays. However, these assays provided limited resolution in distinguishing bacterial lineages, and there was no way to justifiably argue which assays determined the real ecological differences among different strains. For example, Multi-Locus Enzyme Electrophoresis (MLEE) was used to differentiate bacterial lineages based upon different electrophoretic mobilities (EM) of core

metabolic enzymes (10). However, this approach has the drawback of inability to distinguish silent mutations, which does not cause amino acid changes, and also failed to differentiate between enzymes with different amino acids but without sufficiently different EM to generate distinct bands. Serotyping was also established for differentiating bacterial lineages, especially pathogenic strains, based on cell surface antigens (11). But it heavily relies on presence of a few antigenic loci and could be biased due to unpredictability of reactivities of antibodies with different antigenic variants. Furthermore, the promiscuous nature of bacteria, i.e., the ability of harboring genes for various biochemical pathways through HGT from other microorganisms, complicates many of these categorizing efforts (6). Functional features as revealed from biological assays, although still and will always be essential for inferring ecological importance, are not always appropriate or sufficiently resolving for demarcating bacterial lineages.

As molecular typing schemes emerged and were put to use for differentiating bacterial lineages, much weight was given to the genetic coherence among microorganisms. Many of these molecular typing approaches were based on similarity in genomic structures and DNA profiling, determined by methods involving restriction digestion, PCR amplification, DNA-DNA hybridization, etc. For example, both restriction fragment length polymorphism (RFLP) and amplified fragment length polymorphism (AFLP) detection approaches employ digestion of DNA samples by restriction enzymes and visualization of DNA fragment length polymorphism through gel electrophoresis, while the latter method involves an additional amplification step resulting in higher sensitivity (12) (13). The random amplification of polymorphic DNA (RAPD) method, which is also based on PCR amplification, though without enzymatic digestion, can also give genetically distinct individuals different bands from amplified products (14). Microarray

based pair wise DNA-DNA hybridization has also been used for determining genomic similarities among lineages, and was used by systematists for demarcating bacterial species using the 70% in DNA-DNA association criterion (15). However, all these methods lack either resolution, reproducibility and/or accessibility by different labs, whereas the use of Multi Loci Sequence Typing (MLST) overcomes these limitations.

MLST is a method that utilizes multiple gene sequences and can reveal detailed intra-species phylogenetic relationships (16). It directly measures the nucleotide variation in sequences of a set of genes and uses allelic profiles for genetic characterization of the strains. In order to achieve an acceptable level of identification power while being time and cost effective at the same time, sequences of seven or eight genes are commonly used. Since MLST is sequence based, the information is fully reproducible, unambiguous and portable. The sequence information can be easily exchanged between laboratories, as can primer sequences and amplification protocols. MLST also has the potential of being standardized and automated. Combined with bioinformatic techniques for population genetics, MLST can be efficiently used to investigate evolutionary relationships among different bacterial lineages.

The selection of MLST loci has been based on their being unlinked in the genome and containing conserved sites for PCR primer design. Selection of MLST genes has grown up reflecting interests and databases of particular investigators, and varies according to the species studied. The most commonly used MLST loci were of housekeeping genes involved in central metabolism such as aspartate aminotransferase (*aspC*), malate dehydrogenase (*mdh*), sigma factor of RNA polymerase (*rpoS*), DNA primase (*dnaG*), etc (17). However, the disadvantage of past gene selection strategies is that it was unknown how well the

phylogenetic relationships derived from these genes approximated the real phylogeny of the studied strains. In a comparative genomics study that aimed to address this question, genome sequences of *Salmonella*, *Burkholderia*, and *Shewanella* groups were utilized, and the phylogeny based upon each orthologous gene within each group was compared with the phylogeny based upon a concatenation of all the common genes in the whole-genome as well as the ANI (average nucleotide identity) of the genes in common (18). A number of genes were selected from each genus that best reflected the whole-genome based phylogeny. In this study, I followed this MLST strategy using genes predicted to reflect the true phylogeny to reveal intra-species phylogenetic relationship among a *S. baltica* population from the Baltic Sea. Through investigating genetic signatures of the sequences as well as association between the phylogenetic relatedness and environmental parameters, I hope to connect the relationships among the strains with ecology, and get a glimpse into their evolutionary stories.

## Materials and Methods

**Bacterial strains and growth conditions.** Bacterial strains used are described in [Table 2.1]. LB medium (Acumedia, Lansing, MI, USA) was used for their growth. Anaerobic M1 medium was made under oxygen free argon gas. DNA was extracted using a modified CTAB genomic extraction protocol (19).

**Table 2.1. *Shewanella baltica* Strains Used in this Study**

Strain	Isolation Depth /m	Isolation Method	Isolation Year
<i>Shewanella baltica</i> OS106	90	NBNO <sub>3</sub> <sup>a</sup>	1986
<i>Shewanella baltica</i> OS107	120	NBNO <sub>3</sub>	1986
<i>Shewanella baltica</i> OS109	120	NBNO <sub>3</sub>	1986
<i>Shewanella baltica</i> OS110	120	NBNO <sub>3</sub>	1986
<i>Shewanella baltica</i> OS117	130	NBNO <sub>3</sub>	1986
<i>Shewanella baltica</i> OS155	90	ZB <sup>b</sup>	1986
<i>Shewanella baltica</i> OS167	140	ZB	1986
<i>Shewanella baltica</i> OS183	120	ZBan <sup>c</sup>	1986
<i>Shewanella baltica</i> OS185	120	ZBan	1986
<i>Shewanella baltica</i> OS187	120	ZBan	1986
<i>Shewanella baltica</i> OS189	130	ZBan	1986
<i>Shewanella baltica</i> OS190	130	ZBan	1986
<i>Shewanella baltica</i> OS193	140	ZBan	1986
<i>Shewanella baltica</i> OS195	140	ZBan	1986
<i>Shewanella baltica</i> OS223	120	0.1NBNO <sub>3</sub> <sup>d</sup>	1986
<i>Shewanella baltica</i> OS225	130	0.1NBNO <sub>3</sub>	1986
<i>Shewanella baltica</i> OS230	130	0.1NBNO <sub>3</sub>	1986
<i>Shewanella baltica</i> OS250	120	NBNO <sub>3</sub>	1986
<i>Shewanella baltica</i> OS252	120	NBNO <sub>3</sub>	1986
<i>Shewanella baltica</i> OS286	140	0.1NBNO <sub>3</sub>	1986
<i>Shewanella baltica</i> OS288	130	0.1NBNO <sub>3</sub>	1986
<i>Shewanella baltica</i> OS625	80	THNO <sub>3</sub> <sup>e</sup>	1987
<i>Shewanella baltica</i> OS628	80	THNO <sub>3</sub>	1987
<i>Shewanella baltica</i> OS631	80	THNO <sub>3</sub>	1987
<i>Shewanella baltica</i> OS638	120	THNO <sub>3</sub>	1987
<i>Shewanella baltica</i> OS641	120	THNO <sub>3</sub>	1987

**Table 2.1. (Cont'd)**

Strain	Isolation Depth /m	Isolation Method	Isolation Year
<i>Shewanella baltica</i> OS645	130	THNO <sub>3</sub>	1987
<i>Shewanella baltica</i> OS650	130	THNO <sub>3</sub>	1987
<i>Shewanella baltica</i> OS652	130	THNO <sub>3</sub>	1987
<i>Shewanella baltica</i> OS678	110	THNO <sub>3</sub>	1987
<i>Shewanella baltica</i> OS681	110	THNO <sub>3</sub>	1987
<i>Shewanella baltica</i> OS690	80	NBNO <sub>3</sub>	1987
<i>Shewanella baltica</i> OS696	120	NBNO <sub>3</sub>	1987
<i>Shewanella baltica</i> OS697	120	NBNO <sub>3</sub>	1987
<i>Shewanella baltica</i> OS700	120	NBNO <sub>3</sub>	1987
<i>Shewanella baltica</i> OS710	110	NBNO <sub>3</sub>	1987
<i>Shewanella baltica</i> OS712	110	NBNO <sub>3</sub>	1987
<i>Shewanella baltica</i> BA37	80	(m) THNO <sub>3</sub> <sup>f</sup>	1998
<i>Shewanella baltica</i> BA38	80	(m) THNO <sub>3</sub>	1998
<i>Shewanella baltica</i> BA62	120	(m) THNO <sub>3</sub>	1998
<i>Shewanella baltica</i> BA173	120	(m) THNO <sub>3</sub>	1998
<i>Shewanella baltica</i> BA185	175	(m) THNO <sub>3</sub>	1998
<i>Shewanella baltica</i> BA194	175	(m) THNO <sub>3</sub>	1998
<i>Shewanella baltica</i> BA196	175	(m) THNO <sub>3</sub>	1998
<i>Shewanella baltica</i> BA170	120	(m) THNO <sub>3</sub>	1998
<i>Shewanella baltica</i> BA175	120	(m) THNO <sub>3</sub>	1998

<sup>a</sup> Nutrient Broth medium supplemented with nitrate under anaerobic condition

<sup>b</sup> ZoBell agar plate under aerobic condition

<sup>c</sup> ZoBell agar plate under anaerobic condition

<sup>d</sup> 1:10 diluted Nutrient Broth medium supplemented with nitrate under anaerobic condition

<sup>e</sup> Minimal medium supplemented with nitrate and thiosulfate under anaerobic condition

<sup>f</sup> Modified minimal medium supplemented with nitrate and thiosulfate under anaerobic condition

**MLST Analysis.** Seven genes (SO\_0578, SO\_0625, SO\_1771, SO\_2183, SO\_2615, SO\_2706, SO\_4702 from *S. oneidensis* MR-1) from the conserved gene core of nine *Shewanella* genomes were selected for this assay based on a previous study (18). Primers for Polymerase Chain Reactions (PCR) were designed for these genes based on genomic sequences of the four sequenced *S. baltica* strains (*S. baltica* OS155, *S. baltica* OS185, *S. baltica* OS195, and *S. baltica* OS223) [Table 2.2]. Genes were amplified by PCR [Table 2.3]. The products were purified using ExoSap-IT (USB from Affymetrix, Cleveland, Ohio, USA), and were sequenced by the MSU Research Technology Support Facility. Sequences were aligned with MUSCLE (20). A Neighbor Joining tree was built using concatenated sequences of the seven genes by MEGA version 4 software (21). Phylogenetic networks were analyzed using the neighbor-net algorithm (22) and the uncorrelated P distance with the program SplitsTree 4 (23).

**Table 2.2. Marker Genes Used for MLST Analysis**

Locus Tag	Annotation	Length/bp
SO_0578	Periplasmic metalloprotease	696
SO_0625	Periplasmic protein with Sel1-like repeats	537
SO_1771	D-glycerate transporter, GlyT	594
SO_2183	Periplasmic ErfK/YbiS/YcfS/YnhG family protein	321
SO_2615	Aminodeoxychorismate lyase, PabC	474
SO_2706	Succinylarginine dihydrolase, AstB	702
SO_4702	Glutathione reductase, Gor	903

**Table 2.3. Primers Used for MLST Analysis**

	Forward	Reverse	Temperature/°C
SO_0578	ACGCCGCCGARMAAMGA	STCGARGCGGCAAAGTTATGAT	54
SO_0625	CGCTGCATAAATTCATCTGTCTW	CGGCRCACGCATTTTCYA	52
SO_1771	CGGCGTGATYGCATTTATTGTTAT	AAWGGYGCVGAAGGGAAGTTRG	53
SO_2183	GACTMTCGCGGCACTSRTTTCAC	CARCCGTCGGTCCAGTTGTATTG	58
SO_2615	CYYTMGCCATTGAATATCCACATC	ATCGCCAGCACATCCACWA	53
SO_2706	CCCCWCAAGAGCGCCAGAYC	CCDCCGWTTTGCATGCTTTGTTT	61
SO_4702	ATTATGGWTTYGAYGTYWCTGTTA	TTRGTKGCGCCCATTTTCAT	54
<i>gyrB</i>	GAAGTCATCATGACCGTTCTG	AGCAGGGTACGGATGTGCGAGCC	55

Mixed Base Site Code: R=A,g; Y=C,T; M=A,C; K=g,T; S=C,g; W=A,T; D=A,g,T; N=A,C,g,T.

The DnaSP version 5.10 (24) was used to calculate the ratios of synonymous and non-synonymous substitution rates (Ks and Ka, respectively) and to perform Tajima's D test for neutrality. Population mutation and recombination rates were estimated using a likelihood based method implemented in the LDhat2.0 version 2.1 (25) using similar parameters as described in (26). Detection of recombinational breakpoints was conducted through the RDP v3.4.4 (recombination detection program) (27) as well as GARD (genetic algorithm for recombination detection) (28) (29) program using the default settings. Distance matrices from *gyrB* and MLST analysis were generated using maximum composite likelihood model in MEGA 4.0.

## Results

***gyrB* analysis.** 16S rRNA gene analysis already showed that these strains belong to the same species (30). Sequences of the *gyrB* genes were also screened and analyzed. Using these sequences, we were able to make a preliminary determination of the genetic relationships among *the S. baltica* strains. From the total 1122 bp amplicon of the *gyrB* gene, a phylogeny of 10 well supported lineages (bootstrap value > 90%) was revealed [Figure 2.1].



**Figure 2.1. Neighbor Joining tree of 46 *S. maltophilia* strains produced from *gyrB* sequences.** The strain *S. oneidensis* MR-1 was used as the out-group but is not shown in the Figure. Bootstrap values above 90% are shown on branches based on 1000 bootstrap replicates.

**MLST analysis.** Seven genes from the *Shewanella* conserved gene core were chosen for further phylogenetic characterization of these *S. baltica* isolates (Locus tags as in *Shewanella oneidensis* MR-1: SO\_0578, SO\_0625, SO\_1771, SO\_2183, SO\_2615, SO\_2706, SO\_4702) [Table 2.2].

Twenty-six unique sequence types (ST) were revealed from the concatenated sequences of these loci. Eleven well-supported clusters were identified, among which nine were confirmed in the *gyrB* phylogram [Figure 2.2A]. The SO\_4702, SO\_0578 and SO\_1771 genes account for more than half of all polymorphic sites, whereas others provide comparatively less discriminatory power. Ratios of synonymous and non-synonymous substitution rates (Ks and Ka, respectively) were calculated based upon sequence information at each locus, followed by Tajima's test of neutrality [Table 2.4]. All the Ks/Ka ratios (except SO\_2615 and SO\_4702) are much larger than 1, indicating the presence of purifying selection against amino acid changes.

**Table 2.4. Neutrality Test and Estimation of Recombination ( $\rho$ ) and Mutation Rates ( $\theta$ )**

	SO_0578	SO_0625	SO_1771	SO_2183	SO_2615	SO_2706	SO_4702
Polymorphic Sites	97	46	71	23	43	57	131
Ka	0.00667	0.00261	0.00188	0.00129	0.00529	0.00276	0.00528
Ks	0.13958	0.12357	0.08492	0.08012	0.05083	0.08665	0.13912
Ks/Ka	20.9	47.3	45.2	62.1	9.6	31.4	26.3
Watterson $\theta$	22.1	10.5	16.2	5.2	9.8	13.0	29.8
Estimated $\rho$	35.5	10.5	24	4.5	4.5	7.5	21.5
$\theta$ per site	0.0317	0.0195	0.0272	0.0163	0.0206	0.0185	0.0330
$\rho$ per site	0.0510	0.0196	0.0404	0.0140	0.0095	0.0107	0.0238
Recombination to Mutation Ratio	0.804	0.502	0.743	0.430	0.230	0.289	0.361
RDP*	No	No	No	No	No	No	Yes

\* Only recombination signals detected by at least three methods incorporated in the RDP program is displayed.





**Recombination analysis on genes selected for genotyping.** To investigate the influence of recombination on the phylogeny of the ST's, phylogenetic networks were examined using the SplitsTree analysis [Figure 2.2B]. A number of parallel branches were found, indicating the presence of phylogenetic incompatibilities among the MLST loci. Such incompatibilities can result from evolutionary processes such as recombination and recurrent mutation. Based on the amount of parallel branches, which is indicative of the number of shared evolutionary paths among ST's, various levels of recombination were observed among different phylogroups. Specifically, the genotypes in Clade C showed the most extensive amount of recombination, while there also seemed to be recent recombination events between Clade A and B. Moreover, OS625 in Clade B, an isolate from 1987 obtained in a more oxic zone, shared common paths with both OS187 and ST's in Clade A, most of which were from at or below the oxic-anoxic interface, implying the involvement of recombination during possible migration of the *S. baltica* strains.

Population recombination ( $\rho$ ) and mutation ( $\theta$ ) rates were approximated individually for each marker gene. The recombination to mutation ratio calculated for the seven loci ranged from 0.230 to 0.804, indicating the dominance of clonal evolution paths. The sequences were subsequently analyzed in the RDP and GARD program for identification of recombination break points. While RDP failed to detect any recombination signal in any of the MLST loci except for SO\_4702, recombination break points with significance level ( $p < 0.01$ ) were found in all the seven loci by the GARD program [Table 2.4]. However, it has been reported that GARD tends to overestimate the effects of recombination in certain cases.

## Discussion

**Genotyping indicates niche specialization and genetic continuity over time.** When phylogenetic groups were compared to the geographic and temporal distribution of these isolates, significant clustering patterns were found with regard to depths and years ( $p < 0.01$ ). From correlating the MLST phylogeny with the depth from which the strain was isolated, one notable trend was the predominance of *S. baltica* strains in Clade A from the anoxic zone - 80% of strains in Clade A were isolated from water at and below the oxic-anoxic interface. Specifically, 75% of strains isolated from the anoxic zone in 1986 and 100% of strains isolated from the anoxic zone in 1987 are members in this clade. Interestingly, while all strains in the Clade A isolated in 1986 were from at and below the oxic-anoxic interface, this genotype seemed to have expanded in 1987 and was recovered from more diverse depths (including above the oxic-anoxic interface). Although none of the strains isolated from the anoxic water in 1998 were grouped in this clade, we speculate that different strategies of adaptation to the oxygen depleted environment were developed in the 1998 strains considering that some of these strains were isolated from a distinct depth of 175m. In addition, the MLST phylogeny indicates the persistence of certain clades over time in these *S. baltica* lineages. While seven of the eleven clades contain strains isolated from both sampling years of 1986 and 1987, two clades (J, K) contain strains isolated from both 1998 and the years in the 1980's, providing further support for persistence of these particular genotypes in the environment. Strains OS225 and BA173A (as well as the strains in the cluster J) have very high levels of sequence similarities (99.9% and 100%, respectively), which implies that the later strains are very likely descendants of those from the 1980's. Influence of the isolation conditions had been reported on selection

of specific genotypes based on initial genotyping using low-molecular-weight (LMW) RNA profiling with a larger population of *Shewanella baltica* strains (31). While comparing phylogenetic relatedness from the MLST analysis among the selected 46 strains with the corresponding isolation procedures, this trend was also present yet not as significant ( $p = 0.1$ ), which is likely due to the smaller sample size. Strains in the Clade E were mostly isolated using a nitrate supplemented nutrient broth medium, while strains in Clade I and J, mostly isolated in 1998, were isolated from a modified minimum medium supplemented with nitrate and thiosulfate. In summary, phylogenetic inference from the MLST analysis displayed strong correlation with the water geography and temporal distribution of these isolates, and also showed some correlation with isolation procedures.

**Recombination was detected from sequences of the marker genes.** Effects of recombination have been noticed in phylogenetic analysis of other microorganisms (17), and may result in inaccurate representation of evolutionary relationships of the studied species. Yet this limitation could partly be overcome by using concatenated sequences from multiple genes, according to a previous study where significant correlation was found between lengths of marker genes and the quality of phylogram based upon whole-genome-sequence analysis (18). A recent study on comparative genomics of four sequenced *S. baltica* strains revealed significant levels of homologous recombination (32). Similar processes were thought to occur among more *S. baltica* strains, supported by the comparative genomic hybridization profiles. The recombined gene pool included the conserved core genes as well as ecologically important genes and operons. Specifically, one of the genes used in the MLST analysis, SO\_2706, was

shown to be part of the recombined gene pool between OS185 and OS195, suggesting effects of recombination present in the phylogenetic inference among these strains. Thus, sequences of all marker genes were subjected to recombination analysis and checked for signals of recombination [Table 2.4]. As a result, while SO\_0578 has the highest recombination to mutation ratio (0.804), SO\_2706 has a much lower recombination to mutation ratio of 0.289. This low  $\rho/\theta$  ratio for SO\_2706 may be due to lack of polymorphic sites from the sequences, or rather, that recombination was not as important in the evolutionary processes among these 46 strains as it was between the strains OS185 and OS195. Nonetheless, despite the detection of recombination, recombination to mutation ratios of lower than 1 in all the marker genes, suggested dominance of clonal paths in the *S. baltica* population. Furthermore, high agreement between the MLST phylogeny and comparative genomic hybridization (CGH) clustering pattern (see Chapter 3) also indicates that recombination did not heavily influence the accuracy of the MLST genotyping method. In conclusion, although it has been reported that recombinational/sexual rather than clonal propagation dominated evolutionary paths of a few *S. baltica* strains; this result does not appear to be the case for the larger *S. baltica* population studied here, at least based on sequence information of these seven MLST loci.

## REFERENCES

## References

1. Christophe Fraser EJA, Martin F. Polz, Brian G. Spratt, William P. Hanage (2009) The bacterial species challenge: making sense of genetic and ecological diversity. *Science* 323:6.
2. Cohan FM (2002) What are bacterial species? *Annual Review of Microbiology* 56(1):457-487.
3. Doolittle WF & Papke RT (2006) Genomics and the bacterial species problem. *Genome Biology* 7(9):116.
4. Cohan FM (2001) Bacterial Species and Speciation. *Systematic Biology* 50(4):12.
5. Frost LS, Leplae R, Summers AO, & Toussaint A (2005) Mobile genetic elements: the agents of open source evolution. *Nature Reviews Microbiology* 3(9):722-732.
6. Gogarten JP & Townsend JP (2005) Horizontal gene transfer, genome innovation and evolution. *Nature Reviews Microbiology* 3(9):679-687.
7. Hanage WP, Fraser C, & Spratt BG (2006) The impact of homologous recombination on the generation of diversity in bacteria. *Journal of Theoretical Biology* 239(2):210-219.
8. Fraser C, Hanage WP, & Spratt BG (2007) Recombination and the Nature of Bacterial Speciation. *Science* 315(5811):5.
9. Vos M (2009) Why do bacteria engage in homologous recombination? *Trends in Microbiology* 17(6):226-232.
10. Selander RK, *et al.* (1986) Methods of Multilocus Enzyme Electrophoresis for bacterial population genetics and systematics. *Applied Environmental Microbiology* 51(5):873-884.
11. Lancefield RC (1933) A serological differentiation of human and other groups of hemolytic streptococci. *The Journal of Experimental Medicine* 57(4):571-595.
12. RK Saiki SS, F Faloona, KB Mullis, GT Horn, HA Erlich and N Arnheim (1985) Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* 230(4732):1350-1354
13. Pieter Vos RH, Marjo Bleeker, Martin Reijans, Theo van de Lee, Miranda Hornes, Adrie Friters, Jerina Pot, Johan Paleman, Martin Kuiper and Marc Zabeau (1995) AFLP: a new technique for DNA fingerprintin. *Nucleic Acids Research* 23(21):4407-4414.
14. J G Williams ARK, K J Livak, J A Rafalski, and S V Tingey (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research* 18(22):6531-6535.

15. Cho JC & Tiedje JM (2001) Bacterial Species Determination from DNA-DNA Hybridization by Using Genome Fragments and DNA Microarrays. *Applied and Environmental Microbiology* 67(8):3677-3682.
16. Martin C. J. Maiden, *et al.* (1998) Multilocus sequence typing: A portable approach to the identification of clones within populations of pathogenic microorganisms. *Proceedings of the National Academy of Sciences* 95(6):3140-3145.
17. Walk ST, *et al.* (2009) Cryptic Lineages of the Genus *Escherichia*. *Applied and Environmental Microbiology* 75(20):6534-6544.
18. Konstantinidis KT, Ramette A, & Tiedje JM (2006) Toward a More Robust Assessment of Intraspecies Diversity, Using Fewer Genetic Markers. *Applied and Environmental Microbiology* 72(11):7286-7293.
19. J G Williams ARK, K J Livak, J A Rafalski, and S V Tingey (2004) Bacterial genomic DNA isolation using CTAB. *DOE Joint Genome Institute*.
20. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32(5):1792-1797.
21. Tamura K, Dudley J, Nei M, & Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) Software Version 4.0. *Molecular Biology and Evolution* 24(8):1596-1599.
22. Moulton DBaV (2004) NeighborNet: An agglomerative method for the construction of planar phylogenetic networks. *Mol Biol Evol* 21:255-265.
23. Huson DH (2006) Application of Phylogenetic Networks in Evolutionary Studies. *Molecular Biology and Evolution* 23(2):254-267.
24. Librado P & Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25(11):1451-1452.
25. Gil McVean PAaPF (2002) A coalescent-based method for detecting and estimating Recombination from gene sequences. *Genetics* 160:1231-1241.
26. Konstantinidis KT & DeLong EF (2008) Genomic patterns of recombination, clonal divergence and environment in marine microbial populations. *The ISME Journal* 2(10):1052-1065.
27. Rybicki DMaE (2000) RDP: detection of recombination amongst aligned sequences. *Bioinformatics* 16:562-563.
28. Kosakovsky Pond SL (2006) Automated Phylogenetic Detection of Recombination Using a Genetic Algorithm. *Molecular Biology and Evolution* 23(10):1891-1901.

29. Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, & Frost SDW (2006) GARD: a genetic algorithm for recombination detection. *Bioinformatics* 22(24):3096-3098.
30. Frank Ziemke MGH, Jorge Lalucat, and Ramon Rosselló -Mora (1998) Reclassification of *Shewanella putrefaciens* Owen's genomic group II as *Shewanella baltica* sp. nov. *International Journal of Systematic Bacteriology* 48:179–186.
31. Höfle MG & Brettar I (1996) Genotyping of heterotrophic bacteria from the central Baltic Sea by use of Low-Molecular-Weight RNA profiles. *Applied and Environmental Microbiology* 62(4):1383–1390.
32. Caro-Quintero A, *et al.* (2010) Unprecedented levels of horizontal gene transfer among spatially co-occurring *Shewanella* bacteria from the Baltic Sea. *The ISME Journal* 5:131–114.

## CHAPTER III

### Comparative Genomic Study of the *Shewanella baltica* Genomes

#### Using Microarrays

The data from expression arrays were contributed by Jennifer Auchtung, a former post-doc in the lab.

Part of the work from this chapter has been published in:

Caro-Quintero, A., J. Deng, et al. (2010). "*Unprecedented levels of horizontal gene transfer among spatially co-occurring Shewanella bacteria from the Baltic Sea.*" *The ISME Journal* 5: 131–114.

## Abstract

Comparative Genomic Hybridization (CGH) was employed to probe the genome plasticity as well as potential evolutionary processes that act on the genomes of *Shewanella baltica* strains. Microarrays contained probes from all genes and intergenic regions longer than 200 bp of four sequenced *S. baltica* genomes. DNA from 46 *S. baltica* strains was hybridized to the arrays, through which heterogenic gene distribution in *S. baltica* strains was observed. Genes differentially distributed among strains from microaerophilic vs. anoxic conditions, as well as among strains from the 1980s vs. 1998 were extracted and contrasted to reveal genes potentially important for redox specialization and changes in gene content over time. Core genes of strains from several lineages were also analyzed and compared with each other to reveal genomic differences among lineages. In conjunction with expression arrays of two model strains (OS185 and OS195) under different redox growth conditions, where genes and operons of potential functional importance were determined, these CGH profiles allowed us to draw more confident conclusions about which genes are likely important in specialization of the *S. baltica* population. Extensive horizontal gene transfer, reflected through genes linked to mobile genetic elements, was revealed from CGH profiles. Recombination between OS185 and OS195 lineages was also supported.

## Introduction

Comparative genomics provide unprecedented opportunities to probe into the plasticity, dynamics, and evolution of genomes of interest. Through exploring similarities and differences in genome contents and exploiting relationships of genome structures, one can gain insight into the signatures of selection and better understand evolutionary processes that act on genomes. In conjunction with ecological investigations, comparative genomics provides a platform for studying directional evolution and niche adaptation (1) (2).

Oligonucleotide arrays (oligoarrays) are a sequence-based approach and a powerful tool for high throughput characterization of many biological systems (3). Through hybridizing target DNA or cDNA pools to arrays that contain synthesized probes designed from genes of interest, investigators are allowed to obtain quantitative measurement of tens of thousands of genes at the same time. Oligoarrays are still under continuing development to lower cost and improve performance. They have been commonly applied to comparative expression studies (4) (5) (6), but they also can be used to assay the presence of particular genes in comparative genomic and metagenomic studies of pure culture and environmental samples (1) (7) (8). The popularity of this approach in today's microbial ecological studies is largely due to its relative accessibility, cost-efficiency, highly parallel analysis capacity and flexibility of design.

Lately, oligoarrays have been employed for genomic comparisons between different strains in both inter- and intra- species studies (1). Through hybridizing DNA of tester strains with probes designed from genomic sequences of reference strains, differences in gene contents can be revealed. Such oligoarrays provide a promising and convenient tool to uncover gene-content of unknown strains and also are useful in tracking the dynamics of natural

populations over time and over environmental fluctuations. In a recent comparative genomic study, oligoarrays were successfully applied to a collection of over 199 *Burkholderia cenocepacia* isolates from both epidemic and environmental origins and were used to evaluate the distribution of virulence-associated and antibiotic resistance related genes (unpublished data). Comparative genomic hybridization analysis of environmental and clinical *Escherichia coli* strains also provided gene-level evidence of niche adaptation which was further utilized to infer *E. coli* pathogenic transmission and expansion in natural loci (9).

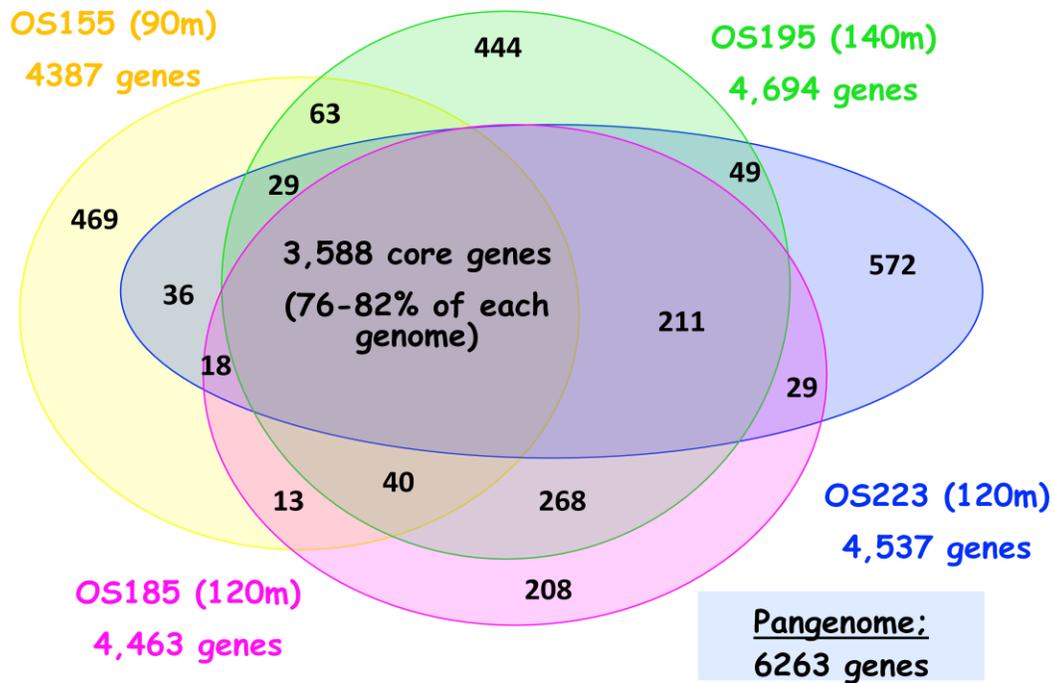
The accuracy of results from microarray experiments is subject to multiple sources of variation including those introduced during the array manufacture processes, sample preparation, array hybridization, and measurement of spot signal intensity. Thus, normalization plays an important role in removing these variations. Different normalization methods may lead to drastically different outcomes of microarray data analysis (3).

I describe here the design and implementation of an oligonucleotide microarray that targets the pan-genome of the *Shewanella baltica* species. Through sets of probes designed to target the genome fragments of four sequenced reference genomes, gene contents of 42 unsequenced strains were determined. A novel normalization method was developed, and the hybridization specificity and sensitivity of the arrays were also tested. From the comparative genomic analysis of these *Shewanella baltica* strains, I sought to answer these questions: (i) what is the extent of gene diversity in the *S. baltica* species, and how is this related to their phylogenetic relationships; (ii) is there a correlation between the particular environment of these isolates and their gene content and, if so, which genes are potentially important for niche adaptation, (iii) during the twelve years' span between time of isolation, how has their gene

content changed and (iv) what is the composition of the core genome of the species and how is it related to the general ecology of the species?

## Materials and Methods

**Array Design.** Microarrays were made by MYcroarrays, LLC (Ann Arbor, MI, USA), using a proprietary light-directed oligonucleotide synthesis method to synthesize 44-48 nucleotide long probes on glass slides. Probes were designed from the genomic sequences of the four sequenced *S. baltica* strains. For the purposes of probe design, those genes that shared 90% sequence identity over 90% of their length between sequenced genomes were considered homologous and treated as a single gene, even when represented by more than one sequence. When homologous gene sequences diverged, probes were designed using the following sequence preference: OS185> OS195> OS223> OS155 [Figure 3.1] [Table 3.1]. This strategy resulted in the generation of 30,000 probes, with up to 7 probes designed for each gene, and up to 2 probes for each unique intergenic region. An additional 720 negative control probes were included for quality control.



**Figure 3.1.** Spheres of the diagram represent the gene content of the four sequenced *S. baltica* strains (OS155, OS185, OS195, and OS223). Numbers present in regions of overlap between spheres indicate the total number of shared orthologous genes (as determined by reciprocal best-BLAST hits). The numbers of genes unique to each strain are also indicated in the regions of each sphere that do not overlap. Summary of analysis was done by A. Caro-Quintero and K. Konstantinidis.

**Table 3.1. Number of Probes Designed from Each Reference Strain**

	Number of Probes	Number of Genes
OS185	23192	4228
OS195	2459	642
OS223	2112	532
OS155	2236	580
Total	30000	5982

**DNA microarray hybridization.** For DNA-DNA hybridization studies, genomic DNA was extracted as previously described (10), and sonicated to produce DNA fragments less than 3 kbp

in size. DNA samples were labeled with the fluorescent Cy5 dye by incorporation of amino-allyl-dUTP through extension from random primers using *E. coli* DNA polymerase Klenow fragment I (Invitrogen, Grand Island, NY), followed by addition of amine-reactive Cy5. Microarray slides were pre-hybridized in buffer containing 0.1% SDS, 5 X SSC buffer (Sigma-Aldrich Corp., St. Louis, MO) and 1 mg/mL bovine serum albumin (Sigma-Aldrich) at 50 °C for 90 min, and washed with 0.5 X SSC and water. Cy5-labeled DNA samples were mixed with the same volume of 2 X hybridization buffer (10xSSC, 0.2% SDS, 0.2 mg/mL herring sperm DNA, and 46% formamide), heated at 95 °C for 5 min and then transferred to 68 °C. Samples were applied to pre-hybridized slides, which were then incubated at 50 °C for 18 h before being washed and scanned using an Axon GenePix 4000B scanner (one-channel hybridization).

**Gene expression studies.** Cells were inoculated into 25 mL of trypticase soy broth and incubated at 22°C with agitation until the cells reached mid-exponential phase (Optical density at 600 nm = 0.4-0.7). Experiments were repeated in triplicate. Cells were pelleted by centrifugation and resuspended in 350 ml anaerobic HEPES medium (11). After 18-19 hours of growth at 22 °C, cells were pelleted anaerobically and resuspended in 10 mL of anaerobic HEPES medium lacking an electron acceptor. Out of this cell suspension, 2 mL was added to 22.5 ml of HEPES medium containing either 10 mM sodium fumarate, 5 mM sodium nitrate, 10 mM sodium thiosulfate or 10 mM sodium chloride (aerobic culture). All cultures were incubated at 22 °C. The aerobic cultures were aerated by shaking under air on an orbital shaker at 150 rpm. RNA was extracted from cell pellets using a Qiagen RNeasy kit following the optional protocol for better recovery of low molecular weight RNA. RNA from three

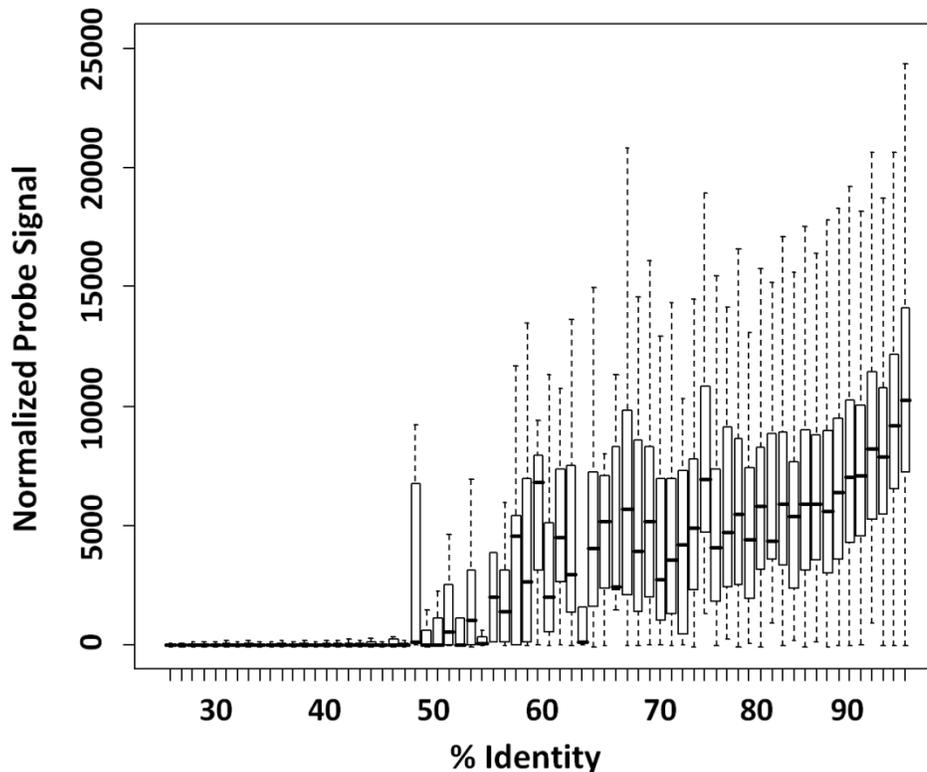
independent cultures of OS185 and OS195 grown in the presence of oxygen, nitrate, and thiosulfate was used as experimental samples in hybridization experiments. RNA from all four strains grown in each condition (oxygen, nitrate, thiosulfate, and fumarate) was used to construct a reference RNA pool, composed of 45 µg of RNA from each condition for strains OS185, OS195 and OS223 and 15 µg of RNA from each condition for strain OS155. RNA (10 µg) from each experimental condition and a parallel aliquot of reference RNA were reverse transcribed with 9 µg of random primers (Invitrogen). Reactions were incubated at 25 °C for 10 min, 42 °C for 70 min, and 70 °C for 15 min. Remaining RNA was hydrolyzed by adding sodium hydroxide to 33 mM and incubation at 70 °C for 10 min. Labeled cDNA was purified using a QiaQuick MinElute PCR purification column following the manufacturer's protocol with the exception that the sample was eluted in 12 µl of RNase-free water (Qiagen, Valencia, CA). For each hybridization, 10 µl of labeled experimental cDNA was mixed with an equal volume of labeled reference cDNA and was applied to the oligoarray as described above for DNA-DNA studies. Cy3 and Cy5 signals for each array were normalized to the arithmetic mean of ratios for each array using the GenePix software version 5.1. Features that had fewer than 50% of pixels with signal more than two standard deviations above background in both Cy5 and Cy3 channels were excluded from further analysis. Genes showing significantly increased anaerobic gene expression (nitrate and thiosulfate) relative to aerobic growth were identified using Significance Analysis of Microarrays (12). All experiments were repeated in triplicates.

**Correlation analysis on *gyrB*, MLST and CGH based distance matrices.** Distance matrices of the *gyrB* gene and concatenated sequences from MLST analysis were generated by maximum

composite likelihood model in Mega4.0. The distance matrix of the CGH profile is based on Jaccard distances. All matrices were standardized before calculation of the mean squared error.

## Results and Discussion

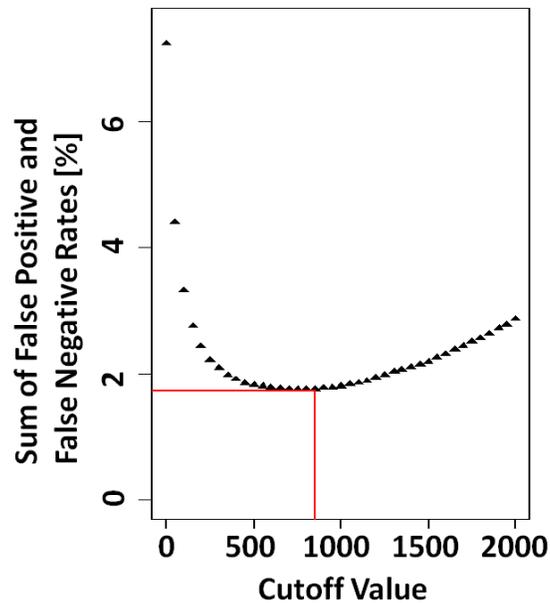
**Evaluation of probe binding efficiency and specificity.** I estimated to what extent DNA-probe hybridization would occur if sequences were not identical by analyzing the hybridization profiles of the four reference genomes [Figure 3]. I plotted the normalized signal of all probes in each CGH profile against the percent identity of their best BLAST hit in the corresponding genome. Data from all four profiles were combined and displayed by box-plot [Figure 3.2a]. The intensity of signals decreased with sequence similarity and dropped to very low levels for sequence similarities lower than 50%. Although there were some high signals from probes with very low sequence identity, they are likely due to short stretches of high similarity between probes and genome sequences which cross-hybridized in the absence of a full-length match. Further statistical analysis showed that the relative number of these probes was very low.



**Figure 3.2a. Box-Plot of probe signals from CGH profiles of the four reference strains against percent identity of the probes matched to their genomes.**

**Array normalization and setting cutoff value for determining gene presence/absence.** For array data processing and normalization, mean signal intensity from the negative control probes was subtracted from the signal intensity of all spots. Subsequently, the median signal from the core probes with 100% identical matches in all four genomes was calculated for each microarray dataset. Based on these calculations, a normalization factor was generated that would adjust the median signal of the 100% identity core probes to the same value for each slide. This normalization factor was then applied to all measured spot intensities minus background. The signal value used for each gene was the median value of normalized signals from all probes designed for that gene. Statistical approaches were used to determine the cut-

off to evaluate presence/absence of genes based upon CGH profiles. False positive values were defined as the percentage of probes of less than 40% similarity but with signals of higher than the proposed cutoff value, and false negative values were defined as the percentage of probes with more than 80% identity but having signals lower than the cutoff. The best cutoff value was determined to be 850 units, at which the sum of false positive and false negative percentages are the lowest (<2%) [Figure 3.2b].



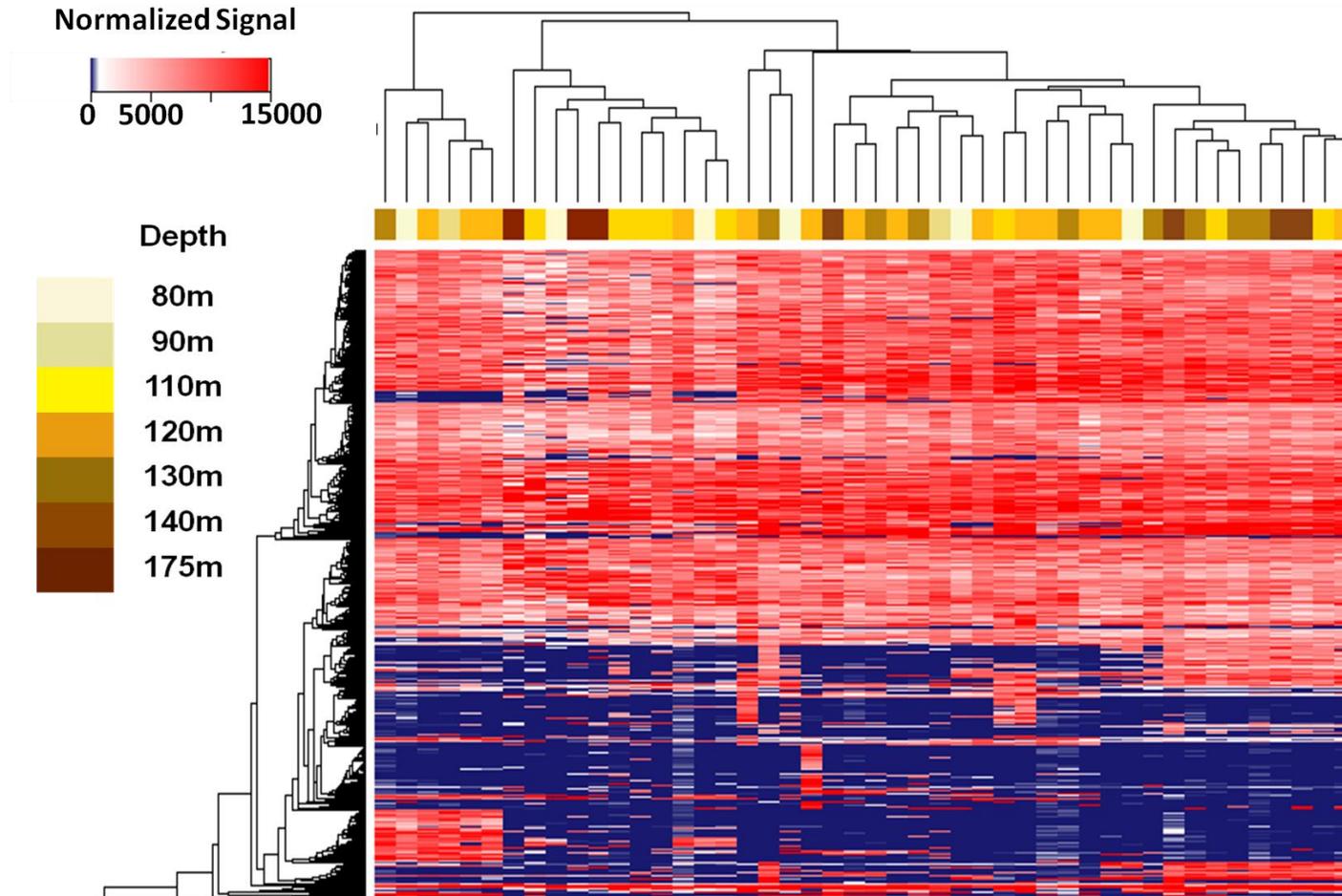
**Figure 3.2b.** Sum of false positive and false negative rates under different cut-off values. Red line indicates the final cutoff value chosen. X-axis is the cutoff values; Y-axis shows the sum of false positive and false negative rates under each cutoff value.

**General statistics of CGH.** Among the 5635 genes targeted by the 28712 probes on the array, 4075 genes (72.3%) on average showed significant signals when tested against *S. baltica* genomes [Figure 3.3]. The genome of *S. baltica* OS195 had the largest number of genes (4397) detected, while only 3331 genes were detected in *S. baltica* BA38. Among all genes represented

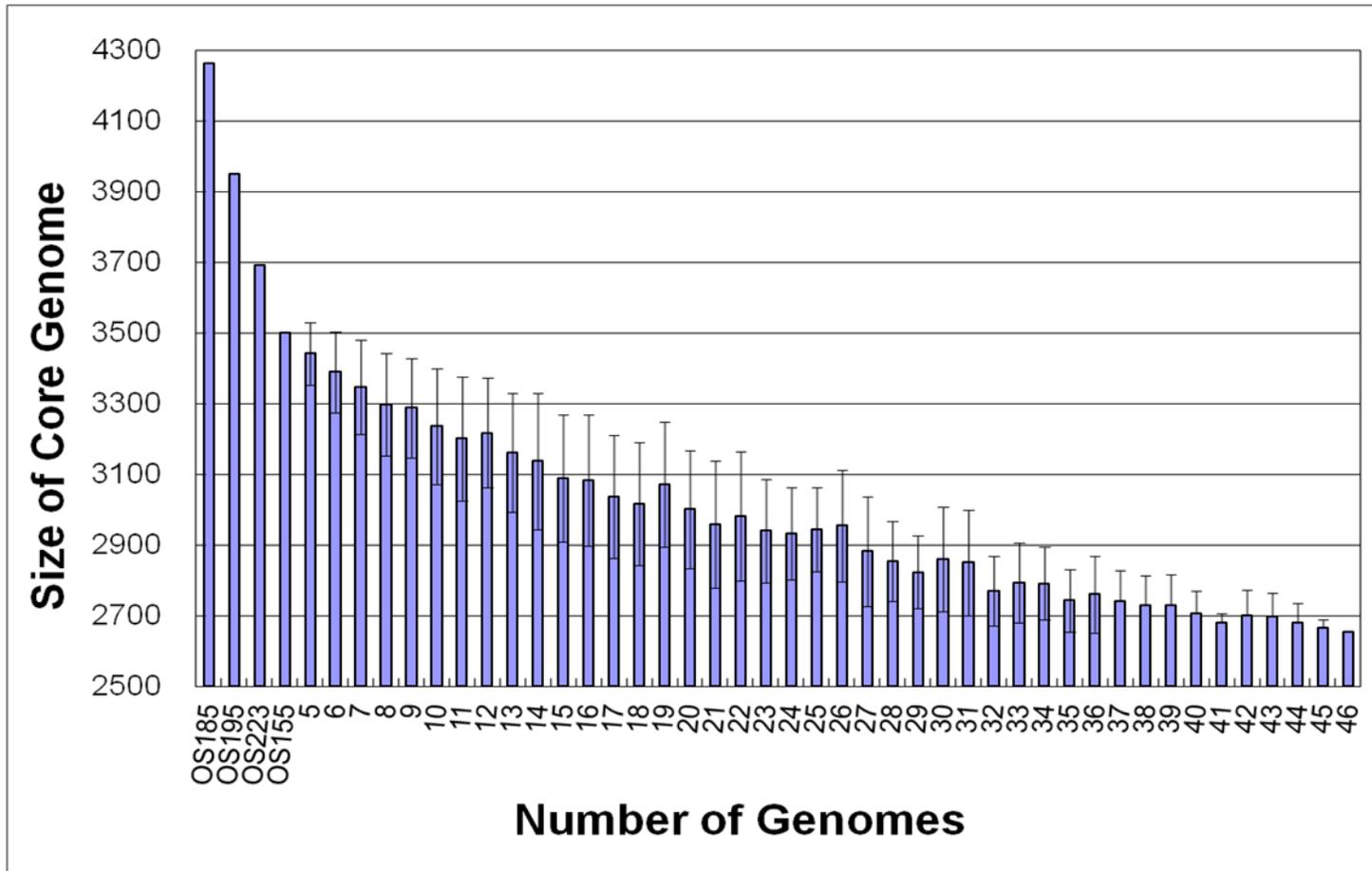
on the array, 32 were not detected in any strains, all of which targeted one plasmid in OS195, implying loss of this plasmid during laboratory culturing of the strain. Moreover, 3395 (60.2%) genes were shared among over 90% of all tested strains and are regarded as the core genome of the species, while the remaining 2208 genes comprise the variable part of the pangenome [Figure 3.4]. A slightly larger proportion of variable genes (37%) was from the OS185 genome based-probes while about equal contributions from the four reference genomes were found in the pool of variable genes [Table 3.2]. As expected, the number of core genes of the *S. baltica* species was 28% larger than the previously estimated size of the core genome (2654 genes) of the *Shewanella* genus which was determined from a comparative genomic study of ten genomes from different species in the *Shewanella* genus (13).

**Table 3.2. Composition of Variable Genes**

	# of variable genes
OS185	817
OS195	453
OS223	504
OS155	434
Total	2208



**Figure 3.3.** Heatmap of CGH signals from all genes and intergenic regions of the 46 *S. baltica* strains. Blue color indicates the genes and intergenic regions with signals lower than the cut-off value. Colors from white to red represent the signal intensity from weak to high. The color bar on the right of the heatmap represents the depths from which strains were isolated. Hierarchical clustering was performed both by strains (vertical) and by genes (horizontal).

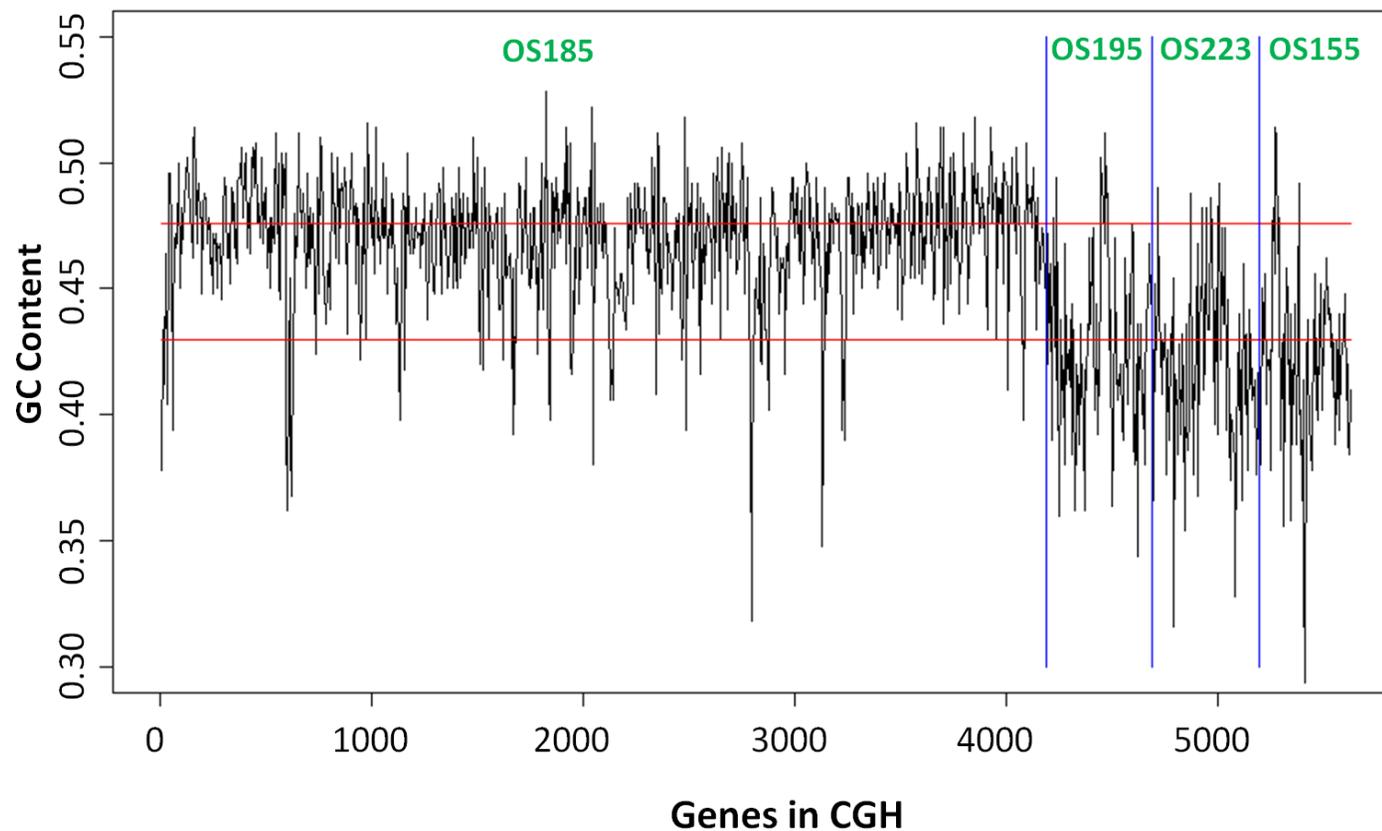


**Figure 3.4. The sizes of core genome calculated from the CGH dataset.** The average numbers of core genes with increasing numbers of genomes were plotted. Error bars represent the standard deviation of core genome size when genomes were sampled randomly.

**GC bias in the core and variable genome.** While the average GC content in the four reference *S. baltica* genomes range between 46.1% and 46.3%, further investigation into the GC content of the core and variable genomes revealed much greater differences [Table 3.3] [Figure 3.5]. Genes belonging to the core genome showed a slightly higher GC content (0.48) comparing to those that vary among genomes (0.43). This may indicate codon usage bias in some variable genes which likely originated from more distantly related organisms. In fact, GC bias has been used as one of the indicators for identifying horizontally transferred genomic islands in many comparative genomic studies. A closer look into the genes with low GC content also shows their connection to mobile genetic elements including phages and plasmids, demonstrating again the importance of horizontal gene transfer in the evolution of the *S. baltica* genomes.

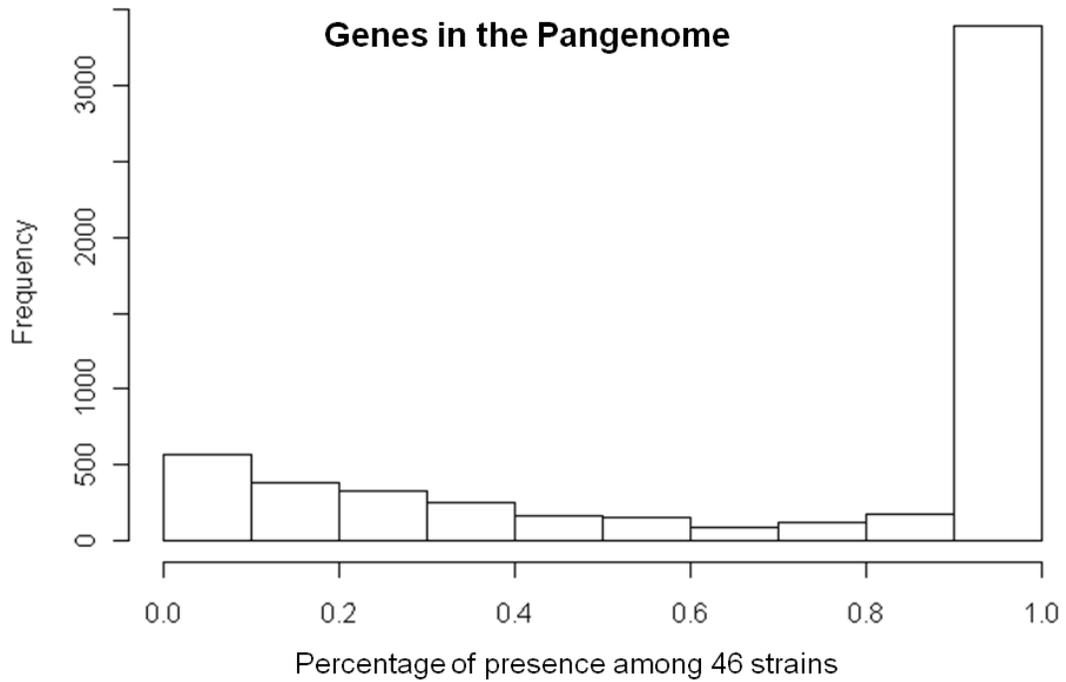
**Table 3.3. Comparison of GC Content in Core Genome and Variable Genome**

GC%	All genes	Core	Var
Mean	45.8	47.6	43.0
Standard Deviation	4.7	2.8	5.5
Median	47.0	48.0	43.0

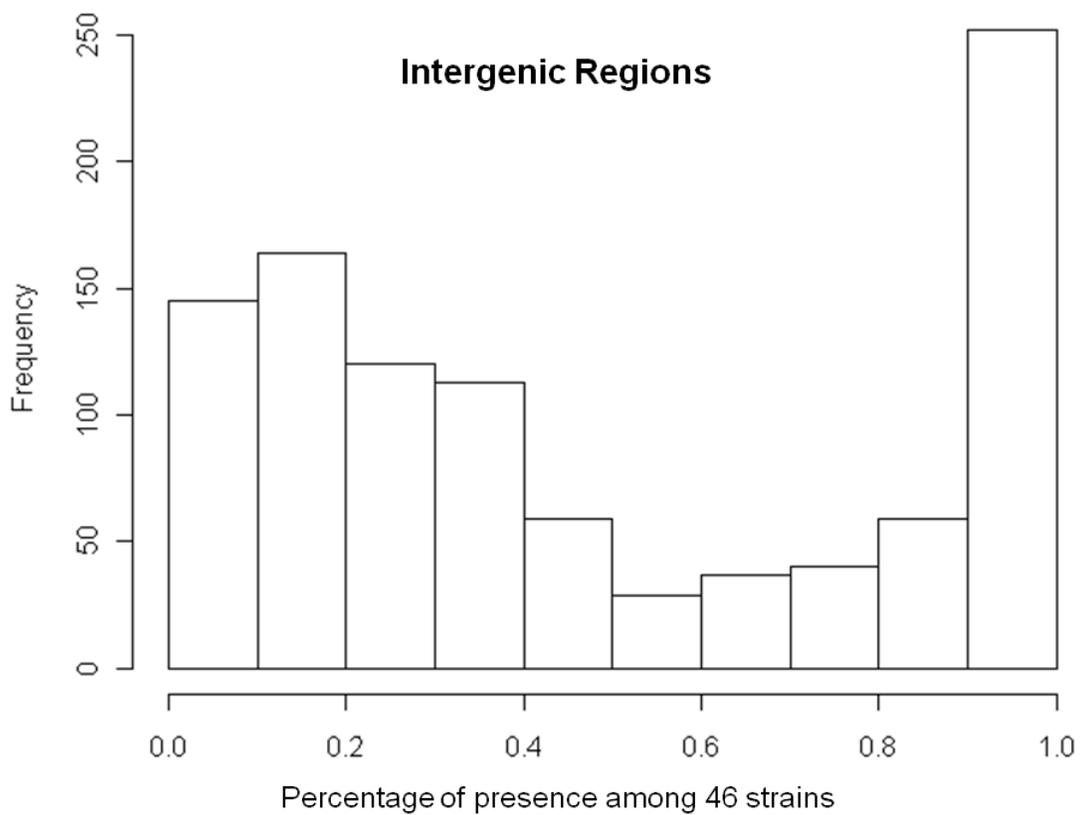


**Figure 3.5. GC contents of genes that are represented in the microarray.** Genes were aligned based on their original order in the four reference genomes. The upper red line across the figure ( $y=0.476$ ) marked the average GC content of the core gene pool, while the lower red line ( $y=0.430$ ) marked the average GC content of the variable core gene pool.

**Less conservation of intergenic regions.** One thousand eighteen intergenic regions were also targeted in the microarray. An average of 493 of these intergenic regions showed positive signals in the 46 tested genomes, while the number of intergenic regions present in each genome varied from 287 (BA38) to 578 (OS195). Distribution of conservation of intergenic regions was further analyzed and plotted in [Figure 3.6]. Compared with the CGH signals from genes, a much lower level of conservation was revealed in intergenic regions, where probed regions shared by fewer than 50% of strains accounted for a relatively larger portion. A smaller number of intergenic regions were observed in the core genome of *S. baltica* strains. Only 252 intergenic regions (24.8%) were shared by greater than 90% of all *S. baltica* strains, which is much lower than the 60.2% of core genes that are shared.



(a)



(b)

**Figure 3.6. Histograms illustrating frequency of presence among 46 *S. baltica* strains of (a) all genes and (b) intergenic regions represented in the microarray.**

**Functional bias in the core genome and variable genome.** To further probe the functional differences between core genes and variable genes, Clusters of Orthologous Groups (COG) of the two gene pools were analyzed [Figure 3.7]. Although most genes in the core genome (85%) were successfully annotated, around half of the variable genes failed to be assigned to any COG categories. Through comparing the annotated genes in both gene pools, I observed that among the seven COG categories that are most enriched in the variable gene pool (L, M, N, K, V, R, U, see [Table 3.4] for description of COG groups), four categories were associated with cellular processes and signaling, two were associated with information storage and processing, and the other one fell into the poorly characterized class. In contrast, among the seven categories most enriched in the core gene pool (J, E, C, H, O, F, I, P), only one category was associated with cellular processes and signaling or information storage and processing, while five of the seven categories were metabolism associated. Thus, there was clearly functional bias in the core and variable gene pool, which supports the idea that, while the core genome encoded essential genes required for basic metabolism and were under higher selective pressure, the variable gene pool tended to be associated with specific life styles of different strains, which could potentially be linked to the differential niche environments of individual strains.

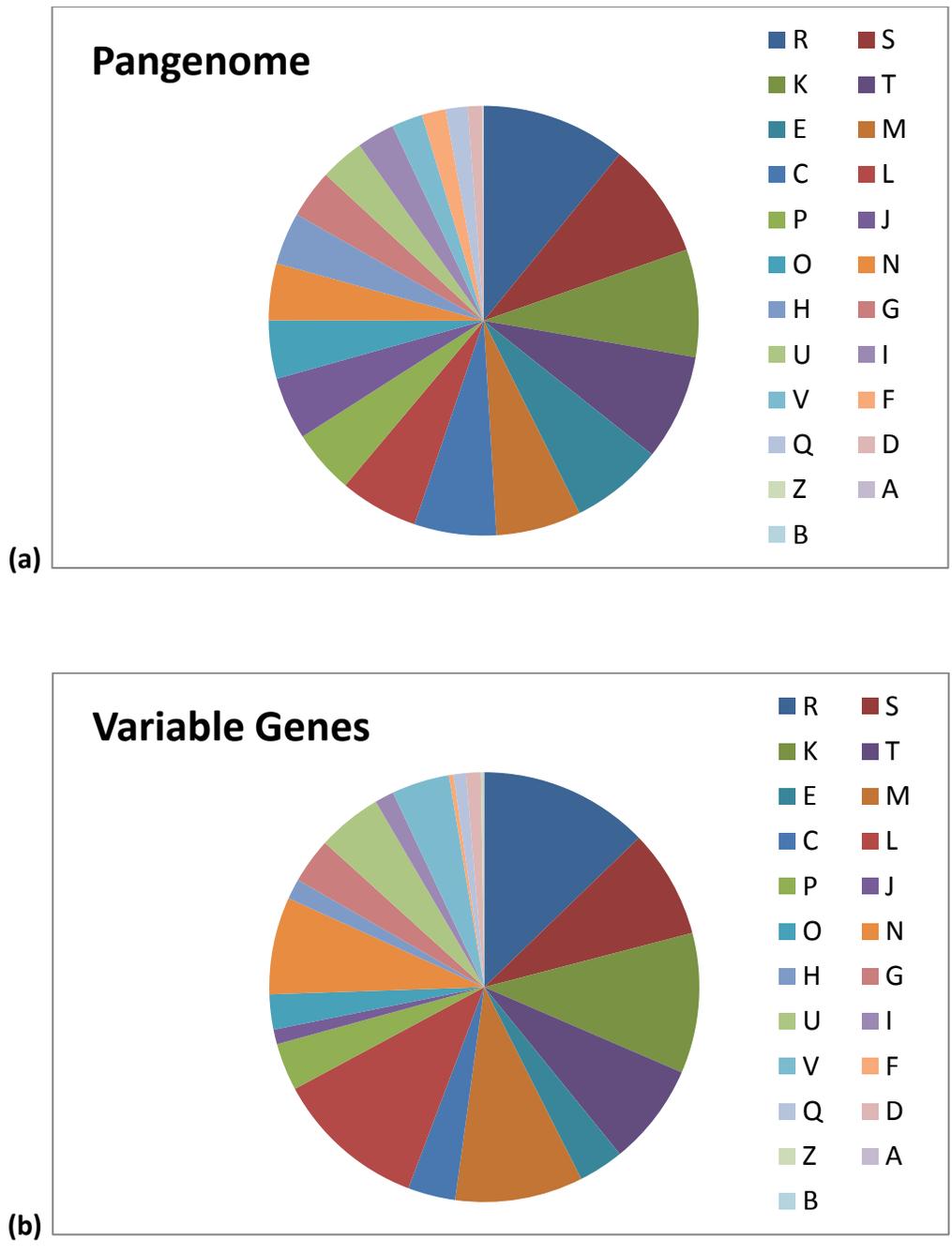


Figure 3.7.

Figure 3.7. (Cont'd)

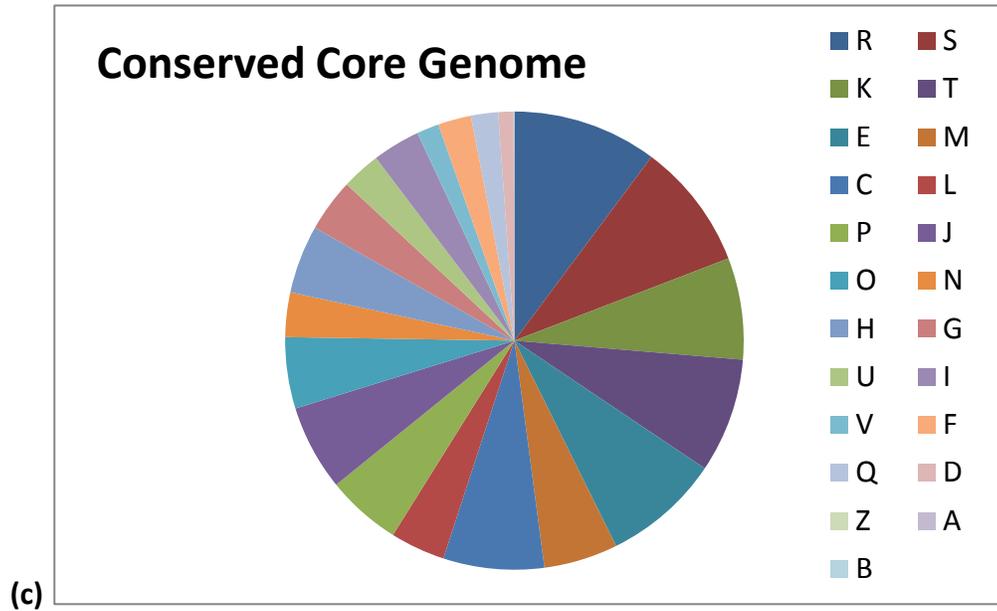


Figure 3.7. Distribution of genes in different COG functional categories in (a) pangenome (b) genes whose presence varies among the 46 *S.baltica* strains (c) genes that are present in greater than 90% of the strains. Genes without COG annotation were not included.

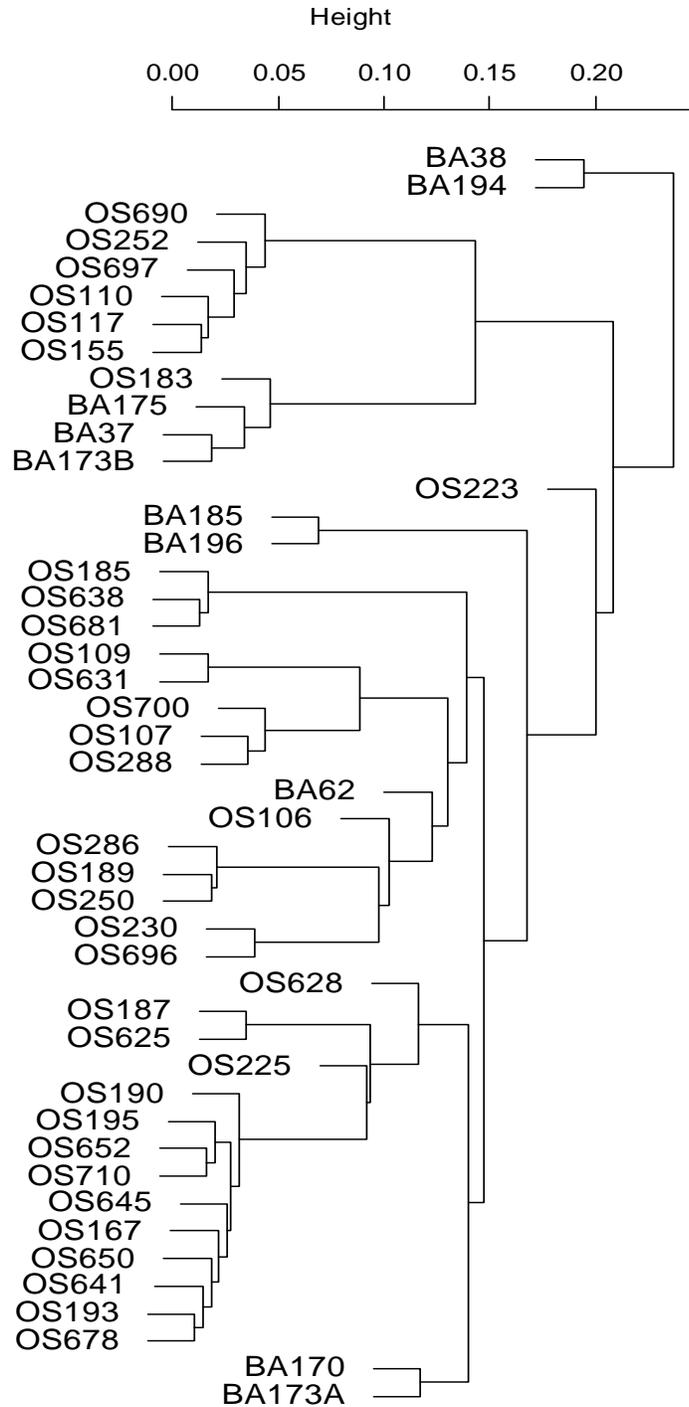
**Table 3.4. Description of COG Categories**

<b>Code</b>	<b>Description</b>
<b>Information storage and processing</b>	
J	Translation, ribosomal structure and biogenesis
A	RNA processing and modification
K	Transcription
L	Replication, recombination and repair
B	Chromatin structure and dynamics
<b>Cellular processes and signaling</b>	
D	Cell cycle control, cell division, chromosome partitioning
Y	Nuclear structure
V	Defense mechanisms
T	Signal transduction mechanisms
M	Cell wall/membrane/envelope biogenesis
N	Cell motility
Z	Cytoskeleton
W	Extracellular structures
U	Intracellular trafficking, secretion, and vesicular transport
O	Posttranslational modification, protein turnover, chaperones
<b>Metabolism</b>	
C	Energy production and conversion
G	Carbohydrate transport and metabolism
E	Amino acid transport and metabolism
F	Nucleotide transport and metabolism
H	Coenzyme transport and metabolism
I	Lipid transport and metabolism
P	Inorganic ion transport and metabolism
Q	Secondary metabolites biosynthesis, transport and catabolism
<b>Poorly characterized</b>	
R	General function prediction only
S	Function unknown

**Similar clustering patterns with MLST phylogeny.** In order to study the genomic relatedness among these *Shewanella baltica* isolates, a dendrogram was constructed based upon the gene presence/absence matrix from the CGH dataset [Figure 3.8], and was compared with the sequence based phylogeny from MLST analysis [Figure 2.2] and *gyrB* gene characterization [Figure 2.1]. Pairwise mean squared error between distance matrices from the *gyrB*, MLST, and the CGH datasets showed higher similarity between the MLST and CGH datasets [Table 3.5]. Specifically, ten out of eleven lineages from the MLST phylogeny were confirmed by the CGH clustering, suggesting high agreement between these two independent phylogenetic methods. This observation supported our hypothesis that the rate of gene flow is positively correlated with the genetic relatedness among strains. The higher agreement observed between the MLST phylogeny and CGH dendrogram also indicates the advantage of the MLST analysis over the single gene-based genotyping method.

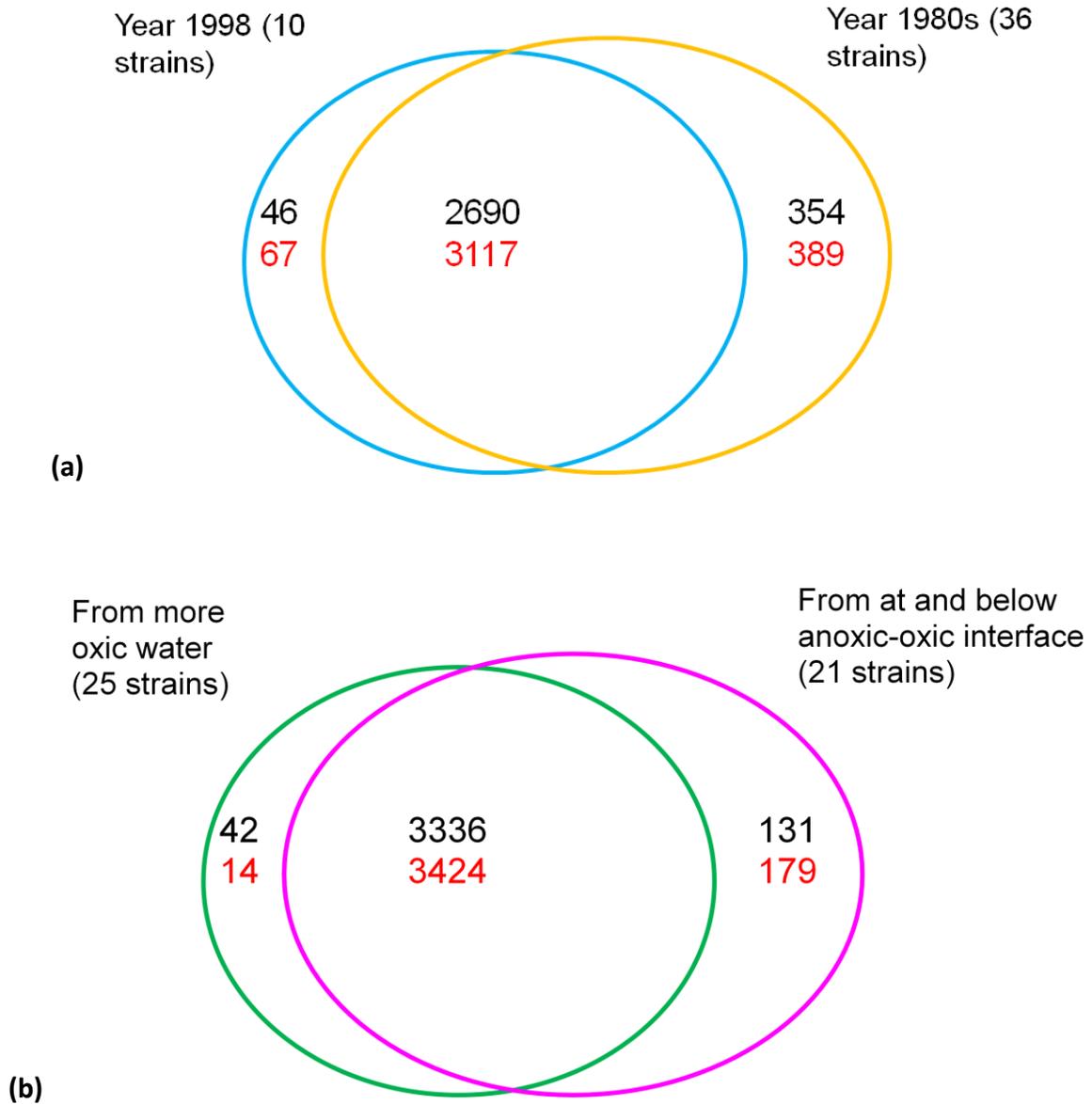
**Table 3.5. Correlation among Three Distance Matrices for Phylogenetic Inference**

	MLST	<i>gyrB</i>	CGH
MLST	-	-	-
<i>gyrB</i>	0.669	-	-
CGH	0.757	0.647	-

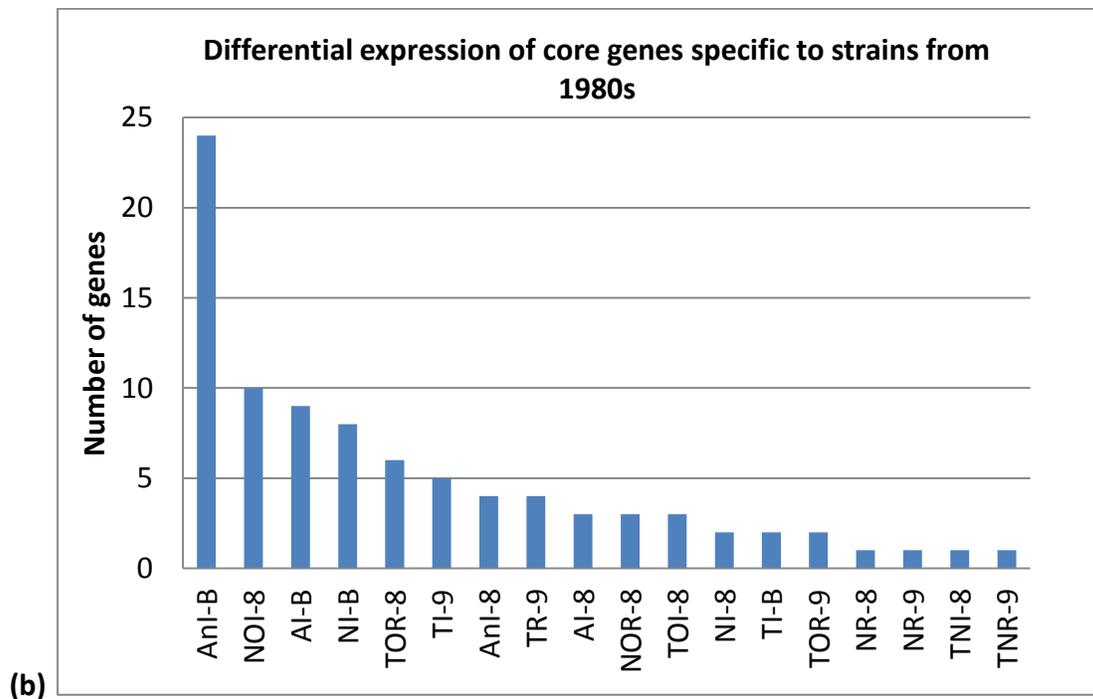
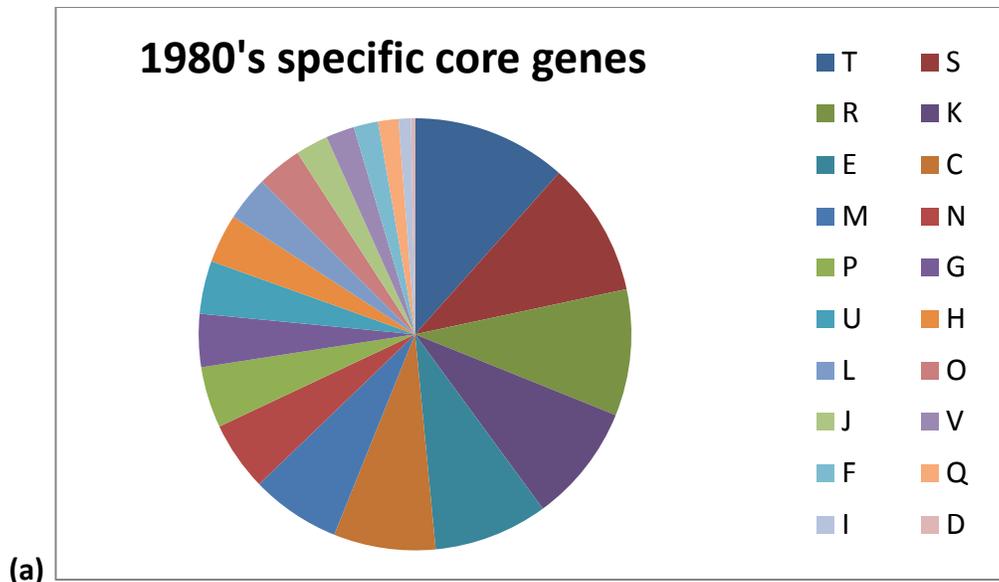


**Figure 3.8. Hierarchical clustering of the CGH profile.** Gene presence and absence information extracted from the CGH profile was used for constructing the distance matrix using the Jaccard index. Hierarchical clustering was performed using the average linkage agglomeration method in R software version 2.9.2.

**Genes related to specialization over time.** In order to track the changes in the conserved core genome over time, we analyzed the core genes of strains isolated from the 1980s versus the strains from 1998 [Figure 3.9a]. Although the 1980's strains and 1998 strains shared the great majority of genes, over 300 genes were apparently lost in the 1998 strains while 46 to 67 genes (computed based on conservation in greater than 90% or 85% of strains, respectively) were likely fixed over this time, most of which encoded hypothetical proteins or were phage related. Within the 389 genes specifically shared among 1980s strains, 99 encoded hypothetical proteins, and among those successfully annotated and better understood, the most enriched categories were those related to signal transduction (T), transcription (K), amino acid transport and metabolism (E), energy production and conversion (C) and cell wall biogenesis (M) [Figure 3.10a]. Furthermore, 89 out of these 389 genes were differentially expressed in either OS185 or OS195 under different redox conditions [Figure 3.10b] [Table 3.6]. Specifically, 24 of these genes were induced under anaerobic growth condition in both OS185 and OS195 strains; nine were induced under aerobic condition and eight were induced during nitrate respiration in both strains. Loss of these genes in 1998 strains could imply a shift in selective pressure over time. But it needs attention that other factors such as genetic drift or mobile element transfer could also lead to gene loss during genome evolution.



**Figure 3.9. Comparison of conserved core genes of (a) strains isolated from the 1980s and 1998 (b) strains recovered from at or below the anoxic-oxic interface and from more oxic water sample.** Note that different levels of conservation were used for defining core genes. The numbers in black are based on greater than 90% conservation while the numbers in red are based on greater than 85% conservation among 46 *S. baltica* strains.



**Figure 3.10. (a) COG functional categories of annotated core genes of strains isolated from the 1980s. (b) Differential expression patterns of core genes of strains from the 1980s in OS185 and OS195. Core genes here were defined as genes shared among over 85% strains from the 1980s or 1998.**

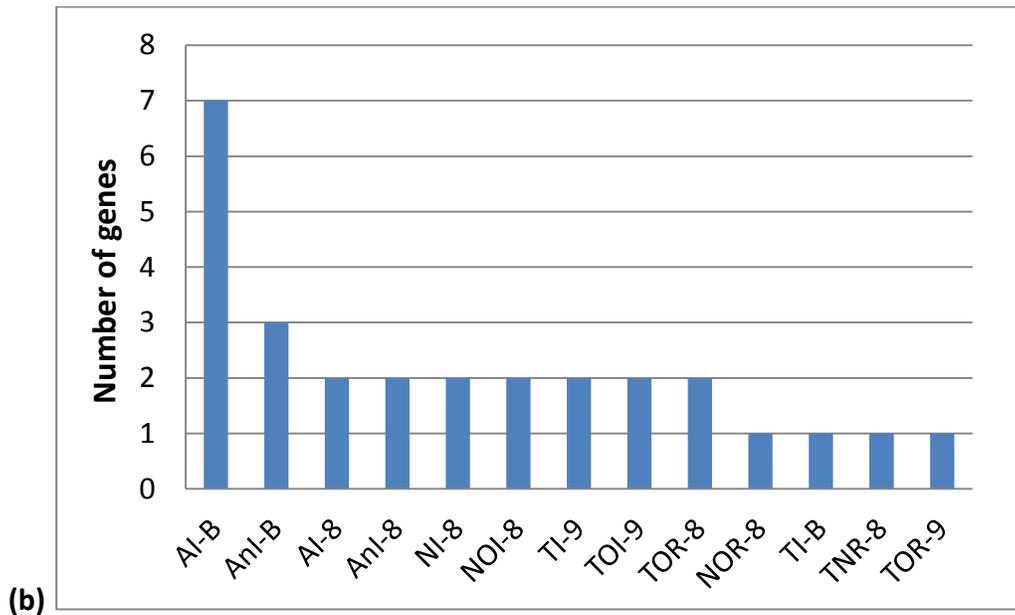
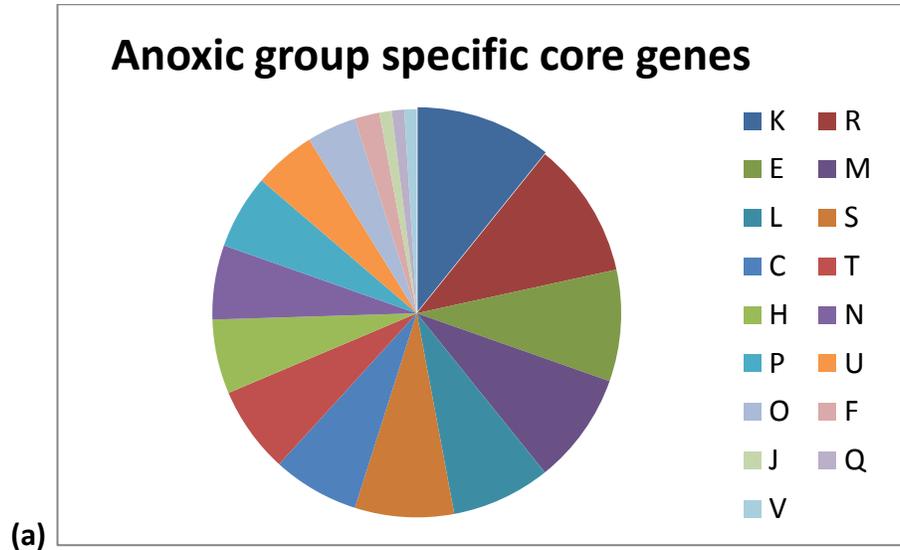
**Table 3.6. Types of Differential Expression**

<b>Code</b>	<b>Description</b>
AI	aerobic induced (higher in O2 relative to nitrate or thiosulfate)
AnI	anaerobic induced (higher in nitrate and thiosulfate relative to O2)
NI	nitrate induced (higher in nitrate than in thiosulfate or O2)
NR	nitrate repressed (higher in O2 and thiosulfate relative to nitrate)
TI	thiosulfate induced (higher in thiosulfate relative to O2 and nitrate)
TR	thiosulfate repressed (higher in O2 and nitrate relative to thiosulfate)
NOI	nitrate over oxygen induced (higher in nitrate)
NOR	nitrate over oxygen repressed (higher in O2)
TOI	thiosulfate over oxygen induced (higher in thiosulfate)
TOR	thiosulfate over oxygen repressed (higher in O2)
TNI	thiosulfate over nitrate induced (higher in thiosulfate)
TNR	thiosulfate over nitrate repressed (higher in nitrate)
-8	Only in OS185
-9	Only in OS195
-B	In both strains

**Genes related to specialization across the redox gradient.** Core genes were extracted from strains isolated at and below the oxic-anoxic interface (anoxic group) as well as from strains isolated from above the oxic-anoxic interface (microaerophilic group). Comparing to the core gene analysis between strains from 1980s and 1998, the microaerophilic and anoxic group shared a larger number of core genes and possessed fewer group-specific genes [Figure 3.9b]. Given that relatively fewer genes (14 under 90% conservation threshold, or 42 under 85% threshold) were specifically shared within the microaerophilic group, specialization to the microaerophilic condition at the level of genomic content was not well supported. In contrast, a larger number of genes were shared among the anoxic group. Furthermore, when the threshold of computing core genes was relaxed from 90% to 85%, the number of core genes specific to the anoxic group increased drastically, probably due to the clustering of strains from anoxic water from more distantly related groups.

Using the 85% threshold, 179 core genes specific to the anoxic group were revealed, with 87 of them encoding hypothetical proteins. Among the genes successfully annotated, many were associated with transcription regulation (K, but mostly mobile genetic elements related), cell wall biogenesis (M), as well as amino acid transport and metabolism (E) [Figure 3.11a]. It is worth noting that a number of dehydrogenases and oxidoreductase (C), as well as membrane ion receptors and transporters (P) were also found in the core gene set of these strains, which suggests they may play essential roles in specialization for the anoxic zone. However, only a small number (28) of genes in the anoxic set showed differential expression in either OS185 or OS195. Seven of these 28 genes were up-regulated under aerobic growth conditions, while only three were induced under anaerobic conditions in both strains [Figure

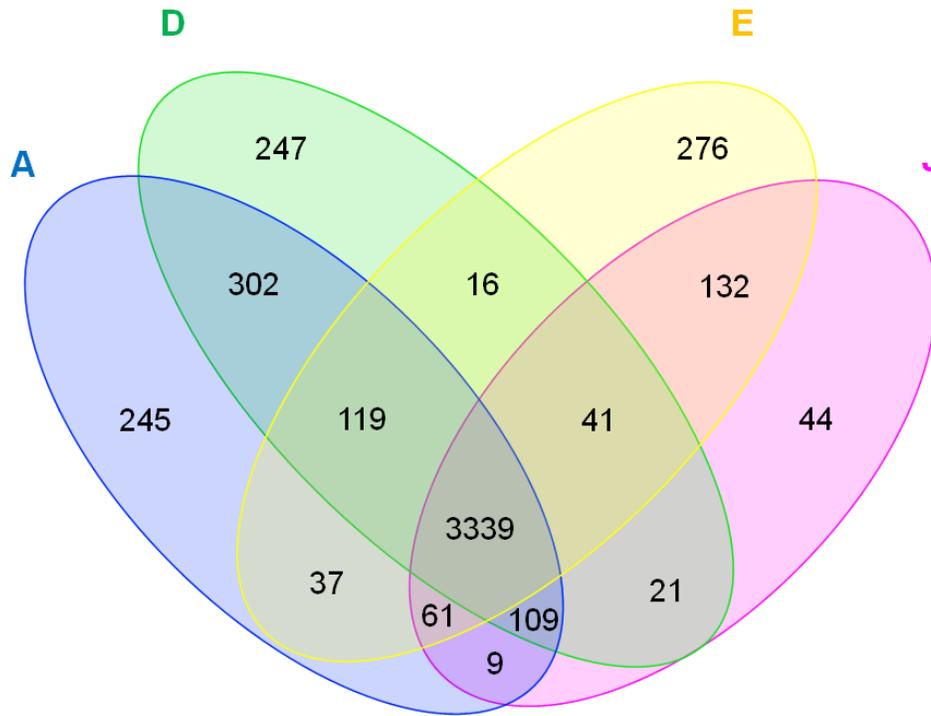
3.11b].



**Figure 3.11. (a) COG functional categories of annotated core genes of strains isolated from anoxic waters. (b) Differential expression patterns of core genes of strains from anoxic waters in OS185 and OS195. Core genes here were defined as genes shared among over 85% strains from the anoxic or microaerophilic group.**

### **Conserved gene core in specific lineages.**

Based upon the strong correlation with the MLST phylogeny and the CGH clustering pattern, tracking the conserved gene core of individual lineages was possible and may provide detailed information on the genetic basis for specialization into different environmental niches by strains from different clusters. Expression profiling of the strains OS185 and OS195 under different redox conditions also provided additional support for whether genes differentially distributed among the *S. baltica* population were likely functionally important. Pairwise comparison of the core gene composition of strains in Clade A, D, E and J was performed [Figure 3.12], given that these four clades potentially represent strains occupying different niches. Clade A strains were mostly isolated from anoxic water or near the anoxic-oxic interface, while Clade E originated from more oxic waters. Clade D was thought to have gone through extensive genetic exchange with strains in Clade A. Furthermore, three out of four strains in Clade J were isolated from 1998 rather than the 1980s. Thus, investigating genetic differences in these clades may provide useful information on the impact of environments on micro-evolution of the *S. baltica* genomes.



**Figure 3.12. Core genome comparison of Clade A, D, E and J.** Core genes were calculated based on their presence in more than 85% strains in each clade.

**Genes specific to strains in Clade A and D.** Among the four clades, A and D exclusively shared a comparatively large number of genes - 302- compared to any other pairs of clades (from 9 genes exclusively shared between A and J, to 121 shared between E and J). While the majority of these genes were either mobile element related or encoded hypothetical proteins, a number of proteins classified in the L (22), K (21), P (15), M (15) COG categories were also revealed. The genes shared exclusively between Cluster A and D included several membrane transporters (Shew185\_0604-0608, Shew185\_3340-3345, Shew185\_3346-3347, Shew185\_1621, 1623), transcriptional regulators (Shew185\_4192- Shew185\_4198), CRISPR-associated proteins (clustered regularly interspaced short palindromic repeats, Shew185\_3236-3241), aminosynthases and transferases (Shew185\_2970-2979, Shew185\_2980-2982), as well as an F-

plasmid associated conjugative transfer system (Shew185\_4429-4448). Expression of a number of putative regulators and oxidoreductases (Shew185\_1391-1395) was induced during thiosulfate respiration by both OS185 and OS195 strains, indicating functional importance of these genes.

Strains in Clade A and Clade D also shared a likely nitrate reductase operon (Sbal195\_3991-4002 in OS195, or Shew185\_3866-3877 in OS185, including *NosYFDL*, *NrfDCG*, and periplasmid cytochrome C), which was also found in Clade B. It was shown to be up-regulated in *S. oneidensis* MR-1 under anaerobic growth condition using thiosulfate as electron acceptor, thus was considered to be related with thiosulfate reduction. In agreement with this prediction, genes in this operon were induced during thiosulfate respiration for both OS185 and OS195 strains, and it was also induced by nitrate respiration by OS195.

Furthermore, looking at a broader range of strains, some of the genes shared between these two clades were exclusively missing from both the OS155 clade (E) and OS183 clade (J), including a number of genes potentially important for anaerobic respiration such as anaerobic dehydrogenases (putative nitrate-inducible Se-containing formate dehydrogenase as in *S. oneidensis* MR-1, Shew185\_0099-0108), NADH oxidoreductase (Shew185\_0454-0456), sulfate transporter associated cytochrome *c* family proteins (strongly induced by thiosulfate in *S. MR-1*, Shew185\_0458, Shew185\_0460), as well as other proteins involved in signal transduction and regulation (Shew185\_4192-4198). Contrasting to the isolation of, especially, most Clade A strains from more anoxic waters (120-140 m), strains of the OS155 lineage (Clade E) and OS183 lineage (Clade J) originated from more oxic regions (80-130 m). Together with the genomic

signatures revealed from these strains, these two lineages appeared less adapted to anoxic water environments.

**Genes specific to Clade E.** Among the 276 genes specific to Cluster E, the biggest majority were hypothetical proteins or mobile elements related, such as prophage encoded genes and other insertion segments (Sbal\_1270- Sbal\_1306, Sbal\_4384- Sbal\_4411). Other Clade E specific genes and operons include those encoding proteins involved in phenazine synthesis (Sbal\_1937- Sbal\_1938), polysaccharide biosynthesis (Sbal\_2876-2893), as well as a few oxidoreductases (Sbal\_0621-0627) likely involved in sugar metabolism, etc.

**Genes specific to Clade A.** Although extensive genetic exchange is proposed to have occurred between the OS185 and OS195 lineages (14), strains in the OS195 cluster (Clade A) possessed 352 specific genes that were not shared with the OS185 lineage (Clade D), and 245 of them were exclusively found in OS195 lineage compared with more strains in Clade E and J. These Clade A-exclusive genes include the insertion of a P2 phage (Sbal195\_2888-2917), an aerobic-respiration induced polysaccharide synthesis operon (Sbal195\_3017-3035), as well as a number of cytochromes, transcriptional regulators and oxidoreductases that are likely related to anaerobic respiration. Some of the cytochromes and transcriptional regulators (Sbal195\_0256, 0257, 0258), as well as a Type I anaerobic DMSO reductase operon (*dmsEFABGH*, Sbal195\_2225-2230), were induced under anaerobic growth condition by OS195, indicating their contribution to specialization of the OS195 lineage in the anoxic environment. Interestingly, The DMSO reductase operon was flanked by integrase proteins and phage

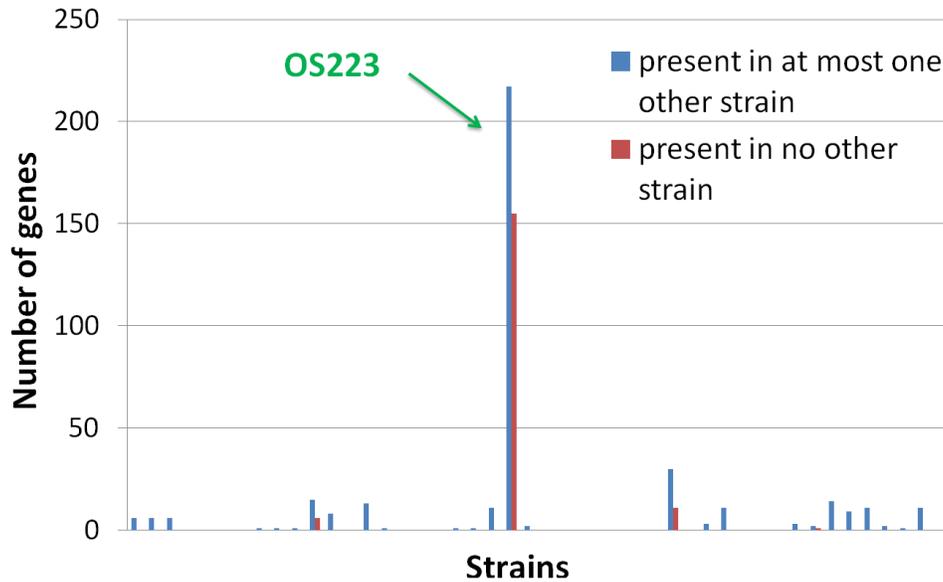
elements, reflecting a potentially important role of marine phages in environmental adaptation of *S. baltica* strains.

**Genes specific to Clade D.** Compared to strains in Clade A, E and J, strains in Clade D specifically shared 247 core genes, 64 of which were differentially expressed by either strain OS195 or OS185 under various respiration conditions, indicating that Clade D strains were likely well adapted to its own redox environment. A polysaccharide synthesis operon (Shew185\_2888-2903) shared exclusively among Clade D strains and OS631 were up-regulated under aerobic growth in OS185; while a few genes and regulators likely related with anaerobic respiration, i.e., an operon potentially related to chemotaxis (Shew185\_2544-2551) and cytochrome c protein-containing operon (Shew185\_2534-2538) were found to be up-regulated or at least partially up-regulated under anaerobic condition in OS185. Genes potentially involved in phenazine biosynthesis (Shew185\_2181-2184), cobyrinic acid a,c-diamide synthase (Shew185\_2564-2567) and galactose metabolism (Shew185\_3331-3339) were also found. Particularly, Clade D strains exclusively shared a number of genes involved in DNA restriction-modification (Shew185\_1976-1981, Shew185\_2188-2198), which may be related to specific phage activities.

**Core genes in Clade B strains.** Strains in Clade B were closely related to those in Clade A based on both MLST phylogeny and CGH analysis. No Clade B-specific genes were found compared with those in Clade A, which was probably due to the lack of reference genomes from Clade B strains for the array design. Compared with core genes revealed in Clade A, 243 genes were seemingly lost in Clade B, including not only some of the Clade A-specific genes described above,

such as the anaerobic DMSO reduction operon, the CRISPR protein operon, polysaccharide synthesis and the P2 phage insertion, but also a number of membrane transporters and signal transduction proteins (Shew185\_0604-0608, Shew185\_4231-4240), as well as a potential pathogenic island that contained a Type III secretion system (Sbal195\_2248-2281).

**High versatility in carbon source utilization was reflected from the OS223 genome.** While comparing gene content of single strains, the strain OS223 has the largest number of strain-exclusive genes [Figure 3.13], the greatest majority of which encoded either hypothetical proteins or were phage related. In addition, OS223 possessed a unique polysaccharide biosynthesis operon (Sbal223\_1470-1487), as well as a short-chain dehydrogenase operon (Sbal223\_2173-2177). OS223 also possessed a number of genes and operons related to sugar metabolism. For example, it shared an operon involving metabolism of multiple sugar substrates (Sbal223\_2138-2162) exclusively with Clade F and BA170, and a potential arabinose metabolism operon (Sbal223\_3608-3611) with a number of strains in other clusters. These observations were also supported by the high-resolution Biolog profiling of the four reference strains, where OS223 showed extraordinary versatility in carbon source utilization compared to the other three tested strains. It has been proposed that although OS223 was isolated from the same depth as OS185, it might have occupied a slightly different ecological niche in the water column relative to OS185 or OS195 (14), for example, being associated with sinking particles as opposed to being planktonic (or vice versa) or being transient or allochthonous at the 120–140 m depth (see also Discussion below). In agreement with the latter hypothesis, only one other OS223-like strain was recovered in the 1986 or 1987 isolation efforts.



**Figure 3.13. Number of strain-specific genes revealed from CGH.** Strains not shown had no strain-specific genes according to the threshold used.

**Horizontal gene transfer and genomic islands.** Compared to higher organisms, in which genomic material is transferred to offspring exclusively from parental cells through reproduction, several different mechanisms are known in bacteria that mediate lateral gene transfer through other individuals including DNA transduction, transformation, and conjugation (15). One of the most important means of innovation in bacteria is to acquire genetic material from other cells or organisms, some that are even distantly related, by horizontal gene transfer (HGT) (16) (17) (18) (19). In aquatic environments, phages or viruses are of special importance in mediating HGT among different organisms (20) (21) (22). In cases where functionally associated genes and operons are transferred together and incorporated into the receptor cell's DNA, genomic islands can be formed (17). Being either silent or actively transcribed after incorporation, these horizontally transferred genes are subsequently subjected to a series of

processes including genetic drift or natural selection, and those with adaptive or survival advantages may become fixed through selection sweeps (23).

Comparative genomics has revealed the ecological importance of a number of genomic islands. One of the most common examples is the prevalence of pathogenicity islands in clinical isolates, such as the LEE (locus of enterocyte effacement) islands in enteropathogenic *Escherichia coli* (EPEC) strains (24), and *cci* in *Burkholderia cenocepacia* (25). A number of genomic islands in *Shewanella* were identified from a comparative genomic study of ten *Shewanella* genomes across different species (13). These genomic islands together with the insertion elements dominated the differences in gene content among genomes of the same species, and decreased where more distantly related strains were compared. In a more recent study that compared the genome sequences of the four reference *S. baltica* strains (14), mosaic genomic structures were revealed, and genomic islands potentially associated with ecological fitness were identified such as those with genes encoding anaerobic respiratory complexes and associated transport and cytochromes, consistent with the low redox status of their habitat.

CGH effectively characterized patterns of gene acquisition and loss among the larger collection of *S. baltica* strains studies here, and further supported the mosaic genome structure of *S. baltica* genomes. These genomes were heavily impacted with phages and other mobile elements related insertions, and the so-formed genomic islands dominated the variable part of the *S. baltica* genomes. Therefore, horizontal acquisition of genomic islands was possibly the major means of harboring new functions by these organisms, while acquired genes were subsequently subjected to natural selection through which ecologically important genes were prone to be fixed in the population. Among the genes that were differentially distributed

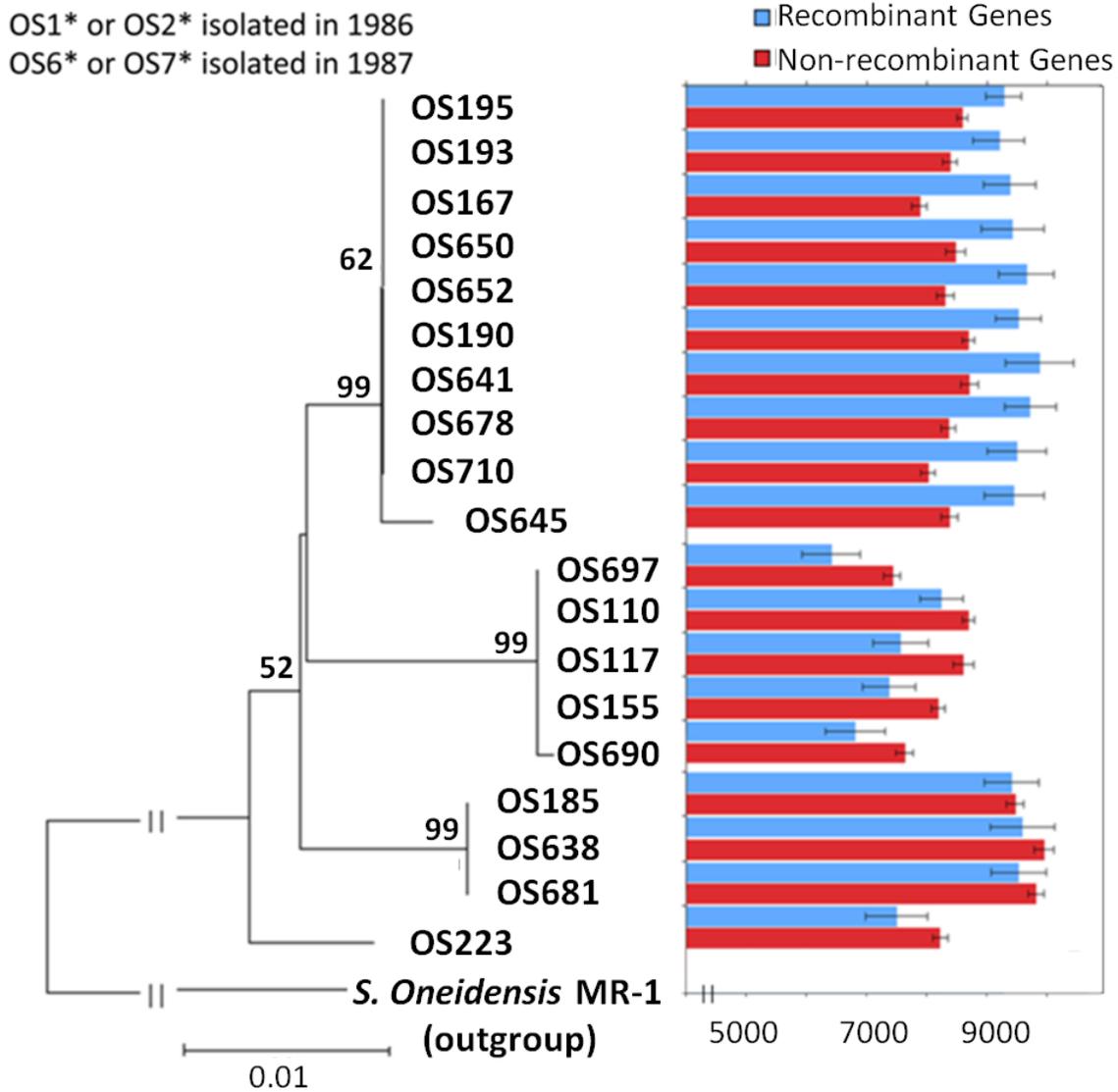
among these *S. baltica* strains, a significant fraction were potentially ecologically important, including those associated with anaerobic life style, nutrient transport and metabolism, phage defense, as well as those involved in signaling and transcriptional regulation, providing further evidences for the importance of HGT in the micro-evolution of these *S. baltica* genomes.

**Recombination among specific lineages was supported.** Recombination plays important roles during sexual reproduction of eukaryotic organisms, especially in facilitating chromosomal crossover in the offspring. Recombination can also occur during DNA damage repair in bacteria (26). Due to homologous recombination, the rate of recombination is largely restricted by sequence similarity between donor and receptor DNA (27), hence this process occurs mostly among genetically closely related strains such as within the *S. baltica* population.

Recombination is thought to play an important role in bacterial genome evolution by driving species cohesion (28) (29). It has been proposed that in cases where recombination rates within bacterial lineages are relatively high, convergent evolution is more likely to take place (30). Thus, analysis of recombination may facilitate understanding of the evolutionary trajectory of a studied population. In the case of *S. baltica*, strains OS195 and OS185 have recently undergone extensive recombination (14). In a previous comparative genomic study of the four reference strains, 580 non-core genes were found to be shared between OS185 and OS195, while ~350 of them plus an additional ~10% of their core genes showed 99.5-100% nucleotide identity, contrasting sharply with ~97% identity for the rest of the genes in the genome and <3% of high identity core genes (99.5-100%) among the remaining pairs of genomes. In addition, the density of the *S. baltica* population was estimated to be about  $10^3$

cells per mL of seawater in both sampling years 1986 and 1987 based on most probable number estimates using several liquid media (31) . Therefore, the genetic exchange patterns revealed by the sequenced genomes may apply to a larger collection of strains and may be persistent over the time (1986–1987) in the natural *S. baltica* population.

In order to seek support for this hypothesis, CGH hybridization signals on probes targeting the recombined and non-recombined gene pools from other strains in the same lineages as the four reference strains were extracted and compared. Strains in the OS195 cluster (n=10) had consistently greater hybridization signal to probes that corresponded to recombined than non-recombined genes [figure 3.14]. Thus the larger sampling of strains in the OS195 lineage showed extensive genetic exchange was likely to have occurred between this lineage and the OS185 lineage. It also appeared that the isolated OS195 strain, which probably had migrated to deeper waters after the recombination event(s) with the OS185 lineage, had presumably adapted further to the more anoxic environment of the deeper waters. For instance, its genome encoded additional genomic islands for anaerobic lifestyle, such as a dimethyl sulfoxide (DMSO) reductase containing island. Furthermore, OS195-like strains were more abundant and consistently recovered from this depth in both sampling years suggesting its adaptation to this redox environment.



**Figure 3.14. The patterns of genetic exchange apply to a larger collection of *S. baltica* strains.** Hybridization signal of all strains in the same lineages as the four sequenced strains (left panel) were analyzed. The average normalized signal of all probes that corresponded to non-recombined OS185 core (red) vs. recombined OS185 core genes with OS195 (blue) are shown (left panel). Error bars represent one standard deviation from the mean. Note that the latter probes show consistently greater hybridization signal only in the OS195 like strains, in agreement with the preferential genetic exchange between the OS185 and OS195 lineages.

However, to what extent the patterns of genetic exchange observed between OS195 and OS185 and their sister strains apply to other natural sub-populations of *S. baltica* in the Baltic Sea and what accounts for the genetic flow among those specific strains remain unknown. To address these issues, in situ genomic studies (for example, metagenomics) and sampling of the natural populations over time will be required.

Although microarrays are sometimes limited in terms of specificity and sensitivity, for example, being unable to recognize pseudogenes due to frame-shift or single nucleotide mutations, they proved an efficient, cost effective, sufficiently resolving and highly parallel method for determination of gene contents and levels of gene expression in the many strains important to this study. The CGH data provided information on how genes were distributed among the *S. baltica* genomes, and the transcriptome revealed which genes were sensitive to changes in redox conditions. Furthermore, these data partially overcame the problem of lack of annotation of *S. baltica* genes, because genes differentially expressed were more likely to be functionally important under the particular redox conditions, and hence allowed us to better evaluate genes differentially distributed among lineages and niche environments.

## REFERENCES

## References

1. Willenbrock H, *et al.* (2006) Design of a Seven-Genome *Escherichia coli* Microarray for Comparative Genomic Profiling. *Journal of Bacteriology* 188(22):7713-7721.
2. Dobrindt U, *et al.* (2003) Analysis of Genome Plasticity in Pathogenic and Commensal *Escherichia coli* Isolates by Use of DNA Arrays. *Journal of Bacteriology* 185(6):1831-1840.
3. Oh S, Yoder-Himes DR, Tiedje J, Park J, & Konstantinidis KT (2010) Evaluating the Performance of Oligonucleotide Microarrays for Bacterial Strains with Increasing Genetic Divergence from the Reference Strain. *Applied and Environmental Microbiology* 76(9):2980-2988.
4. Otto M, Yoder-Himes DR, Konstantinidis KT, & Tiedje JM (2010) Identification of Potential Therapeutic Targets for *Burkholderia cenocepacia* by Comparative Transcriptomics. *PLoS ONE* 5(1):e8724.
5. Murphy D (2002) Gene expression studies using microarrays: principles, problems, and prospects. *Advances in Physiology Education* 26(4):256–270.
6. Beliaev AS, *et al.* (2005) Global Transcriptome Analysis of *Shewanella oneidensis* MR-1 Exposed to Different Terminal Electron Acceptors. *Journal of Bacteriology* 187(20):7138-7145.
7. He Z, *et al.* (2007) GeoChip: a comprehensive microarray for investigating biogeochemical, ecological and environmental processes. *The ISME Journal* 1(1):67-77.
8. Rich VI, Konstantinidis K, & DeLong EF (2008) Design and testing of 'genome-proxy' microarrays to profile marine microbial communities. *Environmental Microbiology* 10(2):506-521.
9. Chengwei Luo STW, David M. Gordon, Michael Feldgarden, James M. Tiedje, and Konstantinos T. Konstantinidis (2011) Genome sequencing of environmental *Escherichia coli* expands understanding of the ecology and speciation of the model bacterial species. *Proceedings of the National Academy of Sciences* 108(17):6.
10. Richards E, Reichardt M, & Rogers S (1994) Preparation of Genomic DNA from Plant Tissue. *Current Protocols in Molecular Biology* eds Ausubel M, Brent R, Kingston RE, Moore DD, Seidman JG, John A. Smith JA, & Struhl K (John Wiley, Hoboken, NJ), pp 2.3.1-2.3.7.
11. Cruz-Garcia C, Murray AE, Klappenbach JA, Stewart V, & Tiedje JM (2007) Respiratory nitrate ammonification by *Shewanella oneidensis* MR-1. *J Bacteriol* 189(2):656-662.
12. Tusher VG, Tibshirani R, & Chu G (2001) Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A* 98(9):5116-5121.

13. Konstantinidis KT, *et al.* (2009) Comparative systems biology across an evolutionary gradient within the *Shewanella* genus. *Proceedings of the National Academy of Sciences* 106(37):15909-15914.
14. Caro-Quintero A, *et al.* (2010) Unprecedented levels of horizontal gene transfer among spatially co-occurring *Shewanella* bacteria from the Baltic Sea. *The ISME Journal* 5:131–114.
15. Wozniak RAF & Waldor MK (2010) Integrative and conjugative elements: mosaic mobile genetic elements enabling dynamic lateral gene flow. *Nature Reviews Microbiology* 8(8):552-563.
16. Thomas CM & Nielsen KM (2005) Mechanisms of, and Barriers to, Horizontal Gene Transfer between Bacteria. *Nature Reviews Microbiology* 3(9):711-721.
17. Dobrindt U, Hochhut B, Hentschel U, & Hacker J (2004) Genomic islands in pathogenic and environmental microorganisms. *Nature Reviews Microbiology* 2(5):414-424.
18. Frost LS, Leplae R, Summers AO, & Toussaint A (2005) Mobile genetic elements: the agents of open source evolution. *Nature Reviews Microbiology* 3(9):722-732.
19. Gogarten JP & Townsend JP (2005) Horizontal gene transfer, genome innovation and evolution. *Nature Reviews Microbiology* 3(9):679-687.
20. Weinbauer MG & Rassoulzadegan F (2003) Are viruses driving microbial diversification and diversity? *Environmental Microbiology* 6(1):1-11.
21. Paul JH (2008) Prophages in marine bacteria: dangerous molecular time bombs or the key to survival in the seas? *The ISME Journal* 2(6):579-589.
22. Corina PD Brussaard SWW, Frede Thingstad, Markus G Weinbauer, Gunnar Bratbak, Mikal Heldal, Susan A Kimmance, Mathias Middelboe, Keizo Nagasaki, John H Paul, Declan C Schroeder, Curtis A Suttle, Dolores Vaquer and K Eric Wommack (2008) Global-scale processes with a nanoscale drive: the role of marine viruses. *International Society for Microbial Ecology* 2:575–578.
23. Cohan FM (2001) Bacterial Species and Speciation. *Syst. Biol.* 50(4):12.
24. Nisan I WC, Hanski E, Rosenshine I. (1998) Interaction of enteropathogenic *Escherichia coli* with host epithelial cells. *Folia Microbiol (Praha)* 43(3):247-252.
25. Malott RJ, Baldwin A, Mahenthiralingam E, & Sokol PA (2005) Characterization of the *ccII*R Quorum-Sensing System in *Burkholderia cenocepacia*. *Infection and Immunity* 73(8):4982-4992.

26. SC Kowalczykowski DD, AK Eggleston, SD Lauder, WM Rehrauer (1994) Biochemistry of homologous recombination in *Escherichia coli*. *Microbiology and Molecular Biology Reviews* 58(3):401–465.
27. Fraser C, Hanage WP, & Spratt BG (2007) Recombination and the Nature of Bacterial Speciation. *Science* 315(5811):476-480.
28. Vos M (2009) Why do bacteria engage in homologous recombination? *Trends in Microbiology* 17(6):226-232.
29. Brown EW, LeClerc JE, Kotewicz ML, & Cebula TA (2001) Three R's of bacterial evolution: How replication, repair, and recombination frame the origin of species. *Environmental and Molecular Mutagenesis* 38(2-3):248-260.
30. Hanage WP, Fraser C, & Spratt BG (2006) The impact of homologous recombination on the generation of diversity in bacteria. *Journal of Theoretical Biology* 239(2):210-219.
31. Ziemke F, Brettar I, & Höfle MG (1997) Stability and diversity of the genetic structure of a *Shewanella putrefaciens* population in the water column of the central Baltic. *Aquatic Microbial Ecology* 13:63-74.

## **CHAPTER IV**

### **Characterization of Carbon Source Utilization and Respiratory Capabilities of the *Shewanella baltica* Strains**

The Biolog dataset was a contribution from Ingrid Brettar.

## Abstract

*Shewanellae* are known for their remarkable versatility in coupling organic matter turnover with a wide range of electron acceptors. In this chapter, we provide a high throughput approach to effectively screen the metabolism of 95 carbon sources by 46 *S. baltica* isolates obtained from different depths of the Baltic Sea over a 20 year period using the Biolog phenotypic microarray system. Positive phenotypes were obtained for 76 substrates, with 11 being metabolized by all isolates. The carbon source metabolism patterns correlated with genetic relatedness, and were ecologically associated with the time of isolation, i.e. strains isolated in 1998 were able to utilize significantly more carbon sources compared to those isolated in the 1980's. Differential respiratory capabilities of several potential alternative electron acceptors were observed in strains isolated from zones in the water column of different redox status. Strains capable of reducing dimethylsulfoxide (DMSO) were mostly from anoxic water, while the inability to reduce thiosulfate and trimethylamine-N-oxide (TMAO) was limited to a cluster where most strains were from more oxic water. Furthermore, although these strains were characterized as respiratory denitrifiers, denitrification activity was not observed in any isolate; but nitrate reduction via ammonification occurred instead. Our results indicate that phenotypic evaluation of the physiology of populations can provide a better understanding on the ecological framework of a species.

## Introduction

Bacteria of the *Shewanella* genus are facultative anaerobic organisms that are known for their versatility in coupling decomposition of organic matter with reduction of a variety of terminal electron acceptors. Species in this genus were reported to be able to utilize organic as well as inorganic electron acceptors such as nitrate, thiosulfate, uranium [U(VI)], chromium [Cr(VI)], technetium, neptunium, selenite, and nitroaromatic compounds(1) (2). Thus, *Shewanellae* are of great potential for bioremediation of many environmental contaminants and have also been used in a number of microbial fuel cell studies (3). *Shewanellae* are commonly found in complex ecosystems such as fresh and marine water and sediments, especially where there is chemical stratification of redox-controlling chemicals. Due to their respiratory versatility, they play important roles in turnover of organic matter during redox fluctuations and near oxic-anoxic interfaces in aquatic environments where oxygen is deficient and alternative electron acceptors become available (4). Owing to their diverse metabolic capabilities, *Shewanellae* are important players in the biogeochemical cycles of carbon, nitrogen, and sulfur.

The remarkable respiratory versatility of *Shewanellae* is reflected in the genomes of sequenced *Shewanella* strains, which are generally rich in cytochromes, reductases, quinones, and iron-sulfur proteins (1) (5). As revealed by the model strain *Shewanella oneidensis* MR-1, a highly diverse electron-transport system was predicted given a total of 42 putative c-type cytochromes, 211 one-component signal transduction systems and 47 two-component systems, as well as a number of novel transcription regulatory systems. System-level studies of multiple

*Shewanella spp.* strains further revealed a diverse range of metabolic processes in signal sending, transduction and regulation.

The *Shewanella baltica* strains in this study were isolated from nutrient enriched Baltic Sea, especially with overloads of nitrogen (6). Vigorous denitrification activity was observed from the oxic-anoxic interface based on *in vitro* growth experiments, which was considered the major factor counteracting eutrophication (7). Methods targeting denitrifying bacteria were used to isolate bacteria possibly responsible for this process, where the *S. baltica* was found to represent the major fraction of culturable denitrifying bacteria (8) (9). The depths of isolation spanned the chemocline of oxygen, nitrate and hydrogen sulfide in the water column, and agreed with the previous notion that *Shewanellae* were favored in complex redox aquatic environments. Thus, we examined the utilization of several potential electron acceptors relevant to the environment in the Baltic Sea, to look for correlation between electron-accepting capability and niche differentiation among these *S. baltica* strains.

In addition to the electron accepting abilities, capabilities in carbon resource utilization is also essential for thriving in complex environments. *Shewanella spp.* are among the bacteria commonly found during food spoilage, especially anaerobic spoilage of seafood. They were shown to be able to use a range of carbon sources including various amino acids and peptides, fatty acids, and even nucleotides, although their ability in degrading organic acids was mostly limited to C<sub>1-3</sub> compounds. Phenotypic characterization of carbon source utilization has been used to distinguish closely related *Shewanella spp.* (10) (8), and was shown to be correlated to ecological conditions. Moreover, *Shewanella spp.* were often considered as 'opportunists' in the environment, which typically use the r-strategy in unpredictable environments. As the ability to

reproduce quickly during nutrition rich periods is crucial for r-strategists, the metabolic capacity of *S. baltica* is probably essential for their survival in fluctuating aquatic environments.

Biolog Phenotype Microarrays (PMs) provide a high throughput phenotyping system with the capacity to effectively measure large numbers of phenotypes at once. It incorporates redox chemistry into growth assays, using cell respiration as a universal reporter for metabolic activity. Tetrazolium reduction is coupled with cells' respiration activity, reflected through color change, which can be monitored and quantified to reveal the extent of respiration. This system has been applied to examine metabolic differentiation of bacteria. In a recent study of several *Shewanella* strains of different species, comparative Biolog analysis has been shown to reflect niche differentiation among the strains (11). Furthermore, with the aid of genomic information, comparative Biolog analysis also provided insights into the functional clarification of ambiguous genes, as well as important evaluation for genome predictions. In this study, I employed the Biolog PM system to screen the utilization of 95 carbon sources by 46 *S. baltica* strains. Through investigating phenotypic traits of these strains, we expect discovery of environmental signatures of the *S. baltica* species, and further understanding of the outcome of selection on this population.

## Materials and Methods

**Bacterial strains and growth conditions.** LB medium and modified M1 medium (12) were used for bacterial growth. Anaerobic M1 medium was made under oxygen free argon gas. Sodium nitrate (2 mM) was supplied for testing denitrification. Durham tubes were inserted into the serum tubes for gas collection. Nitrate levels were tested after 3 days of incubation at 22°C using nitrate test strips (Merckoquant, Darmstadt, Germany, limit of detection >160 mM). Thiosulfate reduction was tested in Triple Sugar Iron agar (TSI agar, Acumedia, Lansing, MI). Strains were inoculated into the bottom of test tubes containing TSI agar and incubated at 22°C. Three replicates were made for each strain. Strains were classified as incapable of reducing thiosulfate into hydrogen sulfide (H<sub>2</sub>S) if the agar did not turn black within a week.

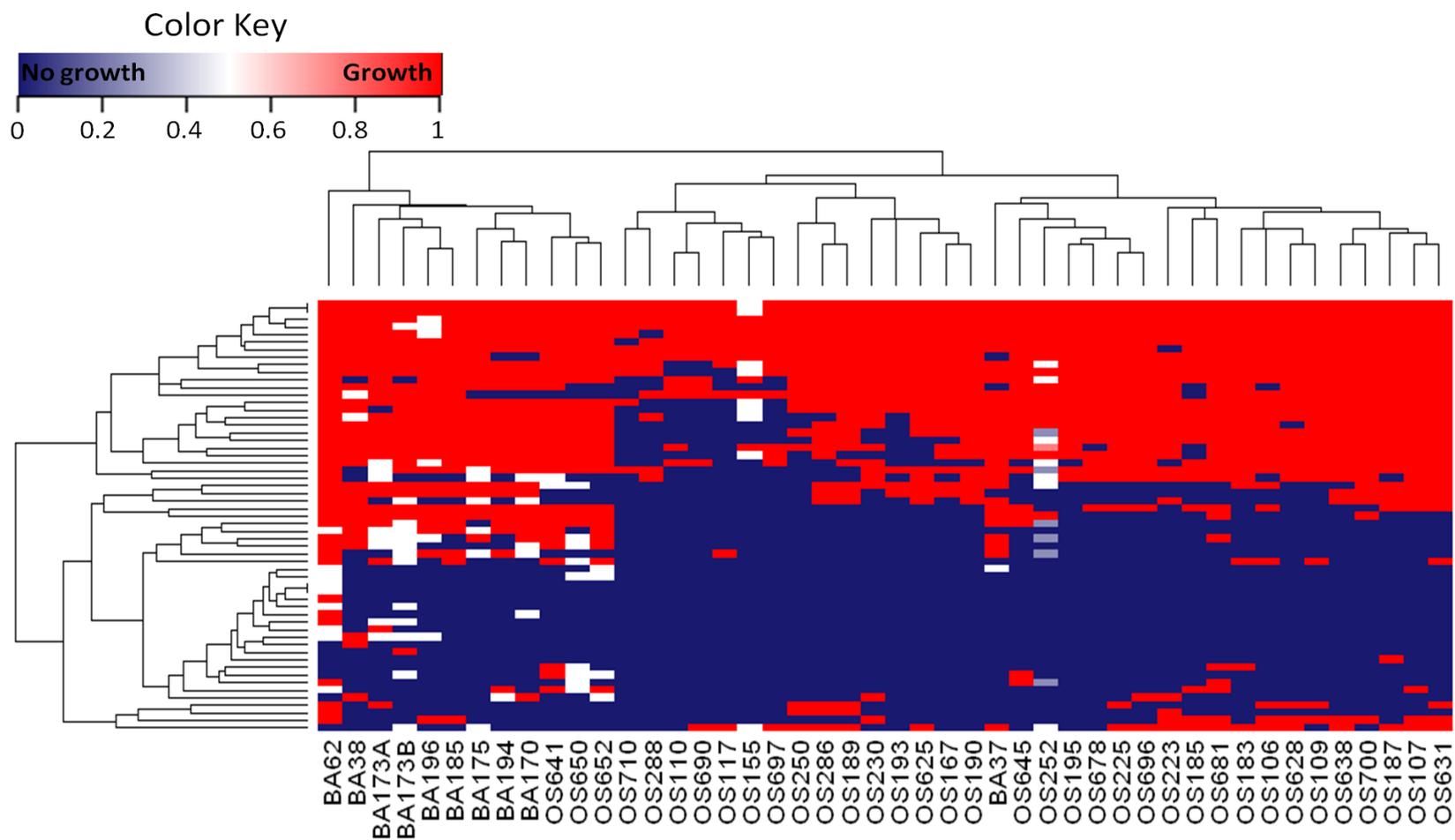
Dimethylsulfoxide (DMSO) and trimethylamine-N-oxide (TMAO) reduction was tested in anaerobic M1 medium supplemented with 10 mM of DMSO and TMAO as electron acceptors, respectively. OD<sub>600</sub> was monitored for 3 days. Strains were characterized as not capable of reducing DMSO or TMAO if the OD<sub>600</sub> did not exceed 0.02 at the end of incubation. *Shewanella oneidensis* MR-1 was used as the positive control.

**Regression analysis on Biolog datasets.** Pair-wise similarities among isolates were calculated for the gene presence and absence matrix from the CGH profile, the Biolog dataset, as well as the environmental parameter profile. Regression analysis was performed on Biolog similarities with CGH similarities to estimate correlation between metabolic capability and evolutionary relatedness. A secondary regression was then performed to estimate the effects from

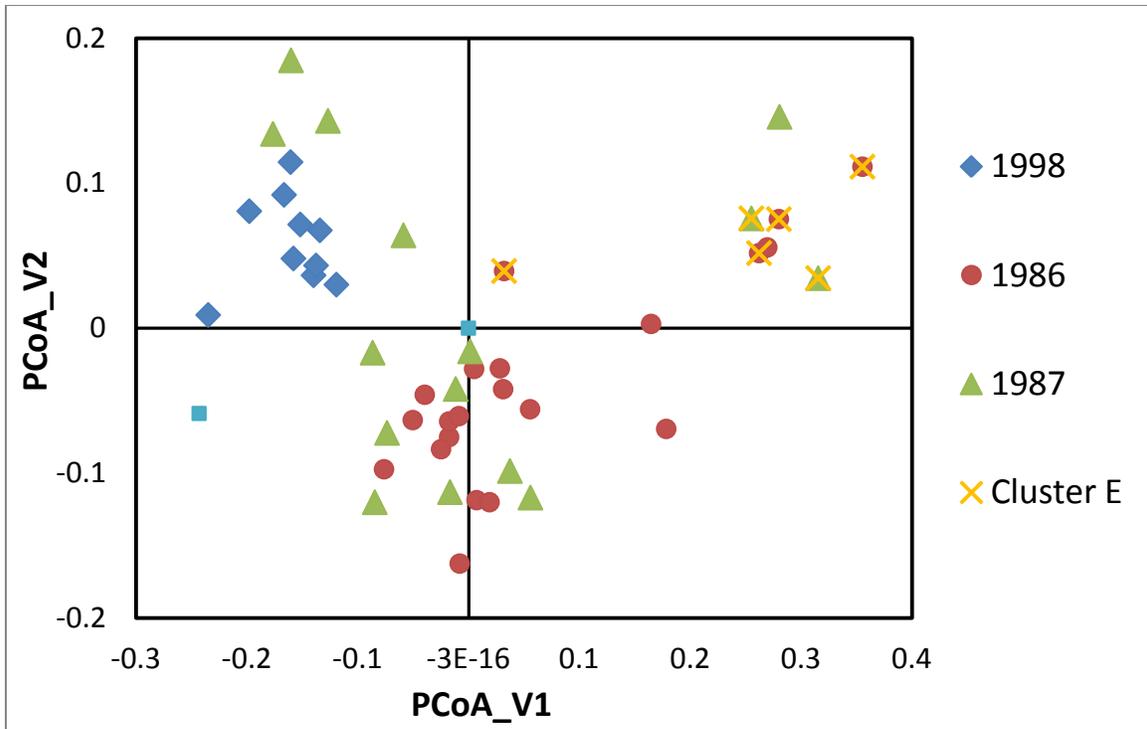
environmental factors using the phylogenetic independent residuals against similarities from environmental profiles, which eliminated the phylogenetic influence on metabolic capabilities.

## Results

**Examining diversity of carbon source utilization using the Biolog system.** Use of 95 different carbon sources by the *S. baltica* strains was screened using the Biolog-GN system. Metabolic activities were measured and assays were scored as either growth or no growth [Figure 4.1]. The *S. baltica* strains were able to utilize an average of 36 compounds under aerobic condition. Among the carbon sources tested, 19 were not used by any *S. baltica* strain, while the following 11 substrates were used by all strains: cis-aconitic acid, sucrose, D-gluconic acid, L-glutamic acid, dextrin, maltose,  $\alpha$ -D-glucose, L-serine, N-acetyl-D-glucosamine, lactic acid, and inosine. These composed the metabolic core of the *S. baltica* population. Interestingly, compared to *S. oneidensis* MR-1, whose carbon source utilization capability was mostly limited to 3 carbon compounds (13), these *S. baltica* strains were capable of using compounds containing 5 or 6 carbons. A principle coordinate analysis of these Biolog profiles [Figure 4.2] showed that all the 1998 strains clustered together, while, strains isolated from 1986 and 1987 were not well separated in this visualization method. However, among those from the 1980s, strains in Clade E [Figure 2.2] appeared to share a similar pattern of Biolog profile, while clustering of Clade A strains was not apparent.



**Figure 4.1. Heatmap of 46 *S. maltica* strains based on Biolog profiles.** Blue color indicates no metabolic activity detected with respect to utilization of the tested substrates; red color indicates significant level of metabolic activity and white indicates weak metabolic activity. Hierarchical clustering was performed both by strains (vertical) and substrates (horizontal).



**Figure 4.2. Principle Coordinate Analysis (PCoA) of Biolog metabolic profiles of the *S. baltica* strains.** Distance matrix is calculated using Jaccard distances. Strains isolated from different years are marked with different labels.

**Examining diversity in utilization of different electron acceptors.** Although these *S. baltica* strains were initially considered as potential in situ denitrifiers in the Baltic Sea, formation of nitrogen gas during nitrate respiration was not detected from any of these strains. However, all of them were able to grow with nitrate as the sole electron acceptor and decrease the concentration of nitrate. Out of the 46 strains tested, 13 of them (all from 1986 and 1987) were able to grow with DMSO as the only electron acceptor [Table 4.1]. Thiosulfate and TMAO can be used as electron acceptors by most of these isolates, except for the strains contained in Clade E, implying the possibilities of mutation or gene loss in their recent common ancestor.

**Table 4.1. Use of Different Electron Acceptors by *S. baltica* Strains**

Strain	Nitrate	Thiosulfate	DMSO	TMAO
MR1	+	+	+	+
OS106	+	+	-	+
OS107	+	+	-	+
OS109	+	+	+	+
OS110	+	-	-	-
OS117	+	-	-	-
OS155	+	-	-	-
OS167	+	+	+	+
OS183	+	+	-	+
OS185	+	+	-	+
OS187	+	+	-	+
OS189	+	+	-	+
OS190	+	+	+	+
OS193	+	+	+	+
OS195	+	+	+	+
OS223	+	+	-	+
OS225	+	+	+	+
OS230	+	+	-	+
OS250	+	+	-	+
OS252	+	-	-	-
OS286	+	+	-	+
OS288	+	+	-	+

Table 4.1. (Cont'd)

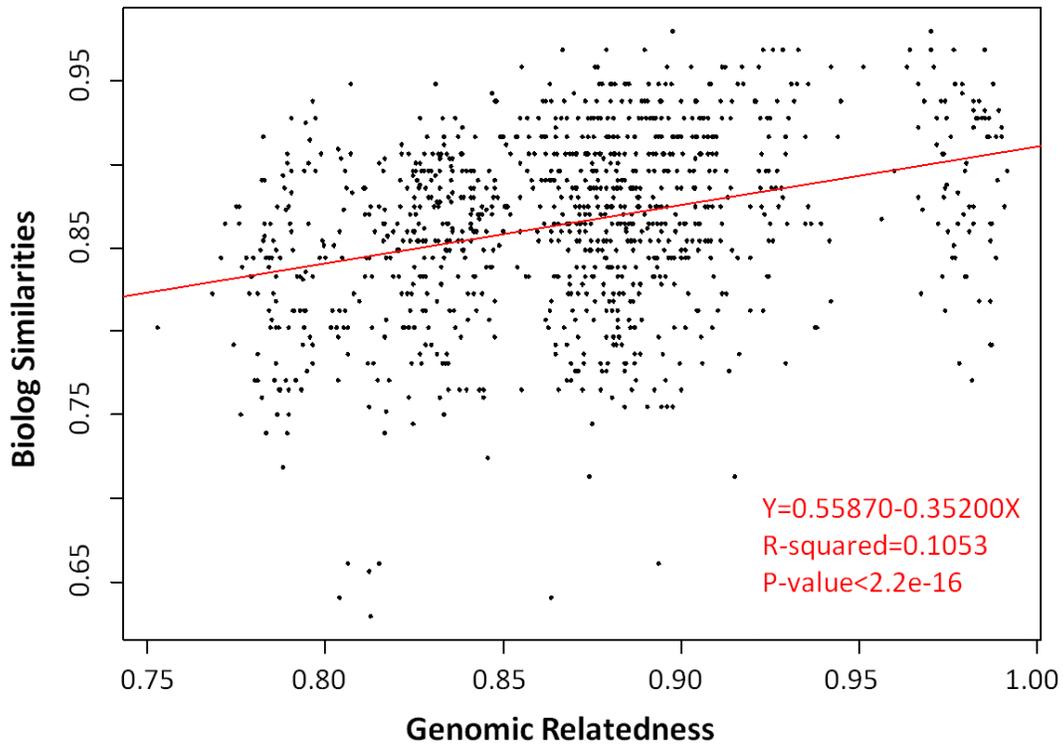
OS625	+	+	-	+
OS628	+	+	+	+
OS631	+	+	+	+
OS638	+	+	-	+
OS641	+	+	+	+
OS645	+	+	-	+
OS650	+	+	+	+
OS652	+	+	+	+
OS678	+	+	+	+
OS681	+	+	-	+
OS690	+	-	-	-
OS696	+	+	-	+
OS697	+	-	-	-
OS700	+	+	-	+
OS710	+	+	+	+
BA37	+	+	-	+
BA38	+	+	-	+
BA62	+	+	-	+
BA170	+	+	-	+
BA173A	+	+	-	+
BA173B	+	+	-	+
BA175	+	+	-	+
BA185	+	+	-	+
BA194	+	+	-	+
BA196	+	+	-	+

## Discussion

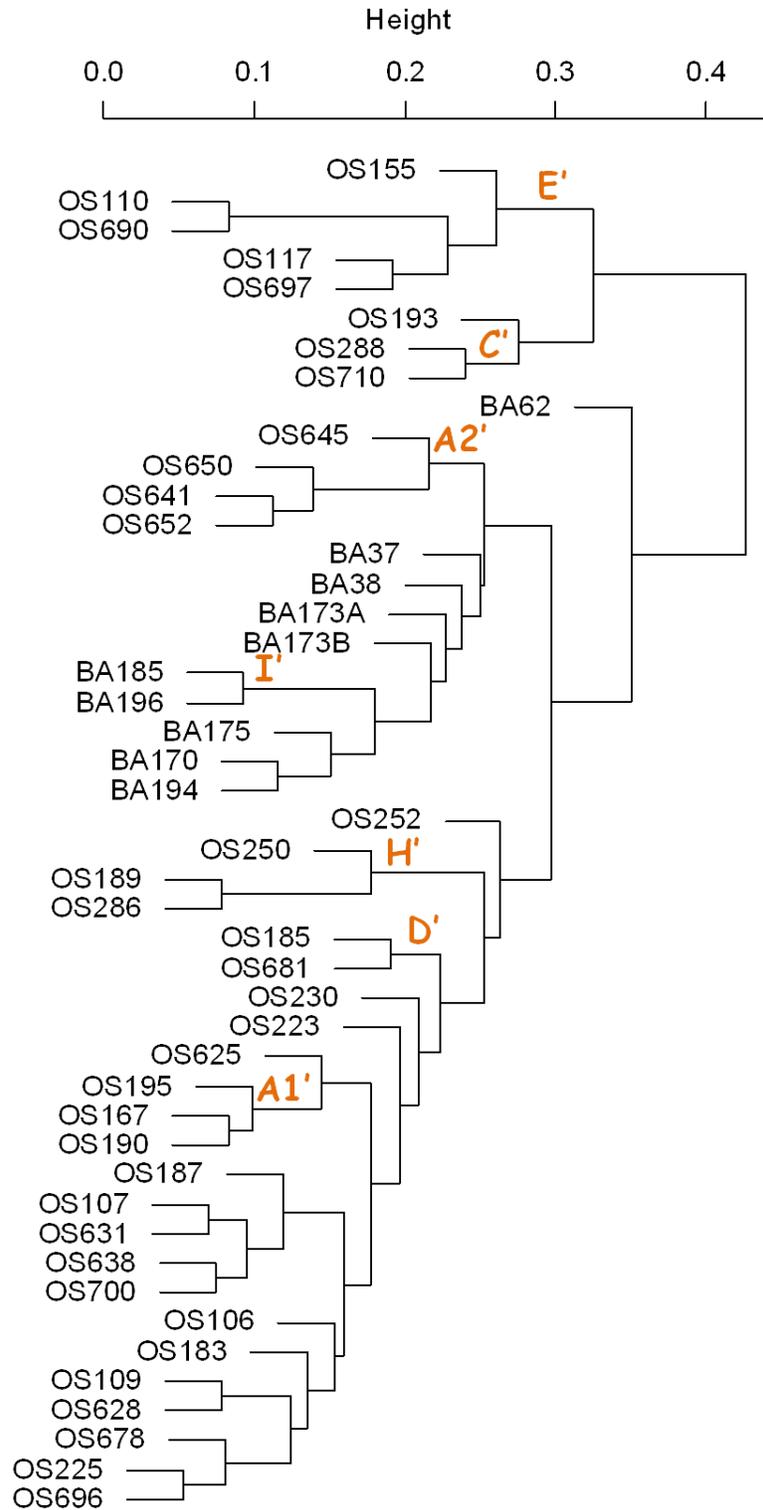
### **Metabolic similarity in carbon source utilization is correlated with evolutionary relatedness.**

One of the important goals in comparative studies of genomic and phenotypic diversity among bacterial species is to discern relationships between the genetic content and phenotypic capability [Figure 4.3]. We found a strong and positive correlation ( $P < 0.001$ ;  $b = 0.3520$ ) between these two parameters for the 46 *S. baltica* strains based on the Biolog dataset and their genomic relatedness from CGH. Thus, the signals from these rather coarse level methodologies were strong enough to detect what must be underlying genetic-phenotype relationships. This correlation can also be visualized through comparison of the hierarchical clustering structures of the CGH and the Biolog profiles [Figure 4.4], where six clades were at least partially shared between the two dendrograms. It is interesting to note that, Clade A in the CGH (or MLST) dendrogram can be matched to two clusters, A1' and A2', from the Biolog dendrogram. While A1' contains only strains isolated from 1986, A2' includes only the strains isolated from 1987, which have metabolic patterns that are more similar to those revealed from the 1998 isolates. Thus, the Biolog profiling seems to have provided further discrimination among these strains beyond the resolution of MLST genotyping. Meanwhile, metabolic patterns of specific clusters can be inferred. For instance, strains in Clade E used a significantly fewer number of carbon sources compared strains in other clusters. In particular, a number of 4 carbon acids such as succinate acid, aspartic acid, and  $\alpha$ -ketobutyric acid failed to stimulate growth of any strains in this cluster, implying the lack of certain catalytic enzymes involved in the pathways. Moreover, metabolic profiling of strains in Cluster A2', comparing to those in Cluster A1', differs mainly at utilization of an additional eight carbon sources (D-psicose, mono-

methyl succinate, alaninamide,  $\alpha$ -ketovaleric acid, glucose-1-phosphate, D-fructose, bromo-succinic acid and propionic acid, with  $p < 0.1$ ), possibly due to selective sweeps that favored capability in utilizing these compounds in the Baltic water environment.



**Figure 4.3. Regression of Biolog similarity and genomic relatedness.** Pair wise genomic relatedness was calculated from gene presence and absence matrix from the CGH profile using formula  $\sum(|X_i - Y_i|)$  divided by length of the vector. Pair wise Biolog similarity was calculated using the same formula.



**Figure 4.4. Hierarchical clustering of the Biolog profile.** Pair wise Biolog dissimilarities were calculated using the Jaccard index. Hierarchical clustering was performed using the average linkage agglomeration method.

**Temporal distribution has a significant impact on evolution of metabolic capabilities by *S. baltica* strains.** After eliminating phylogenetic influences from the metabolic profile, the phylogenetic independent residuals were used to examine the effects from other isolation and environment related factors. These factors include the sites and time of isolation, culturing methods, as well as water chemistry parameters including concentrations of nitrate, oxygen, hydrogen sulfide, etc. Significant correlation was observed only between these residuals with the time of isolation ( $p < 0.01$ ), i.e., effects of other environmental variables are either not significant on the genetic or phenotypic diversification, or are through influencing the phylogenetic relatedness among these isolates.

**Table 4.2. Two Sample T Test Shows that Strains from 1998 Use Significantly More Carbon Sources**

	All	1980s	1980s from MM*	1998 all from MM
Average number of substrates used	35.7	33.5	38.6	43.6
P-value from t-test against 1998 strains	0.00104	9.07e-05	0.0230	-

\* MM: Minimal Medium

A closer look at the Biolog profile revealed higher metabolic versatility in carbon source utilization among strains isolated in the more recent year. *S. baltica* strains isolated from 1998 grew on 44 compounds while those isolated from the 1980s grew on only 34 compounds. Although all the *S. baltica* strains from 1998 were recovered in minimal medium, the t-test

performed between *S. baltica* strains from 1998 and those isolated from 1980s also using minimal medium confirmed a significant difference in the number of carbon sources used [Table 4.2]. In particular, capability of utilizing certain carbon sources including D-fructose, bromo-succinic acid,  $\alpha$ -ketobutyric acid, glucose-1-phosphate, acetic acid,  $\alpha$ -hydroxybutyric acid, glucose-6-phosphate and L-alanine was found in a significantly larger percentage of strains isolated in 1998 comparing to those isolated in the 1980s ( $p < 0.001$ ), implying a shift in selective pressure in the Baltic Sea environment.

**Respiratory versatility within the *S. baltica* species is correlated with their genotypic**

**relatedness.** *Shewanella* species are among the most anaerobically versatile microorganisms and have been reported to be able to use several organic and inorganic compounds as electron acceptors (1). In particular, *Shewanella baltica* population composed the major fraction of culturable denitrifiers during the time of isolation, suggesting the importance of their roles in marine nitrogen cycling. In the other studies, *Shewanella baltica* was identified as one of the most important H<sub>2</sub>S producing bacteria during iced storage of fish (14), implying its ability to couple organic turnover with reduction of sulfur compounds. Furthermore, in addition to nitrate, the water chemistry profile exhibited a gradient of hydrogen sulfide along the water column, which accumulates from the sediment and diffuses across the anaerobic deep water to the oxic-anoxic interface. We hypothesized that, other than nitrate, respiration of certain sulfur compounds might as well be involved in specialization of these *S. baltica* strains. Thus, other than nitrate, we tested utilization of several sulfur compounds as electron acceptors, including thiosulfate and DMSO. We also assayed these *S. baltica* strains for their abilities to reduce

TMAO, which, in addition to H<sub>2</sub>S production, was regarded as a prominent characteristic of fish spoiling bacteria (15) (16). Respiratory capabilities of DMSO, thiosulfate and TMAO are strongly correlated with specific genotypes ( $p < 0.01$ ). Meanwhile, when these respiratory capabilities were compared against the presence and absence of corresponding genes and pathways from the CGH profiles, high agreement was revealed. These results further established the relationship between genetic background and phenotypic outcome among the *S. baltica* strains.

**Reduction of nitrate likely via ammonification is a common phenotypic characteristic among all *S. baltica* strains.** Although these *S. baltica* strains were initially isolated by methods targeting denitrifiers from the Baltic Sea (8) (9), denitrification genes were not found in the sequenced *S. baltica* genomes. Instead, genes that participate in dissimilatory nitrate reduction to ammonium (DNRA) (*nap* operon and *nrfA* gene) were found. Thus, our results supported the prediction that these *S. baltica* strains cannot perform nitrate reduction through complete denitrification to nitrogen gas and indicate that the respiratory nitrate ammonification mechanisms may be involved instead.

**DMSO reduction may be related to specialization to anoxic deep water environment.** It has been proposed that cooling effects caused by dimethylsulfide (DMS), a primary marine biogenic sulfur compound, may help to counteract human-induced global warming (17) (18).

Dimethylsulfoxide (DMSO), one of the precursors of DMS, is naturally formed from the degradation of phytoplankton in marine environments and can be used as an alternative electron acceptor for many bacterial species where oxygen is deficient. Strains capable of

reducing DMSO are limited in Clade A, G and K. Capabilities of reducing DMSO by these *S. baltica* strains were consistent with gene presence/absence profiles from microarray experiments, with the exception of strain OS645, which, though not capable of reducing DMSO, possesses all genes in the Type I DMSO reductase operon (*dmsEFABGH*) according to the CGH profile. The reason for loss in expression of this operon in OS645 is not clear. Furthermore, the mRNA expression level of this operon in strain OS195 is up-regulated under conditions of both nitrate and thiosulfate respiration. As half (seven) of the strains capable of reducing DMSO were isolated from at and below the anoxic-oxic interface, where both oxygen and nitrate are almost non-detectable and alternative electron acceptors are needed for survival, we predict that the ability of reducing DMSO may be correlated with levels of available electron acceptors in the water environment and is thus potentially important for adaptation to the anaerobic water environment.

**Inability to reduce thiosulfate is limited within Clade E.** As a major product of anoxic sulfide oxidation, thiosulfate is a key intermediate in the dissimilatory sulfur cycle of aquatic environments (19). Since H<sub>2</sub>S, often produced during the reduction of thiosulfate, is only detectable at and below the oxic-anoxic surface, we postulate that the ability to reduce thiosulfate into H<sub>2</sub>S may be one of the factors causing specialization of *S. baltica* strains. While compared with the MLST phylogeny, all six strains incapable of thiosulfate reduction are from the Clade E, indicating that they may have evolved through similar evolutionary paths. Interestingly, although these 6 strains came from different zones within the redox gradient

(from 80 m to the oxic-anoxic interface at 130 m), none were isolated from the anoxic water zone, suggesting that the inability to reduce thiosulfate may select against growth in the anoxic water.

**Inability to reduce TMAO is also limited within Clade E.** TMAO is a nitrogenous osmolyte broadly found in marine organisms (20). Reduction of TMAO results in the production of trimethylamine (TMA), which has been implicated as a major contributor to spoilage of seafood and fish (15) (16). TMAO is also a precursor to a variety of reduced nitrogenous biogases that are important intermediates of the marine biogeochemical cycle of nitrogen and potential regulators of atmospheric pH (21). In previous studies, *Shewanella baltica* has been identified as the major bacterial species involved in fish spoilage during iced storage (14). Our experiments showed that all the *S. baltica* strains except those in the Clade E are capable of respiring TMAO, although the CGH profiles suggest that *TorECAD* (22), the genes encoding TMAO reductases (Shew185\_3211-Shew185\_3214), were present in all *S. baltica* strains. Further comparison of genomic sequences reveal that in strain OS155, *TorA* (Sbal\_3213), encoding a periplasmic protein connected to the quinone pool via the cytochrome *TorC*, appears to be a pseudo gene due to a shift of the open reading frame resulting from a single nucleotide deletion at the 1243<sup>th</sup> position of the gene. No matter whether incapability of reducing TMAO by the other strains in this cluster is or is not attributed to a similar mechanism, we do see that a simple change either in regulation or structural integrity can result in non-activity of the entire operon.

## REFERENCES

## References

1. Fredrickson JK, *et al.* (2008) Towards environmental systems biology of *Shewanella*. *Nature Reviews Microbiology* 6(8):592-603.
2. Scott KHNaJ (2006) Ecophysiology of the genus *Shewanella*. *Prokaryotes* 6:19.
3. Tiedje JM (2002) *Shewanella* - the environmentally versatile genome. *Nature Biotechnology* 20:2.
4. Hau HH & Gralnick JA (2007) Ecology and Biotechnology of the Genus *Shewanella*. *Annual Review of Microbiology* 61(1):237-258.
5. Heidelberg JF, *et al.* (2002) Genome sequence of the dissimilatory metal ion-reducing bacterium *Shewanella oneidensis*. *Nature Biotechnology* 20(11):1118-1123.
6. Elmgren R (1989) Man's Impact on the Ecosystem of the Baltic Sea: Energy Flows Today and at the Turn of the Century. *AMBIO*:326-332.
7. Brettar I & Rheinheimer G (1991) Denitrification in the Central Baltic: evidence for H<sub>2</sub>S-oxidation as motor of denitrification at the oxic-anoxic interface. *Marine Ecology Progress Series* 77:157-169.
8. Brettar I (2002) *Shewanella denitrificans* sp. nov., a vigorously denitrifying bacterium isolated from the oxic--anoxic interface of the Gotland Deep in the central Baltic Sea. *International Journal of Systematic and Evolutionary Microbiology* 52(6):2211-2217.
9. Brettar I, Moore ERB, & Höfle MG (2001) Phylogeny and Abundance of Novel Denitrifying Bacteria Isolated from the Water Column of the Central Baltic Sea. *Microbial Ecology* 42(3):295-305.
10. Stenström IM & G M (1990) Classification of the spoilage flora of fish, with special reference to *Shewanella putrefaciens*. *Journal of Applied Bacteriology* 68(6):601-618.
11. Rodrigues JLM, Serres MH, & Tiedje JM (2011) Large-Scale Comparative Phenotypic and Genomic Analyses Reveal Ecological Preferences of *Shewanella* Species and Identify Metabolic Pathways Conserved at the Genus Level. *Applied and Environmental Microbiology* 77(15):5352-5360.
12. Cruz-Garcia C, Murray AE, Klappenbach JA, Stewart V, & Tiedje JM (2006) Respiratory Nitrate Ammonification by *Shewanella oneidensis* MR-1. *Journal of Bacteriology* 189(2):656-662.
13. Serres MH & Riley M (2006) Genomic Analysis of Carbon Source Metabolism of *Shewanella oneidensis* MR-1: Predictions versus Experiments. *Journal of Bacteriology* 188(13):4601-4609.

14. Vogel BF, Venkateswaran K, Satomi M, & Gram L (2005) Identification of *Shewanella baltica* as the Most Important H<sub>2</sub>S-Producing Species during Iced Storage of Danish Marine Fish. *Applied and Environmental Microbiology* 71(11):6689-6697.
15. Barrett EL (1985) Bacterial Reduction of Trimethylamine Oxide. *Annual Review of Microbiology* 39:19.
16. Gram L & Huss HH (1996) Microbiological spoilage of fish and fish products. *International Journal of Food Microbiology* 33:121-137.
17. Vallina SM & Simo R (2007) Strong Relationship Between DMS and the Solar Radiation Dose over the Global Surface Ocean. *Science* 315(5811):506-508.
18. Vallina SM, Simo R, & Manizza M (2007) Weak response of oceanic dimethylsulfide to upper mixing shoaling induced by global warming. *Proceedings of the National Academy of Sciences* 104(41):16004-16009.
19. Stefan M, Sievert RPK, and Heide N. Schulz-Vogt (2007) The Sulfur Cycle. *Oceanography* 20:7.
20. Walsh BASaPJ (2002) Trimethylamine oxide accumulation in marine animals relationship to acylglycerol storage. *The Journal of Experimental Biology* 205:297-306.
21. AD H & SW G (1999) A technique for the determination of trimethylamine-N-oxide in natural waters and biological media. *Analytical Chemistry* 71(21):4886-4891.
22. Jean-Philippe Dos Santos CI-N, Carole Couillaulta, Gérard Giordanoa and Vincent Méjean (1998) Molecular analysis of the trimethylamine N-oxide (TMAO) reductase respiratory system from a *Shewanella* species. *Journal of Molecular Biology* 284(2):421-433.

## **CHAPTER V**

### **Competition Assays among Selected *Shewanella baltica* Strains while Exposed to Different Electron Acceptors**

## Abstract

A key strategy to understand the ecology of organisms is through determining how they respond to their environment by altering gene expression. In order to determine redox specialization of the *Shewanella baltica* strains, competition assays were performed for identification of competitive advantages in conditions mimicking their own niches, from which potential specialization to waters containing high nitrate level was found. Two model strains, *S. baltica* OS185 and OS195, with full genomes sequenced, sharing high genomic similarity, but isolated from depths with distinct nitrate levels, were chosen to further investigate transcriptional response to the nitrate respiring condition. The two competed genotypes were separated by flow cytometry and their transcriptomes analyzed by Illumina RNA-Seq. Regulatory differences were found in their responses both to the nitrate condition and to the competitor strain. Analysis of differentially expressed genes due to presence of the other competitor strain revealed that OS185 appeared to have induced the expression of phage components in OS195, while OS185 itself actively responded to the high nitrate condition by up-regulation of redox related genes. Similarly, analysis of differentially expressed genes due to the nitrate respiring condition also suggested higher responses to nitrate condition in OS185 compared to that in OS195 through higher expression of redox related genes in OS185. These results collectively provide explanations for the competition outcome between OS185 and OS195, and suggested that OS185 is better adapted to nitrate-rich conditions.

## Introduction

In the realm of microbial ecology, it is equally important to study the microbe itself as well as its connections with the environment. The former task was traditionally realized through various cultivation methods followed by genetic and phenotypic characterization, and in this era, using high throughput metagenomic sequencing to finely probe the genetic composition of the entire community. The latter task is, however, of much higher complexity, and demands a thorough understanding of the environment itself, the biotic activities of microorganisms, their interactions with each other, as well as the dynamics of interactions between microbes and the abiotic environments.

Numerous efforts have been made to highlight the importance of the environmental components associated with each microorganism(1) (2) (3), including the proposal of 'ecotype' theory, which incorporates niche properties into the framework of a microbial species definition (4) (5). As proposed in the ecotype theory, selection sweep is the key component in formation of ecotypes, as well as the major force for species cohesion. Microbial specialization is regarded as the processes of formation of various ecotypes within a population, which facilitates resource partitioning in the ecosystems (6) (7). Thus, differential fitness in different 'micro'- niches is often a direct result of specialization.

Specialization of microorganisms has been observed in many ecosystems (8) (9) (10), among which the most extensively studied cases are those of pathogenic bacteria (11) (12) (13). However, even in many well studied systems, the answer to the question of what makes one genotype a better competitor than another is largely unknown. The experimental proof of local adaptation was often challenged by difficulties including the complexity of the ecosystem itself.

Most bacterial specialization was revealed through characterization and comparative analysis of genetic and phenotypic properties (14) (15) (16), while further evidence for adaptation was often investigated through in situ and in vitro quantification of ecological contributions, and in a few cases, as competition assays (17).

The impact of natural selection on specialization is through influencing every local component of the biosphere and adaptation is expected to have occurred at all different levels of scales. Two important tasks in determining adaptation of species are to determine the attributes of selective forces in particular environments, and to identify the phenotypic properties that are responsible for the selective outcome. As influence of natural selection is reflected through the fitness of individual organisms, identification of the differential fitness of microorganisms under a specific environment is direct evidence for natural adaptation. Thus, studying adaptation would improve the understanding of the processes of natural selection and the underlying mechanisms.

Respiratory versatility is of special importance in the ecological strategy of the genus *Shewanellae* (18). Among the *S. baltica* strains, the capability of respiring nitrate is regarded as being important ecologically because nitrogen was enriched in the Baltic Sea due to heavy contamination by urban coastal areas and because the recovered *S. baltica* are the dominant denitrifying species (19) (20) (21) (22). Given the redox gradient formed by various redox-active chemical substrates including oxygen and nitrate, as well as the spacial distribution of different *S. baltica* strains along the depths of the water column, I propose to test whether the isolates exhibit differential fitness from different depths that correspond to their redox conditions at their position of isolation.

In this chapter, five strains were chosen to represent the *S. baltica* population from different depths to study fitness effects under different redox conditions using pair-wise competition assays. To further investigate the mechanisms underlying differential fitness, I used transcriptional characterization by RNA-Seq to determine how two genetically similar strains, but isolated from different depths (redox conditions), OS185 and OS195, responded to competition under nitrate-respiring conditions. Through these efforts, differential fitness of these strains under particular redox conditions were identified, some of which correspond well to the environment from where they came. Transcriptomic profiles of OS185 and OS195 under competition further revealed different reactions of the two strains towards the same redox condition, providing insights into their adaptive mechanisms.

## Materials and Methods

### Bacterial strains and growth conditions.

**Table 5.1. Strain Description and Growth Conditions**

Strain	Description	Growth conditions
JMA1012	<i>E. coli</i> BW29427 pUX-BF13 {Tn7 transposase genes tnsA-E, amp(R), oriT(IncPalpha), R6Kgamma ori (depends on pir gene in BW29427 for replication) thrB pro thi rpsL hsdS lacZ(delta) M15 RPF-1316 (delta)araBAD567 (delta)dapA134::[erm pir(wt)]}	LB medium with DAP/300 µg/mL Amp/100 µg/mL at 37°C
JMA1018	<i>E. coli</i> BW29427 pURR25 {miniTn7 containing (Plac(A1/04/03)-gfpmut3 kan(R) strp(R)) ampR, IncPalpha oriT, and R6Kgamma ori} thrB pro thi rpsL hsdS lacZ(delta)M15 RPF-1316 (delta)araBAD567 (delta)dapA1341::[erm pir(wt)]	LB medium with DAP/300 µg/mL Amp/100 µg/mL at 37°C
JD08031	<i>E. coli</i> BW29427 pURR25 {miniTn7 containing (Plac(A1/04/03)-ecfp kan(R) strp(R)) ampR, IncPalpha oriT, and R6Kgamma ori} thrB pro thi rpsL hsdS lacZ(delta)M15 RPF-1316 (delta)araBAD567 (delta)dapA1341::[erm pir(wt)]	LB medium with DAP/300 µg/mL Amp/100 µg/mL at 37°C
<i>S. baltica</i> OS185	Strain isolated from 120 m, oxic zone with high nitrate level	LB or M1 medium at 22°C
<i>S. baltica</i> OS195	Strain isolated from 140 m, anoxic zone with no nitrate	LB or M1 medium at 22°C
<i>S. baltica</i> OS223	Strain isolated from 120 m, oxic zone with high nitrate level	LB or M1 medium at 22°C
<i>S. baltica</i> OS167	Strain isolated from 140 m, anoxic zone with no nitrate	LB or M1 medium at 22°C
<i>S. baltica</i> OS631	Strain isolated from 80 m, oxic zone	LB or M1 medium at 22°C

**GFP and CFP tagging of *S. baltica* strains.** The *S. baltica* strains used for examining fitness under redox conditions and the *Escherichia coli* strains carrying necessary plasmids for fluorescent protein tagging are listed in [Table 5.1]. The *gfp* and *cfp* delivery plasmids (pURR25-miniTn7-gfpmut3 and pURR25-miniTn7-ecfp) and the helper plasmids (pUX-BF13) were introduced into the *S. baltica* strains by tri-parental mating as described in (23) (24). The *gfp* and *cfp* tagged

strains were used for competition assays. Green fluorescence can be detected in nearly 100% of *gfp* tagged cells by flow cytometry. Although *cfp* tagged strains appeared dark in flow cytometry, *cfp* tagged strains were used as controls for possible effects on fitness due to expression of Tn7 affiliated genes in *gfp* tagged strains.

**Competition assays.** Three replicates of *gfp* and *cfp* tagged strains were inoculated separately in LB medium and grown aerobically overnight under 22°C, followed by 1:100 dilution into fresh M1 minimal medium and grown aerobically overnight before competition. To start competition, an equal number of cells was taken from each cell line and mixed. Around  $10^6$  cells from the mixture were then inoculated into 10 mL of M1 medium supplemented with different electron acceptors as indicated for the particular competition assays (no supplement for oxic growing condition and 5 mM sodium nitrate in anaerobic nitrate-respiring condition). Cells were propagated daily by 1:100 dilution into fresh medium. Samples were taken every day and analyzed through flow cytometry until one cell line was outcompeted or by the end of 7 days.

**Experimental design for transcriptomic characterization.** Model strains *S. baltica* OS185 and OS195 were selected for transcriptomic characterization during competition under nitrate respiration. The CFP labeled OS195 strain and GFP labeled OS185 strain were competed in M1 anaerobic medium supplemented with nitrate as described above. Around 0.5 million cells from each population were taken, mixed together, and inoculated at the beginning of Day 0 for competition, and cells were taken for sorting by flow cytometry at the same day of inoculation during the exponential growth phase. To provide control for the possible disturbance on the

expression level introduced during flow cytometry sorting and cell collection steps, the two strains were also grown separately in the same medium in pure culture, followed by the same procedures of flow cytometry sorting, cell collection, and all the way through illumina RNA-Seq. Three biological replicates were made for each sample.

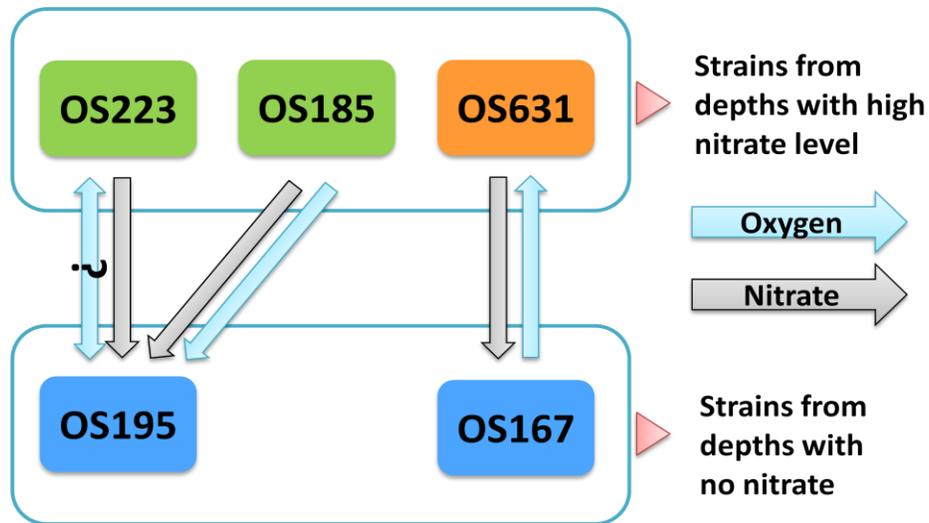
**RNA extraction and amplification.** Fresh cells were taken every 5 min from the anaerobic serum tubes using a needle and syringe, and sorted through the BD FACSVantage SE system directly into collection tubes filled with half the volume of RNAlater (Ambion, Grand Island, NY). An average of 1-2 million cells was collected per sample. Cells were filtered through 13 mm membrane filters with pore size of 0.2  $\mu\text{m}$  followed by total RNA extraction using the NucleoSpin RNA XS kit (Clontech, Mountain View, CA). Eluted RNA was checked for quality by the Agilent BioAnalyzer using the Agilent RNA 6000 Pico kit. Extracted RNA was then amplified with the Ovation<sup>®</sup> RNA-Seq System (Nugen, San Carlos, California), and amplified cDNA was purified using MinElute PCR purification kit (Qiagen, Valencia, CA). Quality and concentration of amplified cDNA was checked with Agilent BioAnalyzer using the Agilent DNA 7500 kit (Santa Clara CA). cDNA samples were subjected to Illumina RNA-Seq library preparation, followed with normalization using duplex-specific nuclease (DSN) for rRNA removal (Evrogen, Moscow, Russia ).

**Mapping, normalization, and statistical analysis of RNA-Seq data.** Demultiplexing was done by the standard Illumina data processing pipeline. Raw reads that passed quality control from Illumina RNA-Seq were mapped to *S. baltica* OS185 and OS195 genomes using the Bowtie

program (25). Reads aligned to unique positions were used in the downstream analysis. Normalization and differential expression analysis were performed through the DESeq package (26) in R. The significance test was based upon assumption of negative binomial distribution followed by the Benjamini-Hochberg procedure included in the package for adjusting P-value for multiple tests and differentially expressed genes were determined at 10% false discovery rate.

## Results and Discussion

**Competition under different electron acceptors.** In order to assess the competitive advantages among *S. baltica* strains from different redox zones, a number of strains isolated from various depths were labeled with constitutively expressed *cfp* or *gfp* reporter genes and competed against each other under aerobic growth condition as well as anaerobic growth with nitrate as the sole electron acceptor. The dynamics of the two competed populations were monitored daily through flow cytometry, which allowed rapid quantification of the ratio of two cell types. Differential fitness was revealed from these competition assays [Figure 5.1]. The strains isolated from nitrate rich waters (OS223, OS185 and OS631) outcompeted those from nitrate-depleted water (OS195, OS167) under the nitrate-respiring condition, indicating possible adaptation to higher nitrate levels in their environment. Also, the strain from the microaerophilic zone (OS185) was able to outcompete the strain from anoxic zone (OS195) under aerobic condition, which agreed with the oxygen levels in the niches from which they were derived. However, one strain isolated from anoxic water (OS167) was shown to have outcompeted another strain from a more oxic zone (OS631) under aerobic condition; while the pattern between the pair of strains OS223 and OS195 under aerobic condition was not clearly resolved.



**Figure 5.1. Competition patterns observed among five *S. Baltica* isolates in minimal medium under aerobic and anaerobic condition with nitrate as the sole electron acceptor.** Each arrow represents a pair of strains competed with each other, and points from the survived strain to the outcompeted one.

Some of the competition patterns did not agree with the niche profiles of certain strains; however, one should note that the experimental conditions tested for competition, though attempting to mimic conditions of the actual environment, do differ from those in nature. Thus, these competition patterns might not represent well the Baltic Sea, but nonetheless offer a better controlled approach to assess fitness effects in these strains by various environmental factors. These assays opened doors to further measurement of relative fitness among a larger collection of strains in a high-throughput manner, while an in-depth case study would promote better understanding of potential mechanisms underlying differential fitness of strains under different environmental conditions.

**Transcriptional characterization of two model strains *S. baltica* OS185 and OS195 during competition with limited nitrate as electron acceptor.** Two of the sequenced strains, OS185 and OS195, were selected for further transcriptional characterization of adaptation to niches with distinct nitrate conditions. OS185 was isolated from an oxic-anoxic transition zone where nitrate was found at the highest concentration along the water column, and OS195 was isolated from an anoxic zone characterized by a depletion of nitrate. The two strains had similar growth kinetics when grown individually with nitrate [Figure 5.2], but when grown in co-culture with nitrate as the limiting electron acceptor, OS185 outcompeted OS195 [Figure 5.3], suggesting the competitive advantage of OS185 under nitrate condition. This observation is consistent with the conditions of environments from which these strains were isolated, which may imply specialization of the OS185 strain for the nitrate-rich conditions. Furthermore, it has been shown that OS185 and OS195 had undergone extensive genetic exchange in the recent past, with over 300 orthologous genes sharing 99.5-100% nucleotide identity (27). Thus, OS185 and OS195 were ideal candidates for further characterization of response to competition at the expression level.

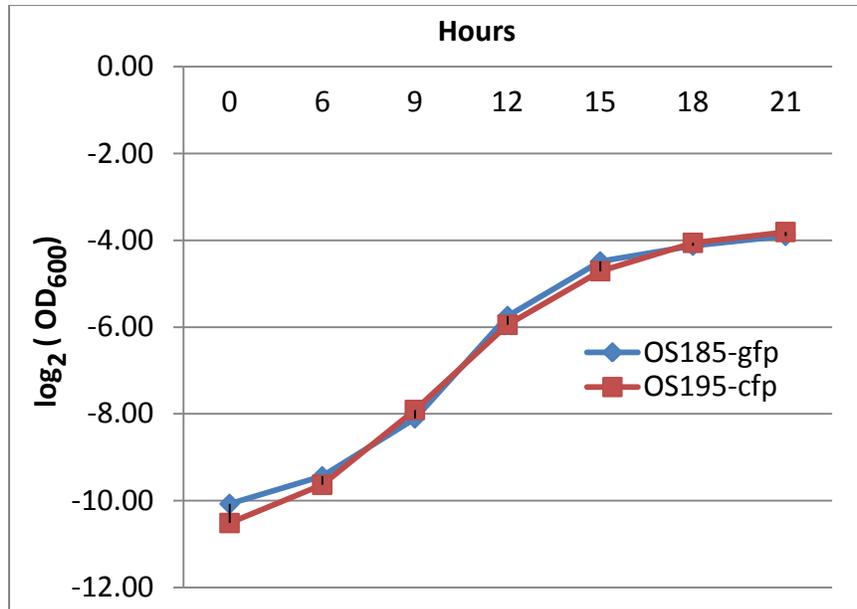


Figure 5.2. Growth curves of fluorescence tagged *S. baltica* OS185 and OS195 strains.

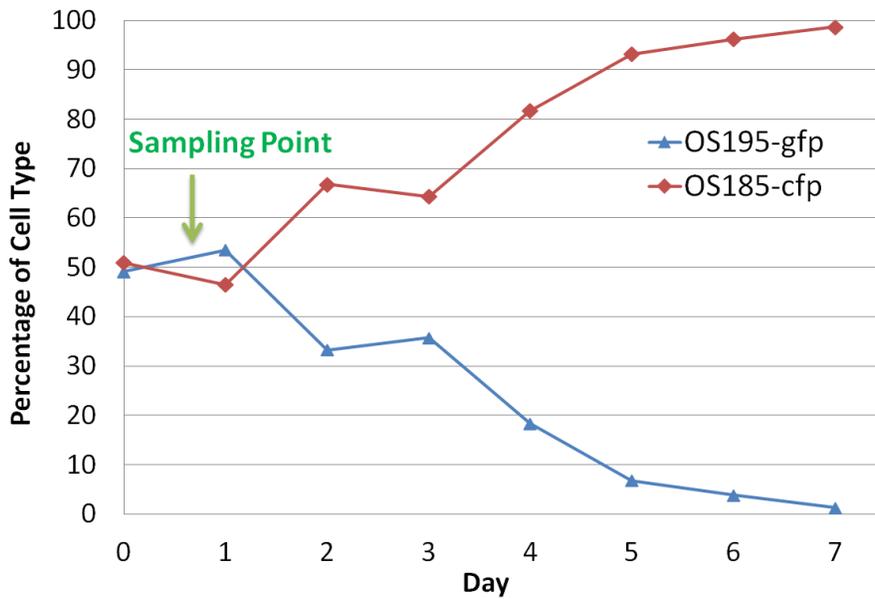


Figure 5.3. Competition pattern between fluorescence tagged *S. baltica* OS195 and OS185 in minimal medium supplemented with nitrate as sole electron acceptor.

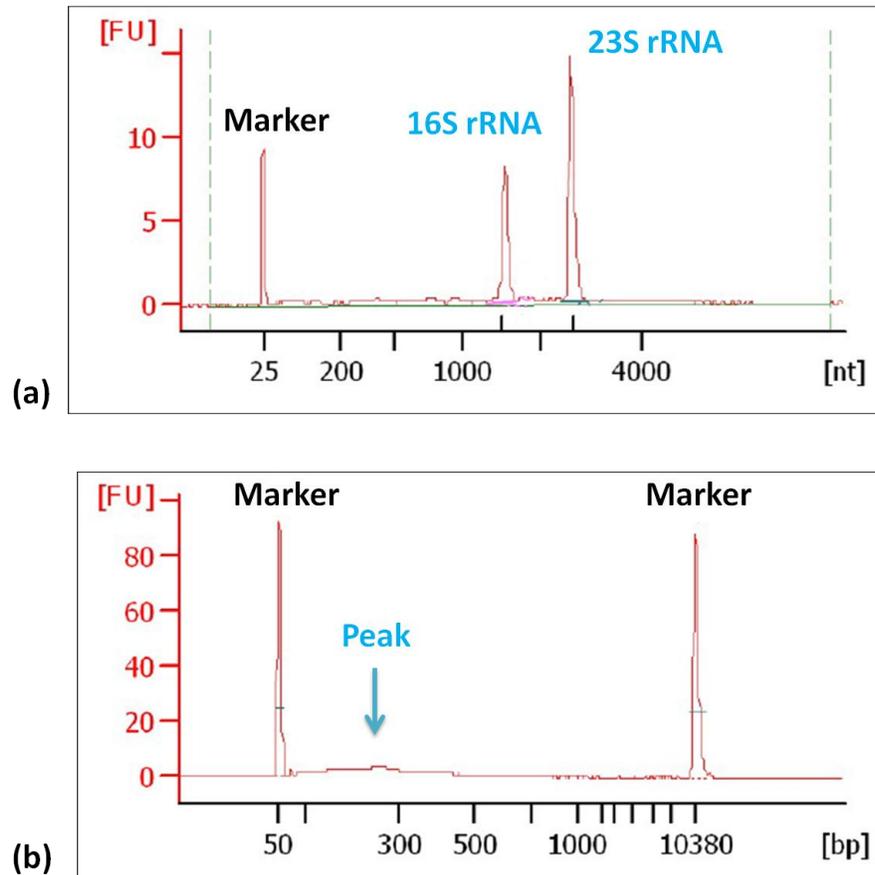
To sense competition at an early stage, we analyzed gene expression of these two competing strains under nitrate-respiring condition by RNA-Seq using the Illumina platform. We used flow cytometry to sort the two populations from the co-culture based on fluorescent signals so that we could characterize each of the transcriptomes. In addition, to provide control for potential bias or altered expression introduced during cell sorting and collecting processes, cells from the two strains grown individually were also processed through exactly the same “sorting” procedure by flow cytometry. A summary of sample description is shown in [Table 5.2]. Four sets of samples were collected, from both OS185-*gfp* and OS195-*cfp* strains grown in both pure culture and co-culture conditions, followed by total RNA extraction, amplification and rRNA removal processes.

**Table 5.2. Sample Description**

Strain	Treatment	
	Grown in Pure Culture and run through Flow Cytometry	Grown in Co-Culture and sorted through Flow Cytometry
OS185- <i>gfp</i>	Pg	Cg
OS195- <i>cfp</i>	Pc	Cc

**RNA processing and Illumina RNA-Seq summary.** The total RNA yield per sample varied from 480 µg to 4295 µg, and no significant RNA degradation was observed [Table 5.3] [Figure 5.4(a)]. Amplification of the total RNA by 1-10 thousand fold was achieved by this amplification method, resulting in 3-6 µg of amplified cDNA from each sample. Lengths of amplified cDNA ranged from 50 bp to 500 bp, with peak at around 250 bp [Figure 5.4(b)] and suitable for Illumina RNA-Seq

library preparation. Illumina sequencing yielded 6.9 to 12.2 million raw reads from each sample and around 80% of raw reads passed the initial quality control. 29.7%-45.9% of these reads were successfully mapped onto coding sequences in the two genomes. mRNA reads accounted for 94.4% to 98.4% of total mapped reads, while only 0.47% to 1.19% of rRNA reads were recovered [Table 5.4].



**Figure 5.4. BioAnalyzer profile of total RNA extracted from sample Pg-2 (a) before amplification with RNA Integrity Number of 9.3 and (b) after amplification. RNA from other samples shared a similar profile.**

**Table 5.3. Total RNA Yield and Amplification Efficiency**

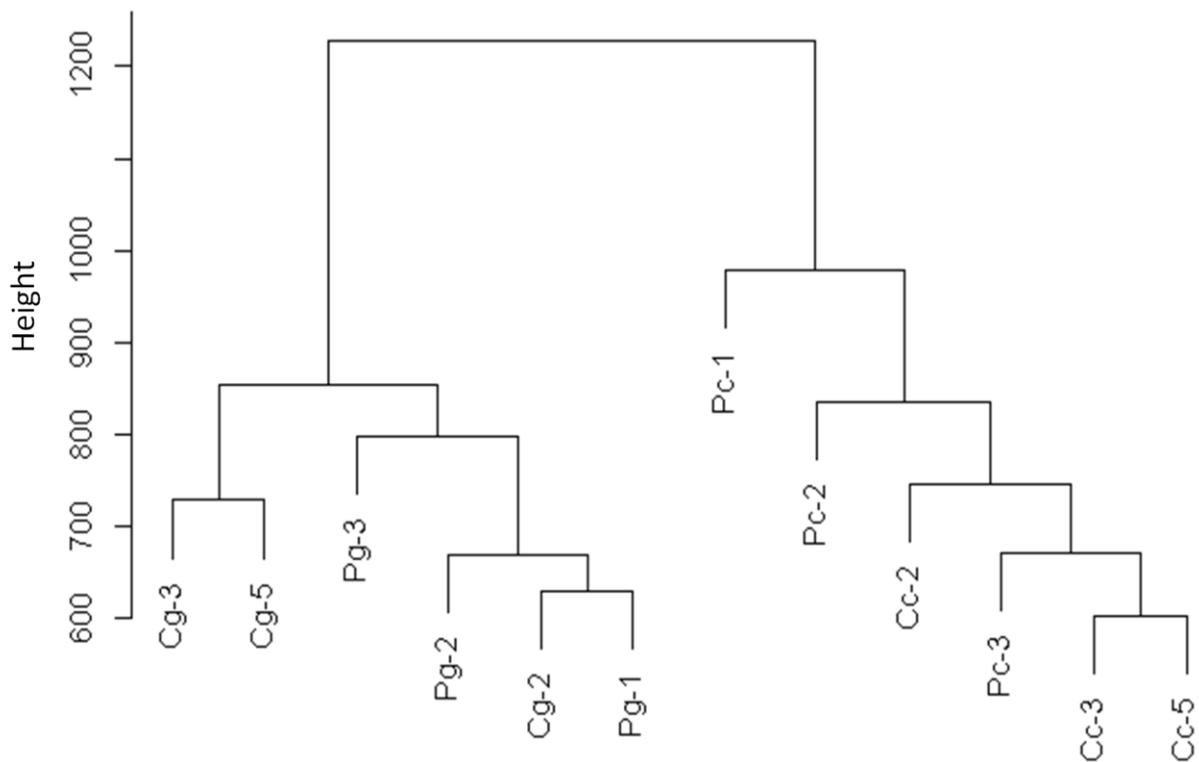
Lane	Sample	Total RNA Input /pg	Amplified cDNA /ng	Fold Amplification	Raw Reads	Reads Passed Filtering (PF)	Sample % PF
1	Cc-2	595	5040	8471	10927941	8675900	79.4
	Cg-2	740	5260	7108	10858169	8625439	79.4
	Pc-1	2560	4302	1680	12676525	10031115	79.1
	Pg-1	4295	5004	1165	7905916	6308766	79.8
2	Cc-3	810	5280	6519	6955956	5670504	81.5
	Cg-3	1115	4880	4377	9210444	7259952	78.8
	Pc-2	1450	5364	3699	15022982	11963035	79.6
	Pg-2	1800	3096	1720	10870701	8641043	79.5
3	Cc-5	1310	5620	4290	9859209	8274841	83.9
	Cg-5	480	5240	10917	12112591	9889348	81.6
	Pc-3	860	6020	7000	8131929	6563987	80.7
	Pg-3	1640	5660	3451	9388052	7610275	81.1

**Table 5.4. Summary of Mapped RNA-Seq Reads**

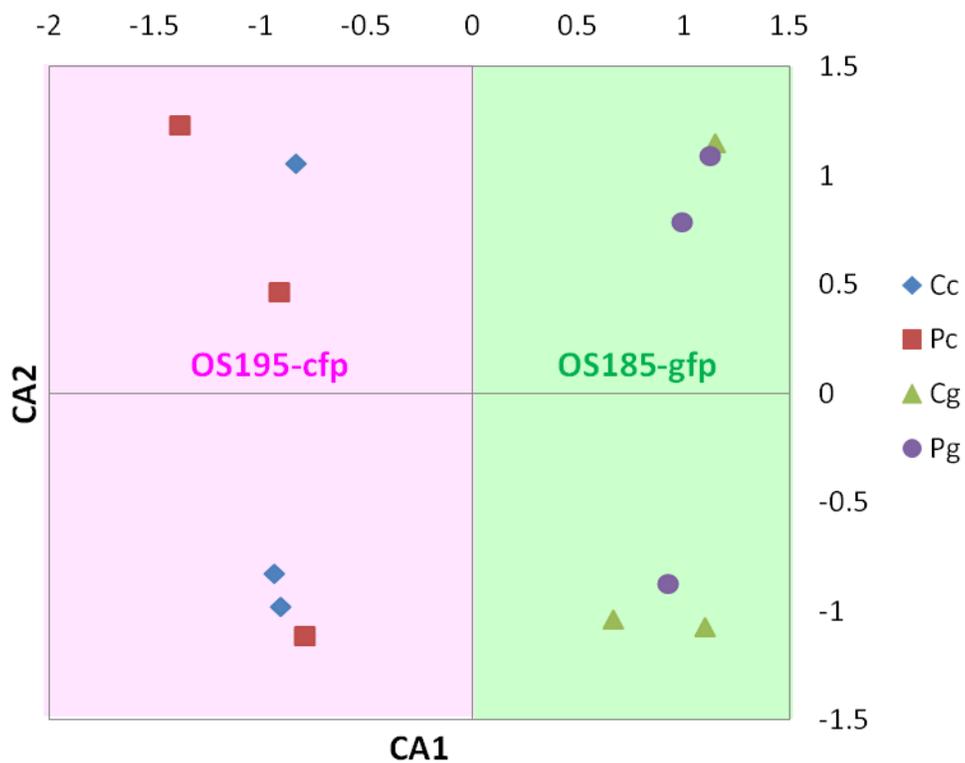
	OS185- <i>gfp</i>					
	Co-culture			Pure Culture		
	Cg-2	Cg-3	Cg-5	Pg-1	Pg-2	Pg-3
All reads	2953204	2617377	4453177	1874399	2814860	3194192
mRNA reads	2884902	2554364	4382618	1821382	2739971	3124396
rRNA reads	13952	15677	21557	12612	20608	18156
% mRNA reads	97.7	97.6	98.4	97.2	97.3	97.8
% rRNA reads	0.47	0.60	0.48	0.67	0.73	0.57

	OS195- <i>cfp</i>					
	Co-culture			Pure Culture		
	Cc-2	Cc-3	Cc-5	Pc-2	Pc-1	Pc-3
All reads	2898594	2121257	3799864	4498426	3656835	2565570
mRNA reads	2735575	2008947	3597115	4287371	3479907	2423778
rRNA reads	34532	23823	30929	46157	40828	23968
% mRNA reads	94.4	94.7	94.7	95.3	95.2	94.5
% rRNA reads	1.19	1.12	0.81	1.03	1.12	0.93

**Overview of transcriptomic profiles.** Hierarchical clustering [Figure 5.5] and canonical correspondence analysis [Figure 5.6] was performed to check the overall similarities among these profiles. As a result, samples were well separated according to strains used; however, they were not distinguishable by the types of growth conditions, suggesting that gene expression patterns were highly similar in co-culture and pure culture conditions.



**Figure 5.5. Hierarchical clustering of normalized mRNA reads from the 12 samples.** Canberra index was used to compute distances among sample profiles.



**Figure 5.6. Canonical Correspondence Analysis of normalized mRNA reads from the 12 samples.**

A summary of pairwise comparisons conducted between samples are depicted [Figure 5.7]. Illumina reads that mapped to all genes in OS185 were used for Cg::Pg comparison, as was done for the Cc::Pc comparison, while reads mapped to only OS185 and OS195 shared genes were used for Cg::Cc and Pg::Pc comparison. Differentially expressed genes were determined at a 10% false discovery rate [Figure 5.8]. Through comparing expression profiles of each strain in co-culture with their expression in pure culture (Cg::Pg and Cc::Pc), we expected to gain more knowledge about which genes were altered due to introduction of the other strain; whereas through comparing expression between these two strains under pure culture or co-culture

condition (Cg::Cc and Pg::Pc), we sought to answer to which genes were differentially expressed between these two strains and how this pattern differed between the two growth conditions.

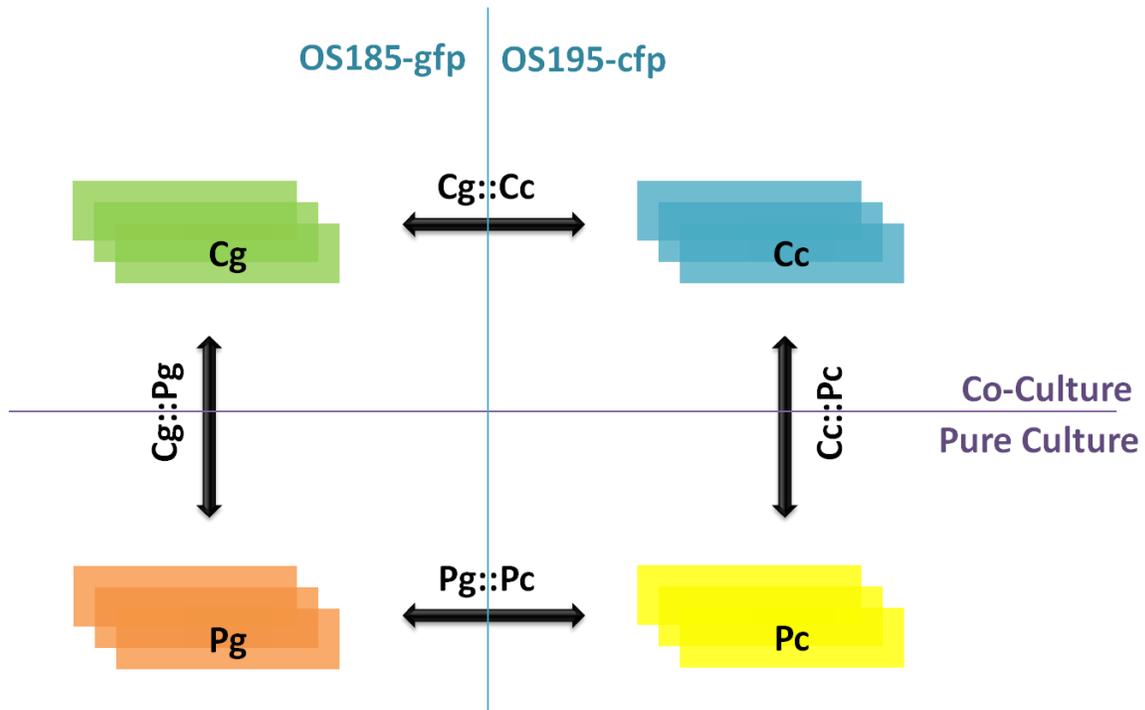
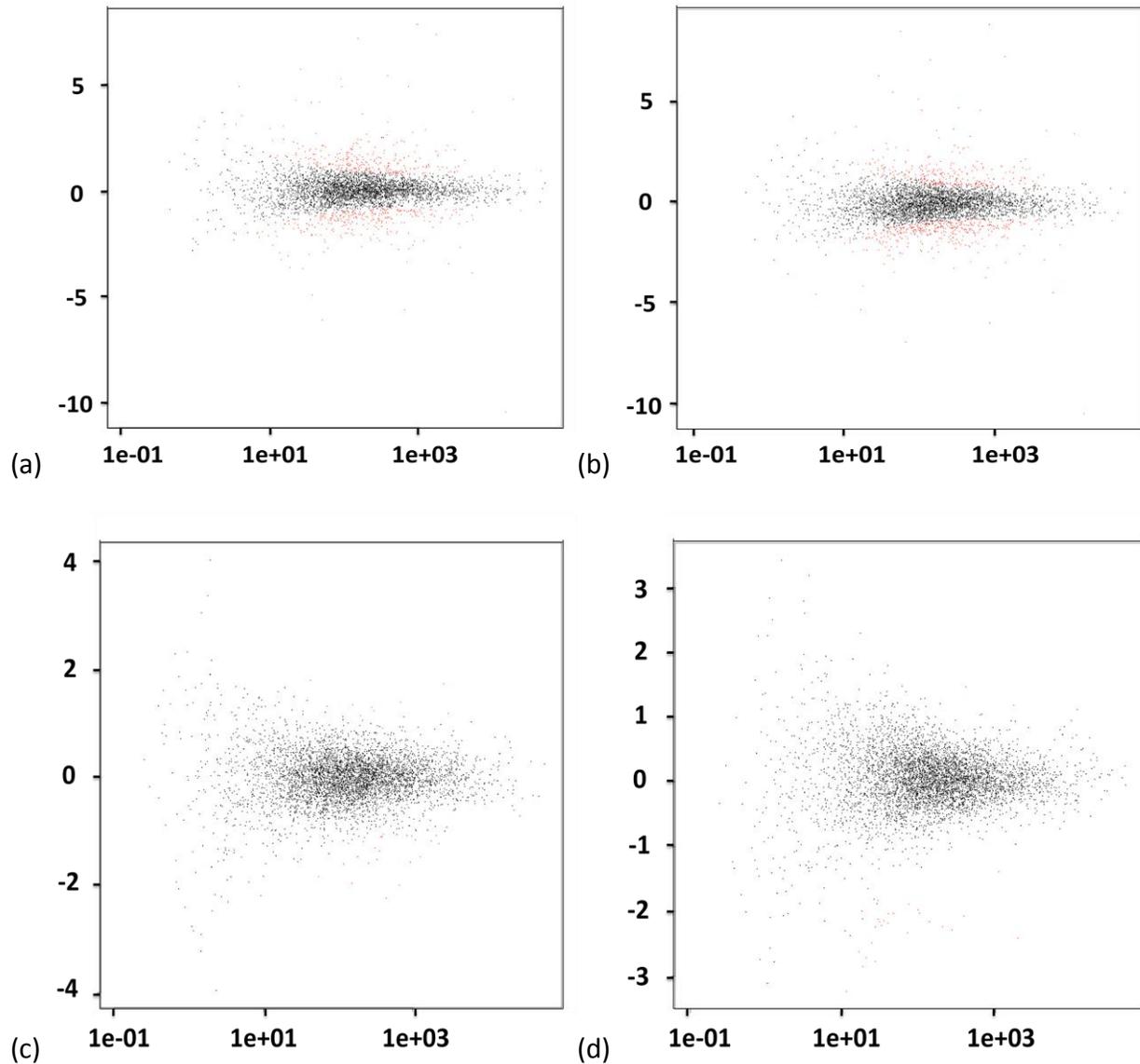


Figure 5.7. Summary of pairwise comparisons performed with RNA-Seq profiles of OS185 and OS195 under nitrate condition.



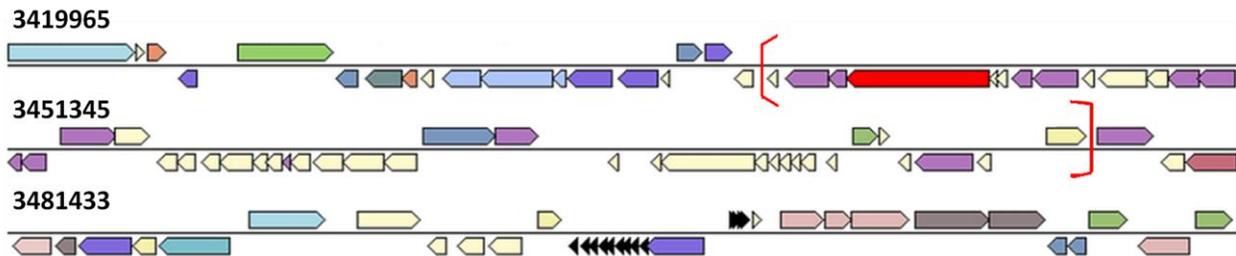
**Figure 5.8. Testing for differentially expressed genes in the four pairwise comparisons. Scatter plots of  $\log_2$  fold changes versus average numbers of reads were drawn for (a) Cc::Cg (b) Pc::Pg (c) Cg::Pg (d) Cc::Pc. The red color marks genes detected as differentially expressed at 10% false discovery rate when the Benjamini-Hochberg multiple testing adjustment was used. X-axis represents the number of reads in each gene, and Y-axis represents the fold change in each comparison. Dots colored in red correspond to genes that are significantly differentially expressed.**

**Differential expression caused by culturing methods (pure culture vs. co-culture) in each strain.**

**Differential expression of OS195-*cfp* genes in pure culture and in co-culture (Pc::Cc).**

Compared to the expression profile of labeled OS195 in pure culture, 27 genes showed differential expression while being co-cultured with OS185 strain. Two of these genes were down-regulated in co-culture, one of which encoded a hypothetical protein, while the other corresponded to a hydroxylamine reductase in a conserved operon, although none of any other genes in the operon were differentially expressed. Among the 25 up-regulated genes, one encoded a tRNA synthetase, with the rest 24 belonging to a phage-like insertion between Sbal195\_2887 (hypothetical protein) and Sbal195\_2933 (putative dehydrogenase), downstream of an operon containing genes encoding membrane transporters [Figure 5.9]. The inserted sequences contained 45 genes (Sbal195\_2888-2932), with the first part resembling components of TP901 family phage (Sbal195\_2888-2902) and the second part (Sbal195\_2905-2915) more similar to components of P2 phage in the *Myoviridae* family, followed by genes involved in phage replication and transcriptional regulation (Sbal195\_2919-2931). These genes were adjacent to a phage integrase loci (Sbal195\_2932), providing further evidence for phage mediated insertion of the sequences. The induced expression of phage components included phage capsid, tail, sheath, as well as baseplate. However, lytic cycle of the phage was not likely initiated, since once initiated, the entire OS195 population would have probably collapsed instantaneously, rather than being outcompeted at the end of a week of competition. Nonetheless, expression of these phage proteins probably simultaneously introduced disadvantage to the growth of OS195 in the co-culture, which was apparently caused by the

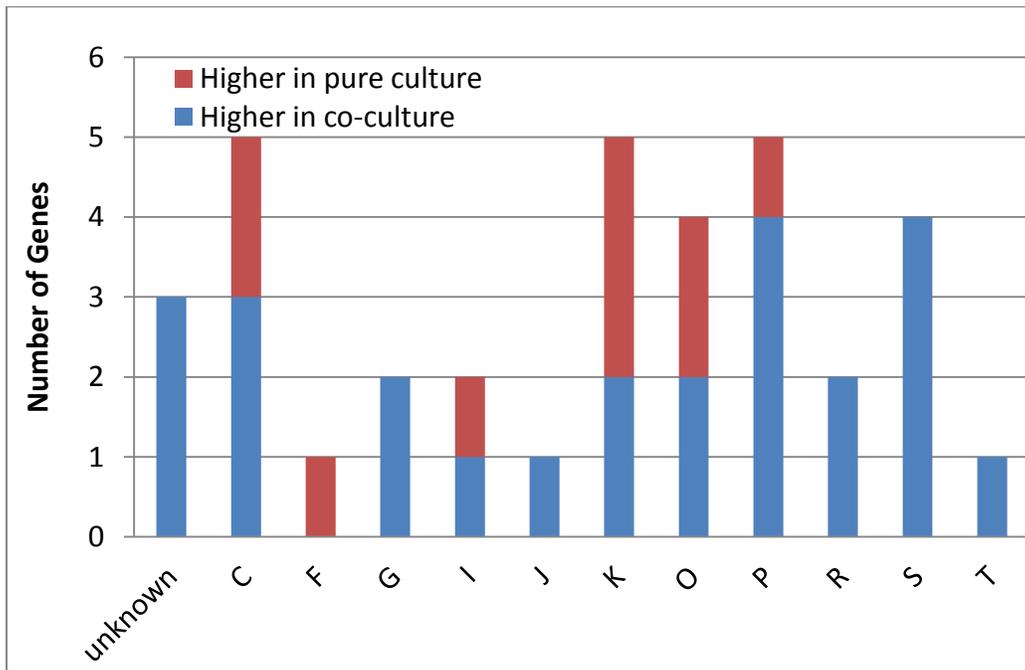
competitor strain OS185. However, at this stage, what factors, i.e., signaling agents or transcriptional regulation pathways, from OS185 might have caused this phenomenon is not clear and needs to be further characterized through additional biochemistry assays.



**Figure 5.9. Insertion of phage genes in OS195 chromosome.** Inserted genes were marked between the red parentheses.

**Differential expression of OS185-*gfp* genes in pure culture and in co-culture (Pg::Cg).** Thirty-five genes were differentially expressed in labeled OS185 grown in co-culture compared to grown in pure culture, among which 25 genes were up-regulated with the other 10 genes down-regulated. Genes encoding proteins involved in energy metabolism, ion transport and metabolism, transcriptional regulation, and posttranslational modification accounted for over half of all differentially expressed genes. The up-regulated genes included those encoding a number of transporter related proteins (Shew185\_2824-2825, Shew185\_1499), transcriptional regulators (Shew185\_1313, Shew185\_1006), outer membrane receptor (Shew185\_0720) as well as thioredoxin (Shew185\_3102), an enzyme possibly involved in reducing oxidative stress; whereas the down-regulated genes in co-culture covered a considerable number of transcriptional regulators such as *TetR*, *CadC*, *ArsR*, (Shew185\_2533, Shew185\_3867, and Shew185\_1117, respectively) and a response regulator associated polyferredoxin domain

protein (Shew185\_0311), as well as enzymes involved in nucleotide biosynthesis (Shew185\_1650-1651). Among these genes, 14 were also shown to be differentially expressed in previous transcriptomic characterization of OS185 and OS195 under various redox conditions using microarrays (Chapter 3) [Figure 5.10], indicating their functional importance in redox fluctuated environments.

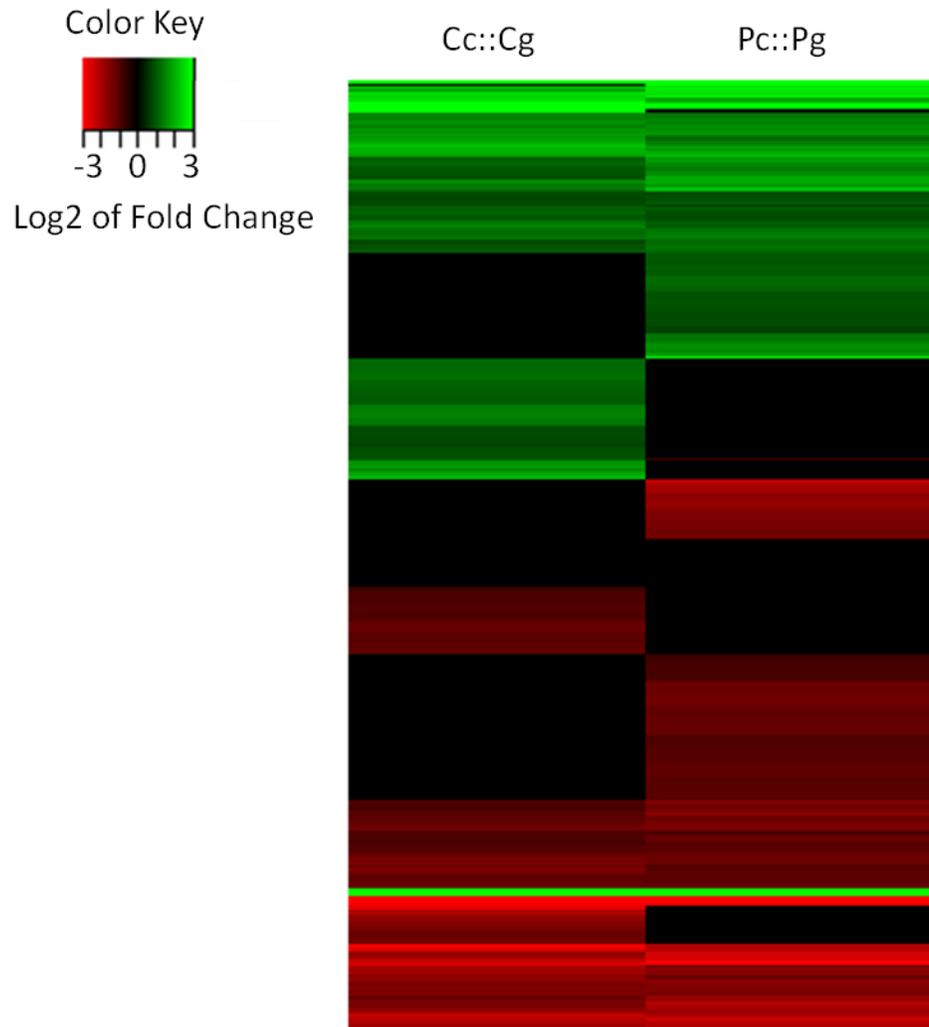


**Figure 5.10. COG categories of differentially expressed genes by OS185 in co-culture vs. pure culture (Cc::Pc).**

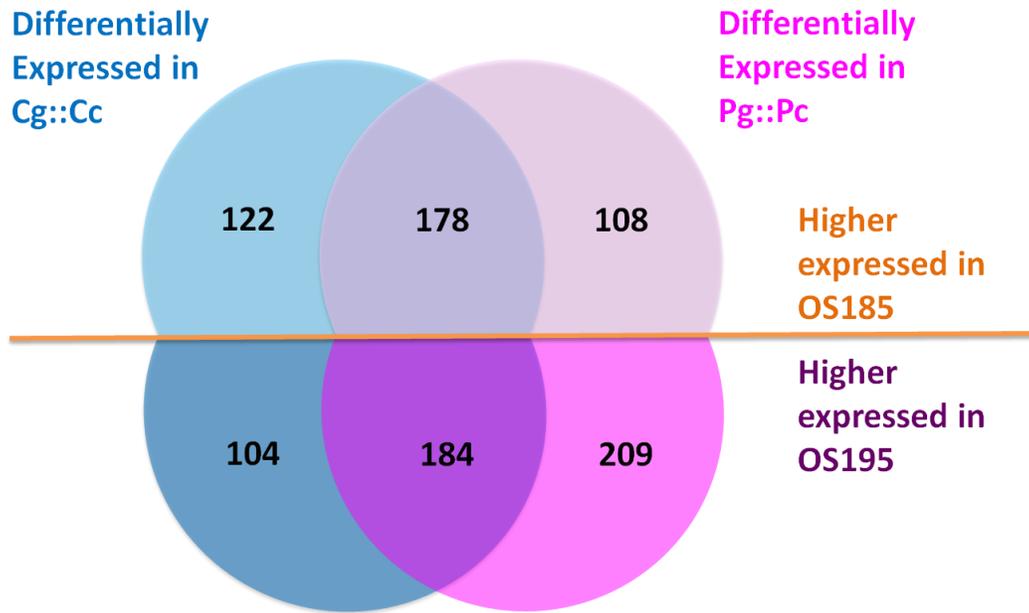
In contrast to those genes differentially expressed in OS195 under pure culture and co-culture conditions (the majority of which were within a potential phage insertion), most of the genes differentially expressed in OS185 encoded functional proteins involved in various pathways, and a number of them were previously shown to be important for anaerobic growth condition. This suggests that OS185 had a much higher response to the nitrate condition in co-

culture compared to OS195, and impacts from these differentially expressed genes could have also contributed to the competition outcome between the two strains.

**Comparison of expression levels of shared genes between OS185 and OS195 in pure culture and in co-culture (Pg::Pc and Cg::Cc).** In order to discern the general expression patterns among the samples, a heatmap was generated illustrating differentially expressed genes between OS185 and OS195 in pure culture and co-culture conditions [Figure 5.11]. The numbers of differentially expressed genes are also summarized in [Figure 5.12]. No conflicting expression patterns were observed for any gene (i.e., having higher expression in one strain under pure-culture condition, while having higher expression in the other strain under co-culture condition). This further supports consistency of the expression profiles.



**Figure 5.11.** Heatmap of differentially expressed genes between OS185 and OS195 in co-culture (Cc::Cg) and in pure culture (Pc::Pg). Green color indicates higher expression in OS185, and red indicates higher expression in OS195.



**Figure 5.12. Summary of numbers of differentially expressed gene shared by both OS185 and OS195 between the two strains in co-culture and in pure culture.**

About 25% of all differentially expressed genes failed to be assigned with any COG functional categories. Among the successfully annotated genes, and out of the three better studied classes of COG functional categories, i.e., information storage and processing, cellular processes and signaling and metabolism, COG categories related to metabolism accounted for a major fraction of the differentially expressed genes [Table 5.5]. Furthermore, other than the poorly characterized COG categories R and S, genes associated with amino acid transport and metabolism (E) were the most frequent, followed by those associated with energy production and conservation (C), signal transduction (T), transcription (K), and cell wall/ membrane/ envelope biogenesis (M) [Table 5.6].

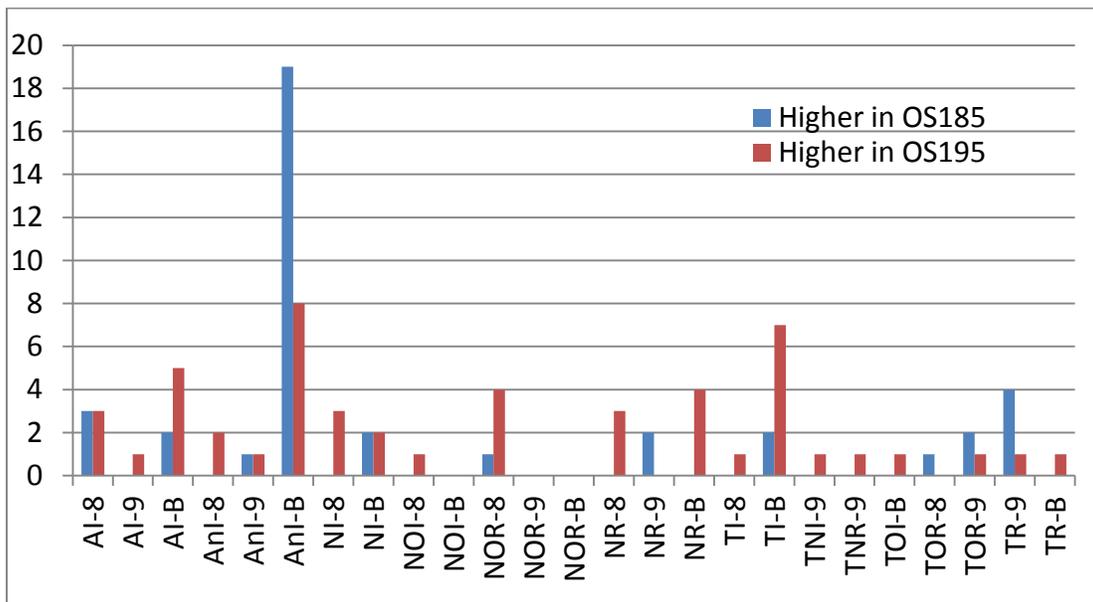
**Table 5.5. Number of Genes Classified in Different Cellular Functions**

	Co-Culture (Cg::Cc)		Pure Culture (Pg::Pc)	
	Higher Expressed in OS185	Higher Expressed in OS195	Higher Expressed in OS185	Higher Expressed in OS195
Information storage and processing	45	32	34	52
Cellular processes and signaling	62	58	65	84
Metabolism	79	93	70	129

**Table 5.6. Number of Genes Classified in Different COG Categories**

COG	Cc::Cg	Pc::Pg
D	3	4
U	9	13
F	11	15
Q	8	10
M	30	37
O	19	32
V	6	9
I	14	13
N	16	17
L	28	25
H	19	12
J	19	18
C	39	43
G	12	16
P	27	40
R	49	65
S	48	52
K	30	43
T	37	37
E	42	50
Unknown	159	179

**Differentially expressed genes in pure culture (Pc::Pg).** Among the 317 genes differentially expressed specifically in pure culture, 108 genes had higher expression in OS185 while the other 209 had higher expression in OS195, out of which 39 and 50 genes, respectively, were previously shown to be differentially expressed under redox conditions (Chapter 3) [Figure 5.13]. Although OS195 has more genes being highly induced, higher expressed genes in OS185 included more of those induced under anaerobic conditions (20 genes in OS185 compared to 11 genes in OS195, as in AnI-8, AnI-9 and AnI-B), while more highly expressed genes in OS195 included those more induced under thiosulfate respiring and aerobic conditions (10 and 14 genes in OS195, compared to 2 and 9 genes in OS185, respectively). These results suggested that, while cultured individually, OS185 reacted more actively towards anaerobic growth compared to OS195.



**Figure 5.13. Differentially expressed genes specific to Pc::Pg comparison.** The types of differential expression were described in [Table 3.6].

**Differentially expressed genes in both pure culture and co-culture (Pc::Pg, Cc::Cg).** One hundred seventy-eight genes showed higher expression in OS185 and 184 showed higher expression in OS195 in both co-culture (Cc::Cg) and pure culture comparisons (Pc::Pg) [Figure 5.12]. Consistent patterns in these differentially expressed genes indicated their importance under nitrate growth condition in both strains. These differentially expressed genes were further compared with the previously characterized differentially expressed gene pool using microarrays, where 38% of the genes with higher expression in OS185 were redox sensitive, whereas a lower percentage (27%) of redox sensitive genes were found in OS195. In addition, more anaerobic respiration related (AnI) genes were found among those with higher expression in OS185 (18 genes), contrasting to only nine in OS195 [Figure 5.14], further supporting a highly active transcriptional response to anaerobic conditions by OS185.

A closer look at functional characterization of differentially expressed genes reveals similar distribution of higher expressed genes by OS185 and OS195 in most COG categories [Figure 5.15]. However, only 4 genes with higher expression in OS185 were considered to be associated with inorganic ion transport and metabolism (P), contrasting to 16 of those with higher expression in OS195, including a number of membrane transporters and antiporter (Sbal195\_1047-1053, Sbal195\_1296), bacterioferritin (Sbal195\_3554-3556), as well as the *NrfD*-like reductase (Sbal195\_3997) in a thiosulfate inducible operon. Meanwhile, more genes (15 genes) associated with energy production and conversion (C) had higher expression in OS185 compared to those of OS195 (11 genes), including a number of proteins involved in mediation of electron transfer, such as cytochromes (Shew185\_4167, Shew185\_3505), ferredoxins (Shew185\_0311, Shew185\_0967), and anaerobic dehydrogenases (Shew185\_0099,

Shew185\_0615). Particularly, the [4Fe-4S] ferredoxin iron-sulfur binding domain protein (Shew185\_0311) was expressed 12.7 fold higher in OS185, suggesting its relative importance under nitrate respiration in strain OS185.

A number of genes were found to be strongly induced in OS185 or OS195. For example, two antibiotic activity-related genes (Shew185\_1137-1138) had 455 and 152 fold higher expression in OS185 compared to that in OS195, respectively. Genes in a type-F conjugative transfer system (Shew185\_4435, Shew185\_4439, and Shew185\_4441-4442) were also strongly induced in OS185, by 13 to 137 fold higher than in OS195. Similar patterns were also observed in genes related to a few cell surface structures and pilus biosynthesis (Shew185\_0491, Shew185\_4442, 6 and 79 fold higher in OS185), as well as several transposases and integrases (Shew185\_4403, Shew185\_4407, and Shew185\_0594, 26 to 360 fold higher in OS185). Compared with OS185, although OS195 has a greater number of more highly expressed genes in total, highly induced genes (more than 6 fold higher) were fewer than those in OS185 (29 genes in OS185, and 17 in OS195). Moreover, the most greatly induced genes in OS195 were mostly mobile genetic element associated, such as a number of transposases (Sbal195\_0589, Sbal195\_0593, 6.8 and 38.6 fold higher in OS195, respectively), putative prophage transcriptional regulator (Sbal195\_2021, 116 fold higher), and a plasmid partitioning protein (Sbal195\_4640, 1350 fold higher).

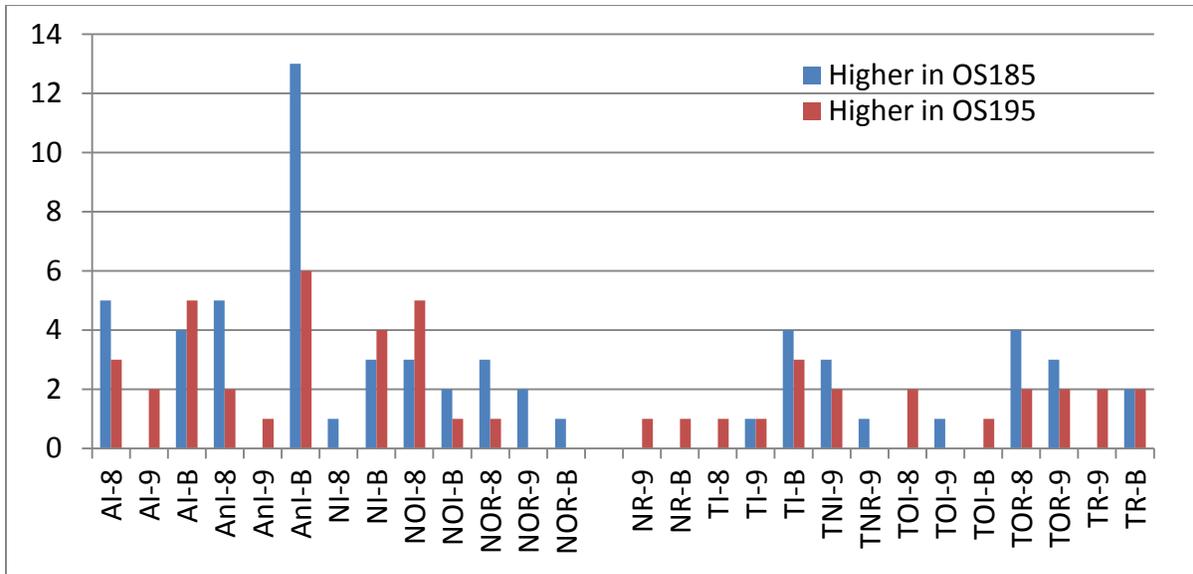


Figure 5.14. Differentially expressed genes from both Pg::Pc and Cg::Cc comparisons.

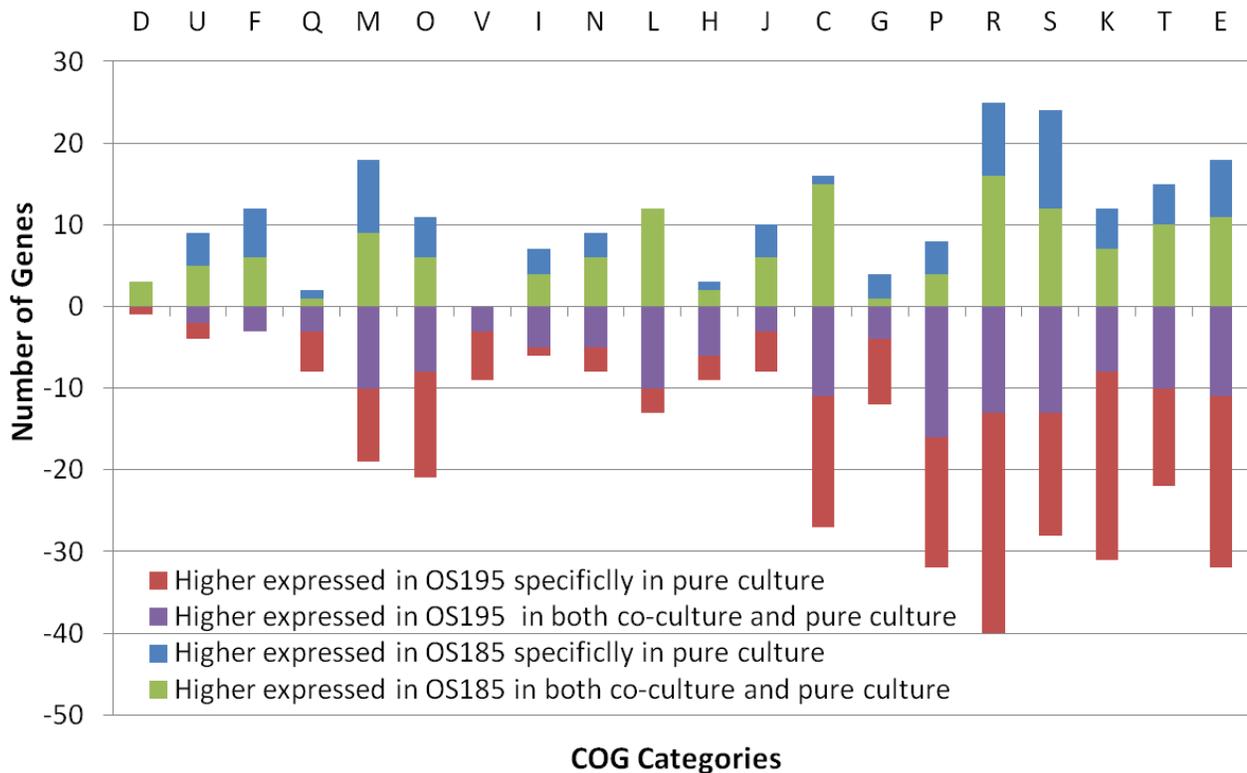
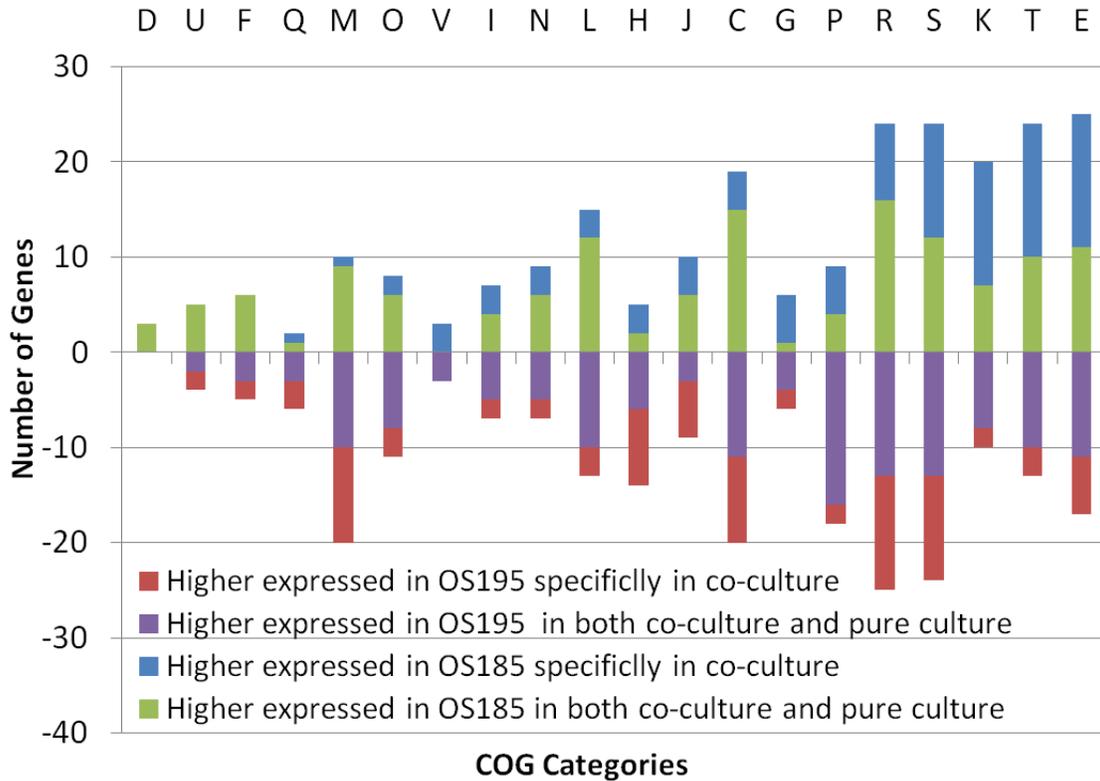


Figure 5.15. COG categories of differentially expressed genes shared by OS185 and OS195 from Pg::Pc comparison. Positive gene numbers correspond to genes with higher expression in OS185; Negative gene numbers correspond to genes with higher expression in OS195.

**Co-culture specific differentially expressed genes (Cc::Cg).** One hundred twelve genes exhibited higher expression in OS185 while 104 exhibited higher expression in OS195 exclusively in co-culture transcriptional comparison of the two strains [Figure 5.12]. The numbers of more highly expressed genes that fell into most COG categories were similar, except for the cell wall/membrane/envelope biogenesis related category (M), where 10 genes with higher expression in OS195 were found, contrasting to only 1 with higher expression in OS185 [Figure 5.16]. Moreover, 4 out of the 10 genes in the (M) category were previously shown to be up-regulated under aerobic condition. The situation was quite the opposite in a few other categories, including those associated with transcription (K), signal transduction mechanisms (T), as well as amino acid transport and metabolism (E), where more genes with greater expression in OS185 were found (13 vs. 2, 14 vs. 3, and 14 vs. 6 genes in OS185 and OS195, in K, T and E categories, respectively) compared to those in OS195. Remarkably, among these higher expressed genes in OS185, 11 transcriptional regulators, 4 periplasmic sensors (Shew185\_0052, 0253, 0259, 1002), as well as 2 chemotaxis sensory proteins (Shew185\_2245, Shew185\_4113) were found. Some of the transcriptional regulators were known to be essential under fluctuating redox conditions (28) (18), and moreover, two of them (Shew185\_0096, Shew185\_2371) were previously shown to be up-regulated by OS185 under nitrate-respiring condition. A number of nitrogen sensors and regulatory proteins were also included, i.e., a sensor histidine kinase for nitrogen response *NtrB* and the corresponding transcriptional regulator *NtrC* (Shew185\_0258-0259), as well as the nitrogen regulatory protein *GlnK\_1* (Shew185\_3708), suggesting highly active response of OS185 to nitrate-respiring condition in co-culture.



**Figure 5.16. COG categories of differentially expressed genes shared by OS185 and OS195 from Cg::Cc comparison.** Positive gene numbers correspond to genes with higher expression in OS185; Negative gene numbers correspond to genes with higher expression in OS195.

**RNA-Seq technology in transcriptomic analysis.** One of the key strategies to understand the ecology of an organism is through determining how it responds to its environment by altering gene expression. High-throughput Illumina sequencing of cDNAs provides a platform for comprehensively and quantitatively assessing the cellular response at the transcription level (29). Studies using this method have already altered our views of the extent and complexity of diverse transcriptomic systems (30) (31) (32). Being known as one of the Next Generation Sequencing (NGS) technologies and dubbed "a revolutionary tool for transcriptomics", Illumina RNA-Seq is capable of generating millions of short reads from little cDNA and offering much greater sequencing depth. In contrast to hybridization-based methods, sequence-based

approaches directly measure cDNA products at base-scale resolution. Compared to microarrays, RNA-Seq produces much lower background noise given that sequencing reads can be unambiguously mapped to specific regions of the genome. Also due to the great sequencing capacity, the RNA-Seq approach can detect a much larger dynamic range of differentiated expression levels.

In the first publication of bacterial expression profiles derived from RNA-Seq by Illumina NGS, transcriptional responses of two *Burkholderia cenocepacia* strains of the same lineage to two different conditions were examined (33). The two *B.cenocepacia* isolates (one from soil and the other from the sputum of a cystic fibrosis patient) were each grown under conditions mimicking soil and CF sputum, followed by sequencing of their transcriptomes. As a result, differences were not only found between the expression profiles of cells grown in two different media, but also found between expression profiles of the two strains under the same conditions, suggesting probable strain specific adaptations to the niches from which they came. A number of previously unknown putative non-coding RNAs (ncRNAs) were also found to be preferentially up-regulated under conditions mimicking the soil environment, indicating potential roles of ncRNAs in specializing to the soil environment.

In this study, Illumina sequencing of whole transcriptomes of the two *S. baltica* strains yielded 15 to 35 times coverage (based on number of mapped reads) of the whole genome of OS185 and OS195, making possible detection of transcripts with low copy numbers. Fold change between differentially expressed genes of up to 1350 times was identified, compared to maximum fold change of 648 times in differential expression analysis using microarrays. Over 96.6% of all genes in OS195 and 97.4% of all genes in OS185 were represented by at least a

single read, while fewer than 100 singletons were revealed in each replicate sample, further demonstrating the great sequencing depth of this method. Due to efficient removal of RNA with secondary structure, reads that mapped to rRNA, tRNA and ncRNAs were very limited (fewer than 6% in total in each replicate sample), rendering biased any quantification of these types of RNAs. Thus, we were not able to quantitatively assess the relative abundance of these RNA types between samples. On the other hand, given a large number of reads from mRNA sequences, these mRNA reads should be able to well represent the true transcriptomic profiles of the samples.

**Mechanisms of differential fitness.** Similar to the RNA-Seq study of the two *B.cenocepacia* isolates, we also aimed to investigate the mechanisms underlying specialization in their own environmental niches by these two *S. baltica* strains. In addition, we took into consideration the relative fitness of the two strains while grown together and competed with each other, which leads to interpreting the specializing outcome of the two strains. Given that OS185 was able to outcompete OS195 by the end of the 7<sup>th</sup> day (45 generations), the mean relative fitness of OS185 over OS195 was estimated to be 1.05 (34), meaning that OS185 was able to grow under the nitrate condition at a rate 5% faster than OS195 on a daily base. Although it is by far not clear which factors caused this difference between the two strains under the nitrate amended condition, several possibilities were proposed based on the transcriptomic profile of the strains. Strikingly, one possibility of OS195 being outcompeted may be associated with induction of expression of phage components by OS185 in the co-culture. However, based upon previous knowledge, phage induction, often associated with initiation of the lytic cycle, was most

commonly triggered by exposure to UV radiation or other physico-chemical stress factors (35) (36), whereas the case of phage induction by another organism is rarely known. Also, although it has long been known that bacteria can produce toxins and release them into their surroundings to inhibit growth of other bacteria (37), introducing competitive disadvantage via altering expression in the other organism, i.e., inducing expression of phage components, has not been reported. Thus, further experimental validation and investigation would provide important evidence to these possibilities, and may contribute to a re-evaluation of the role of bacteriophages in competition in the natural environment. The other possibility that might have contributed to the competition outcome was more effective transcriptional regulation in the strain OS185. From all levels of comparisons, i.e., Cc::Cg, Pc::Pg, as well as Cg::Pg, genes potentially important for redox specialization were found to be differentially regulated in OS185. Moreover, compared with transcription profiles of OS195, OS185 appeared to have more active response to the nitrate respiring condition on the gene expression level, which further explains the adaptation of OS185 to a nitrate-rich environment. Although detailed comparison on the gene level was often limited by the quality and quantity of gene annotation available, and verification of importance of differentially expressed genes through molecular biochemical studies is still needed, this study establishes the groundwork for future mechanistic investigation of niche adaptation of the *S. baltica* population, and meanwhile provided an effective method for studying interaction among microorganisms.

## REFERENCES

## References

1. Cohan FM & Koeppel AF (2008) The Origins of Ecological Diversity in Prokaryotes. *Current Biology* 18(21):R1024-R1034.
2. Koonin EV (2008) Darwinian evolution in the light of genomics. *Nucleic Acids Research* 37(4):1011-1034.
3. Weinbauer MG & Rassoulzadegan F (2003) Are viruses driving microbial diversification and diversity? *Environmental Microbiology* 6(1):1-11.
4. Cohan FM & Perry EB (2007) A Systematics for Discovering the Fundamental Units of Bacterial Diversity. *Current Biology* 17(10):R373-R386.
5. Koeppel A, *et al.* (2008) Identifying the fundamental units of bacterial diversity: A paradigm shift to incorporate ecology into bacterial systematics. *Proceedings of the National Academy of Sciences* 105(7):2504-2509.
6. Hunt DE, *et al.* (2008) Resource Partitioning and Sympatric Differentiation Among Closely Related Bacterioplankton. *Science* 320(5879):1081-1085.
7. Johnson ZI (2006) Niche Partitioning Among Prochlorococcus Ecotypes Along Ocean-Scale Environmental Gradients. *Science* 311(5768):1737-1740.
8. Coleman ML (2006) Genomic Islands and the Ecology and Evolution of Prochlorococcus. *Science* 311(5768):1768-1770.
9. Dufresne A, *et al.* (2008) Unravelling the genomic mosaic of a ubiquitous genus of marine cyanobacteria. *genome biology* 9(5):R90.
10. Swingley WD, *et al.* (2008) Niche adaptation and genome expansion in the chlorophyll d-producing cyanobacterium *Acaryochloris marina*. *Proceedings of the National Academy of Sciences* 105(6):2005-2010.
11. Jacobsen A, Hendriksen RS, Aaresturp FM, Ussery DW, & Friis C (2011) The *Salmonella enterica* Pan-genome. *Microbial Ecology*.
12. Jauregui F, *et al.* (2008) Phylogenetic and genomic diversity of human bacteremic *Escherichia coli* strains. *BMC Genomics* 9(1):560.
13. Mahenthiralingam E, Baldwin A, & Dowson CG (2008) Burkholderia cepacia complex bacteria: opportunistic pathogens with important natural biology. *Journal of Applied Microbiology* 104(6):1539-1551.

14. Kaiser S, Biehler K, & Jonas D (2009) A *Stenotrophomonas maltophilia* Multilocus Sequence Typing Scheme for Inferring Population Structure. *Journal of Bacteriology* 191(9):2934-2943.
15. Stomp M, *et al.* (2004) Adaptive divergence in pigment composition promotes phytoplankton biodiversity. *Nature* 432:104-107.
16. Denev VJ, *et al.* (2010) Inaugural Article: Proteogenomic basis for ecological divergence of closely related bacteria in natural acidophilic microbial communities. *Proceedings of the National Academy of Sciences* 107(6):2383-2390.
17. Belotte D, Curien J-B, Maclean RC, & Bell G (2003) An experimental test of local adaptation in soil bacteria. *Evolution* 57(1):27-36.
18. Fredrickson JK, *et al.* (2008) Towards environmental systems biology of *Shewanella*. *Nature Reviews Microbiology* 6(8):592-603.
19. Elmgren R (1989) Man's Impact on the Ecosystem of the Baltic Sea: Energy Flows Today and at the Turn of the Century. *AMBIO*:326-332.
20. Höfle MG & Brettar I (1996) Genotyping of heterotrophic bacteria from the central Baltic Sea by use of Low-Molecular-Weight RNA profiles. *Applied and Environmental Microbiology* 62(4):1383-1390.
21. Brettar I (2002) *Shewanella denitrificans* sp. nov., a vigorously denitrifying bacterium isolated from the oxic-anoxic interface of the Gotland Deep in the central Baltic Sea. *International Journal of Systematic and Evolutionary Microbiology* 52(6):2211-2217.
22. Ziemke F, Brettar I, & Höfle MG (1997) Stability and diversity of the genetic structure of a *Shewanella putrefaciens* population in the water column of the central Baltic. *Aquatic Microbial Ecology* 13:63-74.
23. Victor De Lorenzo, Marta Herrero, Ute Jakubzik, & Timmis AKN (1990) Mini-TnS Transposon Derivatives for Insertion Mutagenesis, Promoter Probing, and Chromosomal Insertion of Cloned DNA in Gram-Negative Eubacteria. *Journal of Bacteriology* 172:6568-6572.
24. Lambertsen L, Sternberg C, & Molin S (2004) Mini-Tn7 transposons for site-specific tagging of bacteria with fluorescent proteins. *Environmental Microbiology* 6(7):726-732.
25. Langmead B, Trapnell C, Pop M, & Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *genome biology* 10(3):R25.
26. Simon Anders WH (2010) Differential expression analysis for sequence count data. *genome biology* 11(10):R106.

27. Caro-Quintero A, *et al.* (2010) Unprecedented levels of horizontal gene transfer among spatially co-occurring *Shewanella* bacteria from the Baltic Sea. *The ISME Journal* 5:131–114.
28. Beliaev AS, *et al.* (2005) Global Transcriptome Analysis of *Shewanella oneidensis* MR-1 Exposed to Different Terminal Electron Acceptors. *Journal of Bacteriology* 187(20):7138-7145.
29. Wang Z, Gerstein M, & Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics* 10:57-63.
30. Tariq MA, Kim HJ, Jejelowo O, & Pourmand N (2011) Whole-transcriptome RNAseq analysis from minute amount of total RNA. *Nucleic Acids Research* 39(18):e120-e120.
31. Malone JH & Oliver B (2011) Microarrays, deep sequencing and the true measure of the transcriptome. *BMC Biology* 9(34):1741-7007.
32. Croucher NJ & Thomson NR (2010) Studying bacterial transcriptomes using RNA-seq. *Current Opinion in Microbiology* 13(5):619–624.
33. Otto M, Yoder-Himes DR, Konstantinidis KT, & Tiedje JM (2010) Identification of Potential Therapeutic Targets for *Burkholderia cenocepacia* by Comparative Transcriptomics. *PLoS ONE* 5(1):e8724.
34. Lenski RE, M. R. Rose, S. C. Simpson, and S. C. Tadler (1991) Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2,000 generations. *American Naturalist* 138:1315-1341.
35. Baluch J & Sussman R (1978) Correlation between UV dose requirement for lambda bacteriophage induction and lambda repressor concentration. *Journal of Virology* 26(3):595-602.
36. Choi J KS, Goel R (2010) Various physico-chemical stress factors cause prophage induction in *Nitrosospira multififormis* 25196--an ammonia oxidizing bacteria. *Water Research* 44(15):4550-4558.
37. Hibbing ME, Fuqua C, Parsek MR, & Peterson SB (2009) Bacterial competition: surviving and thriving in the microbial jungle. *Nature Reviews Microbiology* 8(1):15-25.

## **CHAPTER VI**

### **SUMMARY AND OUTLOOK**

With the improving ability to differentiate and characterize bacterial strains, as well as the development of powerful genomic technologies, microbiologists today are able to make more significant advances in the fast paced field of environmental microbiology. The work presented in this thesis has taken advantage of past and current technologies, along with the unique biological resources and ecological data, aiming to provide further understanding of bacterial specialization.

The *Shewanella* represent an interesting group of bacteria due to their respiratory versatility, their ability to thrive in many environments, as well as their genetic diversity and genomic flexibility (1). All these properties provide *Shewanella* with greater ability to adapt especially to fluctuating redox environments. While a number of *Shewanella* strains have been well studied and characterized, in part due to their tremendous bioremediating potential, a number of new strains continue to be discovered.

*S. baltica* represents a bacterial population that was persistent in the Baltic Sea where it played a major role in the nitrogen cycle in the last half of the 20<sup>th</sup> century when nitrogen and organic carbon inputs to the Baltic Sea increased greatly (2) (3). With their important roles in the in the Baltic environment and their great potential for use in studying short-term evolution, a number of *S. baltica* genomes were sequenced by DOE's Joint Genome Institute, with four genomes finished and released during 2007-2008, and an additional five released during 2010-2011. Although analysis of the newly available genomes was not included in this study, these genomes will continue to provide valuable insights into mechanistic studies of *S. baltica* evolution.

Chapter 1 framed the comparative genomic, phenotypic and transcriptomic analyses for the subsequent chapters, and provided an overview of *Shewanella* as well as the history behind *Shewanella baltica* species. Chapter 1 also supplied an evolutionary perspective on bacterial specialization, and signified the value of the *S. baltica isolate* collection in advancing understanding of evolutionary strategies and mechanisms.

Chapter 2 examined the genetic relatedness among *S. baltica* lineages, and made the initial correlation of these isolates with the environmental gradients. It also tackled the question of the levels of recombination within the *S. baltica* population, and extended a previous finding on two strains on which extensive recombination has occurred in the recent past. This estimation is important for population genetics studies because it reflects the evolutionary paths, i.e., divergent or convergent evolution, of the population. Thus, the *S. baltica* population in general, can be regarded as undergoing diversification. Further, through comparing MLST phylogeny with the broader scale similarity of CGH presented in Chapter 3, the new gene selection strategy used for this MLST analysis was substantiated, which could benefit MLST study in other microbial systems as well.

Chapter 3 coped with the question of genomic variability within the *S. baltica* species. In addition to the CGH assays, transcriptomic characterization also helped identify genes of functional importance under fluctuating redox conditions. Core genome analyses were performed at levels of both the entire *S. baltica* population and within specific lineages representing isolates from potentially different environmental niches. Although the analysis was to some extent limited by lack of quality gene annotations, the large proportion of genes,

both hypothetical and annotated, that are associated with mobile genetic elements indicated the capacity of these organisms for harboring new genes and functions through means of HGT.

Chapter 4 described the efforts in seeking for evidence of specialization on the phenotypic level. Given that *Shewanella* species are often associated with organic-rich, electron acceptor varying environment, I examined the carbon source utilization and respiratory versatility of these *S. baltica* strains. Importantly, the strains isolated in 1998 used more carbon sources than those from 1986/87, which could reflect a change in nutrient availability with time in the Baltic Sea. One other important task in microbial genetics is to reveal to which extent genomic information can reflect phenotypic properties. Chapter 4 attempted to correlate phenotypic similarity with genetic relatedness at the whole genome level and at the individual gene level, and found a few features that appeared to be explained by particular genes. Beyond this thesis, phenotypic profiles will continue to assist functional discovery of genes and pathways, and to improve gene annotation of *Shewanella* genomes (4).

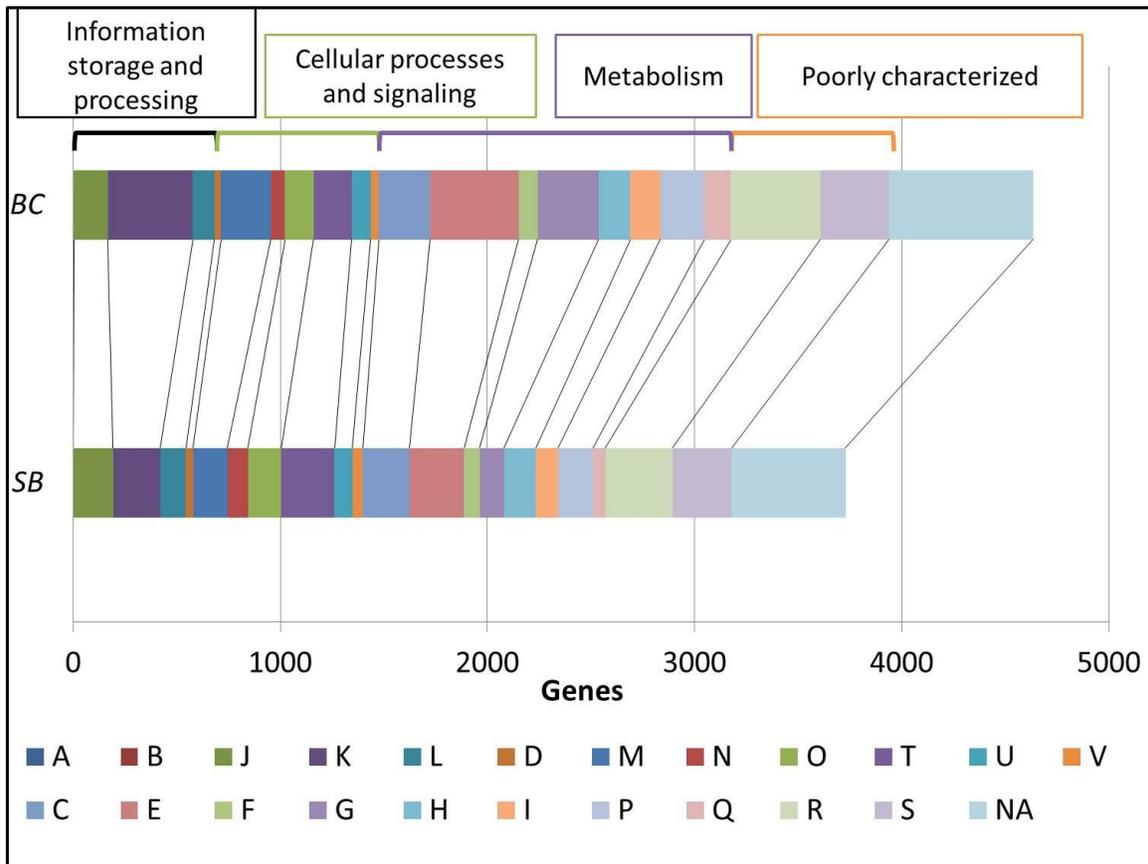
For Chapter 5, I sought to determine the basis of competitive advantage in competing strains under conditions mimicking their own environments, and found differential fitness of strains under aerobic and anaerobic nitrate respiring condition. These model studies of course do not accurately reflect the organisms' natural environment, but they do better control conditions and isolate variables that affect fitness of strains. Probably the most pace-setting aspect of this work was to extend the competition analysis to the transcriptome level, which revealed distinguishable patterns of differential gene regulation in the competed strains. While analysis of these data is not complete, I did find induction of genes for phage production in the out-competed strain and higher expression of genes related to anaerobic growth in the favored

strain, which was consistent with the competitive outcome. In addition, a phage repressor (Sbal195\_2927) and one other transcriptional regulator (Sbal195\_2888) were identified in the phage island. Hence, a next step would be to do bioinformatic searches to find potential regulatory mechanisms, i.e., regulatory proteins or nucleotides that may bind to these regulator genes, if found, their function could then be tested by biochemical assays.

Further work would likely enhance the explanatory connections among the genomics, phenotypes, and the ecology. For examples, in order to further investigate the relation between genomics and phenotypes, one can obtain the Biolog metabolic profile and try to match this physiology with metabolic pathways encoded in the genomes. This was recently successfully accomplished for different species of *Shewanella* (4), and it should be extendable to strains within a species, namely, the *S. baltica* isolates. Meanwhile, the CGH dataset of *S. baltica* strains will provide a useful reference and additional support for functional predictions of genes and pathways based on this analysis. This strategy was shown to accelerate functional discovery of genes and will be also utilized to refine gene annotation of the *S. baltica* species, which will no doubt improve our understanding about to what extent genomic information can be used to infer phenotypic properties.

Investigation in gene composition of populations occupying different niches will provide important insights in evaluating the impact of ecology on bacterial population, as reflected from their genomes, and determining whether we can use genomics to predict the environmental niches of individual strains. To further illustrate this potential, Erick Cardenas (who studies *Burkholderia*) and I compared the conserved gene core of *Shewanella baltica* and *Burkholderia cenocepacia*. *Burkholderia* are Betaproteobacteria with incredible phenotypic

diversity, likely owing to their large, multi-replicon genomes. Although *B. cenocepacia* is primarily known as a pathogen in cystic fibrosis patients, many of *B. cenocepacia* strains are of environmental origin and are able to colonize diverse niches such as plant rhizospheres, soil, water, insects and even mammals (5) (6) (7) (8). The gene core of *B. cenocepacia* was also obtained through CGH in a parallel study (Cardenas, unpublished data), and was used to compare with that of the *S. baltica* strains as described in Chapter 3. The analysis is so far only at the level of comparing COG functional categories, but has already revealed interesting signatures of functional differences [Figure 6.1]. Given the versatility in inhabiting diverse environments by *B. cenocepacia*, it is not surprising that, compared to the gene core of *S. baltica*, *B. cenocepacia* has more genes involved in transcription (K), amino acid transport and metabolism (E), and carbohydrate transport and metabolism (G), as well as secondary metabolism (Q). On the other hand, the core of *S. baltica* has more genes involved in signal transduction (T), energy production (C), coenzyme transport and metabolism (H), cell motility (N), and post-translational modification, protein turnover and chaperones (O), which agreed with the nature of *Shewanella* of being respiratory specialists and occupying niches with rich and fluctuated energy sources. These results indicated that the ecology of a bacterial population can be reflected from particular groups of functional genes in their genomes. To gain perspectives of functional specialization in more bacterial groups that inhabit more diverse niches, we plan to extend this analysis to more bacterial species, such as *Escherichia coli* and certain *Prochlorococcus* groups.



**Figure 6.1. COG functional comparison of *S. baltica* gene core and *Burkholderia cenocepacia* gene core.**

A larger set of genome sequences is needed to further evaluate how ecology impacts the evolutionary paths of the *S. baltica* population. In this regard, the sequenced genomes of five more strains of *S. baltica* are now available. Recombination analysis on the single gene level can be biased in inferring evolutionary paths, while on the whole genome level, it can more conclusively reveal evolutionary trajectories. For example, in a recent comparative genomic study of *Escherichia coli* strains, higher levels of recombination were found among strains from more similar environments (9). Thus, through analyzing the extent of recombination among *S. baltica* strains within similar and between more different redox zones, we may learn more

about how stratification of the water environment in the Baltic Sea impacted the population's genomic structure.

In advancing our understanding of the connections between phenotypes and ecology of *S. baltica*, larger-scale competition assays can be performed to test differential fitness among a larger number of strains. These assays can be conducted in a high throughput manner, for example, by culturing mixed cells in 96-well plates in anaerobic chamber, with medium containing desired electron acceptors. Results from these assays will help us to know about which environmental parameters are of importance in determining differential fitness among the strains.

Finally, as we continue to explore specialization of microorganisms, it is important to remember that microbes do not live by themselves. Rather, communities are composed of coexisting organisms including Bacteria, Archaea, eukaryotes, and viruses. Hence it is possible that through learning more about associated communities, as well as understanding those organisms in the communities, we will come to more comprehensive understanding about the studied microorganisms. For instance, in this case, 16S rRNA pyrotag sequencing of the community DNA from the original water samples can provide information on composition of the bacterial community and an estimation of the proportion of the entire bacterial population that belongs to the *S. baltica* species (10). Meanwhile, pyrotag sequencing of functional genes, for example, those involved in nitrate reduction pathways, should provide insights into the diversity of nitrate-reducing genes and, especially in conjunction with 16S pyrosequencing community structure information, will be informative about which bacterial groups play important roles in nitrate reduction in the Baltic Sea (11) (12).

Shotgun metagenomic sequencing should provide even greater depth of information on the Baltic Sea community, but is more resource demanding both for sequencing and massive data analysis. Nonetheless, it would provide much more sequence and with some assembly, larger fragments of community DNA, enabling further analysis of a larger number of functional genes. In particular, shotgun metagenomic sequencing in combination with bait capture technology (13) using *S. baltica* DNA would enable the sequencing to be targeted to only the *S. baltica* genomes from the community DNA, which should allow further analysis of the diversity of *S. baltica* genes as well as of recombination between *S. baltica* genomes. In summary, metagenomics in its ultimate vision should provide further understanding of the function, structure and ecology of the natural community.

## REFERENCES

## References

1. Fredrickson JK, *et al.* (2008) Towards environmental systems biology of *Shewanella*. *Nature Reviews Microbiology* 6(8):592-603.
2. Elmgren R (1989) Man's Impact on the Ecosystem of the Baltic Sea: Energy Flows Today and at the Turn of the Century. *AMBIO*:326-332.
3. Brettar I & Höfle MG (1993) Nitrous oxide producing heterotrophic bacteria from the water column of the central Baltic: abundance and molecular identification. *MARINE ECOLOGY PROGRESS SERIES* 94:253-265.
4. Rodrigues JLM, Serres MH, & Tiedje JM (2011) Large-Scale Comparative Phenotypic and Genomic Analyses Reveal Ecological Preferences of *Shewanella* Species and Identify Metabolic Pathways Conserved at the Genus Level. *Applied and Environmental Microbiology* 77(15):5352-5360.
5. Coenye T & Vandamme P (2003) Diversity and significance of *Burkholderia* species occupying diverse ecological niches. *Environmental Microbiology* 5(9):719-729.
6. Mahenthiralingam E, Baldwin A, & Dowson CG (2008) *Burkholderia cepacia* complex bacteria: opportunistic pathogens with important natural biology. *Journal of Applied Microbiology* 104(6):1539-1551.
7. Vial L, Chapalain A, Groleau M-C, & Déziel E (2011) The various lifestyles of the *Burkholderia cepacia* complex species: a tribute to adaptation. *Environmental Microbiology* 13(1):1-12.
8. Mahenthiralingam E, Urban TA, & Goldberg JB (2005) The multifarious, multireplicon *Burkholderia cepacia* complex. *Nature Reviews Microbiology* 3(2):144-156.
9. Chengwei Luo STW, David M. Gordon, Michael Feldgarden, James M. Tiedje, and Konstantinos T. Konstantinidis (2011) Genome sequencing of environmental *Escherichia coli* expands understanding of the ecology and speciation of the model bacterial species. *Proceedings of the National Academy of Sciences* 108(17):6.
10. Petrosino JF, Highlander S, Luna RA, Gibbs RA, & Versalovic J (2009) Metagenomic Pyrosequencing and Microbial Identification. *Clinical Chemistry* 55(5):856-866.
11. Demaneche S, *et al.* (2008) Characterization of Denitrification Gene Clusters of Soil Bacteria via a Metagenomic Approach. *Applied and Environmental Microbiology* 75(2):534-537.
12. Iwai S, *et al.* (2009) Gene-targeted-metagenomics reveals extensive diversity of aromatic dioxygenase genes in the environment. *The ISME Journal* 4(2):279-285.

13. Gnirke A, *et al.* (2009) Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nature Biotechnology* 27(2):182-189.