UNIVERSITY OF CALIFORNIA, SAN DIEGO

Genomic Analysis of High Pressure Adaptation in Deep Sea Bacteria

A thesis submitted in partial satisfaction of the requirements for the degree Master of Science

in

Biology

by

Taylor Kristen Stratton

Committee in charge:

Professor Douglas H. Bartlett, Co-Chair
Professor Eric E. Allen, Co-Chair
Professor Milton H. Saier

2008

The Thesis of Taylor Kristen Stratton is approved and it is acceptable in quality and form for publication on microfilm:

_____

_____
Co-Chair

_____
Co-Chair

University of California, San Diego
2008

Table of Contents

List of Figures

List of Tables

Acknowledgements

The completion of this thesis would not have been possible without the help of many amazing people.

First of all, I need to thank my advisor, Doug Bartlett, who was kind enough to give me a position in his lab and patiently lead me through my first foray in to science research. His support and encouragement were a great motivation through both the successes and troubles that I encountered in my two years under his advisement. His constant demonstration of dedication and enthusiasm will continue to inspire me.

Thanks, also, to my other committee members, Eric Allen and Milton Saier, whose guidance and support were a great help.

I would like to thank all members of the Bartlett lab for making it fun to show up everyday. A special thanks must go to Federico Lauro, Emiley Eloe, and Ian Kerman who were integral to the work in this thesis. They have been an immense help in these two years and I would have never survived without them.

Thank you to Sheila Podell and Roger Chastian for their assistance and technical expertise.

Finally, thank you to my friends and family, who have always been a source of love, support, and encouragement.

ABSTRACT OF THE THESIS

Genomic Analysis of High Pressure Adaptation in Deep Sea Bacteria

by

Taylor Kristen Stratton

Master of Science in Biology

University of California, San Diego, 2008

Professor Douglas H. Bartlett, Co-Chair

Professor Eric E. Allen, Co-Chair

Though the extreme conditions of the deep sea cover the majority of the Earth's biosphere, the adaptations of organisms that grow there remain a mystery. Of particular interest is adaptation to high pressure, which acts as a unique stress on many physiological processes. The use of genomics has been vital to the study of piezophilic

microbes, providing detailed information where traditional lab studies are impractical. This thesis describes genome assemblies and analyses related to deep sea microbiology. The genome assembly of *Carnobacterium* sp. strain AT7 is the first sequencing project of a gram positive piezophile. The genome consists of a single circular chromosome and two plasmids, including genes thought to be involved in heavy metal, phage, and antimicrobial resistance. *Psychromonas* sp. strain CNPT3 was the first isolated piezophile and the finishing and annotation of its genome serves as a foundation for the study of deep sea adaptation. The manual annotation of the shallow water strain, *Photobacterium profundum* strain 3TCK, is presented and is used as a piezosensitive comparison to the piezophile model organism *Photobacterium profundum* strain SS9. The comparison shows that the deep sea strain has more genes dedicated to metabolism, cellular processes and mobile elements. Protein adaptation to pressure was explored by examining amino acid usage of piezophilic and piezotolerant strains in proteins of pressure-sensitive processes. The primary structures revealed no universal trends in protein adaptation. This thesis provides a foundation for future comparative analyses, necessary to the understanding of deep sea adaptation.

Introduction

The Deep Sea

Although it comprises more than 70% of the Earth's biosphere, the deep sea and the life within it remain an enigma to those who study it. The unchanging conditions in the deep are characterized by low temperature, the absence of light, little carbon influx, and high hydrostatic pressure (Lauro and Bartlett, 2007). Organisms native to this environment demonstrate adaptations to these harsh conditions. The effects of many of these stresses on the organisms have been well studied. Deep sea microbes lack UV repair systems needed for light exposure (Lauro et al, 2007; Vezzi et al, 2005), have membrane and enzyme adaptations to cold temperature (Delong et al, 1997; Lauro et al, 2007), and multiple metabolic pathways for the spotty nutrient availability (Vezzi et al, 2005; Wang et al, 2008), but the adaptations necessary to life at high pressure are not well understood.

High hydrostatic pressure has a profound impact on the distribution of organisms in the ocean (Lauro and Bartlett, 2007). Pressure can reach as much as 110MPa in the deepest parts of the ocean and its gradients with depth are thought to be responsible for the vertical zonation of the ocean's inhabitants (Delong et al, 2006; Lauro, 2007). Early scientists believed that life was unable to survive the extreme conditions of the deep sea and that microbial life found there arrived via falling particles but are not adapted to growth at high pressure (Simonato et al, 2006). In 1956, Zobell and Morita uncovered bacterial communities living in the deep sea that grew preferentially at high hydrostatic pressure, confirming the existence of deep sea adapted microbes. Not until the work of Yayanos and colleagues nearly 30 years later was a high pressure adapted microbe

isolated in culture (Yayanos et al, 1979). The difficulties in isolation and culturing the organisms make progress toward understanding them slow and expensive. The term "piezophile" (from the greek *piezo* meaning to press, and *philo* meaning love) is now used to describe microorganisms who demonstrate optimal growth at pressures greater than 0.1MPa (Yayanos et al, 1995).

Microbial diversity is oftentimes underestimated and misrepresented due to the difficulty in culturing the majority of microbes in an environmental sample. The "great plate anomaly" expresses this phenomenon by stating that less than 1% of the total bacteria can be enumerated by culturing methods (Staley and Konopka, 1985). Sampling from the deep sea is additionally challenging because it requires specialized and expensive equipment to obtain and maintain samples from extreme depths. Improvements in culturing techniques, including the pressurized culture vessels, have increased the success of isolating and culturing piezophilic microbes. However, the number of piezophilic isolates remains small and physiological knowledge is limited. The slow growth and necessity for high pressure/low temperature culturing equipment limits the field of piezomicrobiology to a small number of individuals. The isolates that are in culture have limited phylogenetic diversity (primarily from the gamma-proteobacteria), though metagenomic analyses show that these deep sea communities are far more populated and diverse than originally anticipated (Sogin et al, 2006).

Non-culture based techniques, such as molecular biology, genetics, and genomics have proven to be useful in studying microbes, particularly those from extreme environments. Genomic and genetic studies give insight to the biology of the prokaryotes and prokaryotic communities, elucidating such information as metabolic capabilities,

regulatory processes, and ecological interactions. A genome refers to all of the genetic material in an organism. Genomics, then, is the study of the genome, or the study of the genes and gene functions of an organism. Genomic analyses are often subdivided in to comparative genomics, which studies the relatedness of complete genomes of different organisms, and functional genomics, which determines the function of genes on a genome-wide scale (Wren, 2006). Analyses of genomic data can provide information about taxonomy, physiology, and metabolic capabilities in organisms and communities that are otherwise impractical to study. These investigations provide a basic understanding of the biological properties of the organism as well as forming a framework for future research. Comparative genomics can be particularly informative as it allows the identification of strain-specific regions which can be used to assess particular phenotypes and adaptations (Simonato et al, 2006). Microarray technology has also been combined with genomes to improve understanding of deep-sea adaptations by elucidating gene regulation in piezophilic and piezosensitive organisms at high pressure (Campanaro et al, 2005). Future studies should lead to increasing the quality, number, and annotation of available genomes as well as expanding in to transcriptomics and proteomics to determine expression and regulation of genes and proteins under varying conditions.

Hydrostatic pressure is a highly influential stressor, yet very little is understood about the adaptations it elicits. High pressure effects are exerted on every cellular process that involves a positive volume change, affecting the genome on a global scale (Campanaro et al, 2005). However, the genomic changes necessary to confer piezophily remain unknown. Several categories of genome changes have been suggested as

necessary for high pressure adaptation. These include the presence of key genes, the absence of genes that inhibit growth at high pressure, select or global gene sequence changes, and appropriate gene regulation (Lauro, 2007). Comparative genomic studies have deduced that no single genes exist that enable high pressure growth (Lauro, 2007). Other genomic investigations on piezopiles have searched for patterns in both global and gene-specific alterations across phenotypic and phylogenetic lines (Bidle and Bartlett, 1999; Saito et al, 2006; Saito and Nakayama, 2004; Welch and Bartlett, 1997). Thus far, universal patterns have not been uncovered, due in part to the lack of available data.

Expansion of the genomic data set is essential to bioinformatics accuracy and future research on everything from deep sea adaptation to general cellular processes. The genomes of deep sea microbial communities have implications that extend beyond piezophily. Microbes in the deep sea are thought to be ancient life forms that shape global biogeochemical processes (Sogin et al, 2006). Understanding the biology of the deep sea is necessary to the understanding ecological cycles as a whole. Knowledge provided by genomic analyses, including that regarding biological capabilities and requirements, can assist in isolation and culturing methods in new strains, thereby leading the way for future studies.

Studying microbial life in the deep sea has many potential applications. High hydrostatic pressure is a unique stress, adaptation to which is an area of much intrigue. Enzyme and cellular function under high pressure have been points of interest in biotechnology fields such as food processing, pharmaceuticals, and medicine (Simonato et al, 2006). As with any environmental niche, deep sea communities have the potential to provide insight into ecological, environmental, and cellular processes. The roles of

these communities in global cycles are currently not well understood. Some have suggested turning to the deep sea as a source of carbon sequestration to alleviate the pressures of climate change, but without further understanding of the biological processes in this region, the consequences cannot be adequately predicted.

In this thesis, I expand the annotation of genomes from organisms relevant to deep sea microbiology. It describes several aspects of genomics on multiple bacterial strains: genome assembly and closure of *Carnobacterium* sp. strain AT7, sequencing and finishing of the *Psychromonas* sp. strain CNPT3 genome, and manual annotation of the *Photobacterium profundum* strain 3TCK genome. Preliminary analyses of the genome features are reported for each, including comparisons to the genomes of closely related piezosensitive strains. This information was then used to investigate trends previously thought to be necessary for adaptation to high pressure, specifically looking at trends in amino acid usage in the primary structure of proteins in known pressure-sensitive processes. The genomic and protein data provided here will provide a foundation for future investigations into deep sea adaptations through genomic comparisons and protein structure analyses.

I

Genome Assembly of *Carnobacterium* spp. AT7

Background

Until now, studies of piezophilic bacteria have been on the Gamma-proteobacteria, because of their high representation in deep sea samples. Only a few bacteria outside of this class are known in the deep-sea environment (Lauro et al, 2007). This makes the genome sequencing project of *Carnobacterium* species AT7 (hereafter referred to as AT7) a unique undertaking. Not only is it the first member of its genus to be sequenced, it is also the first gram positive piezophile to undergo genome sequencing. Investigation into the genome of AT7 not only allows for insight in to a new genus and environment, it enables the investigation of piezophily across a broad phylogenetic spectrum.

The genus *Carnobacterium* was first developed in 1987 as a way to describe the lactobacillus-like organisms being increasingly held accountable for spoilage of processed meats (Hammes and Hertel, 2006). Over time, new genus members were discovered in non-food related habitats, including as fish pathogens as well as environmental samples of marine, arctic, and Antarctic ecosystems (Hammes and Hertel, 2006). In general, microbes belonging to the genus *Carnobacterium* are similar to the *Lactobacillus* in that they are gram positive, catalase-negative, non-sporeforming, rod-shaped (growing singly or in small chains) and facultatively heterofermentative (Hammes and Hertel, 2006). Oftentimes, members of this genus are involved in processes such as fermentation, manganese oxide reduction and bacteriocin compound production (Hammes and Hertel, 2006). Like the closely related *Enterococcus*,

Carnobacterium species typically show resistance to heavy metals (Hammes and Hertel, 2006). However, as AT7 is the first *Carnobacterium* species to undergo a sequencing project, very little is known about their genomic composition.

By 2007, nine species of *Carnobacterium* were known, revealing the diversity of the genus (Leisner et al, 2007). These organisms are now known to inhabit temperate/polar aquatic and terrestrial habitats (Leisner et al, 2007). In addition to their roles in food spoilage and as fish pathogens, their antimicrobial capabilities have implicated them as useful as protective cultures in food processing as well as probiotics in aquaculture (Leisner et al, 2007). Some species have shown resistance to freezing (Katayama et al, 2007) or, like AT7, to high pressure, both with implications in industries such as food processing. With such variation in habitats, there is also diversity in metabolic capabilities. One example is the arginine deiminase pathway, which was previously known in only five of the nine known members of the genus (Leisner et al, 2007). This metabolic and environmental diversity may be reflected in the genome sizes, which are predicted to be larger in the *Carnobacterium* than in the other lactic acid bacteria (Leisner et al, 2007).

AT7 is part of the collection of piezophilic bacteria isolates at the Scripps Institution of Oceanography. It was collected, along with the other culture gram positive piezophile AT22, from a water sample in the Aleutian Trench at a depth of 2500m (1.8°C) (Lauro et al, 2007). AT7 is most closely related to psychrophilic *Carnobacterium* such at those isolated from arctic and Antarctic ice (Lauro et al, 2007). Although native to psychrophilic and piezophilic environments, AT7 is able to grow well at room temperature and atmostpheric pressure. Previous work with AT7 included an

investigation of membrane lipids in which it was shown that 77% of the fatty acids in AT7 are monounsaturated fatty acids that maintain membrane fluidity at high pressure (Delong and Yayanos, 1986). Since there are no sequences available for close relatives of AT7, comparisons were done using *Enterococcus faecalis*, a closely-related, gram positive (order Lactobacillales), intestinal bacterium that share many genome features with the *Carnobacterium* (Hammes and Hertel, 2006; Paulsen et al, 2003).

Genome assembly

Scaffold building

The shotgun sequencing was performed by the J. Craig Venter Science Foundation Joint Technology Center using Applied Biosystems 3730XL DNA sequencers (Applied Biosystems, Foster City, CA) with funding by the Gordon and Betty Moore Foundation Marine Microbial Genome Sequencing project. The initial assembly was achieved using the Celera assembler (Myers et al, 2000) which puts the reads together into contiguous segments (termed contigs) based on overlapping regions in the individual reads. The assembly was imported in to the Phred/Phrap/Consed software package (Ewing and Green, 1998; Ewing et al, 1998; Gordon et al, 1998), enabling visualization and editing of the contigs for closure and finishing. The contigs were complied into scaffolds that form the framework for the completed genome. The scaffolds put the contigs in the correct orientation to assemble the genome. Due to variations in sequencing bias throughout the genome, some regions are left without coverage. These gaps in sequence scaffold fall between the contigs and are filled in during the closure process. The assembly visualization tool, Hawkeye, was used to identify the locations of plasmids (small inserts; 4kb) and fosmids (large inserts; 40kb)

within the scaffolds. Inserts that span multiple contigs were be used to link them together. In many cases, the gaps could be closed with the fosmid template using a process known as genome walking, in which sequences were added to the ends of the contigs until they overlapped and connected. Depending on the length of the gap, several rounds of primer design and sequencing were required. Contigs linked only by plasmids were joined using PCR. Contig ends with no fosmid or plasmid linkages were known as physical ends and served as the end of a scaffold. Physical ends may be the bi-product of clonal bias or difficult to sequence areas such as large repeats or strong secondary structure (Goldberg et al, 2006). Such gaps between scaffolds were closed using combinatorial PCR. As sequences were added, contigs were joined together, decreasing the number of total contigs and scaffolds.

Assembly of ribosomal regions

Gaps that contain ribosomal RNA operons must be assembled separately due to cloning bias, difficulty in sequencing, and high degree of sequence similarity that causes overlaps. Ribosome-specific primers were designed to sequence the ribosomal regions. All sequencing and PCR were done off of fosmid templates to ensure isolation of the correct ribosomal operon. Using knowledge of the beginning and ends of the sequences of the rRNA subunits, the operon was assembled outside of the general assembly. The regions with SNPs (single nucleotide polymorphisms) were avoided and the remaining conserved regions were used to create surrogate sequences for each operon. The surrogate sequences are composite sequences derived from the high number of reads from the initial shotgun assembly. The surrogate sequences are then added back in to the assembly using the fosmid-sequenced reads as a guide for correct alignment.

Current status

The AT7 genome project began as 69 contigs that were built in to 9 scaffolds. With the addition of 39,481bp of new sequence, it is currently in 18 contigs that comprise 6 scaffolds. The remaining genome gaps are regions that are difficult to sequence. Several genome features may lead to such problems including highly repeated regions or DNA secondary structure that interfere primer binding and transcription. A comparison strain may be necessary to map and align the remaining scaffolds.

AT7 analysis

The draft sequence was used in the autoannotation, performed by the Joint Genome Institute of the Department of Energy. Further analyses were performed on the Integrated Microbial Genomes Version 2.5 (March 2008) website hosted by the Joint Genome Institute (Markowitz et al, 2007). Genome comparisons were made using *Enterococcus faecalis* the genome for which is available on GenBank at NCBI.

Genome features

The genome is currently 2,481,599bp organized in a single circular chromosome and two plasmids (approximately 12,000bp and 75,000bp). Data about the AT7 genome is listed in Table 1.1. Autoannotation of the draft sequence found 2489 genes, 96% of which are predicted to be coding. Among the coding regions, 1467 genes (58.94%) have a predicted function, 792 (31.82%) are conserved hypothetical and 129 (5.18%) are hypothetical with no database matches (Figure 1.1). Based on 16S alignments, the genome is most closely related to other *Carnobacterium* isolates from the deep sea and polar regions, including *Carnobacterium pleistocenium* (Lauro et al, 2007), a psychrotolerant bacterium isolated from Alaskan ice (Pikuta et al, 2005). With 9 rRNA

operons, AT7 follows the trend of piezophiles possessing large numbers of rRNA operons, though falls short of SS9, which holds the record at 15 (Vezzi et al, 2005). Other work with SS9 showed a large degree of intragenomic variation among the many rRNA sequences (5.13% in 16S and 2.56% in 23S), due primarily to insertion sequences (Lauro et al, 2007; Vezzi et al, 2005). Vezzi et al (2005) suggested that this variation enables adaptation to pressure and other aspects of deep sea life. AT7 rRNA genes do not demonstrate the same variation. Alignments of both the 16S and 23S rRNA sequences reveal that they are 99% identical to one another. Thus, AT7 lacks the intragenomic variation described in SS9.

The autoannotation of the AT7 draft sequence reveals at least 17 phage genes from at least 6 different sources of phage (Table 1.2), indicating several instances of genome alterations due to horizontal gene transfer. Unlike other members of *Carnobacterium*, AT7 does not appear to contain genes for manganese oxide reduction. This may be an indication of either ecological adaptation or incomplete annotation. Also, AT7 can be included among the *Carnobacterium* that contain arginine deiminase, though its expression is unknown. Repeated sequences within the genome were also investigated. It was determined that there are two repeats that are scattered between genes on the chromosome. They are 82bp and 72bp in length and both appear 18 times within the genome. There were no significant matches to other sequences, and their function is unknown.

Plasmids

The AT7 genome includes two plasmids of very different sizes and composition. The larger, Contig 1009, is about 75,000bp (34% GC) in length with 77 predicted genes

(Figure 1.2; COG categories shown in Table 1.3). Of these 77 genes, 34 (44.2%) have a predicted function and 46 (59.7%) are hypothetical. Genes found in this plasmid are listed in Table 1.4. This plasmid contains several genes necessary for its own dispersal, including 7 genes for conjugation, 2 transposases, a putative integrase and 2 involved in replication. It is possible that the plasmid also confers metal resistance to its host. Specifically, the plasmid contains genes for cadmium and tellurium resistance, copper translocation and processing, as well as for a two-component response regulator and sensory transduction, presumably enabling detection of environmental conditions. Of the hypothetical proteins, 25 are predicted to be conserved and many have domains for membrane insertion or binding to the cell wall. These genes, like those involved in cation and metal transport, may be part of the metal resistance. Another gene of note is the AbiGI, an abortive infection gene originally described on the pCI750 plasmid of *Lactococcus lactis* subsp. *cremoris* UC653 (O'Conner et al, 1996). The AbiGI gene product is shown to aid in host resistance of psi712 and psic2 phage (O'Conner et al, 1996). This gene has a GC content of 34%, which is much higher than the 27% and 29% described previously for AbiGI (O'Conner et al, 1996), but still may be the result of horizontal gene transfer.

The smaller plasmid, Contig 927, is approximately 12,000bp in length, encoding 18 genes, 7 of which have predicted function (Figure 1.3). Like Contig 1009, this plasmid contains the replication and mobilization proteins necessary for plasmid dispersal. It also contains two ABC-type transport systems, one for multi-copper enzyme maturation and one for multidrug transport. This plasmid may also be involved in metal resistance for its host. Without genetic experiments, the functions of these plasmids are

predicted solely on the putative types of genes they possess.

Genome comparisons

The AT7 genome was compared to that of *E. faecalis* to identify potential adaptational differences. The *E. faecalis* genome is 3,359,974bp with a GC content of 37.38% and encoding 3385 genes (Table 1.1) (Paulsen et al, 2003).

Metabolic pathways

The AT7 genome, though much smaller in size, has a much higher number of genes that fall in to KEGG (Kyoto Encyclopedia of Genes and Genomes; Kanehisa and Goto, 2000) categories (256 genes, compared with 192 genes in *E. faecalis*), indicating more genes involved in metabolism. Specifically, Figure 1.4 shows that AT7 exceeds *E. faecalis* in genes involved in amino acid, lipid and glycan metabolism. This variation is understandable considering the variation in native environment, since nutrients in the deep sea are far more limiting than in the human gut. A more detailed review of the KEGG genes found in each genome is listed in Table 1.6. Unlike *E. faecalis*'s deoxyribodipyrimidine photolyase, AT7 does not contain any genes annotated as a photolyase. This is common among deep sea bacteria, which are not exposed to light in their natural environment (Vezzi et al, 2005; Wang et al, 2008). Interestingly, AT7 does contain two putative genes for UV damage repair (*usvE* and *uvdE*). Since both are endonucleases with putative UV association, it is likely that they function in DNA repair of damage caused by mutagens other than UV. In terms of COG (Clusters of Orthologous Groups; Tatusov et al, 1997) functional categories, AT7 has a larger percentage (74.05%) of its genome that falls in to these categories compared to *E. faecalis* (65.29%). However, since the analysis was performed using an unfinished

sequence, this result may be skewed by the total number of genes available. *E. faecalis* has a larger number of total genes in COG categories, so it is not surprising that it also has more genes in nearly all categories (Figure 1.5).

Modification of RNA

A previous study on *Shewanella* found that strains from the deep sea and other extreme environments contain large numbers of genes pertaining to tRNA and rRNA modification (Wang et al, 2008). These genes, which encode pseudouridine synthase and pseudouridylate synthase, typically have 1-4 copies within a single bacterial genome and are involved in maintaining RNA flexibility at low temperature (Wang et al, 2008). Wang et al (2008) believed that this large number of RNA modification genes allows RNA to function in many environments and is essential to the wide distribution of the genus. The AT7 genome has 5 genes encoding pseudouridine synthase and one gene for pseudouridylate synthase. It should be noted that AT7 has fewer of those genes than many of the *Shewanella*, which have up to 19 genes (Wang et al, 2008). When compared to *E. faecalis*, which had no genes for either psequdouridine synthase or pseudouridylate synthase, AT7 has far greater capacity for RNA modification and potentially for survival in environments of high pressure or low temperature.

Bacteriocins

Analyses of *Carnobacterium* genes thus far have been on bacteriocin-related sequences (Leisner et al, 2007). No bacteriocins have yet been identified in the AT7 genome, though there is evidence of a putative bacteriocin-resistance gene. Bacteriocin-resistance genes are common in bacteriocin-producing operons, to protect the host from its own antimicrobial production (Leisner et al, 2007). The bacteriocin-resistance gene

may be connected with yet to be uncovered bacteriocin genes. Furthermore, the genes for antimicrobial production may have been lost through time if they no longer confer a necessity in AT7's unique environment. This is plausible since the bacterocins produced by gram positive bacteria are often gram positive-specific (Jack et al, 1995), and in the deep-sea, gram positive bacteria are rare (Sogin et al, 2006). Comparatively, *E. faecalis* has no matches to bacteriocin genes.

Conclusion

This study increases the current knowledge of life in the deep sea by expanding taxonomic representation of genome sequences from this environment. As the first gram positive piezophile genome sequencing project, AT7 offers a unique perspective on deep sea adaptation. This initial work reveals that AT7 differs from features found in previous piezophile studies. AT7 appears to lack the variation and redundancy noted in other deep sea species, such as the rRNA operons of SS9 and RNA modification proteins of *Shewanella piezotolerans* WP3. However, it possesses high numbers of genes involving metabolic processes, relative to comparison strains. This is a feature common of deep sea organisms and is thought to be an adaptation to the poor nutrient availability in the ocean trenches. AT7 also shows unique features of ecological adaptations beyond that of high pressure. There is reason to believe, for example, that AT7 is capable of heavy metal resistance as well as antimicrobial compound production. The roles of each of these properties in AT7's natural habitat are thus far unknown, but are of particular interest for further lab work. The plasmids appear to confer heavy metal resistance to the host. This could be tested by inserting the plasmids in to an organism known not to be resistant to metals, and examining the resultant phenotype. It would also be interesting to determine

if AT7 can be grown without the plasmids and how their loss effects its survival. Similarly, genetic and biochemical studies should be performed in search of bacteriocin production. The discovery of such a compound could have interesting ecological and biotechnological applications. In general, transcriptional analyses would be useful in understanding gene regulation under a variety of conditions, notably at optimal and suboptimal pressures. Further investigation in to the genome is required to elucidate more features related both to phylogeny and deep sea adaptation.

Table 1.1: List of genome statistics about *Carnobacterium* species AT7 and *Enterococcus faecalis*, including genome size, number of genes, and number of RNA operons.

	AT7	E. faecalis
Size	2481599	3359974
G+C ratio	35.28%	37.38%
Percentage coding	86.63%	86.93%
Number of ORFs	2489	3385
Protein coding genes	95.94%	96.48%
With function prediction	58.94%	58.91%
Conserved hypothetical	31.82%	31.96%
Hypothetical	5.18%	5.61%
RNA genes	101	119
rRNA operons	9	4
tRNA genes	74	68

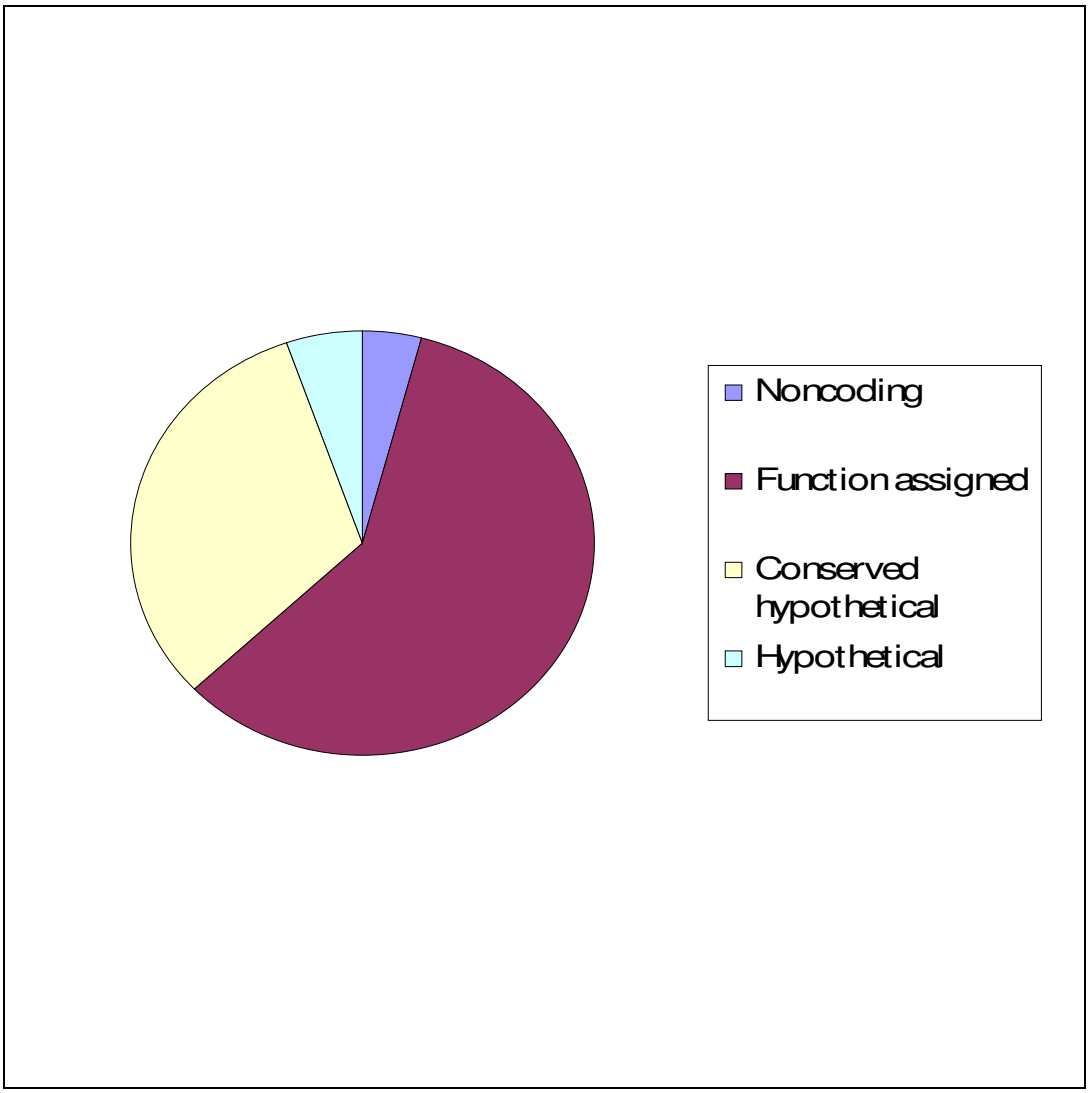

Figure 1.1: Graph displaying the relative amounts of the AT7 genome devoted to noncoding sequences, and ORFs with predicted function, no predicted function but with similarity (*conserved hypothetical*), and no predicted function and no similarity (*hypothetical*).

Table 1.2: Lists the gene products of genes predicted to originate from phage, including their phage of origin (when known) and the length of the gene product (in number of amino acids).

Product Name	**Origin**	**Length**
Predicted xylosidase/arabinosidase	CP4-6 prophage	195
Putative phage envelope protein		130
Lj965 prophage protein	Lj965 prophage	80
PblB, putative	prophage LambdaSa2	1435
Tail protein, putative	prophage LambdaSa04	236
ORF003-like protein, phage terminase, large subunit	phi77	552
Phage protein		162
putative integrase	bacteriophage 370.1	477
Phage endonuclease		126
Phage protein		101
Phage protein		145
Uncharacterized phage related protein		63
DNA polymerase, bacteriophage-type		216
tyrosine recombinase XerC subunit (IMGterm)		299
Protein 45	prophage pi2	682
Phage protein		110
Phage endonuclease		122
Phage protein		101

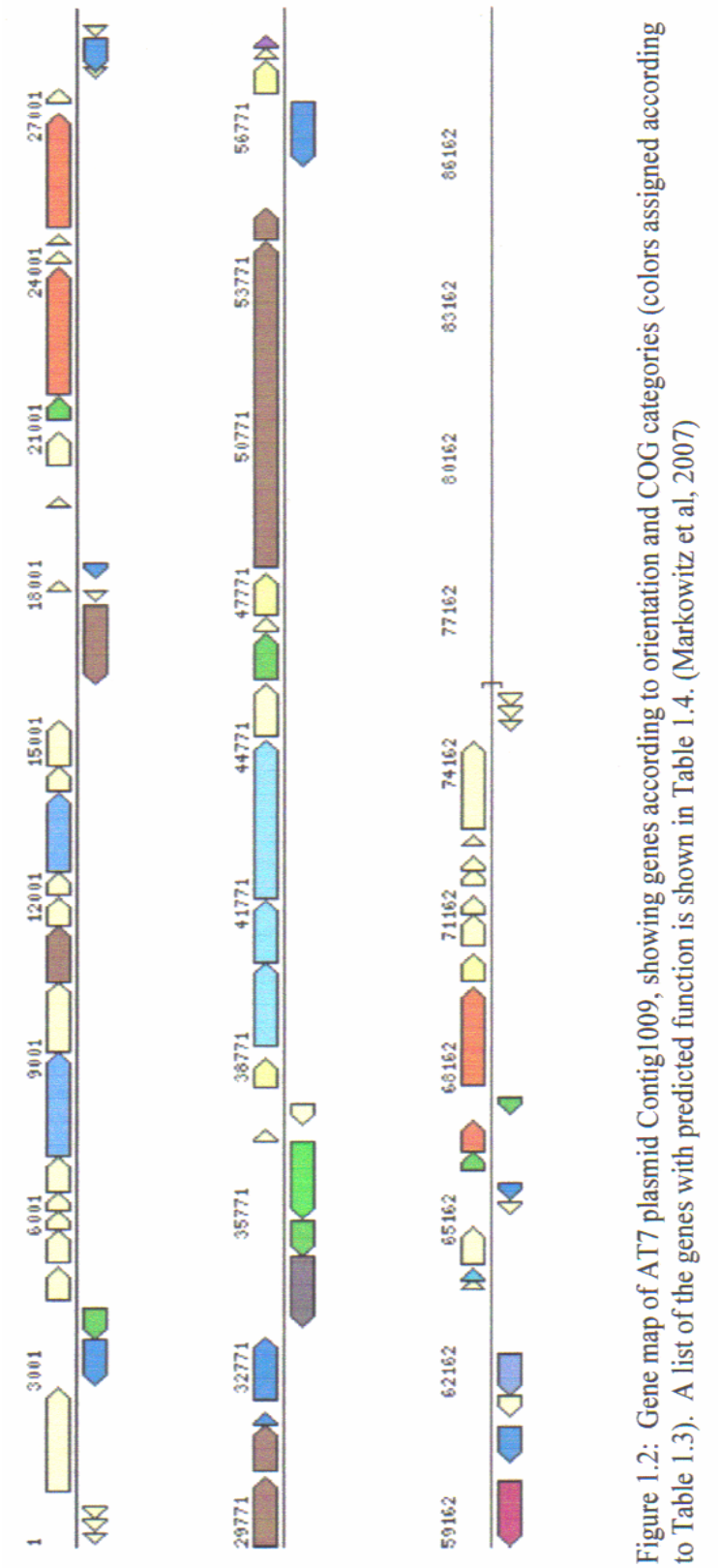

Figure 1.2: Gene map of AT7 plasmid Contig1009, showing genes according to orientation and COG categories (colors assigned according to Table 1.3). A list of the genes with predicted function is shown in Table 1.4. (Markowitz et al, 2007)

Table 1.3: Key to the colors assigned to COG categories as shown in Figures 2 and 3.

COG Code	COG Function Definition
[A]	RNA processing and modification
[B]	Chromatin structure and dynamics
[C]	Energy production and conversion
[D]	Cell cycle control, cell division, chromosome partitioning
[E]	Amino acid transport and metabolism
[F]	Nucleotide transport and metabolism
[G]	Carbohydrate transport and metabolism
[H]	Coenzyme transport and metabolism
[I]	Lipid transport and metabolism
[J]	Translation, ribosomal structure and biogenesis
[K]	Transcription
[L]	Replication, recombination and repair
[M]	Cell wall/membrane/envelope biogenesis
[N]	Cell motility
[O]	Posttranslational modification, protein turnover, chaperones
[P]	Inorganic ion transport and metabolism
[Q]	Secondary metabolites biosynthesis, transport and catabolism
[R]	General function prediction only
[S]	Function unknown
[T]	Signal transduction mechanisms
[U]	Intracellular trafficking, secretion, and vesicular transport
[V]	Defense mechanisms
[W]	Extracellular structures
[Y]	Nuclear structure
[Z]	Cytoskeleton
NA	Not assigned

Table 1.4: Name and description of protein products of the genes found on AT7 plasmid Contig1009 (hypothetical proteins not included).

Name	Description
plasmid mobilization system relaxase	MobA/MobL family
abortive infection protein AbiGI	DNA polymerase III, alpha subunit
conjugation protein	
conjugation protein	
conjugation protein	Type IV secretory pathway, VirB4 components
conjugation protein	ATPase involved in DNA repair
Antigen	Soluble lytic murein transglycosylase and related regulatory proteins (some contain LysM/invasin domains)
conjugation protein	Type IV secretory pathway, VirD4 components
conjugation protein	Membrane protein TerC, possibly involved in tellurium resistance
UDP-N-acetylmuramoylalanyl-D-glutamate--2, 6-diaminopimelate ligase	UDP-N-acetylmuramyl tripeptide synthase
transcriptional repressor CopY	Predicted transcriptional regulator
copper-translocating P-type ATPase	Cation transport ATPase
copper transport protein CopZ	Copper chaperone
cation-transporting ATPase, E1-E2 family protein	Cation transport ATPase
prolipoprotein diacylglyceryl transferase	Prolipoprotein diacylglyceryltransferase
beta-propeller fold protein	Uncharacterized conserved protein
Sensory transduction histidine kinase	Signal transduction histidine kinase
two-component response regulator	Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain
Putative multicopper oxidase	Secondary metabolites biosynthesis, transport and catabolism
putative resolvase	Site-specific recombinases, DNA invertase Pin homologs
type I restriction-modification system methyltransferase subunit	
HsdS specificity protein of type I restriction-modification system	Restriction endonuclease S subunits
type I restriction-modification system restriction subunit	Type I site-specific restriction-modification system, R (restriction) subunit and related helicases
transposase	Transposase and inactivated derivatives
transposase	Transposase and inactivated derivatives
Uncharacterized protein related to Endonuclease III	
N-carbamoyl-L-amino acid amidohydrolase	
N-carbamoyl-L-amino acid amidohydrolase	Acetylornithine deacetylase/Succinyl-diaminopimelate desuccinylase and related deacylases
replication-associated protein	ATPases involved in chromosome partitioning
Predicted transcriptional regulators	Bacterial regulatory protein, arsR family
cadmium resistance transporter, putative	Predicted permease, cadmium resistance protein
transcriptional regulator (ArsR family) protein	Predicted transcriptional regulators
copper-transporting ATPase copA	Cation transport ATPase
integrase/recombinase, fragment (putative)	Integrase
Membrane-fusion protein	

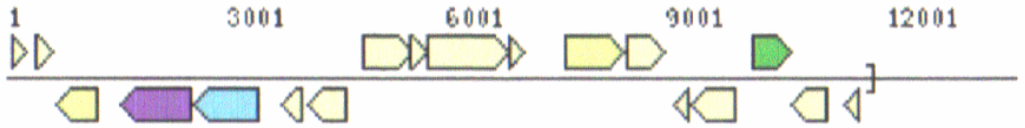

Figure 1.3: Gene map of AT7 plasmid Contig927, showing genes according to orientation and COG categories (colors assigned according to Table 1.3). A list of the genes with predicted function is shown in Table 1.5. (Markowitz et al, 2007)

Table 1.5: Name and description of protein products of the genes found on AT7 plasmid Contig927 (hypothetical proteins not included).

Name	Description
integrase/recombinase, fragment (putative)	
YhcI	ABC-type transport system involved in multi-copper enzyme maturation, permease component
YhcH	ABC-type multidrug transport system, ATPase component
mobilization protein	ATPase involved in DNA repair
rlx protein, putative	Type IV secretory pathway, VirD2 components (relaxase)
replication protein	Protein involved in initiation of plasmid replication
DNA-binding protein	Predicted transcriptional regulators

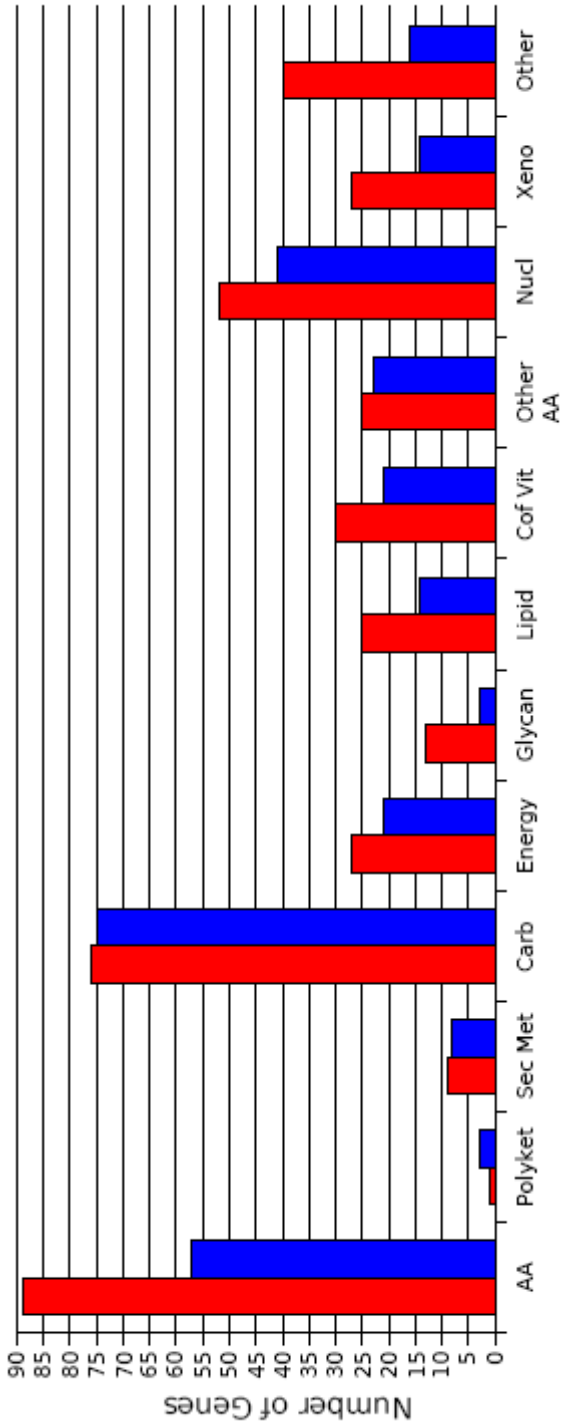

Figure 1.4: KEGG Categories for AT7 (red) and *E. faecalis* (blue). Categories, from left to right, are: amino acid metabolism, AA; biosynthesis of polyketides and nonribosomal pepetides, Polyket; biosynthesis of secondary metabolites, Sec Met; carbohydrate metabolism, Carb; energy metabolism, Energy; glycan biosynthesis and metabolism, Glycan; lipid metabolism, Lipid; metabolism of cofactors and vitamins, Cof Vit; metabolism of other amino acids, Other AA; nucleotide metabolism, Nucl; xenobiotics biodegredation and metabolism, Xeno; Other.

Table 1.6: KEGG category comparison between *Carnobacterium* AT7 and *Enterococcus faecalis*, showing function name of the category and the number of genes in each genome.

Function name	AT7	E. faecalis
Glycolysis / Gluconeogenesis	20	21
Citrate cycle (TCA cycle)	3	2
Pentose phosphate pathway	13	10
Inositol metabolism	2	1
Pentose and glucuronate interconversions	8	8
Fructose and mannose metabolism	8	10
Galactose metabolism	11	13
Ascorbate and aldarate metabolism	2	2
Fatty acid biosynthesis	8	5
Fatty acid elongation in mitochondria	2	0
Fatty acid metabolism	4	1
Synthesis and degradation of ketone bodies	2	0
Biosynthesis of steroids	5	1
Bile acid biosynthesis	2	2
Ubiquinone biosynthesis	2	2
Androgen and estrogen metabolism	2	0
Oxidative phosphorylation	11	2
ATP synthesis	10	0
Photosynthesis	10	0
Urea cycle and metabolism of amino groups	9	3
Purine metabolism	32	32
Pyrimidine metabolism	34	16
Glutamate metabolism	11	12
Alanine and aspartate metabolism	12	10
Tetracycline biosynthesis	3	1
Glycine, serine and threonine metabolism	19	8
Methionine metabolism	5	5
Cysteine metabolism	7	8
Valine, leucine and isoleucine degradation	4	1
Valine, leucine and isoleucine biosynthesis	4	4
Lysine biosynthesis	10	3
Lysine degradation	6	2
Arginine and proline metabolism	11	9
Histidine metabolism	3	3
Tyrosine metabolism	2	1
Phenylalanine metabolism	1	1
Benzoate degradation via hydroxylation	1	0
Bisphenol A degradation	1	0
Tryptophan metabolism	8	3
Phenylalanine, tyrosine and tryptophan biosynthesis	9	7
Novobiocin biosynthesis	0	1
beta-Alanine metabolism	1	4
Taurine and hypotaurine metabolism	3	2
Aminophosphonate metabolism	2	0
Selenoamino acid metabolism	8	6
Cyanoamino acid metabolism	1	2
D-Glutamine and D-glutamate metabolism	3	2
D-Arginine and D-ornithine metabolism	1	0
D-Alanine metabolism	2	3
Glutathione metabolism	7	4

Table 1.6, Continued: KEGG category comparison between *Carnobacterium* AT7 and *Enterococcus faecalis*, showing function name of the category and the number of genes in each genome.

Function name	AT7	E. faecalis
Starch and sucrose metabolism	14	10
Nucleotide sugars metabolism	6	7
Streptomycin biosynthesis	0	3
Polyketide sugar unit biosynthesis	0	3
Aminosugars metabolism	10	11
Lipopolysaccharide biosynthesis	1	0
Peptidoglycan biosynthesis	8	3
Glycerolipid metabolism	3	3
Glycosylphosphatidylinositol(GPI)-anchor biosynthesis	1	0
Glycerophospholipid metabolism	3	4
Linoleic acid metabolism	1	0
Glycosphingolipid biosynthesis - neo-lactoseries	3	0
Glycosphingolipid biosynthesis - globoseries	3	0
Glycosphingolipid biosynthesis - ganglioseries	3	0
Pyruvate metabolism	18	13
1- and 2-Methylnaphthalene degradation	1	0
Tetrachloroethene degradation	1	0
Nitrobenzene degradation	3	0
Glyoxylate and dicarboxylate metabolism	3	3
1,2-Dichloroethane degradation	0	1
Benzoate degradation via CoA ligation	5	0
Propanoate metabolism	10	6
Butanoate metabolism	7	5
One carbon pool by folate	6	5
Methane metabolism	1	2
Carbon fixation	6	7
Reductive carboxylate cycle (CO2 fixation)	1	0
Thiamine metabolism	0	2
Riboflavin metabolism	2	0
Vitamin B6 metabolism	4	0
Nicotinate and nicotinamide metabolism	6	4
Pantothenate and CoA biosynthesis	5	6
Folate biosynthesis	0	1
Porphyrin and chlorophyll metabolism	5	1
Terpenoid biosynthesis	3	1
Limonene and pinene degradation	2	1
Nitrogen metabolism	5	6
Sulfur metabolism	4	4
Caprolactam degradation	1	0
Stilbene, coumarine and lignin biosynthesis	0	1
Alkaloid biosynthesis I	0	1
Alkaloid biosynthesis II	1	0
Aminoacyl-tRNA biosynthesis	18	13
Glycan structures - biosynthesis 2	3	0
Biosynthesis of vancomycin group antibiotics	0	1
Biosynthesis of type II polyketide products	1	0
Neurodegenerative Disorders	1	1
ABC transporters - General	3	3
Two-component system - General	7	2
Bacterial chemotaxis - General	1	0

Table 1.6, Continued: KEGG category comparison between *Carnobacterium* AT7 and *Enterococcus faecalis*, showing function name of the category and the number of genes in each genome.

Function name	AT7	E. faecalis
Bacterial chemotaxis - Organism-specific	1	0
Flagellar assembly	10	0
Phosphotransferase system (PTS)	3	4
RNA polymerase	4	1
DNA polymerase	2	1
Protein export	1	0
Type III secretion system	10	0
Type II secretion system	2	0
PPAR signaling pathway	3	1
Calcium signaling pathway	1	0
Apoptosis	1	0
Insulin signaling pathway	6	3
Type II diabetes mellitus	1	1
Amyotrophic lateral sclerosis (ALS)	1	2
Epithelial cell signaling in Helicobacter pylori infection	10	0

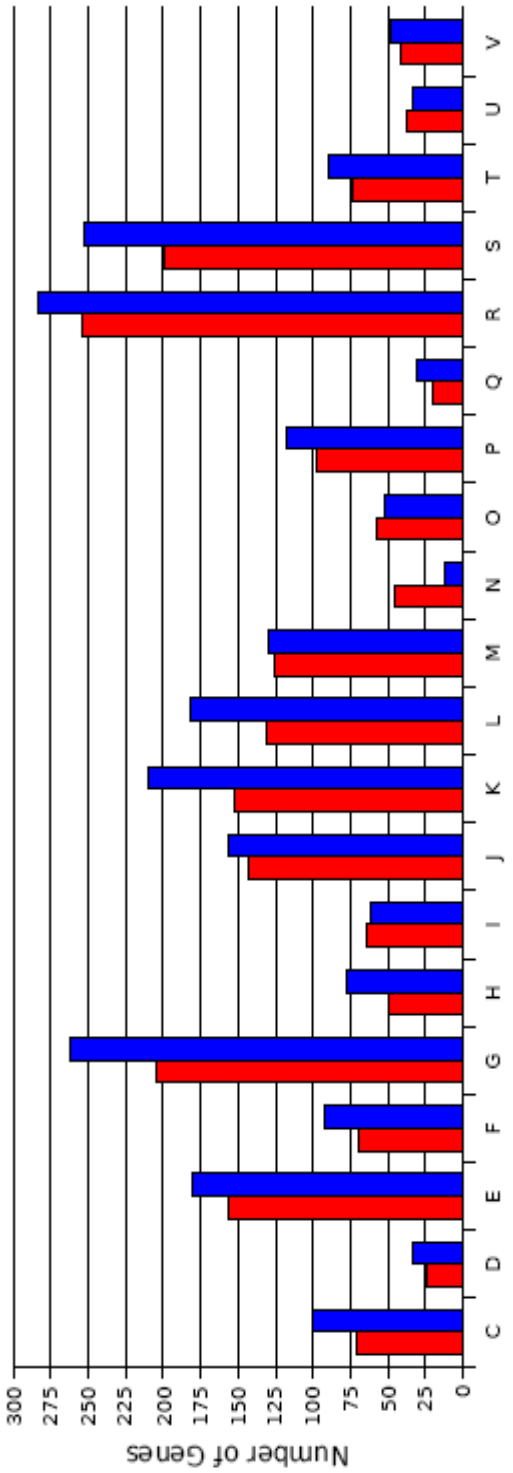

Figure 1.5: COG categories for AT7 (red) and *E. faecalis* (blue). The categories are as follows: energy production and conversion, C; cell cycle control, cell division, chromosome partitioning, D; amino acid transport and metabolism, E; nucleotide transport and metabolism, F; carbohydrate transport and metabolism coenzyme transport and metabolism, H; lipid transport and metabolism, I; translation, ribosomal structure and biogenesis, J; transcription, K; replication, recombination and repair, L; cell wall/membrane/envelope biogenesis, M; cell motility, N; posttranslational modification, protein turnover, chaperones, O; inorganic ion transport and metabolism, P; secondary metabolites biosynthesis, transport and catabolism, Q; general function prediction only, R; function unknown, S; signal transduction mechanisms, T; intracellular trafficking, secretion, and vesicular transport, U; defense mechanisms, V.

.

II

Genome Closure and Finishing of *Psychromonas* spp. CNPT3

Background

The genus *Psychromonas* was first described in 1998 upon the classification of its type species, *Psychromonas antarcticus* (Mountfort et al, 1998). This organism, isolated from a salt concentrated lake in Antarctica, was found to be closely related to *Moritella marinus*, but had enough distinction in its 16S rRNA gene, GC content, morphology, and physiological and biochemical properties to suggest a separate genus (Mountfort et al, 1998). Members of the *Psychromonas* genus have the following characteristics: gram-negative, aerotolerant anaerobes, motile with polar flagellum, growth temperature range 2-22 degrees C, and carbohydrates are not the main energy source (Mountfort et al, 1998). Currently, 34 species of *Psychromonas* are known from several extreme environments, including those of low temperature (as low as -12ºC) and high salt concentration (Auman et al, 2006). Others have since been identified with varying degrees of piezophily (Lauro et al, 2007; Nogi et al, 2002; Yayanos et al, 1979). Currently, genome sequences exist for only two species, of *Psychromonas,* including the piezophilic sp. CNPT3.

Psychromonas sp. strain CNPT3 (hereafter referred to as CNPT3) was isolated in 1979 (Yayanos et al, 1979), nearly two decades before being classified among the *Psychromonas*. CNPT3 was cultured from a decaying amphipod, captured at 5700 meters in the Central North Pacific Basin, and maintained at temperature and pressure conditions of the collection site (Yayanos et al, 1979). Its closest relative, based on 16S rRNA sequence, is *Psychromonas* sp. HS11, another piezophilc psychrophile (Delong

and Yayanos, 1986). Unlike the initial description of *Psychromonas*, which describe its species as rod-shaped or spherical, CNPT3 has a spirillium-like morphology (Yayanos et al, 1979). As the first cultured piezophilic bacterium, early lab work with CNPT3 gave important insight in to high pressure physiology. Its preferential growth at high pressure and low temperature allow insight into deep sea lifestyles. Tests with this organism enabled the receipt of data regarding growth characteristics and morphologies (Yayanos et al, 1979). Importantly, Yayanos et al (1979) noted that piezophilic organisms are neither inactivated by periods of decompression, nor are high pressure adaptations lost by decompressions. Yayanos and Dietz (1982), showed that thermal inactivation of CNPT3 was possible at temperatures above 0ºC. These discoveries have been vital to laboratory studies of deep sea bacteria. Prior to the isolation of CNPT3, high pressure studies were performed by exposing mesophilic bacterial strains to high hydrostatic pressure to elucidate pressure-sensitive processes (Yayanos et al, 1979). While these investigations have had significant impact on the understanding of high pressure physiology, they give little insight in to the growth characteristics and lifestyle of bacteria native to the deep sea.

The CNPT3 genome project was also a first in terms of sequencing technology trials. As genome sequencing gains popularity, technological advances attempt to improve the quality and lower the costs of whole genome shotgun sequencing. The pyrosequencing-based 454 sequencing method (454 Life Sciences, Branford, CT) is one such technology suggested as a high output, low cost alternative to traditional 3730*xl* Sanger sequencing techniques and nucleotide repeat errors (Goldberg et al, 2006). Early work with 454 technology indicated that, although beneficial in its lack of cloning bias

and decreased costs, there are difficulties in assembly due to short read lengths (Goldberg et al, 2006). The CNPT3 genome was one of 6 bacterial genomes in the Gordon and Betty Moore Marine Microbial Sequencing Project used to examine the effectiveness of a hybrid Sanger and pyrosequencing approach as a way to combine the benefits of both technologies (Goldberg et al, 2006). Due to its involvement in this study, the CNPT3 genome was sequenced with a combination of Sanger and pyrosequencing, though the additional sequences for closing and finishing were entirely Sanger.

Genome Closure/Finishing

Sequence preparation

Two genomic libraries were prepared in GC10 cells as described in Goldberg et al (2006). The libraries consisted of a large inserts (fosmids) and the other of small inserts (plasmids). Libraries were stored as glycerol stocks in 384-well plates at -80ºC. Genome closure and finishing involved enhancing regions of low read coverage through sequencing from specific fosmid clones. Regions in need of added sequences (i.e. areas of low coverage or with low quality consensus sequence) were found by scanning the sequence in Consed, a graphical tool for visualizing assemblies at the sequence level (Gordon et al, 1998). Appropriate fosmid clones were chosen using Hawkeye, an assembly visualization tool that enables the user to view the scaffolds of assembled reads (Schatz at al, 2007).

Fosmid-containing *E. coli* were grown overnight in 5 ml Luria-Bertani (LB) medium with chloramphenicol (30 µg/ml) at 37ºC with shaking. Following the overnight incubation, 400 µl of the culture was transfered to 6 ml fresh LB media with chloramphenicol and incubated at 37ºC with shaking. After incubating for an hour,

induction solution was added to a final concentration of 1% L-arabinose and the cultures continued to incubate overnight. The tubes were centrifuged at 5000 rpm for 6 minutes at 4ºC. The supernatant was discarded and the remaining pellet resuspended in Buffer P1 (50mM Tris-Cl pH 8.0, 10mM EDTA, 100 µg/ml RNase A) by vortexing. The solution was transferred to a 1.5 ml microcentrifuge tube. The cells were lysed using 250 µl lysis buffer (200mM NaOH, 1% SDS) and neutralized within five minutes by neutralization buffer (3.0M potassium acetate pH 5.5). The solutions were centrifuged in a tabletop centrifuge for 10 minutes at 14,000 rpm, after which the supernatant, containing the solublized DNA, was transferred to a new tube and the pellet discarded. To the supernatant, 640 µl of isopropanol was added and mixed thoroughly by inversion. The solution was then centrifuged for 10 minutes at 14,000 rpm. The pelleted DNA was washed twice with 600 µl of 70% ethanol and then dried in a SpeedVac (Savant) for approximately 15 minutes. The dried pellets were stored at -2ºC. The DNA was resuspended in 10 µl Buffer EB (10mM Tris) and the concentration determined on an agarose electrophoresis gel.

Regions of the genome not covered by at least two independent fosmid clones were covered using a combination of long-range and short-range PCR. Contigs with unknown positions in the scaffold were linked using combinatorial PCR, in which primers at the ends of contigs were paired at random. Primer combinations that resulted in a PCR product indicate linkages between contigs. In cases where multiple PCR products result, those PCRs are rerun with a gradient of annealing temperatures to reduce nonspecific binding between the primer and template and ensure proper scaffold assembly. In preparation of sequencing, the PCR products were cleaned using

UltraClean PCR Clean-up Kit (Mo Bio Laboratories, Inc., Carlsbad, CA)

Primers were specifically designed to locations in the genome using Consed. The primer sequences were downloaded and ordered from Integrated DNA Technologies (Coralville, IA). The primers were resuspended in Buffer EB and then diluted to a concentration of 2.5 µM.

Sequencing was performed by SeqXcel, Inc (Sorrento Valley, CA) using ABI Prism Capillary Electrophoresis Genetic Analyzers and BigDye Terminator Chemistry. Sequencing reactions were set up with the primer and template concentrations indicated by SeqXcel and submitted in a 96-well PCR plate.

Sequence data

The finished sequences were then added in to the assembly. The returned sequences were collected into a file of files (.fof) and converted from ab1 to scf format and moved to the Consed directories where they could be added to the program. This format conversion allows the formation of the chromatogram, from which Consed can assign quality values, as well as allow the user to visualize the results of the sequencing reactions for manual judgment calls. Each added sequence was checked to ensure correct assembly. In the case of closure, the sequences were used to extend the length of the contig and combine it with neighboring contigs. For finishing, additional primers were designed and sequences added to improve read coverage and confidence as well as to correct the consensus sequence in regions of poor read quality. Autoannotation was performed by The Institute for Genomic Research (TIGR; Rockville, MD) and manual annotation was completed using MANATEE via TIGR. Further analyses were performed on the Integrated Microbial Genomes Version 2.5 (March 2008) website hosted by the

Joint Genome Institute (Markowitz et al, 2007) and the Comprehensive Microbial Resource hosted by the J. Craig Venter Institute (Peterson et al, 2001).

Comparison strain

Genome comparisons were made using *Psychromonas ingrahamii 37*, a psychrophilic bacterium isolated from Arctic sea ice (Auman et al, 2006; Breezee et al, 2004). It is the piezosensitive strain with a sequenced genome that is most closely related to CNPT3. *P. ingrahamii*, like many other polar sea ice isolates, have gas vacuoles that are thought to be involved in motility, though their exact function has yet to be determined (Auman et al, 2006). However, the presence of two morphologically distinct gas vacuoles makes *P. ingrahamii 37* unusual (Auman et al, 2006). The *P. ingrahamii* genome is a single circular chromosome 4,559,598bp in length with 10 rRNA operons, and 40% GC content (Table 2.1). The genome is available via GenBank at NCBI.

Genome comparisons

The initial assembly of the CNPT3 genome was 2,945,265bp in 175 contigs. This large number of contigs was due to the reliance on short 454 sequences that resulted in very few regions of overlap. With the addition of 107,145bp, the finished genome is 3,052,410bp in a single circular chromosome (Figure 2.1). CNPT3 has a 38.6% GC content and 10 rRNA operons (Table 2.1). Of its 2752 protein coding genes, 72.71% have predicted function, 12.86% are hypothetical genes with similarity to other known sequences, and 14.42% are hypothetical genes with no known sequence similarity (Table 2.1, Figure 2.2). Few phage proteins were found (three phage integrases and one prophage regulatory homolog) and no conjugation proteins, suggesting few opportunities for horizontal gene transfer.

Approximately 74% of the protein coding genes fell in to COG (Clusters of Orthologous Groups; Tatusov et al, 1997) categories that describe their predicted function (Figure 2.3). Not surprisingly, the most highly represented categories are those of hypothetical (13.65%), conserved hypothetical (12.24%) and genes of unknown function (15.08%). Of those with predicted function, transport and binding proteins are the most highly represented with 273 genes making up 9.38% of the functional gene products. The proportion of gene functions in the CNPT3 genome is not drastically different from that of other organisms, shallow water or deep sea. The COG representation of the CNPT3 genome was compared to that of *Psychromonas ingrahamii 37*. With a much larger genome, it is not surprising to find that *P. ingrahamii* has a higher number of genes assigned to COG categories (Figure 2.4). CNPT3 exceeds *P. ingrahamii* in cell motility (primarily genes involved in chemotaxis), cell cycle control, cell division and chromosome partitioning (a difference of two genes), and intracellular trafficking, secretion and vesicular transport (a difference of 3 genes). There was no significant difference between the organisms in terms of percent representation of each category. Specific breakdown of number of genes in COG categories for CNPT3 and *P. ingrahamii* is shown in Table 2.2.

Specific COG classifications were compared to show the differences in number of genes. Figure 2.5 is a heat map that illustrates the number of genes assigned to each COG function (CNPT3 annotation in this comparison used the draft sequence, so gene numbers are most likely underestimates). In total, 2041 clusters were examined, though only the portions with the greatest differences are shown. Although adjusted for genome size, *P. ingrahamii* still demonstrates a higher abundance of specific genes. In general,

P. ingrahamii has a larger proportion of genes encoding transposases, signal transduction, and stress response. CNPT3 dominant genes are primarily those involved in outer membrane interaction. One notable gene classification is COG0840, which encodes methyl-accepting chemotaxis protein. CNPT3 has 19 genes in this classification, whereas *P. ingrahamii* has zero. These genes are listed in Table 2.4. The implications of these genes are not yet known, though, with one being an aerotaxis receptor, it is believed that CNPT3 contains a system for detection of oxygen. Many members of the *Psychromonas* genus are known to be anaerobic (Mountfort et al, 1998), so these environment-detecting proteins may be important in ecological adaptation. In other forms of signal transduction, genes found in *P. ingrahamii* outnumber those in CNPT3. Like other deep sea microorganisms, CNPT3 does not contain genes for photolyases or light activated processes. *P. ingrahamii*, on the other hand, has four photolyase genes.

Metabolic and biosynthetic capabilities were compared by investigating gene abundances in KEGG categories (Kyoto Encyclopedia of Genes and Genomes; Kanehisa and Goto, 2000). The results are shown in Figure 2.6. CNPT3 had a greater number of genes involved in the biosynthesis of polyketides and nonribosomal peptides, biosynthesis of secondary metabolites, glycan biosynthesis and metabolism, metabolism of cofactors and vitamins, and xenobiotics biodegredation and metabolism. In the category of polyketide and nonribosomal peptides, CNPT3 had enzymes that fell in to the processes involved in the biosynthesis of ansamycins, biosynthesis of siderophore group nonribosomal peptides, biosynthesis of type II polyketide backbone, and biosynthesis of type II polyketide products, but none constituted a complete, functional pathway. The same is true for *P. ingrahamii*, which contained genes involved in biosynthesis of 12-,

14-, and 16-membered macrolides, biosynthesis of ansamycins, biosynthesis of vancomycin group antibiotics, and polyketide sugar unit biosynthesis. The diversification of metabolic capabilities is necessary in the deep sea, where nutrient resources are often variable in type and quantity. *P. ingrahamii* had a higher number of genes in the remaining pathways, as well as a larger total number of genes categorized. Since *P. ingrahamii* has a much larger genome, these results are unsurprising. CNPT3, though, has a larger percentage of its genome dedicated to KEGG pathways. Again, this may be an indication of the need for metabolic diversity in the deep sea.

Membrane lipids are a point of interest in deep sea bacteria. Under high pressure and low temperature, lipids become more crystalline, decreasing the fluidity of biological membranes (Bartlett, 2002; Braganza and Worcester, 1986). Like psychrophiles, piezophiles demonstrate homeoviscous-like adaptation in which the solidifying effects of high pressure on lipid membranes is offset by increasing the proportion of unsaturated fatty acids (Delong and Yayanos, 1985). Research on CNPT3 growth at various pressures demonstrated this phenomenon, as it was shown that unsaturated fatty acid content increases with increasing pressure (Delong and Yayanos, 1985). The fatty acid genes in CNPT3 included *plsX*. The PlsX/Y pathway is the most widely distributed pathway for the initiation of phospholipid formation, and serves as a phosphate:acyl-ACP acyltransferase (Lu et al, 2006). Other genes for fatty acid synthesis, specifically for unsaturated fatty acid synthesis, are those in both the *pfa* and *fab* gene clusters. The *pfa* gene cluster products, the omega-3 polyunsaturated fatty acid synthases, introduce double bonds into a single acyl chain in the absence of desaturation reactions (Allen and Bartlett, 2002). These polyunsaturated fatty acids (PUFAs) are thought to be involved in

maintenance of membrane fluidity in high pressure and cold temperature (Allen and Bartlett, 2002). Their importance in human health has made their production in bacterial membranes a point of much interest (Allen and Bartlett, 2002). Likewise, the *fabA* and *fabZ* genes found in CNPT3 are involved in unsaturated fatty acid biosynthesis. Both identified as encoding beta-hydroxyacyl-(acyl-carrier-protein) dehydratases, one of FabA's functions is to initiate the biosynthesis of unsaturated fatty acids, while FabZ is the primary dehydratase in the unsaturated fatty acid pathway (Heath and Rock, 1996). No genes were identified as encoding fatty acid desaturases.

It has been suggested that bacteria adapted to high pressure counteract the effects of high external hydrostatic pressure with high internal osmolytic pressure (Bartlett, 2002). High concentrations of intracellular osmolytes are believed to assist in protein folding and maintenance of three dimentional conformation (Bartlett, 2002). No genes in the CNPT3 genome were implicated in the production or import of the common osmolytes: trimethylamine-N-oxide (TMAO), betaine, proline, ectoine, and polybetahydroxybutyrate (PHB).

Previous studies on piezophiles noted that deep sea bacteria have high numbers of genes involved in RNA modification (Wang et al, 2008). These genes include pseudouridine synthase and pseudouridylate synthase. Unlike most bacteria, which have only 1-4 copies per genome, deep sea *Shewanella piezotolerans* WP3 was the first to demonstrate large numbers of these genes with seven pseudouridine synthases, nine pseudouridylate synthases, and three involved in other RNA modifications (Wang et al, 2008). CNPT3 follows this trend with thirteen pseudouridine synthases and 2 other RNA modification genes. No pseudouridylate genes were found in CNPT3. *P. ingrahamii*

also had a large number of these genes, with eleven pseudouridine synthases, as well as one pseudouridylate synthase and one RNA modification gene. The presence of these genes was investigated in other piezophiles and comparison strains with sequenced genomes (Table 2.5), revealing that high numbers of RNA modification genes are found in the gamma-proteobacterial strains, all of which are psychrophilic. These gene products, then, may be an adaptation that enables the maintenance of RNA flexibility at low temperatures.

Conclusion

Research with CNPT3 has made important contributions to piezophilic investigations. As the first cultured deep sea bacterium, it enabled characterization of piezophilic growth characteristics and physiology. In genome sequencing, its genome was used in the analysis of sequencing technology. The fully sequenced genome adds to the availability of genome sequence data in general, as well as increasing piezophilic database, enabling the investigation of patterns in piezophilic adaptation. It is believed that further investigations in to the CNPT3 genome will continue to reveal characteristics of deep sea lifestyles. As with other genomic studies, transcriptional analyses will be useful in determining the effect of pressure on high pressure adapted gene regulation.

Table 2.1: Data about the genomes of *Psychromonas* sp. CNPT3 and *Psychromonas ingrahamii 37*, including information about size (in number of base pairs), GC content, and gene coding.

	CNPT3	P. ingrahamii
Size	3052410	4559598
G+C ratio	38.60%	40.09%
Percentage coding	85.93%	80.79%
Number of ORFs	2870	3861
Protein coding genes	95.88%	96.92%
With function prediction	72.71%	71.72%
Conserved hypothetical	12.86%	24.35%
Hypothetical	14.42%	0.85%
RNA genes	99	119
rRNA operons	10	10
tRNA genes	75	86

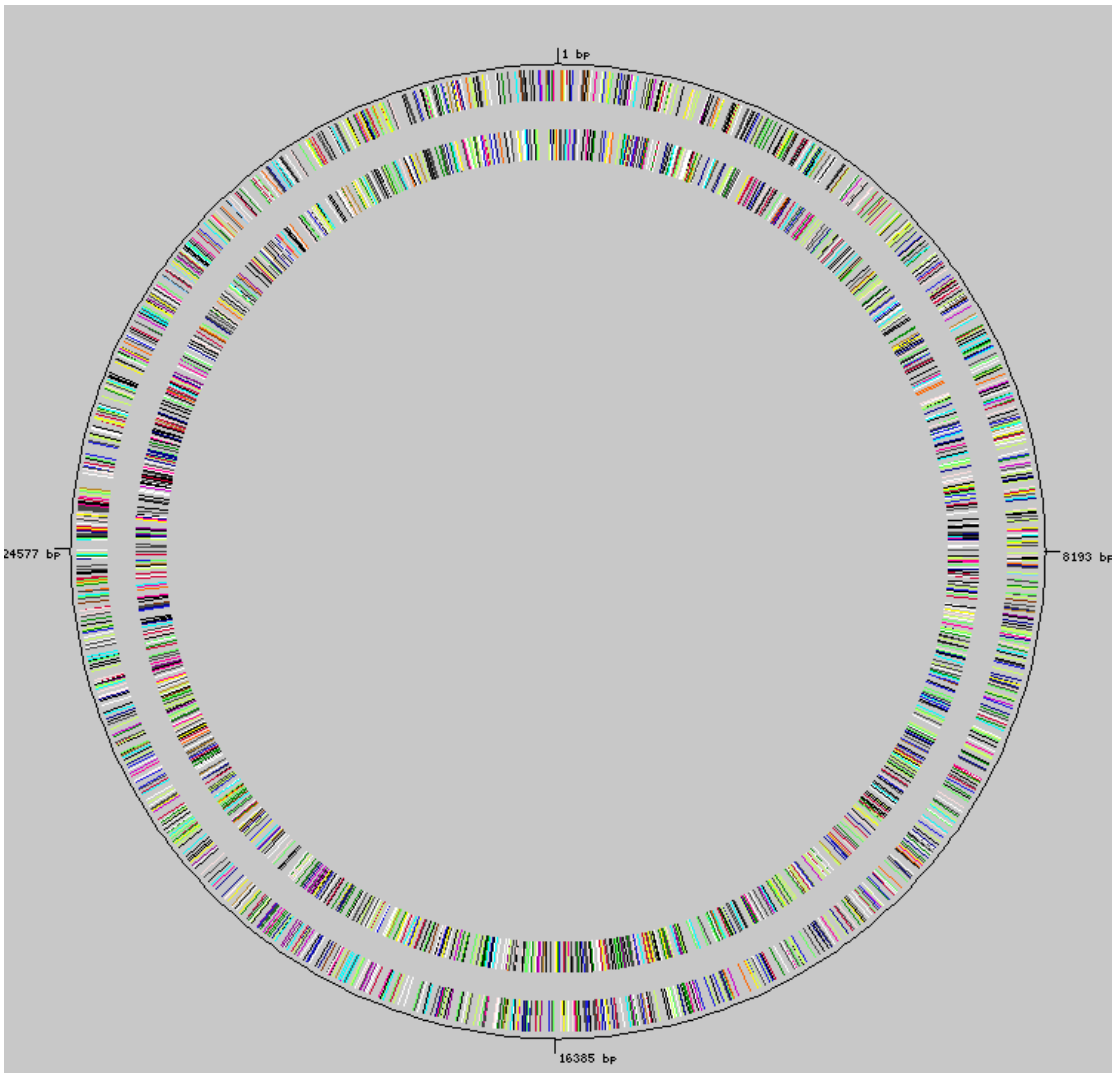

Figure 2.1: Circular map of the completed CNPT3 genome. Lines represent gene locations with colors assigning COG categories. Outer circle is the forward strand and inner circle is the reverse strand. The numbers represent approximate number of base pairs.

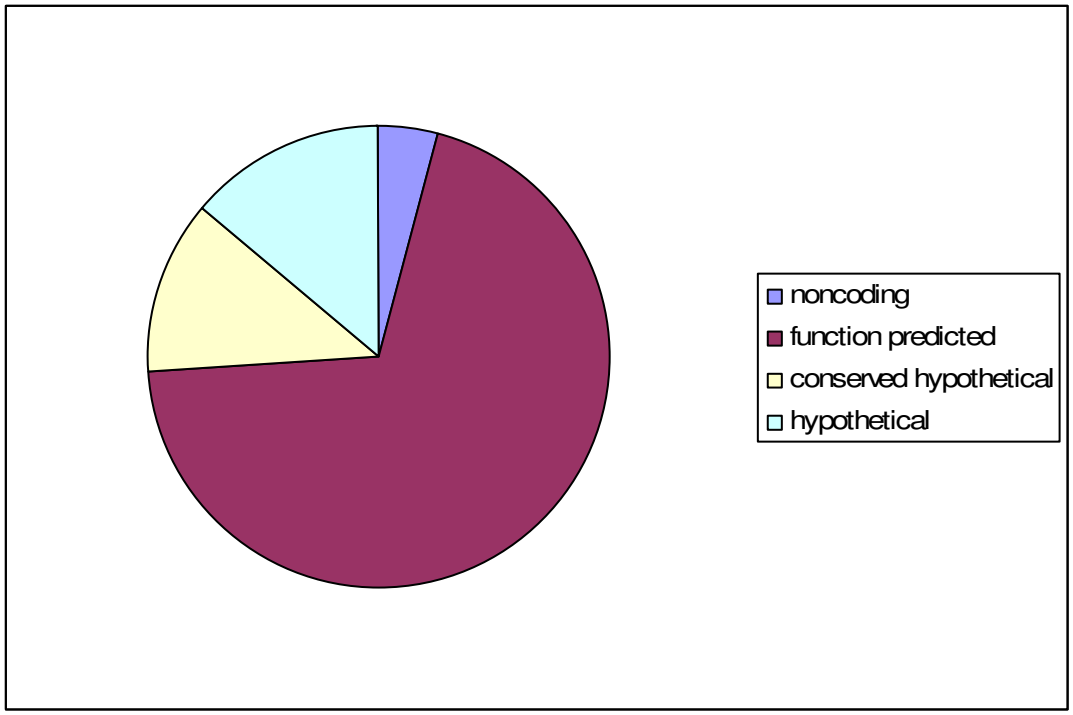

Figure 2.2: Pie chart representing the relative genome coding usage including portions of the genome that are noncoding, have functional prediction, no functional prediction but sequence similarity, or no functional prediction and no sequence similarity.

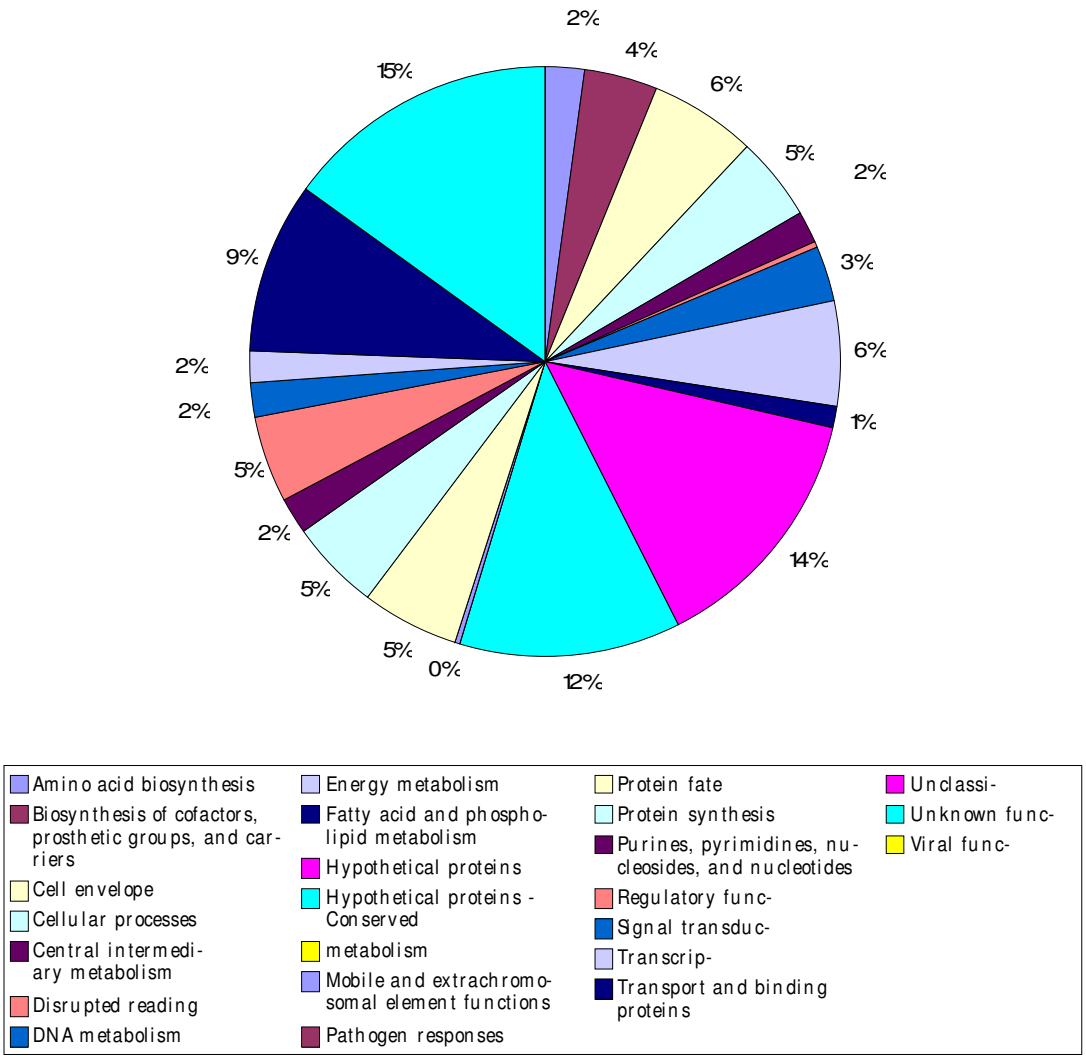

Figure 2.3: Role category breakdown for CNPT3 genome, in terms of COG category.

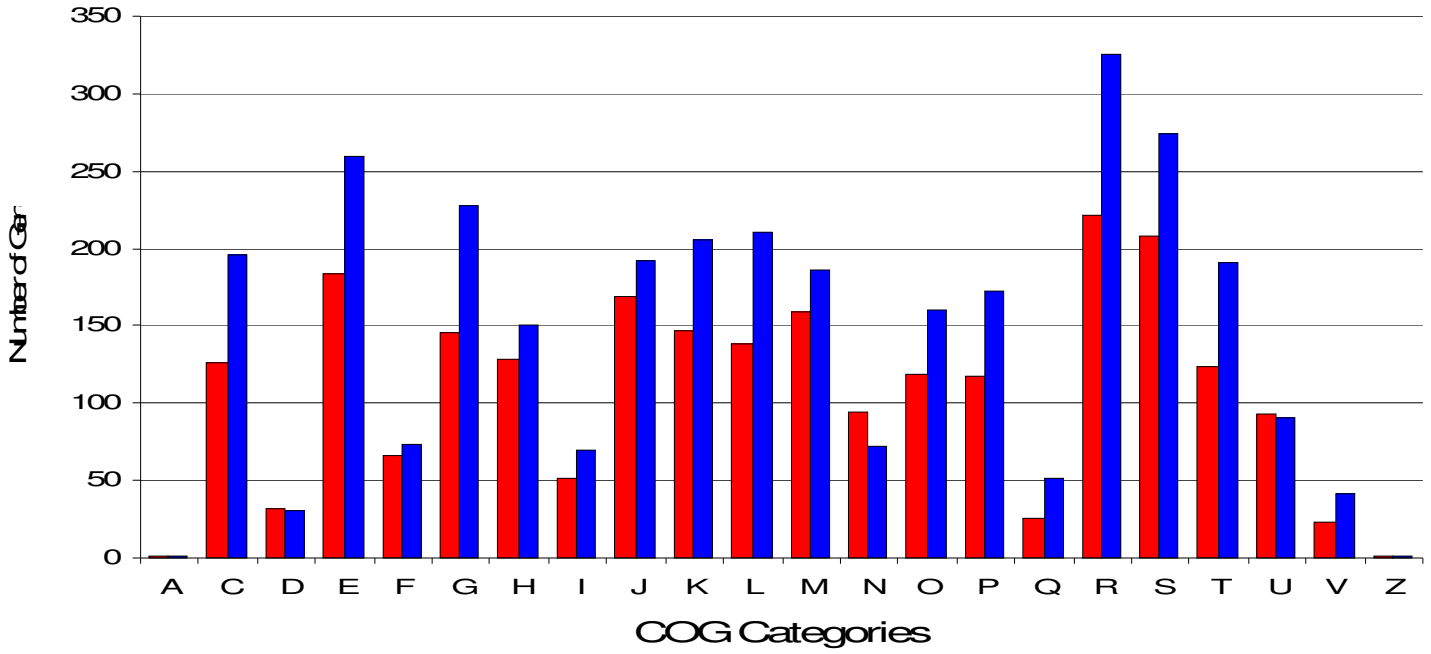

Figure 2.4: COG categories for CNPT3 (red) and P. ingrahamii (blue). The categories are as follows: RNA processing and modification, A; energy production and conversion, C; cell cycle control, cell division, chromosome partitioning, D; amino acid transport and metabolism, E; nucleotide transport and metabolism, F; carbohydrate transport and metabolism, G; coenzyme transport and metabolism, H; lipid transport and metabolism, I; translation, ribosomal structure and biogenesis, J; transcription, K; replication, recombination and repair, L; cell wall/membrane/envelope biogenesis, M; cell motility, N; posttranslational modification, protein turnover, chaperones, O; inorganic ion transport and metabolism, P; secondary metabolites biosynthesis, transport and catabolism, Q; general function prediction only, R; function unknown, S; signal transduction mechanisms, T; intracellular trafficking, secretion, and vesicular transport, U; defense mechanisms, V; cytoskeleton, Z.

Table 2.2: Comparison of CNPT3 and *P. ingrahamii* in terms of number of genes assigned to COG categories. Name of COG categories are listed as well as the number of genes from each genome.

ID	Name	CNPT3	P.ingrahamii
1	Archael/Vacuolar-type H+ ATPase subunits	0	1
2	F0F1-type ATP synthase subunits	7	15
3	Glyoxylate bypass	0	2
4	NA+-transporting NADH:Ubiquinone oxireductase subunits	8	7
5	NADH:Ubiquinone oxidreductase subunits	3	3
6	Pyruvate decraboxylation	6	8
7	TCA cycle	18	27
8	Arginine biosynthesis	7	16
9	Histidine biosynthesis	12	11
10	Isoleucine biosynthesis	5	11
11	Leucine biosyntehsis	9	15
12	Methionine biosynthesis	4	10
13	Phenylalanine/tyrosine biosynthesis	13	14
14	Proline biosynthesis	5	9
15	Threonine biosynthesis	6	6
16	Trypotophan biosynthesis	17	17
17	Valine biosynthesis	5	10
18	Purine biosynthesis	13	16
19	Purine salvage	5	4
20	Pyrimidine biosynthesis	12	12
21	Pyrimidine salvage	13	10
22	Thymidylate biosynthesis	10	8
23	Entner-Doudoroff pathway	3	8
24	Gluconeogenesis	14	14
25	Glycolysis	15	17
26	Pentose phosphate pathway	6	12
27	Biotin biosynthesis	6	7
28	Cobalamin biosynthesis	4	4
29	Coenzyme A biosynthesis	9	10
30	FAD biosynthesis	5	9
31	Heme biosynthesis	13	17
32	Menaquinone biosynthesis	16	21
33	NAD biosynthesis	5	6
34	Pyridoxal phosphate biosynthesis	7	8
35	Riboflavin biosynthesis	4	6
36	Thiamine biosynthesis	10	11
37	Ubiquinone biosynthesis	22	27
38	Deoxyxylulose pathway of terpenoid biosynthesis	5	5
39	Fatty acid biosyntheis	15	21
40	Aminoacyl-tRNA synthetases and alternate systems for amino acid activation	24	26
41	Ribosomal proteins - large subunit	31	32
42	Ribosomal proteins - small subunit	17	24
43	Translation factors and enzymes involved in translation	19	21
44	Basal transription factors	6	7
45	DNA-dependent RNA polymerase subunits	9	13
46	Transcriptional regulators	75	103
47	Basal replication machinery	26	27
48	DNA polymerase III subunits	11	11
49	Lipid A biosynthesis	10	10
50	Flagellum structure and biogenesis	35	28
51	Preprotein translocase subunits	9	9
52	Multisubunit NA+/H+ antiporter	2	6

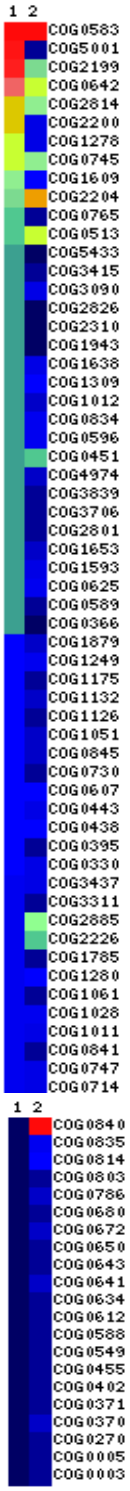

Figure 2.5: Heat map depicting relative abundances of genes in specific COG categories. Colors represent gene numbers as follows: red>yellow>green>blue>dark blue. Column 1 represents *P. ingrahamii*. Column 2 represents CNPT3. Key to some of the COG labels are found in Table 2.3. The colors have been adjusted for genome size.

Table 2.3: Key to the COG categories depicted in Figure 2.5. COG IDs are shown on the left, with their common name on the right. Categories highlighted in yellow are those with a higher number of genes in *P. ingrahamii*, categories highlighted in green are those with a higher number of genes in CNPT3, the categories that are white have the same number of genes in each genome, relative to genome size.

ID	func_name
COG0366	Glycosidases
COG0451	Nucleoside-diphosphate-sugar epimerases
COG0513	Superfamily II DNA and RNA helicases
COG0583	Transcriptional regulator
COG0589	Universal stress protein UspA and related nucleotide-binding proteins
COG0596	Predicted hydrolases or acyltransferases (alpha/beta hydrolase superfamily)
COG0625	Glutathione S-transferase
COG0642	Signal transduction histidine kinase
COG0745	Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain
COG0765	ABC-type amino acid transport system, permease component
COG0834	ABC-type amino acid transport/signal transduction systems, periplasmic component/domain
COG0840	Methyl-accepting chemotaxis protein
COG1012	NAD-dependent aldehyde dehydrogenases
COG1278	Cold shock proteins
COG1309	Transcriptional regulator
COG1593	TRAP-type C4-dicarboxylate transport system, large permease component
COG1638	TRAP-type C4-dicarboxylate transport system, periplasmic component
COG1653	ABC-type sugar transport system, periplasmic component
COG1943	Transposase and inactivated derivatives
COG2199	FOG: GGDEF domain
COG2200	FOG: EAL domain
COG2204	Response regulator containing CheY-like receiver, AAA-type ATPase, and DNA-binding domains
COG2226	Methylase involved in ubiquinone/menaquinone biosynthesis
COG2310	Uncharacterized proteins involved in stress response, homologs of TerZ and putative cAMP-binding protein CABP1
COG2801	Transposase and inactivated derivatives
COG2814	Arabinose efflux permease
COG2826	Transposase and inactivated derivatives, IS30 family
COG2885	Outer membrane protein and related peptidoglycan-associated (lipo)proteins
COG3090	TRAP-type C4-dicarboxylate transport system, small permease component
COG3415	Transposase and inactivated derivatives
COG3706	Response regulator containing a CheY-like receiver domain and a GGDEF domain
COG3839	ABC-type sugar transport systems, ATPase components
COG4974	Site-specific recombinase XerD
COG5001	Predicted signal transduction protein containing a membrane domain, an EAL and a GGDEF domain
COG5433	Transposase

Table 2.4: List of genes found in CNPT3 (and not in *P. ingrahamii*) categorized as methyl-accepting chemotaxis proteins, including length of gene product (in number of amino acids).

Product Name	Length
aerotaxis receptor Aer	572
putative methyl-accepting chemotaxis protein	459
methyl-accepting chemotaxis sensory transducer (IMGterm)	628
methyl-accepting chemotaxis protein	518
methyl-accepting chemotaxis sensory transducer (IMGterm)	621
methyl-accepting chemotaxis protein	396
methyl-accepting chemotaxis protein	433
methyl-accepting chemotaxis protein	362
methyl-accepting chemotaxis protein	564
putative methyl-accepting chemotaxis protein	514
hypothetical protein	656
hypothetical protein	629
methyl-accepting chemotaxis protein	673
methyl-accepting chemotaxis transducer	693
hypothetical methyl-accepting chemotaxis protein	636
hypothetical methyl-accepting chemotaxis protein	636
putative methyl-accepting chemotaxis protein	625
hypothetical methyl-accepting chemotaxis protein	243
hypothetical methyl-accepting chemotaxis protein	690

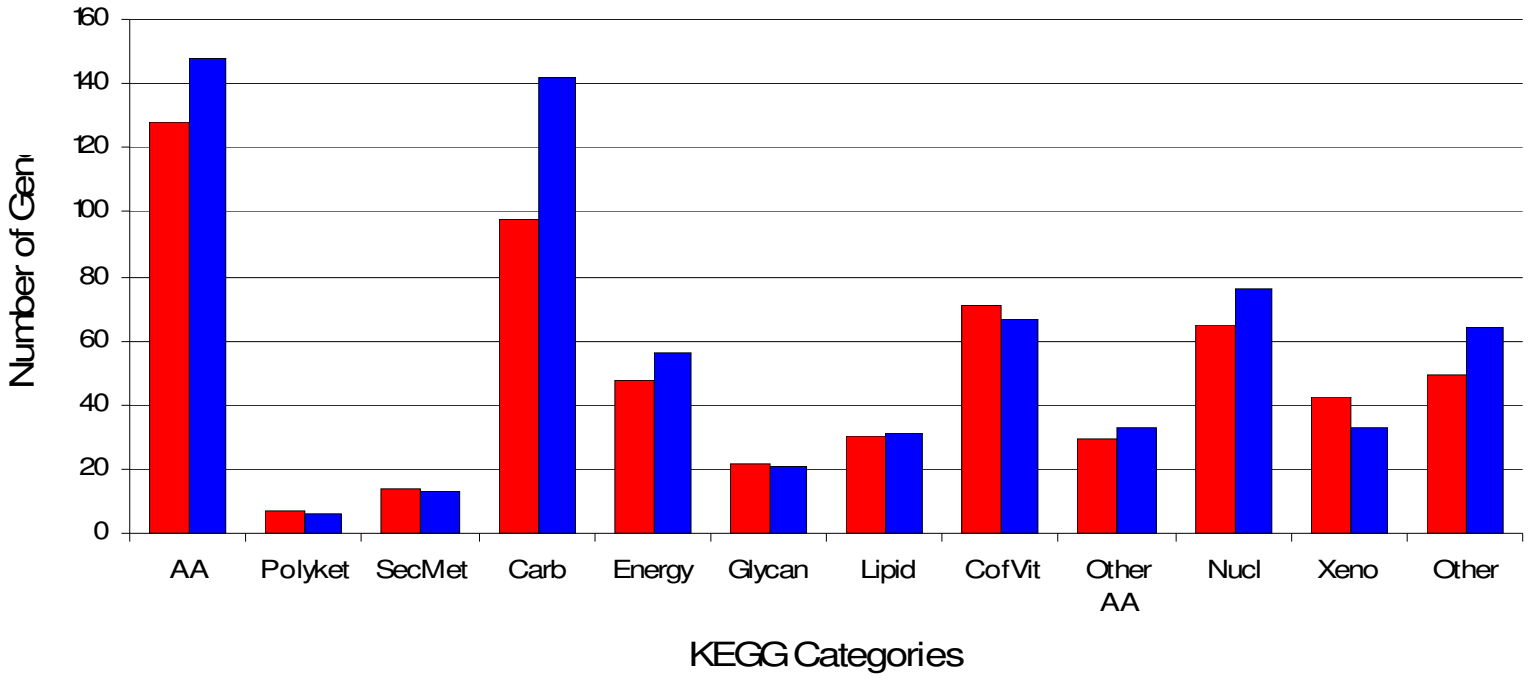

Figure 2.6: KEGG Categories for CNPT3 (red) and P. ingrahamii (blue). Categories, from left to right, are: amino acid metabolism, AA; biosynthesis of polyketides and nonribosomal pepetides, Polyket; biosynthesis of secondary metabolites, Sec Met; carbohydrate metabolism, Carb; energy metabolism, Energy; glycan biosynthesis and metabolism, Glycan; lipid metabolism, Lipid; metabolism of cofactors and vitamins, Cof Vit; metabolism of other amino acids, Other AA; nucleotide metabolism, Nucl; xenobiotics biodegredation and metabolism, Xeno; Other.

Table 2.5: Table depicting the number of genes involved in RNA modification found in piezophilic and comparison strains with sequenced genomes. The piezophilic strains are: *Psychromonas* sp. strain CNPT3, *Photobacterium profundum* strain SS9,

	Pseudouridine synthase	Pseudouridylate synthase	Other RNA modification	Total
CNPT3	13	0	2	15
P. ingrahamii	11	1	1	14
SS9	10	3	1	14
3TCK	14	0	2	16
AT7	4	1	1	6
E. faecalis	0	0	1	1
PE36	10	4	1	15
KT99	11	3	1	15
S. frigidimarina	17	3	2	22

Carnobacterium sp. strain AT7, *Moritella* sp. strain PE36, and *Shewanella benthica* strain KT99. All strains except *Enterococcus faecalis* are psychrophilic. *Carnobacterium* sp. strain AT7 and *E. faecalis*, are gram positive, all others are members of the gamma-proteobacteria.

III

Annotation of *Photobacterium profundum* strain 3TCK

Introduction

Photobacterium profundum strain SS9 (hereafter referred to as SS9) has been used as a model organism for piezophilic studies (Vezzi et al, 2005). It has been the subject of many piezophile analyses both because it is relatively easy to work with (due to its tolerance to a large range of pressures) and, as a member of the *Vibrionaceae* family, is related to a large number of species for which much is known (Vezzi et al, 2005). The completion of the SS9 genome sequence in 2005 was the first look at the molecular adaptations necessary for piezophily. However, distinguishing between genomic characteristics that are adaptations to environmental stresses and those simply due to shared phylogenetic traits requires genome comparisons between closely related species that differ in native environment. In this case, SS9 needed to be compared to a closely related, shallow water species. These comparisons would be used to investigate genomic perspectives on piezophilic adaptations. The chosen organism was *Photobacterium profundum* strain 3TCK (hereafter referred to as 3TCK), a piezosensitive bacterium from shallow water sediments (Campanaro et al, 2005).

The genus *Photobacterium*, as a member of the family *Vibrionaceae*, contains species from a diverse array of environments, and are, therefore, useful in studying many aspects of microbiology (Farmer, 2006)). The term "*Photobacterium*" was coined in 1889 to describe a group of bioluminescent bacteria (Farmer and Hickman-Brenner, 2006), though the genus currently contains many species that are not bioluminescent. The *Photobacterium* genus is comprised of gram negative, rod-shaped, motile,

fermentative bacteria from the marine environment (Farmer and Hickman-Brenner, 2006). The species *Photobacterium profundum* was described in 1998 to characterize the strains SS9 and DSJ4 (Nogi et al, 1998). Today, the species consists of four cultured strains, three with some degree of piezophily (SS9, DSJ4, and 1230) and one piezosensitive strain (3TCK).

As a piezosensitive strain of the species *Photobacterium profundum*, 3TCK provides a useful tool for comparisons with the well-studied strain SS9. Already, the similarities and differences between SS9 and 3TCK have been presented in several publications (Campanaro et al, 2005; Lauro et al, 2007). These studies, though, were performed using an incomplete 3TCK genome with an autoannotation. The finished genome with manual annotation is necessary to ensure the completeness and accuracy of the genes being compared.

Annotation

Annotation, when used in genomics, is the process of determining the names and functions of gene products within a genome. The process of annotation involves three main steps: gene finding, homology searches, and functional assignment. Annotation can be performed automatically, in which a computer using statistical models determines the identity of the putative gene, as well as manually, in which the gene name and function is determined by a human. The annotation of 3TCK utilized a combination of both methods. Automated annotation was initially completed by The Institute for Genomic Research (TIGR) which performs the necessary statistical analyses on the completed genome sequence and presents the data through the MANATEE program for manual review and annotation. MANATEE was downloaded from SourceForge

(manatee.sourceforge.net) and stored on a local server.

The annotation first performs "gene finding" using Glimmer to determine which open reading frames are predicted to be real genes. Open reading frames (ORFs) are regions of the genome between stop codons and of adequate length to contain a gene. Glimmer uses Interpolated Markov Models (IMMs) to compare nucleotide patterns of the organism's known genes to the entire genome. ORFs with similar patterns are deemed real genes. The function of the predicted gene is determined via homology searching which, relying on the assumption that shared sequence means a shared function, compares the unknown genes to those in other organisms that are known due to experimental characterization. Evidence for sequence similarity comes from pairwise alignments (via BLAST, Smith-Waterman, and Needleman-Wunsch), multiple alignments (via ClustalW and Muscle), protein families and motifs. TIGR searches against the NIAA (Non-Identical Amino Acid), a protein file comprised of several databases including Swiss-Prot, Omnium, NCBI, and PIR. The search and results are presented in MANATEE's BER list. The BER (BLAST-Extend-Repraze) creates a minidatabase of similar proteins as determined by BLAST, extends the query protein, and then uses a Smith-Waterman alignment to compare the extended protein to those in the minidatabase. The resulting matches are presented based on percent identity, percent similarity and the status of characterization of the matched protein. Another source of evidence is the HMMs (Hidden Markov Models). The HMMs are statistical models of amino acid patterns in a group of functionally related proteins that compare the query protein to a multiple alignment of proteins known to have the same function. MANATEE uses the statistical scores to interpret gene function to varying degrees of

specificity. Similar statistical models are also used to give information about membrane spanning regions, secreted proteins tags and lipoprotein characteristics.

A common problem in genome annotation is genome rot due to passing on the annotation of one gene to another based purely on sequence similarity. These transitive annotation errors were prevented through manual annotation, in which all of the lines of evidence were analyzed and interpreted. Further analyses were performed on the Integrated Microbial Genomes Version 2.5 (March 2008) website hosted by the Joint Genome Institute (Markowitz et al, 2007) and the Comprehensive Microbial Resource hosted by the J. Craig Venter Institute (Peterson et al, 2001).

Genome Characteristics and Analyses

The completed and fully annotated 3TCK genome is 6,167,633bp encoding 5584 genes. Like SS9 and other members of the *Vibrionaceae*, the 3TCK genome is divided between two circular chromosomes (Vezzi et al, 2005). Chromosome 1 (Figure 3.1) is the larger of the two, with 3,836,305bp and 3474 genes, including four rRNA operons (Table 3.1). The distribution of protein coding genes in chromosome 1 is shown in Figure 3.2. Other than hypothetical genes or genes of unknown function, the largest proportion of this chromosome is dedicated to transport and binding proteins, cell envelope, and energy metabolism. Compared to chromosome 2, chromosome 1 contains the majority of the stable, highly conserved genes (Figure 3.3). The more variable gene regions are found on chromosome 2 (Figure 3.4), which is smaller at 2,332,328bp and 2110 genes. Like chromosome 1, the largest proportions of this chromosome are dedicated to hypothetical proteins, proteins of unknown function, transport and binding proteins, and energy metabolism (Figure 3.5), though these proportions are higher than in

chromosome 1. Overall, there is less variation in the gene properties found in chromosome 2. Both chromosomes also have slightly different GC content, with chromosome 1 at 41.7% (Figure 3.6) and chromosome 2 at 40.8% (Figure 3.7).

Previous genome comparisons with SS9 have focused on genes that are present in SS9, but not found in 3TCK (Campanaro et al, 2005). SS9 has a larger genome (6,403,284bp) comprised of two circular chromosomes and one plasmid. Campanaro and colleagues (2005) determined that much of the additional sequence is due to horizontal gene transfer in the form of both phages and conjugated plasmid integration. 3TCK, though, is not without genes encoding phage or transposases. On chromosome 1, there are 13 genes predicted to belong to phage function, plus three that are putative phage shock proteins, involved in cellular response to infection (Table 3.2). Similarly, chromosome 2 contains six phage proteins (five of which are putative integrases) and three transposon genes (Table 3.3). 3TCK appears to lack genes for conjugation.

3TCK was compared to SS9 in terms of the representation of genes in the TIGR role categories (Figure 3.8). The proportion of genes in the 3TCK genome categorized as hypothetical genes and genes of unknown function is higher than the proportion in SS9. In many categories, the SS9 and 3TCK genomes have a similar composition. SS9 has a much higher percentage of genes dedicated to cellular processes, metabolism, and mobile and extrachromosomal elements. 3TCK genes with predicted function exceeded SS9's in only the cell envelope, signal transduction, and transport and binding proteins categories. The specific genes present or absent in these categories and their implications need to be investigated further. SS9 also shows a large number of unclassified genes. This is thought to be due to discrepancies in annotation, as many appear to be hypothetical or

merely not present in 3TCK.

An investigation of KEGG (Kyoto Encyclopedia of Genes and Genomes; Kanehisa and Goto, 2000) metabolic pathways revealed a set of 33 enzymes found in 3TCK, but not in SS9. Many of the enzymes belong to multiple metabolic pathways. These enzymes are listed in Table 4 along with the KEGG pathways in which they are found, their TIGR role category, and their chromosome of origin. The Glycerophopholipid Metabolism and Glycine, Serine, and Threonine Metabolism pathways are the ones in which 3TCK has the greatest number of enzymes not present in SS9 (Figure 3.9). These pathways are involved in structural lipids in membranes and amino acid usage, respectively. The implications of these extra genes in 3TCK are unknown, though it could be due to the differences in nutrient availability in its shallow water habitat.

3TCK shares some features investigated in the SS9 genome. One is genes for the Stickland reaction, an amino acid fermentation pathway found only in the Clostridiales, Spirochaetales (all anaerobic bacteria) and SS9 (Vezzi et al, 2005). 3TCK possesses several genes for selenocysteine containing proteins and the production of selenoproteins, which are necessary for the Stickland reaction. Further analyses would be required to determine if 3TCK possesses the genes necessary for the complete pathway. 3TCK also contains the genes necessary for the production of polyunsaturated fatty acids (PUFAs). Like SS9, 3TCK has genes *pfaA-D* (PUFA synthases) and a *pfaA* with five acyl carrier protein (ACP) domains (Allen and Bartlett, 2002). The number of ACP domains appears to vary between genera, though the implication of this variability is unknown (Allen and Bartlett, 2002). A comparison of all the genes in both genomes will be necessary to

determine the similarities and differences between the two, and if this variability constitutes deep sea adaptation.

Conclusion

Because of the previous work done with SS9, the completed genome of 3TCK will serve as a useful tool for comparing deep-sea organisms to shallow-water relatives. Since they are two strains of the same species, SS9 and 3TCK are the most closely related pair of piezophile/piezosensitive comparison strains. Having the completed genomes for both strains will allow for thorough comparisons of each. Future research should include an alignment of the genomes as a whole. The necessary tools for these analyses will soon be available on the Comprehensive Microbial Resource (Peterson et al, 2001). Using a whole genome homology, it will be important to determine the genes that differ between SS9 and 3TCK. This will allow for identification of specific similarities between the two genomes, such as metabolic pathways and regulatory processes. Perhaps more importantly, it will identify specific differences between the two, information that may be vital to understanding their differential adaptation to high pressure. Analyses, such as those found in Campanaro et al (2005), should be produced to determine gene clusters present in 3TCK, but missing in SS9. Furthermore, there are hypotheses that indicate that 3TCK evolved from a piezophilic species (Lauro et al, 2007). The completed genome sequence and annotation may help to confirm or deny these hypotheses and others regarding deep-sea adaptation.

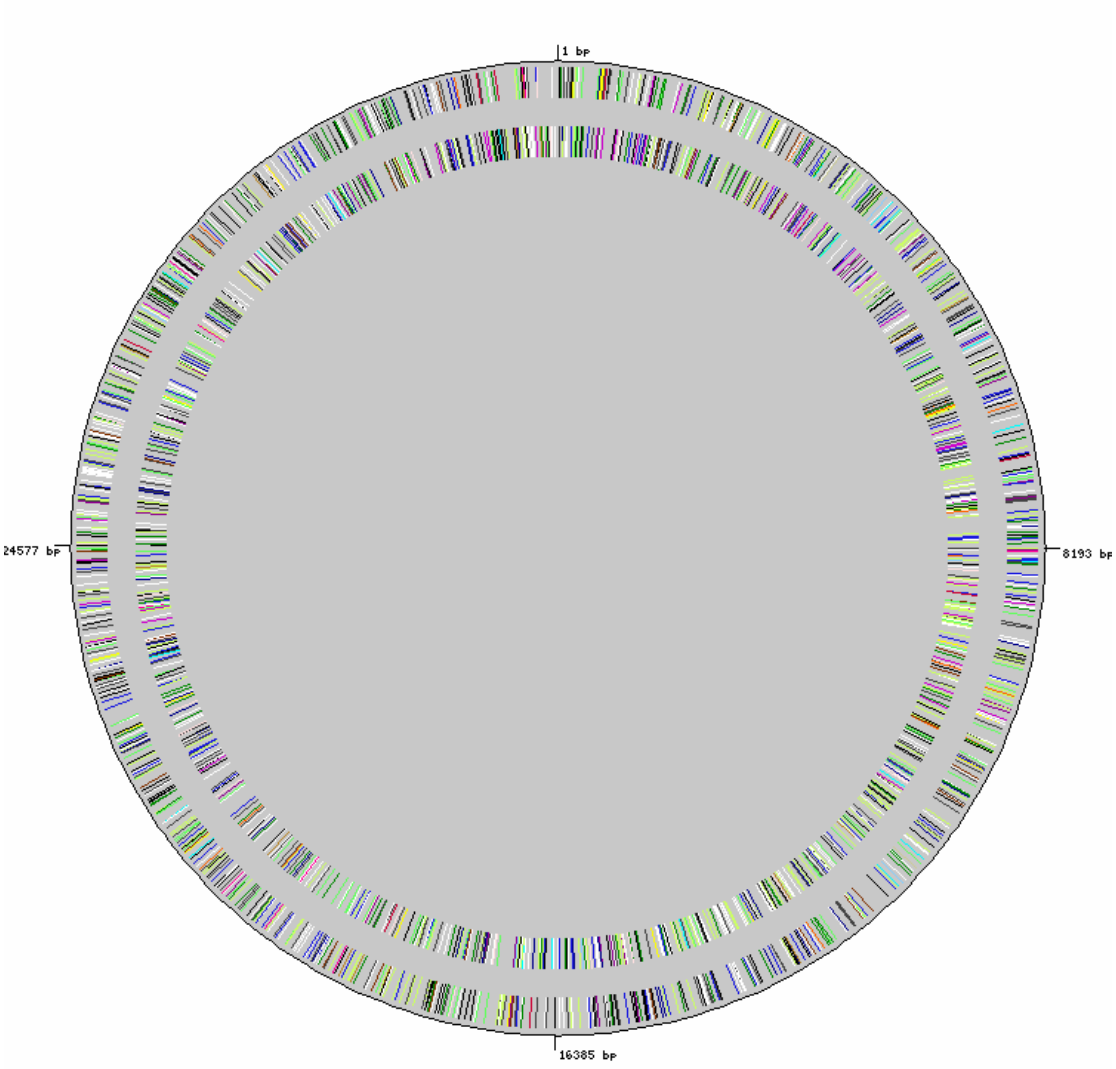

Figure 3.1: Circular map of *Photobacterium profundum* 3TCK chromosome 1. The outer circle is the forward strand, the inner circle is the reverse strand and the lines represent individual genes with colors to identify COG categories of the genes.

Table 3.1: Data about the DNA molecules of *Photobacterium profundum* 3TCK and *Photobacterium profundum* SS9, including information about size (in number of base pairs), GC content, and gene coding (Peterson et al, 2001; Vezzi et al, 2005).

	3TCK		SS9		
	Chrom 1	Chrom 2	Chrom 1	Chrom 2	Plasmid
Size	3,836,305	2,331,328	4,085,304	2,237,943	80,033
GC	41.7%	40.8%	42.0%	41.2%	44.0%
% coding	84.5%	84.2%	82.1%	80.2%	73.3%
# ORFs	3474	2110	3531	2079	69
Protein coding					
Function predition	61.9%	53.8%	76.6%	64.1%	39.1%
Conserved hypothetical	26.3%	35.4%	14.9%	22.8%	26.1%
Hypothetical	12.3%	11.5%	4.5%	8.9%	21.7%
RNA operons					
rRNA	4	3	14	1	0
tRNA	130	21	145	32	0

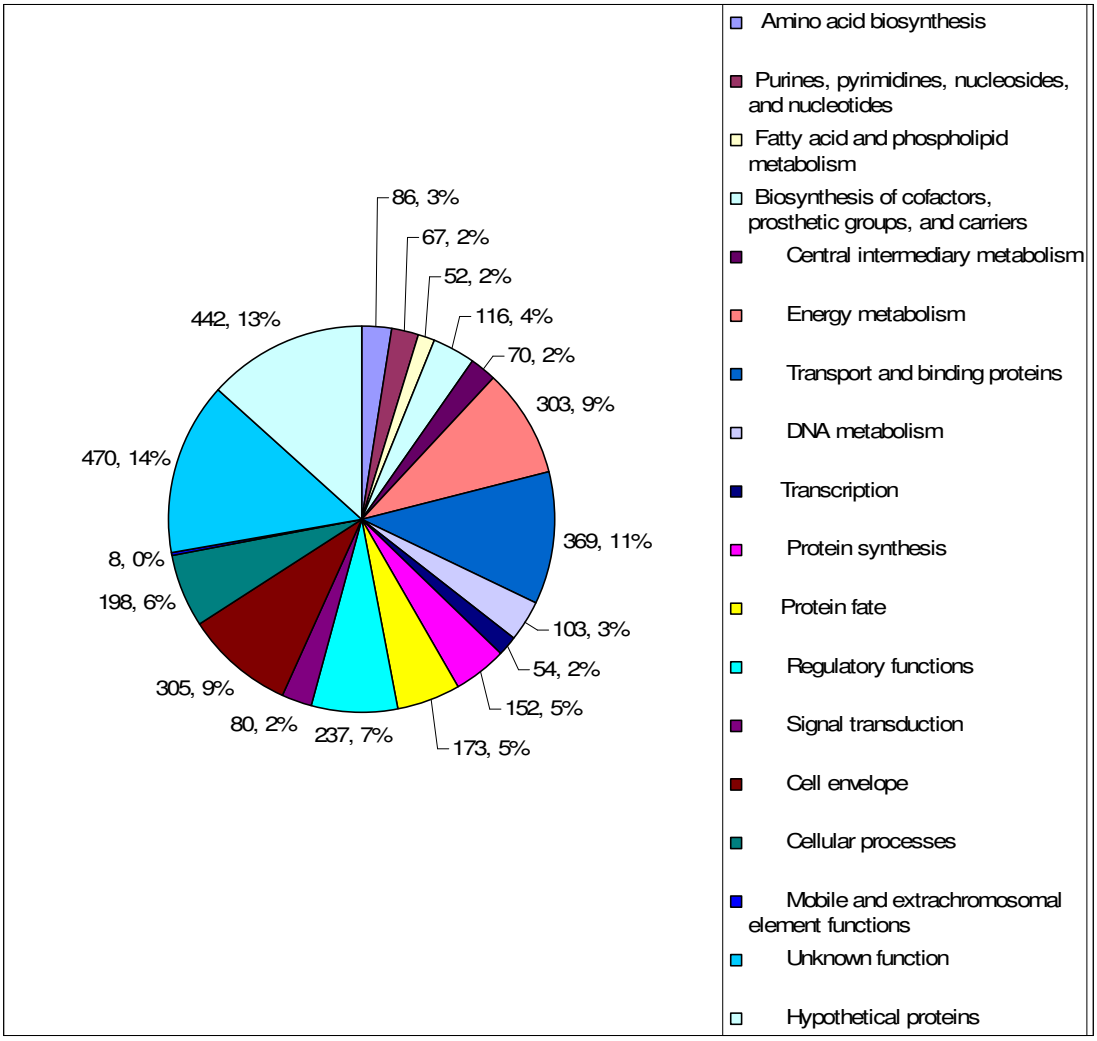

Figure 3.2: Pie chart showing the percentage of 3TCK chromosome 1 genes in TIGR role categories. Colors represent role categories as shown in the legend on the right, numbers represent the number of genes in each category and the percentage of total genes in the role category.

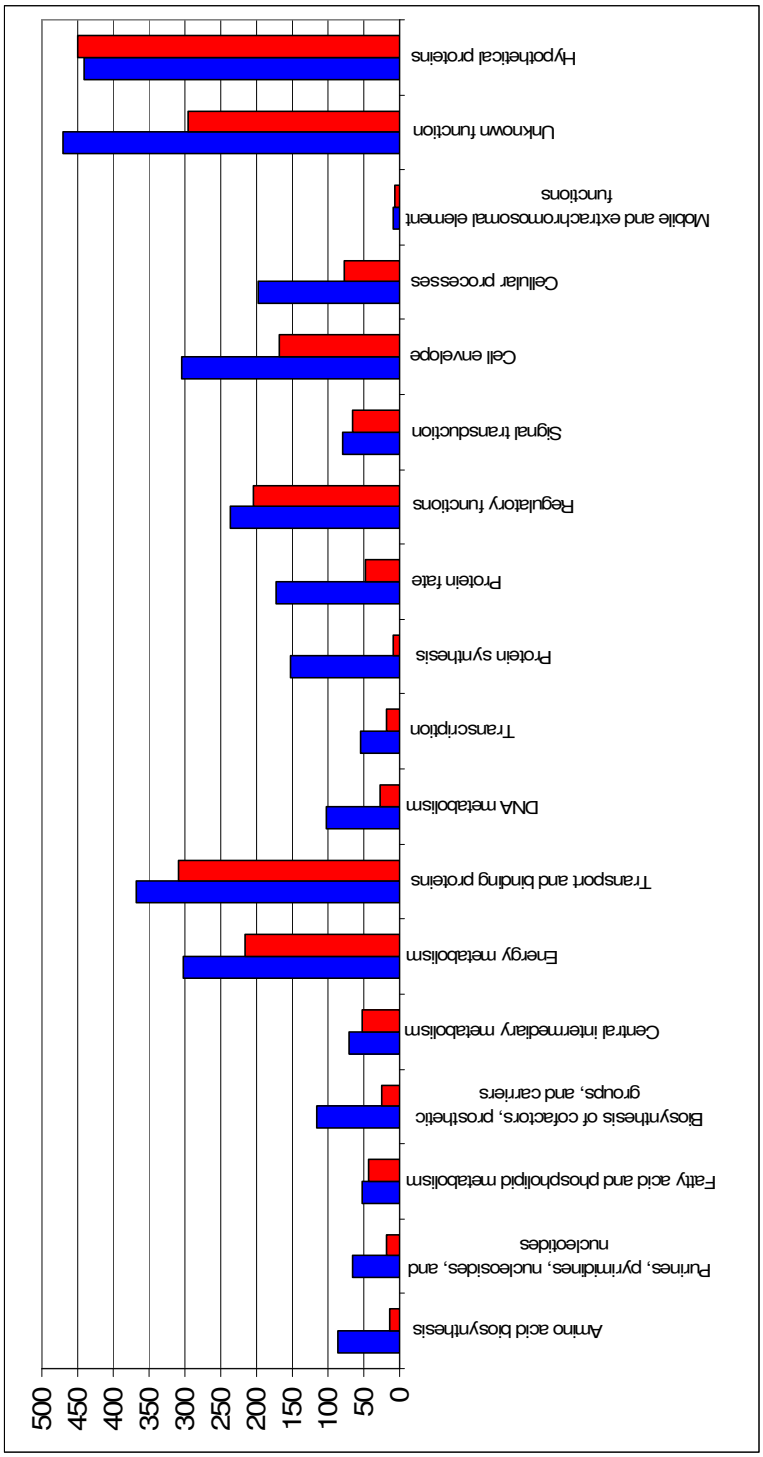

Figure 3.3: Bar chart showing the number of genes in TIGR role categories in chromosome 1 (blue) and chromosome 2 (red). It compares the number of genes in each role category between the chromosomes showing that chromosome 1 has a greater number of genes in all categories except hypothetical proteins.

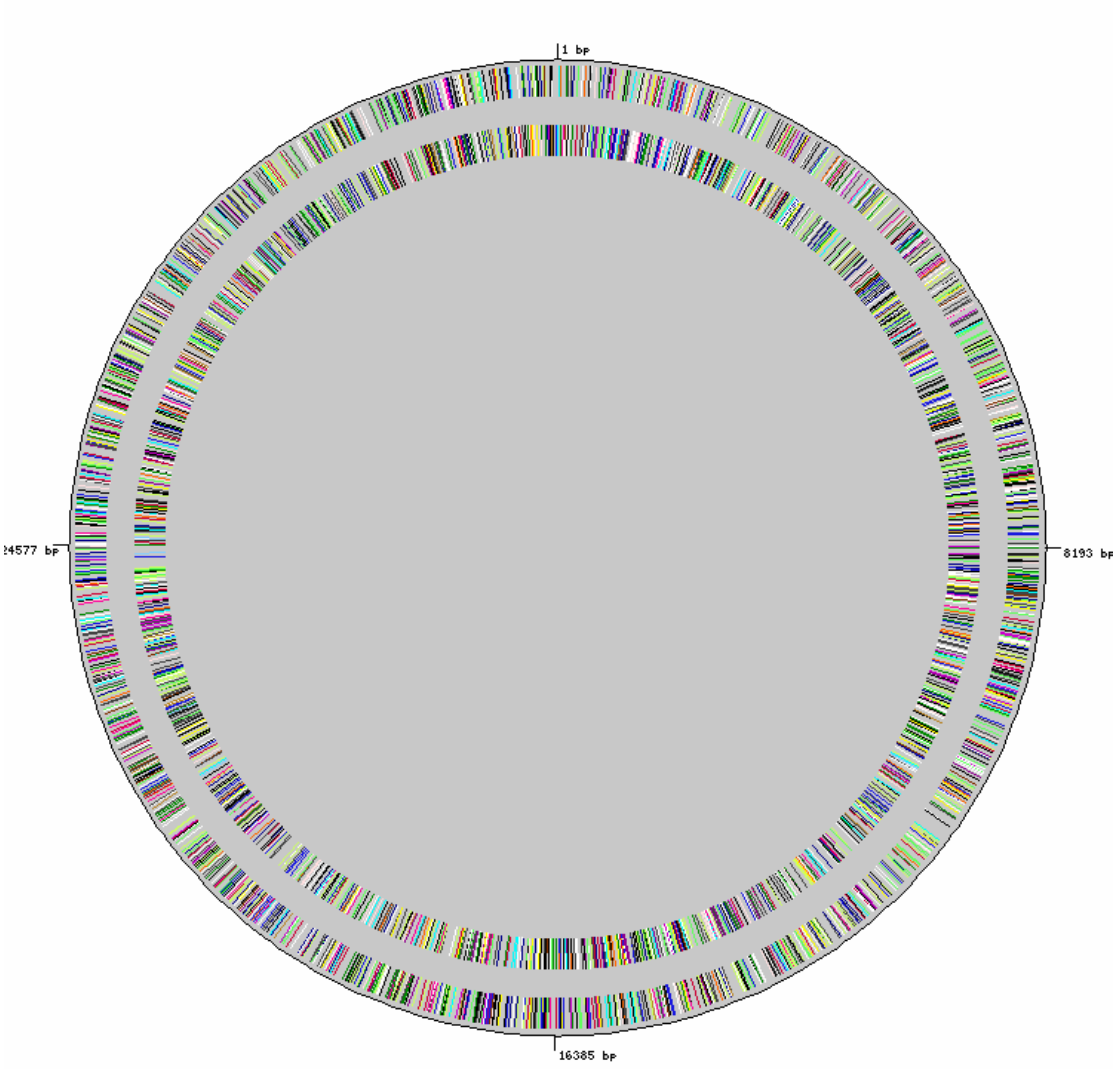

Figure 3.4: Circular map of *Photobacterium profundum* 3TCK chromosome 2. The outer circle is the forward strand, the inner circle is the reverse strand and the lines represent individual genes with colors to identify COG categories of the genes.

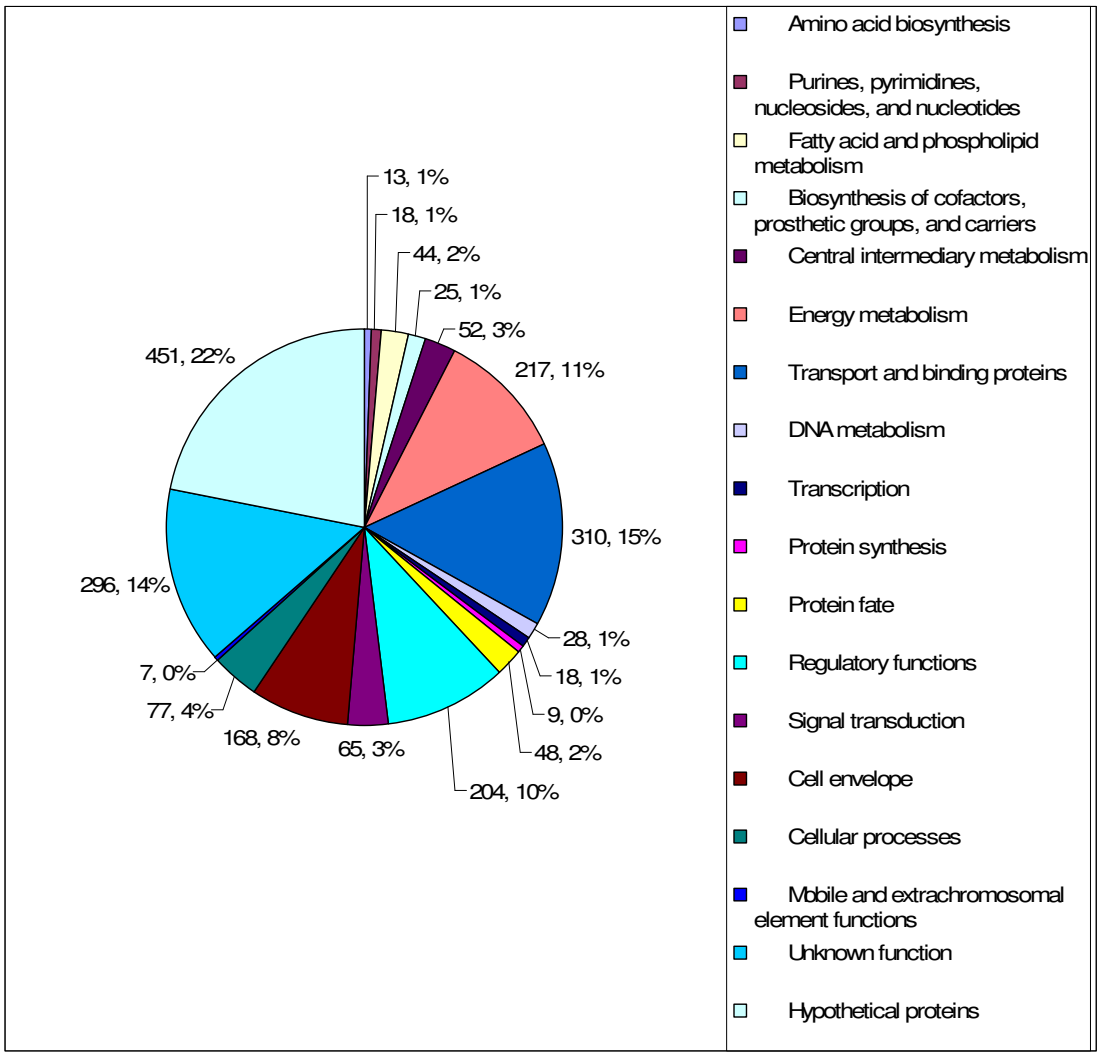

Figure 3.5: Pie chart showing the percentage of 3TCK chromosome 2 genes in TIGR role categories. Colors represent role categories as shown in the legend on the right, numbers represent the number of genes in each category and the percentage of total genes in the role category.

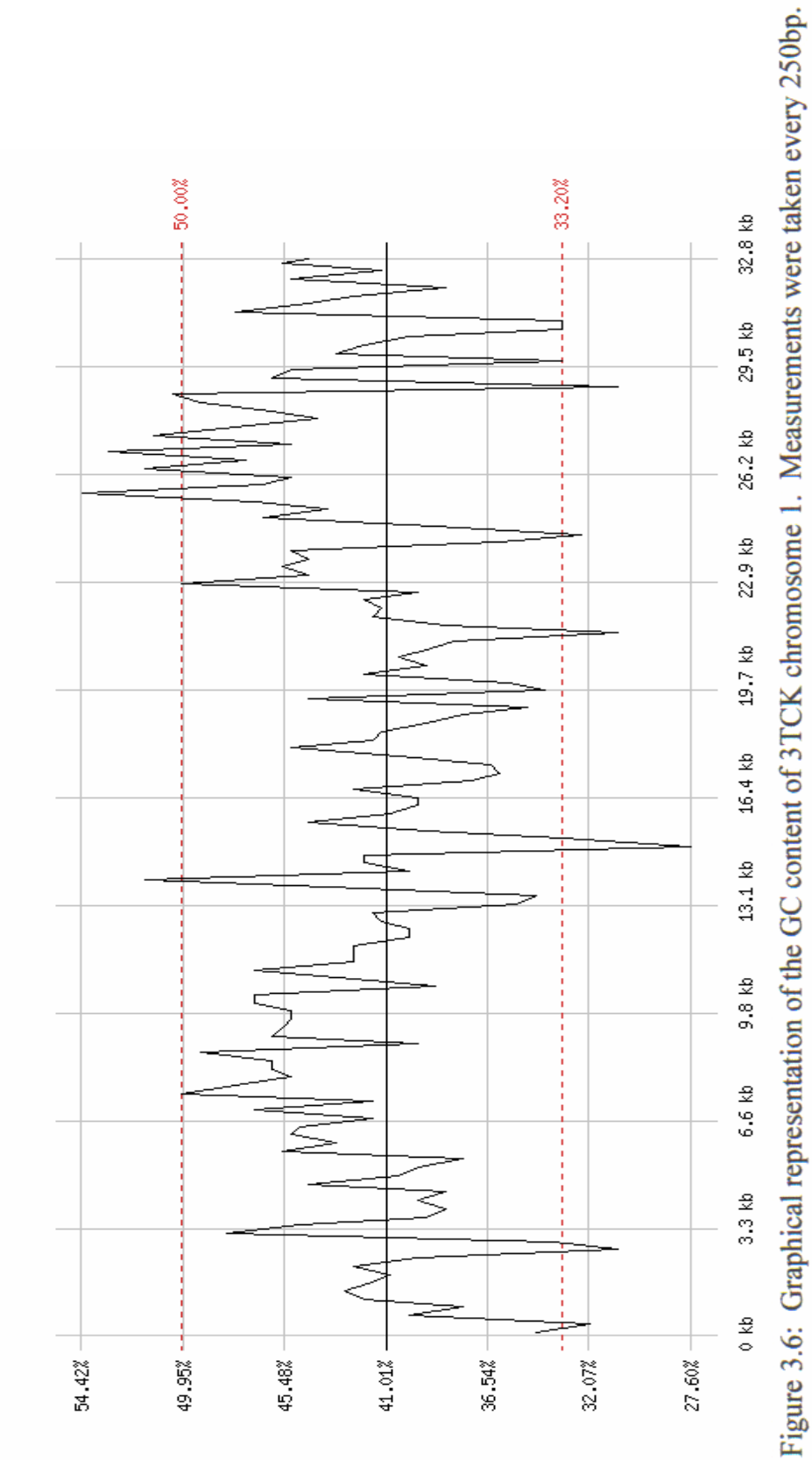

Figure 3.6: Graphical representation of the GC content of 3TCK chromosome 1. Measurements were taken every 250bp.

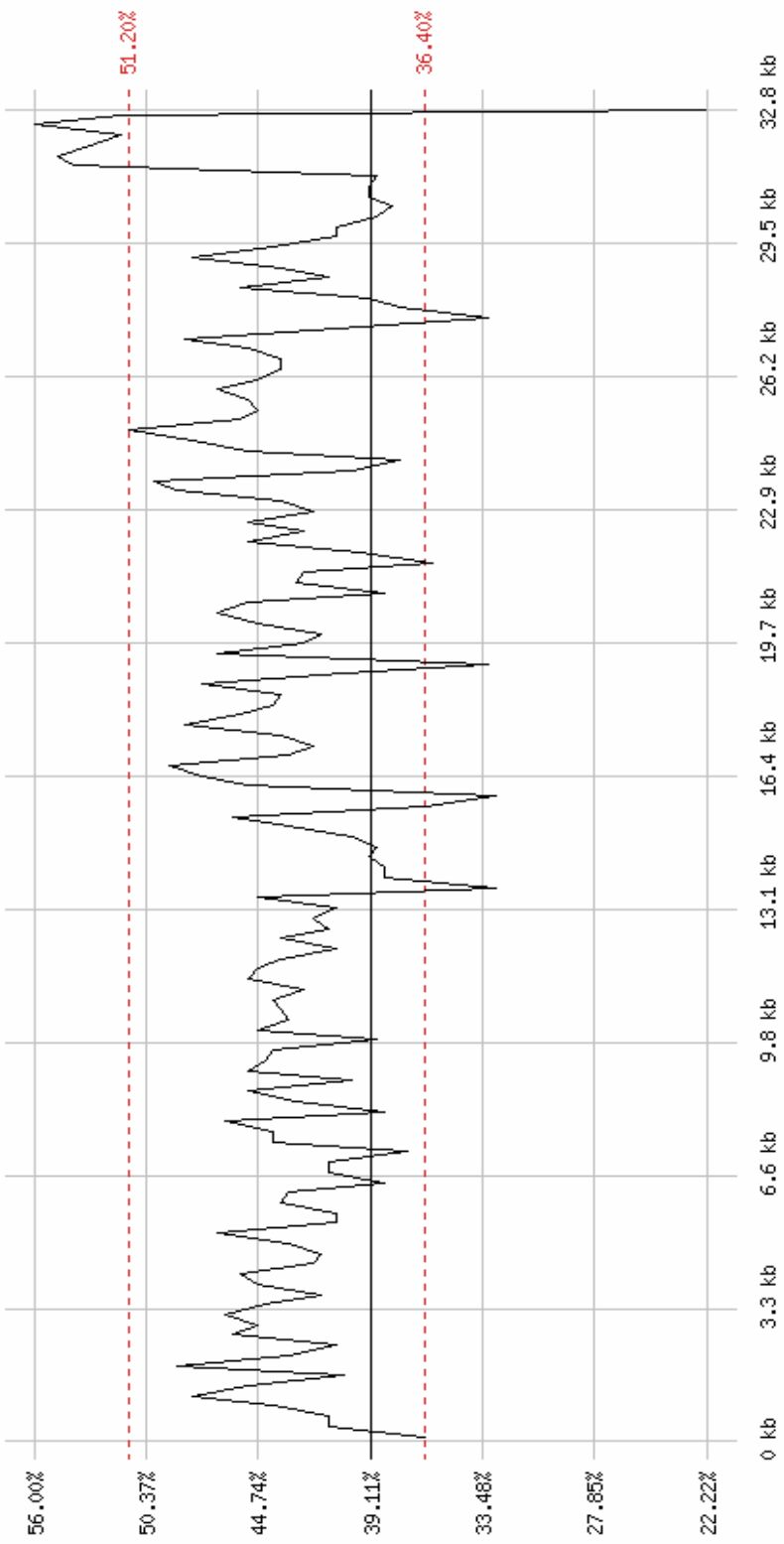

Figure 3.7: Graphical representation of the GC content of 3TCK chromosome 2. Measurements were taken every 250bp.

Table 3.2: Table of the genes found in chromosome 1 thought to be involved in phage and phage response functions.

gene id	locus	gene name	gene symbol
ORF00856	NT04PP0932	putative phage integrase family	
ORF00887	NT04PP0964	putative phage integrase family domain	
ORF01124	NT04PP1206	putative phage shock protein A	pspA
ORF01125	NT04PP1207	phage shock protein B	pspB
ORF01126	NT04PP1208	putative phage shock protein C	pspC
ORF01741	NT04PP1839	phage major coat protein	Gp8
ORF01743	NT04PP1841	bacteriophage replication gene A protein (GPA) domain homolog	
ORF01760	NT04PP1858	bacteriophage replication gene A protein (GPA) domain homolog	
ORF02202	NT04PP2304	prophage tail length tape measure protein putative	
ORF02208	NT04PP2310	phage head-tail adaptor putative	
ORF02211	NT04PP2313	phage major capsid protein HK97 family	
ORF02213	NT04PP2315	phage portal protein HK97 family	
ORF02214	NT04PP2316	phage terminase large subunit putative	
ORF02249	NT04PP2351	phage integrase family	
ORF03065	NT04PP3181	putative phage integrase family	
ORF03069	NT04PP3185	bacteriophage replication gene A protein (GPA) domain homolog	

Table 3.3: Table of the genes found in chromosome 2 thought to be involved in phage and phage response functions.

gene id	locus	gene name
ORF01028	NT03PP_1039	phage major coat protein
ORF01033	NT03PP_1044	Phage integrase family protein
ORF01034	NT03PP_1045	Phage integrase family protein
ORF01067	NT03PP_1078	Phage integrase family protein
ORF01497	NT03PP_1512	Phage integrase family protein
ORF01505	NT03PP_1520	Phage integrase family protein
ORF00806	NT03PP_0817	Transposase (IS4 family)
ORF01044	NT03PP_1055	Putative transposase IS116/IS110/IS902 family
ORF01055	NT03PP_1066	Transposase (IS4 family)

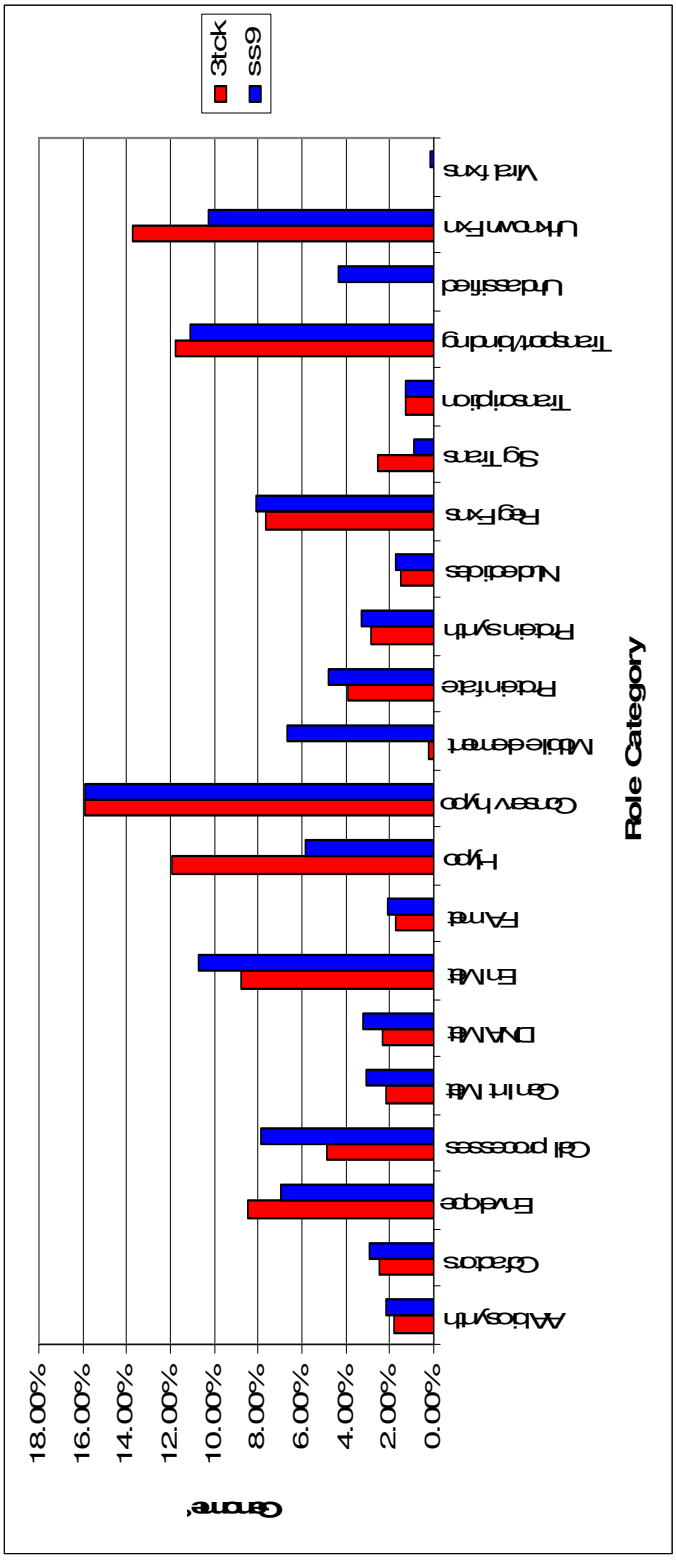

Figure 3.8: Bar chart comparing the percentage of the genes in the genomes of 3TCK (red) and SS9 (blue) assigned to TIGR role categories. Role categories are listed as follows: Amino acid biosynthesis, "AA biosynthesis"; Biosynthesis of cofactors, prosthetic groups and carriers, "Cofactors"; Cell envelope, "Envelope"; Cellular processes, "Cell processes"; Central Intermediary Metabolism, "Cen Int Met"; DNA Metabolism, "DNA Met"; Energy Metabolism, "En Met"; Fatty Acid and Phospholipid Metabolism, "FA met"; Hypothetical proteins, "Hypo"; Conserved Hypothetical Proteins, "Conserv hypo"; Mobile and Extrachromosomal Element Functions, "Mobile element"; Protein Fate, "Protein fate"; Protein Synthesis, "Protein synth"; Purines, Pyrimadines, Nucleosides, and Nucleotides, "Nucleotides"; Regulatory Functions, "Reg Fxns"; Signal Transduction, "Sig Trans"; Transcription; Transport and Binding Proteins, "Transport/binding"; Unclassified, "Unclassified"; Unknown Functions, "Unknown Fxns"; Viral Functions, "Viral fxns".

Table 3.4: Table of enzymes found in 3TCK, but not in SS9, including their TIGR role category, KEGG metabolic pathway and chromosome of origin.

Enzyme name	Chromosome	TIGR role category	KEGG Pathway
2-methylcitrate synthase	1	Energy metabolism: fermentation	Propanoate metabolism
5,10-methylevetetrahydrofolate reductase (metF)	2	Amino acid biosynthesis: aspartate family	Methane metabolism
acetate CoA-transferase alpha subunit	1	Energy metabolism: Other	Propanoate metabolism
			Butanoate metabolism
aminoacyl-histidine dipeptidase	2	Protein fate: degradation of proteins, peptides and glycopeptides	Alanine and aspartate metabolism
			Arginine and proline metabolism
			Histidine metabolism
			Beta-alanine metabolism
arginine decarboxylase (speA)	1	Central intermediary metabolism: polyamide biosynthesis	Glutamate metabolism
			Arginine and proline metabolism
aspartate aminotransferase	2	Amino acid biosynthesis: aspartate family	Glutamate metabolism
			Alanine and aspartate metabolism
			Arginine and proline metabolism
			Tyrosine metabolism
			Phenylalanine metabolism
			Phenylalanine, tyrosine, and tryptophan biosyı
beta-glucosidase A	2	Energy metabolism: biosynthesis and degradation of polysaccharides	Starch and sucrose metabolism
			Cyanoaminoacid metabolism
carbonic anhydrase	2	Central intermediary metabolism: other	Nitrogen metabolism:reduction and fixation
CDP-diacylglycerol-serine O-phosphatidyltransferase	2	Fatty acid and phospholipid metabolism: biosynthesis	Glycerophospholipid metabolism
			Glycine, serine, and threonine metabolism
cytosine deaminase (codA)	1	Purines, pyrimidines, nucleosides, and nucleotides: salvage of nucleosides and nucl	Pyrimidine metabolism
deoxycytidine triphosphate deaminase (dcd)	2	Purines, pyrimidines, nucleosides, and nucleotides: 2'-deoxyribonucleotide metaboli	Pyrimidine metabolism
deoxyuridine 5-triphosphatenucleotidohydrolase	2	Purines, pyrimidines, nucleosides, and nucleotides: 2'-deoxyribonucleotide metaboli	Pyrimidine metabolism
diacylglycerol kinase	1	Fatty acid and phospholipid metabolism: biosynthesis	Glycerolipid metabolism
			Glycerophospholipid metabolism
diaminobutyrate-2-oxogluterate aminotransferase	1	Cellular processes: adaptation to atypical conditions	Glycine, serine, and threonine metabolism
	2	Central intermediary metabolism: other	Glycine, serine, and threonine metabolism
ethanolamine ammonia-lyase, light chain (eutC)	1	Energy metabolism: amino acids and amines	Glycerophospholipid metabolism
ferredoxin-dependent glutamate synthase I	2	Amino acid biosynthesis: glutamate family	Nitrogen metabolism:reduction and fixation
fructokinase	2	Energy metabolism: sugars	Starch and sucrose metabolism
GDP-L-fucose synthase	2	Biosynthesis and degradation of surface polysaccharides and lipopolysaccharides	Fructose and mannose metabolism
glutaminase homolog	2	Unknown function: general	Nitrogen metabolism:reduction and fixation
			Glutamate metabolism
			D-glutamine and D-glutamate metabolism

Table 3.4, Continued: Table of enzymes found in 3TCK, but not in SS9, including their TIGR role category, KEGG metabolic pathway and chromosome of origin.

Enzyme name	Chromosome	TIGR role category	KEGG Pathway
guanine deaminase (guaD)	2	Purines, pyrimidines, nucleosides, and nucleotides: other	Purine metabolism
inositol-1-monophosphatase	1	Central intermediary metabolism: other; Unknown function: Other	Inositol phosphate metabolism
lactonizing lipase precursor	1	Fatty acid and phospholipid metabolism: degradation	Glycerolipid metabolism
long-chain-fatty-acid CoA ligase	1	Fatty acid and phospholipid metabolism: degradation	Fatty acid metabolism
low-specificity L-threonine aldolase	1	Energy metabolism: amino acids and amines	Glycine, serine, and threonine metabolism
maltodextrin glucosidase	2	Energy metabolism: sugars; Biosynthesis and degradation of polysaccharides	Galactose metabolism
			Starch and sucrose metabolism
methionyl-tRNA synthase	2	Protein synthesis: tRNA aminoacylation	Methionine metabolism
			Selenoamino acid metabolism
N-acetyl-D-glucosamine kinase	2	Central intermediary metabolism: amino sugars	Aminosugars metabolism
			Glutamate metabolism
N-acetylglucosamine 2-epimerase (GlcNAc 2-epimeras	1	Central intermediary metabolism: amino sugars	Aminosugars metabolism
NAD-dependent glycerol-3-phosphate dehydrogenase (	2	Energy metabolism: other	Glycerophospholipid metabolism
O-acetylhomoserine sulfydrylase	1	Amino acid biosynthesis: aspartate family	Methionine metabolism
phosphate acetyltransferase (pta)	1	Central intermediary metabolism: other; Energy metabolism: fermentation	Pyruvate metabolism
			Propanoate metabolism
			Taurine and hypotaurine metabolism
phosphatidylglycerophosphatase A (pgpA)	2	Fatty acid and phospholipid metabolism: degradation	Glycerophospholipid metabolism
phosphogluconate dehydratase	1	Entner-Duodoroff pathway	Pentose phophate pathway
phosphomannomutase (yhxB)	1	energy metabolism: sugars	Fructose and mannose metabolism
phosphomannomutase (yhxB)	2	energy metabolism: sugars	Fructose and mannose metabolism
polyribonucleotide nucleotityltransferase	2	Transcription: degradation of RNA	Purine metabolism
			Pyrimidine metabolism
putative agmanitase	1	Central intermediary metabolism: polyamide biosynthesis	Arginine and proline metabolism
putative glycosyl transferase (rfaCFKQ)	1	Cell envelope: biosynthesis and degradation of surface polysaccharides and lipopoly	Lipopolysaccharide biosynthesis
putative hydroxypyruvate reductase	2	Central intermediary metabolism; other	Glyoxylate and dicarboxylate metabolism
			Glycine, serine, and threonine metabolism
putative sucrose phosphorylase	1	Energy metabolism: biosynthesis and degradation of polysaccharides	Starch and sucrose metabolism
putative UDP-N-acetylglucosamine 2-epimerase	2	Cell envelope: biosynthesis and degradation of surface polysaccharides and lipopoly	Aminosugars metabolism
ribose-phosphate pyrophosphokinase	2	Purines, pyrimidines, nucleosides, and nucleotides: purine ribonucleotide biosynthes	Pentose phophate pathway
			Purine metabolism
S-formylglutatione hydrolase (fghA)	2	Cellular processes: detoxification	Methane metabolism
sulfoacetaldehyde acetyltransferase	1	Central intermediary metabolism: other	Taurine and hypotaurine metabolism
Urease (ureA, ureB, ureC)	1	Central intermediary metabolism: nitrogen metabolism	Purine metabolism
			Urea cycle and metabolism of amino groups

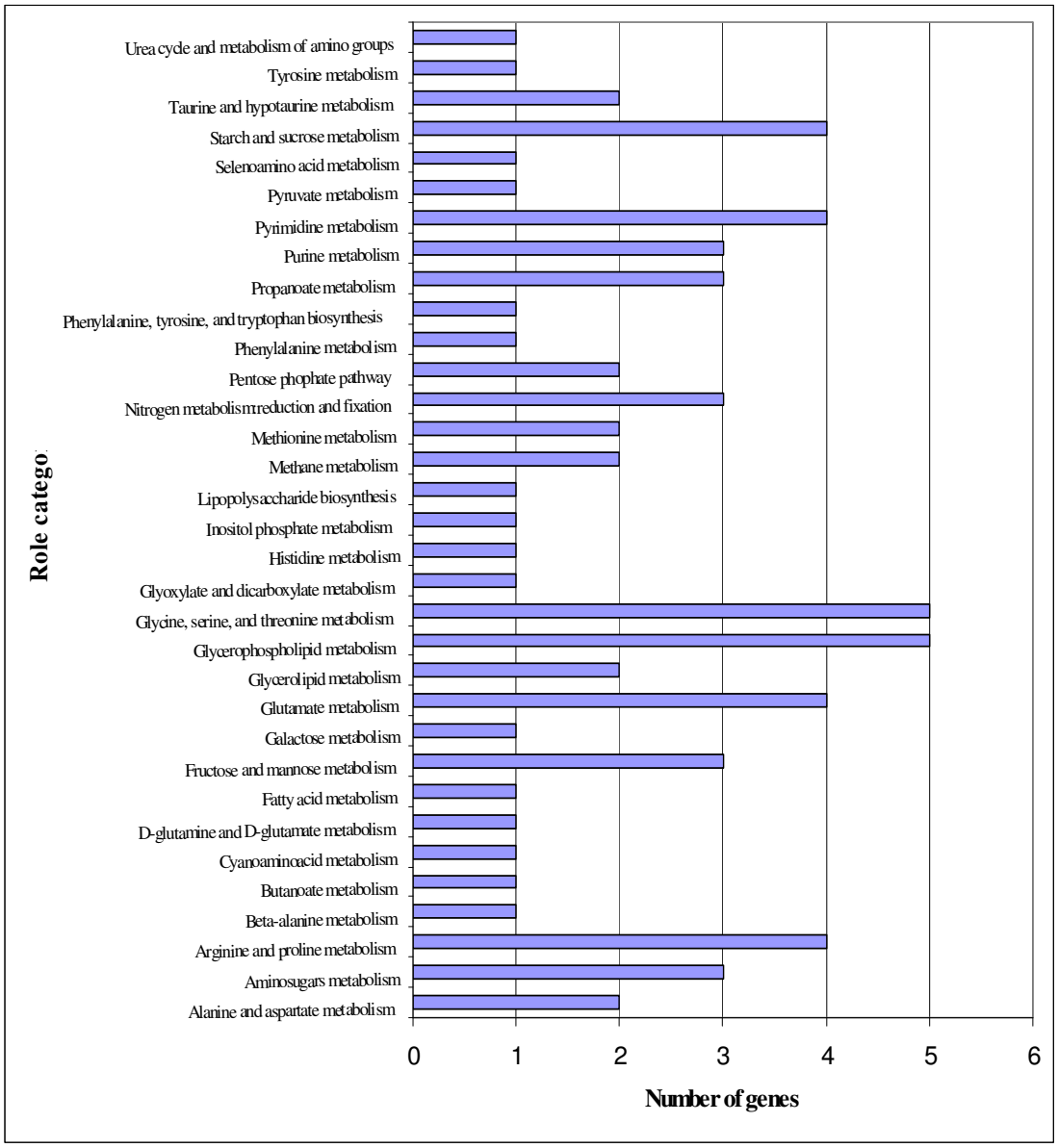

Figure 3.9: Graph showing the number of genes encoding enzymes categorized in KEGG metabolic pathways and present in 3TCK but not SS9.

IV

Analysis of Pressure Adaptation in the Primary Structure of Selected Proteins

Introduction

The importance of studying the effects of pressure on proteins was realized in the last half of the last century (Mozhaev et al, 1996). Exploring the effects of biosystems under pressure not only gives valuable insight to cellular regulation and bioprocesses in general (Gross and Jaenicke, 1994), but can also have useful biotechnological applications, such as in the food processing and pharmaceutical industries (Masson et al, 2001; Mozhaev et al. 1994). Ecologically speaking, pressure effects all aspects of organism functioning, and with over 70% of the Earth's surface is at high pressure, it is important in the distribution of life in aquatic environments (Somero, 1990).

The effects of pressure on cell physiology have primarily been studied with *E. coli*. These studies show that processes such as growth, cell division, and reproduction are greatly impacted by increases in pressure (Yayanos and Pollard, 1969; ZoBell and Cobet, 1962). Mesophilic cells subjected to high pressure initiate both heat shock and cold shock stress responses, indicating that a variety of physiological processes are impacted by this stress (Welch et al, 1993). Proteins involved in DNA replication, cell division, and stress response are believed to be the most likely sources of piezophilic adaptation. Additionally, other studies have found piezosensitive processes in the areas of signal transduction, motility, substrate transport and membrane biosynthesis (Bartlett, 2002; Lauro, 2007). Pressure, it seems, hinders the performance of proteins in many aspects of cellular functioning.

Hydrostatic pressure is known to impact protein structure and function. In fact, as

little as 50MPa is required to reduce enzyme flexibility, and 100MPa for protein degradation (Gross and Jaenicke, 1994). The difficulty that deep sea proteins face can be described by Le Chalier's principle. At high pressure, processes accompanied by positive volume change have increased energetic costs (Somero, 1990). Protein folding, the assembly of subunits, and conformational changes can be effected by high hydrostatic pressure because of the positive volume changes associated with some of these processes (Gross and Jaenicke, 1994; Somero, 1990). Final protein conformation, and its resulting volume fluctuations, is inherent in the protein's primary structure. It follows, then, that adaptations that require alterations in protein structure would be evident in the amino acid sequence. In fact, it has been reported that the development of pressure adaptation requires only minor changes in primary structure (Siebenaller, 1984; Somero, 1990). Investigations of piezophily in enzymes such as lactate dehydrogenase indicate that as few as a single amino acid change may be necessary for adaptation to high pressure (Siebenaller, 1984), especially when the substitutions influence secondary structure or surface hydration (Somero, 1990).

The advances in sequencing technology allow for the rapid determination of protein sequences from organisms from many environments and phylogenies. Trends in primary and secondary structure were studied using sequences from piezophiles from six genera and isolated from various parts of the deep sea (Table 4.1, Figure 4.1). This study sought to uncover protein trends that may be involved in high pressure adaptation.

Protein alignments

Fifteen proteins were investigated, chosen from processes known to be pressure-sensitive or that resulted in a pressure-sensitive phenotype when mutated as determined

by previous studies (Table 4.2). The protein alignments were performed using MUSCLE (Multiple Sequence Comparison by Log-Expectation), a multiple sequence alignment program for DNA and amino acid sequences, with output in the ClustalW2 format. Colors were used to identify amino acids based on the physio-chemical properties thought to be important in protein structure. The color groupings are identified in Table 4.3. The alignments compared six piezophiles and their close relatives in pairings as follows: *Photobacterium profundum* SS9 (Delong et al, 1997) and *Photobacterium profundum* 3TCK (Campanaro et al, 2005); *Psychromonas spp.* CNPT3 (Yayanos et al, 1979) and *Psychromonas ingrahamii* (Breezee et al, 2004); *Carnobacterium spp.* AT7 (Lauro et al, 2007) and *Enterococcus faecalis* (Paulsen et al, 2003); *Shewanella betnthica* KT99 (Lauro et al, 2007) and *Shewanella frigidimarina* (Bowman et al, 1997); *Colwellia spp.* MT41 (Yayanos et al, 1981) and *Colwellia psychrerythraea* (Methe et al, 2005)*; Moritella* spp. PE36 (Yayanos et al, 1986) and *Moritella viscosa* (Benedikstoddir et al, 2000). The location and depth of origin of all organisms is listed in Table 4.1. The amino acid sequences for proteins from published genes were retrieved from NCBI's GenBank either through a search for annotated gene name or through a BLAST of the query gene against a single organism's genome. Genes for the unreleased genomes *Colwellia* MT41 (Methe, JCVI, Rockville, Maryland) and *Moritella viscosa* (Grove, National Veterinary Institute, Oslo, Norway) were determined using amino acid sequence from the genes of their comparison genome and the tBLASTn function in command-line BLAST to search within the completed DNA sequence. Data about each protein sequence, including length, molecular weight, and isoelectric point (pI) was determined using the Pepstats program in the EMBOSS software suite (Rice et al, 2000).

FASTA files of the amino acid sequences were uploaded in to MUSCLE for alignment. The results were analyzed first as pairwise alignments, looking for amino acid differences between piezophiles and their piezosensitive relatives. The genes were then realigned comparing all high-pressure genes (in order of increasing piezophily) against their comparison strains to investigate trends in the effects of increasing pressure. The alignments were then compared, looking for patterns across the piezophile-piezosensitive relationships. Alignments were numbered according to the alignment as a whole, they include gaps in individual sequenes and do not necessarily reflect the actual amino acid number.

Sequences aligned more closely across phylogenetic lines than environmental groupings, with very few revealing piezophilic patterns. Where amino acid substitutions did occur, most fell within functionally similar amino acid subdivisions, and were not predicted to have resulted significant in protein structure or function. Several dissimilar amino acid substitutions occurred between the proteins of the comparison strains, but the substitution patterns were not conserved among the orthologs of piezophile origin, and were not deemed universal piezophile signatures.

Amino acid usage for individual amino acids and residue properties were calculated for each protein using Pepstats of the EMBOSS software suite (Rice et al, 2000). To verify the effects of amino acid substitutions on, secondary structure predictions (Jones, 1999) were performed on proteins showing the greatest degree of amino acid divergence. Previous studies have shown that β-sheet conformations are more stable under high hydrostatic pressure than α-helices (Gross and Jaenicke, 1994; Somero, 1990), leading to the belief that an increase in the use of β-sheets in protein

structure may be a piezophilic adaptation.

Possible trends in amino acid substitutions are discussed below. Additional investigation will be necessary to determine the biochemical consequences of these substitutions. It must also be acknowledged that strategies that enable survival in high pressure environments may be different between species, and may not be expressed through similarities in single amino acid substitutions.

Cell Division/Cell shape

FtsZ

The filamentous growth of mesophiles, such as *E. coli*, at high hydrostatic pressure revealed that cell division is disrupted by pressure (Ishii et al, 2004; Zobell and Cobet, 1962; Zobell and Cobet, 1964). Cell shape is determined by the cytoskeleton, which in bacteria, consist of the proteins FtsZ and MreB (Ishii et al, 2004). FtsZ polymerizes in to a ring localized at the division site and plays a vital role in the process of cell division (Bi and Lutkenhaus, 1991). Where FtsZ is non-functional, the result is filamentous growth from a lack of cell division (Bi and Lutkenhaus, 1991; Ishii et al, 2004). It has been shown that high pressure inhibits FtsZ ring formation in mesophiles, thereby preventing normal cell division (Ishii et al, 2004). As piezophilic bacteria do not demonstrate filamentous phenotypes at high pressure, their FtsZ is believed to be adapted to functioning at high pressure. PE36 did not have any significant matches to FtsZ. This could be because the alignment used an unfinished sequence and FtsZ had not yet been added to the assembly, or because its FtsZ sequence is highly divergent. The aligned sequences varied greatly in length (from 330 to 418 amino acids) and percent identity (as low as 38% between AT7 and CNPT3 or 51% just within the gamma-proteobacteria).

The rest of the aligned sequences failed to reveal any notable trends in amino acid substitutions (Figure 4.2).

MreB

MreB, the other major bacterial cytoskeletal protein, is an actin-like protein that is important in width control and essential for shape maintenance in non-spherical bacteria (Ishii et al, 2004; Jones et al, 2001). When mutated, bacteria demonstrate a loss of cell shape that is, in some cases, lethal (Jones et al, 2001). While MreB has not been investigated under pressure, studies on other cytoskeletal proteins have shown an increased incidence of depolymerization at high hydrostatic pressure (Bourns et al, 1996). As cell shape is a process commonly disrupted by exposure to high pressure (Zobell and Cobet, 1962; Zobell and Cobet, 1964), MreB was investigated among the rod-shaped bacteria for alterations in amino acid sequence. Since the spherical *E.* faecalis does not have the MreB protein, neither it nor AT7 were included in the alignment. The remaining ten sequences showed a high degree of conservation (percent identity 81-99%). The exceptions were those belonging to the genus *Moritella*, whose sequences were shorter than the rest by approximately 100 amino acids. There were no significant amino acid substitutions between the piezophilic and non-piezophilic sequences (Figure 4.3).

Membrane lipid regulation

FabF/FabB

Due to their high compressibility, lipids, and the membranes they comprise, are particularly sensitive to the solidifying effects of high pressure (Bartlett, 2002; Braganza and Worcester, 1986). Organisms adapted to the deep-sea, maintain membrane fluidity

through the addition of unsaturated fatty acids in their cell membranes (Kamimura et al, 1993). Specifically, microbes increase the proportion of *cis*-vaccenic acid, the production of which is catalyzed by beta-ketoacyl-acyl carrier protein synthases I (FabB) and II (FabF)(Garwin et al, 1980). Previous research has shown that mutations in either one of these genes results in temperature sensitive mutants in *E. coli* (Garwin et al, 1980).

FabF has been studied in the past for its role in piezophilic adaptation (Allen and Bartlett, 2000). In *P. profundum* strain SS9, it was shown that FabF, *cis*-vaccenic acid, is required for the piezoregulation and piezophilic growth (Allen and Bartlett, 2000). FabB has only been recently implicated in high pressure adaptation, when a *fabB* mutation was discovered among piezosensitive SS9 mutants (Lauro, 2007). The FabF alignment (Figure 4.4) showed a high degree of conservation both within phylogeny as well as across all twelve sequences (at least 50% identity). An exception is position 328, which shows a great deal of divergence among all species. Positions 281 and 331 have amino acid substitutions that were found in the three most piezophilic species. At 281, the deep sea strains showed a preference for acidic residues. At 331, all deep-sea strains differed from their shallow-water relatives, but these alterations were not conserved. Another interesting series of substitutions is from 379-383, where five amino acids are altered only in the piezophilic *Shewanella*. The region is otherwise conserved among the other gamma-proteobacteria, and may be an indication of a *Shewanella*-specific adaptation. Similar results were shown for FabB (Figure 4.5). Sequence length ranged from 400 to 405 amino acids, with sequence identity as low as 23%. As shown in position 84, amino acid substitutions tended to take place in one or a few of the tested genera. In this case, both CNPT3 and KT99 have acidic aspartates in that position as opposed to the non-polar

alanine and proline of their comparison strains. While substitutions in FabF and FabB cannot be deemed universal to piezophiles, they may be indications of different evolutionary strategies adopted by various genera.

Transcription

RpoE/RseB

The *rpoE* gene encodes σ^E, one of two σ factors that regulates the temperature shock response (Chi and Bartlett, 1995; Rouviere et al, 1995). The *rpoE* gene product is thought to respond to extracytoplasmic stresses, such as temperature and pressure (Chi and Bartlett, 1995; Rouviere et al, 1995). In *E. coli*, it is necessary for survival at high temperature as it controls the transcription of several genes involved in cell integrity (Rouviere et al, 1995). Additionally, Chi and Bartlett (1995) determined that a mutation in one gene of the *rpoE* locus, which regulates outer membrane protein OmpH, results in a phenotype that is sensitive to cold temperature and high pressure. It turns out, that this operon is comprised of *rpoE* followed by three genes of regulatory proteins *rseABC* (De Las Penas et al, 1997; Missikas et al, 1997) and it is the mutation in *rseB* that results in pressure-sensitive mutants (Chi and Bartlett, 1995; Lauro, 2007). RseB is found in the periplasm where it negatively regulates the activity of RpoE (De Las Penas et al, 1997; Missikas et al, 1997).

RpoE and RseB are not found in the gram positive strains (i.e. not in AT7 or *E. faecalis*). The RpoE alignment (61% identity or greater on sequence length ranging from 192-208 amino acids) showed no notable trends in amino acid substitution (Figure 4.6). The RseB alignment (310-347 amino acids with at least 25% identity) revealed three locations in which substitutions are notable (Figure 4.7). At position 43, most of the

piezophilic proteins favored the small, hydrophobic amino acids such as alanine and isoleucine, whereas the shallow water proteins were less consistent. At position 93, most of the piezophilic proteins contained charged residues such as aspartate and glutamate, while their comparison sequences had neutral amino acids including glutamine and asparagine. Finally, at position 115, the four most piezophilic proteins differed from the comparions, though there was little consistency in their substitutions.

ToxR

The ToxR protein was first described in *Vibrio cholerae* where it was shown to be a multimeric transmembrane protein that acts as a transcriptional regulator of several genes involved in virulence (Miller *et al*, 1987). In particular, ToxR regulates the expression of *ompU*, a gene that encodes for the production of an outer membrane protein (Miller *et al*, 1987). Homologues of the ToxR protein have been found throughout the Vibrionaceae family, where its range of functions varies with environment (Welch and Bartlett, 1998). It is of note that ToxR has a role in pressure sensing in the deep sea bacterium *Photobacterium profundum* SS9, where it regulates the expression of genes encoding outer membrane proteins (Welch and Bartlett, 1998).

The ToxR protein sequence was investigated due to its previously reported role in high-pressure adaptation. Among the piezophilic species available, only those of the genus *Photobacterium* and *Psychromonas* contained genes for ToxR. The most closely aligned sequences in other genera come from genes with DNA-binding response regulator motifs. Because ToxR was initially described in *V. cholerae*, it was also included in the alignment. The ToxR protein sequences ranged in length from 271-294 amino acids. There was little conservation between genera (12%-19%), with the

exception of SS9 and *V. cholerae,* which were 46% identical. The majority of the amino acid substitutions fell within phylogenetic lines or within amino acid classifications and were not deemed significant. There was one position, though, of potential interest. At *V. cholerae* position 207, the shallow water, non-pathogenic representatives both have alanines, which is a small and non-polar amino acid (Figure 4.8). In contrast, the deep water species SS9 and CNPT3 have threonine and serine in that position, respectively. These amino acids both contain hydroxyl groups and are polar, neutral residues. *V. cholerae*, while still a marine species, is the only human pathogen represented and has lysine, an amino acid with basic properties, in this position. This pattern of amino acid substitution may be an indication of adaptation to environmental conditions.

Stress induced

DnaK(hsp70)

Studies on *E. coli* have shown that high pressure is a unique stress in that it induces the production of both heat shock proteins (HSPs) and cold shock proteins (CSPs), in addition to other proteins which are found solely in response to pressure (Welch et al, 1993). Based upon current knowledge of the effects of pressure on cellular functioning, these results are not surprising. Pressure, like low temperature, has a negative effect on membrane fluidity, whereas high temperature and pressure share their destabilizing effects on macromolecules (Bartlett, 2002). One high temperature and high pressure stress-induced protein is DnaK (Welch et al, 1993), a member of the hsp70 family (heat shock protein of approximately 70 kD in size), which functions as a chaperone, aiding in the correct folding of denatured proteins (Rothman, 1989). DnaK levels have been known to increase with temperature, or under conditions in which

protein denaturing occurs (McCarty and Walker, 1991). Experiments with *E. coli* show that DnaK is induced by high hydrostatic pressure (Welch et al, 1993). It is thought that DnaK is able to directly sense the environment and modulate its activity (McCarty and Walker, 1991). It has further been postulated that pressure could directly influence its phophorylation state or ATPase activiy (Welch et al, 1993). Due to its potential for detecting changes in pressure, DnaK was investigated for amino acid alterations that allow for adaptation to stresses. The strains within *Photobacterium* and *Colwellia* had two copies of the *dnaK* gene that differed only slightly from one another (greater than 80% identity) (Figure 4.9). The 16 sequences examined were highly conserved between organisms (greater than 50% identity), including the threonine at position 199 which is thought to be the site of autophosphorylation and which is necessary to ATPase activity (McCarty and Walker, 1991). One substitution of note is at position 67 (Figure 4.9). There, the deep-sea bacteria appear to have a preference for tyrosine, though this trend is not universal.

H-NS (Csp)

Similar to the heat shock protein DnaK, a cold shock protein was investigated for amino acid sequences suggesting adaptation to high pressure. H-NS, histone-like protein H1, is a prevalent protein in the bacterial nucleoid, involved in DNA supercoiling and regulation of gene expression (Hulton et al, 1990). This protein has been implicated in regulatory responses to environmental conditions, particularly to cold temperature (Dersch et al, 1994; Hulton et al, 1990). More recently, H-NS mutants have been shown to have pressure sensitivity and the protein has been implicated as a possible transcriptional regulator for adaptation to high pressure (Ishii et al, 2004). Interestingly,

hns mutation in SS9 resulted in a cold-sensitive, yet pressure-enhanced phenotype, suggesting that H-NS may have evolved specific traits in this organism (Lauro, 2007). Genetic tests on the strains in this study would be required to determine their growth patterns with H-NS mutations and under high pressure. However, the conflicting results in these previous studies imply H-NS as a potential source of piezophilic adaptation. In this study, only members of the *Photobacterium, Moritella,* and *Shewanella* genera had genes identified as H-NS. These genes ranged from 130-136 amino acids in length (at least 35% identity), though the piezophilic sequences were typically shorter by a few amino acids. Only genera-specific substitutions were found (Figure 4.10).

DNA replication/recombination/repair

RecD

RecD, a component of Exonuclease V, is involved in DNA recombination and repair (Finch et al, 1986). Along with RecB and RecC, it participates in several aspects of homologous recombination such as exonuclease activity, single-stranded endonuclease activity as well as helicase activity (Bidle and Bartlett, 1999). RecD influences DNA replication and DNA synthesis, which is among the most pressure-sensitive of all of the macromolecular synthetic processes (Yayanos and Pollard, 1969). Previous studies on RecD, have shown that it is necessary for normal growth and morphology, as well as plasmid maintenance at high pressure in both SS9 and *E. coli* (Bidle and Bartlett, 1999). While amino acid substitution has not been studied in piezophile RecD, Bidle and Bartlett (1999) did investigate the sequence characteristics of SS9 RecD. They noted the addition of a 64 amino acids that were not found in other reported sequences, with the exception of *V. cholerae*. They added that those species with the increased protein length did not

demonstrate filamentous growth at high pressure, and therefore, may be a source of high pressure adaptation. When comparing *E. coli* RecD to the species in this study, it was found that only *E. coli* lacked the inserted sequence (Figure 4.11). Although the effects of high pressure on the growth and morphology of the shallow water species is unknown, it is likely that the increased length is due to some other environmental factor. In fact, most of the piezophilic strains had RecD sequences that were shorter than their shallow-water relatives. Figure 4.11 shows a series of pairwise alignments with sequence gaps highlighted. The sequence length differential ranged from 87 amino acids (MT41/*C. psychrerythraea*) to being equal in length (SS9/3TCK and PE36/*M. viscosa*)(Figure 4.11). These gaps in piezophile sequence do not appear to show a pattern in either location or in types of missing amino acids. To further explore these gaps and their significance, secondary structure was predicted and compared (Figure 4.12). SS9 structure was aligned with *E. coli* to investigate the large gaps found in *E. coli*. The SS9 structure shows the insertion of a large helix between positions 159-181 and of several coil regions in 187-236, not found in *E. coli.* Various smaller gaps (up to six amino acids) occur throughout. The significance of these additional features on tertiary structure are unknown. Likewise, MT41 predicted secondary structure was compared to that of *C. psychrerythraea* (Figure 4.13). These strains had the largest difference in size, with MT41 having 87 fewer amino acids than its shallow-water relative, and have the largest difference in depth of isolation. This alignment (Figure 4.13) shows three large gaps (at least 12 amino acids in length) which are occupied primarily by helical regions in *C. psychrerythraea*. Since helical structures are believed to be the least adapted to high pressure (Gross and Jaenicke, 1994; Mozhaev et al, 1996), these missing amino

acids may be an adaptation to the deep sea. The other piezophiles, though, did not display such a high degree of variation from their shallow-water relatives.

Similarly, amino acid substitutions showed very little consistency across all phylogenetic lines (Figure 4.14). The RecD protein sequence ranged from 673-865 amino acids and from 35-46% identity between genera. However, there were several instances where changes were consistent among all members of the most piezophilic species. Perhaps the most striking example is at position 278, in which the proteins derived from piezophiles in the five gamma-proteobacteria species favor residues of the hydroxyl/amine group. In two locations (57 and 532), the amino acids of the three most piezophilic species lack the conservation demonstrated in their comparison strains. In both cases, glutamine or other hydroxyl/amine residues are exchanged for other amino acids in the piezophiles. The implications of these exchanges are unknown. Five other positions (162, 729, 763, 804 and 805) demonstrate substitutions specific to the *Psychromonas* and *Shewanella* species. Other amino acid substitutions were not conserved across phylogenetic lines.

DiaA

DiaA, a DnaA initiator-associated protein, is required for the initiation of DNA replication at the correct time in the cell cycle (Ishida et al, 2004). It directly binds to DnaA, the DNA replication initiation factor, and stimulates its function (Ishida et al, 2004). Mutations in this gene in SS9 have resulted in both cold temperature-sensitive and pressure-sensitive phenotypes (Lauro, 2007), though studies with *E. coli* have shown that it can suppress temperature-sensitive phenotypes caused by mutations in DnaA (Ishida et al, 2004). The protein contains an SIS (sugar isomerase) domain (Ishida et al,

2004), whose catalytic site is formed by the dimerization of identical subdomains (Teplyakov et al, 1998). The alignments showed that the sequences display a broad range of conservation across genera (9-70% identity) and are highly conserved within genera (53-99% identity). The gram positive strains showed the highest degree of divergence, both in amino acid composition and in length (Figure 4.15). Also, strains CNPT3 and KT99 have two copies of the *diaA* gene whose products differ greatly from one another. The second copies were very similar to each other (88% identity). No amino acid substitutions were identified as potential piezophilic adaptations.

SeqA

Like DiaA, SeqA is involved in the regulation of chromosome replication initiation (Lu et al, 1994). However, SeqA is a negative regulator, in that it prevents the initiation of replication through the process of sequestration (Lu et al, 1994). This process helps ensure that DNA replication does not occur too frequently and that it remains coordinated with the cell cycle (Lu et al, 1994). Additionally, SeqA may have a role in membrane properties as a negative regulator of membrane permeability (Wegrzyn et al, 1999). Interestingly, the mutation of the SS9 *seqA* gene results in a pressure-enhanced phenotype (Lauro, 2007). AT7 and *E. faecalis* did not have significant matches to the SeqA sequence and were not included in the alignment. The rest of the sequences ranged from 178-213 amino acids and have greater than 26% identity with one another. The alignments showed five sites in which piezophiles differed from their comparison strains (Figure 4.16). At position 33, most of the non-piezophiles have polar neutral residues (all from the hydroxyl/amine classification), whereas the piezophiles are far more divergent. At position 35, PE39, KT99 and MT41 have amino acid substitutions,

with the orthologs form the most piezophilic bacteria having charged residues at this site. This is in contrast to the neutral residues of the shallow water strains. Finally, positions 54, 112, and 220 show substitutions among PE36, CNPT3 and KT99, though these changes are not consistent.

Respiration

CydD

CydD, first identified in E. coli, is necessary for the formation of a functional cytochrome d oxidase, a component of the aerobic respiratory chain (Bebbington and Williams, 1993). The work of Kato et al (1996) demonstrated that the respiratory system is impacted by high hydrostatic pressure. Specifically, they noted that the *cydD* gene product is required for growth at high pressure. In piezophilic shewanellas, it appears downstream of a pressure-responsive protein required for gene expression at high pressure (Bartlett, 2002). Strains AT7, MT41, PE36, and CNPT3 do not have *cydD* genes. This could be due to access to only unfinished genomes, or because this gene is not present in these organisms. *E. faecalis* and *C. psychrerythraea* possessed sequences for CydD, but they aligned very poorly to the other five sequences. *E. faecalis* differed so greatly that it was not included in the alignment. The alignment, then, consists of two piezophilic and four piezosensitive strains ranging from 567-597 amino acids and having at least 30% identity. Although Kato et al (1996) investigated the percent alignment of piezophilic CydD to that of E. coli, they did not look at individual amino acid substitutions. Of the many amino acid substitutions, a few were specific to KT99 (Figure 4.17). Most likely, this is due to either genus- or species-specific substitutions. However, because KT99 is an obligate piezophile it is more piezophilic than the

facultative piezophile SS9 and these substitutions could be an indication of high pressure adaptation. More sequences from obligate piezophiles are necessary for further comparison. Other substitutions appeared to be due to phylogenetic differences.

Enzyme activity

MDH

Malate dehydrogenase (MDH) catalyzes the interconversion of malate and oxaloacetate in several metabolic pathways, including the tricarboxylic acid cycle (Welch and Bartlett, 1997). Since MDH has been well studied in terms of kinetics, structure, and sequence, it has been useful in the investigation of enzyme adaptation to extreme environments, such as high hydrostatic pressure (Saito et al. 2006; Saito and Nakayama, 2004; Welch and Bartlett, 1997). Enzyme activity under high pressure is a topic of great interest, yet remains very poorly understood. The structure and function of dehydrogenases (such as lactate and malate dehydrogenases) are particularly well studied due to the large conformational changes and volume increases that they undergo during ligand binding and processing (Gross and Jaenicke, 1994; Somero, 1990). These enzymes possess biochemical adaptation to high pressure in comparison to their shallow water homologs (Saito et al, 2006; Saito and Nakayama, 2004; Welch and Bartlett, 1997). Both malate dehydrogenase and lactate dehydrogenase have been the focus of several studies involving the amino acid sequence of high pressure-adapted proteins (Saito et al, 2006; Saito and Nakayama, 2004; Welch and Bartlett, 1997). Unlike the previous studies, which used a different assortment of organisms, this investigation revealed no patterns of amino acid substitutions between piezophilic and piezosensitive strains. A comparison of their findings to the findings of this analysis will be discussed further. The

MDH proteins compared here were all of similar size (most were 312 amino acids, but actual length ranged from 311 to 319 amino acids) and showed a high degree of sequence conservation (greater than 60% identity) (Figure 4.18). Where amino acid variation occurred, it tended to stay within phylogenetic lines or within amino acid physio-chemical classifications. Occasionally, residue substitutions appeared, but were not shared within the environmental groupings.

Welch and Bartlett (1997), investigated the psychrophilic and piezophilic adaptation of MDH by comparing the MDH amino acid alignment from the moderate piezophile *P. profundum* SS9 to those of three mesophilic bacteria and one deep-sea *Vibrio*. The authors noted a few positions with amino acid substitutions that were thought to be involved in environmental adaptation. They described two SS9-specific substitutions, including leucine to isoleucine at position 6 and alanine to arginine at position 219, as well as a valine to isoleucine substitution shared by the deep-sea bacteria in position 146, and an alanine to serine substitution in the *Vibrio* at position 159. Alignments from this study show that the substitution at position 6 is specific to *P. profundum*, but is shared by both strains SS9 and 3TCK, indicating a phylogenetic relationship. Additionally, both leucine and isoleucine fall into the same physio-chemical-category. Interestingly, SS9's arginine at position 219 does not appear in the sequence used in this study. The sequence used in this analysis shows that it is still an alanine and, therefore, does not differ from the other sequences used in either study. In position 146, this analysis reveals an isoleucine in all examined proteins except those of the *Colwellia* genus. This implies that the substitution is not a wide-spread adaptation to the deep sea, nor is it universal to psychrophiles. The substitution thought to be *Vibrio*-

specific is also seen in the two *Moritella* species. This substitution at position 159, alanine to serine, is a change from a non-polar to a polar residue. The impact of this substitution is unknown, though because it is found in both deep-sea and shallow water species, is most likely not due to high pressure.

More recent studies from Saito and Nakayama (2004) and Saito et al (2006), have further explored the amino acid substitutions of MDH from deep-sea bacteria. Saito and Nakayama (2004), investigated the sequences of obligate piezophile *Moritella* sp. strain 2D2 and the psychrophilic, piezotolerant *Moritella* sp. strain 5710 by aligning them with six bacteria from various environments and genera. The authors found two substitutions specific to the piezophilic strain: a valine to alanine substitution at position 180 and a glutamine to histidine substitution at position 229. Neither of these changes were found in this study. Instead, I found only a *Psychromonas*-specific substitution of glutamine to arginine at position 229. All other amino acid differences were found in both members of the phylogenetic comparisons. Saito et al (2006), expanded upon their earlier findings by further investigating the Ala-180 and His-229 substitutions in both *Moritella* and *Shewanella* piezophilic bacteria. While the authors rejected the importance of the Ala-180 substitution, they reemphasized the claim that a His-229 substitution is useful in obligately and facultatively piezophilic *Moritella* species. While many of their strains showed this substitution, it could not be confirmed by this study as the facultative piezophile *Moritella* sp. PE36 contained a glutamine at this position like many of the shallow-water species. Saito et al also found substitutions in piezophilic *Moritella* at positions 103, 111, and 283. Only the Glu-Asp substitution at position 283 was seen in this study. Because these are the same type of amino acid, it is not predicted to have

much effect, though further investigation is required. In fact, the opposite substitution was found in the *Shewanella*s. In the *Shewanella* strains, the authors identified five amino acid substitutions that were conserved among only the piezophilic strains (positions 61, 65, 107, 161, and 202). The alignment data for *Shewanella* in this study agree with the changes reported by Saito et al for positions 65, 107, and 161, although they are not found in any other genus. Only the substitution at position 65 in the comparison shows an alteration in the kind of amino acid residue where an aspartate, an acidic residue, is exchanged for an alanine, a non-polar residue. This is not true for the results of Saito et al, which had a proline to alanine exchange, meaning that both residues were of the same type. Again, this type of exchange was only found among the *Shewanella* at this position. In positions 61 and 202, the reported amino acid residues agreed with those found in *Shewanella benthica* KT99, but did not differ from those of the shallow-water *Shewanella frigidimarina*, indicating that they are not piezophilic adaptations.

MDH was also investigated to determine if these amino acid substitutions had an effect on secondary structure, using strongly and weakly piezophilic examples. The secondary structure were predicted and compared for the *Colwellia* and the *Photobacterium* pairs (not shown). These comparisons show longer coil regions for the deep-sea bacteria, however, these differences occurred in regions of low confidence in the prediction.

The analysis of MDH, both in this investigation and through comparisons with previous studies, did not reveal universal trends in piezophilic adaptations. In fact, the results of this study discounted many of the claims made in previous reports. More

investigation, including a larger sample size, is needed to further elucidate the potential of these trends.

AsnA

Transposon mutagenesis of SS9 resulted in several pressure-sensitive mutants, including those with mutations in the *ansA* gene (Lauro, 2007). *ansA* encodes for L-asparaginase I, which catalyzes the hydrolysis of L-asparagine to L-aspartic acid and ammonia (Del Casale et al, 1983). This enzyme enables bacteria to utilize L-asparagine when it is the only available nitrogen source (Del Casale et al, 1983; Willis and Woolfolk 1974). Although the reason for the pressure-sensitivity of this mutant is unknown, it is believed that this pathway is more important at high pressure than at atmospheric pressure (Lauro, 2007). The sequence for L-asparaginase is approximately 300 amino acids long for all twelve organisms. Those of the genus *Colwellia* were the most divergent from the rest of the group, both in terms of gaps and amino acid usage (maximum of 8% identity with other genera). AT7 and *E. faecalis* also differed (about 20% identity), but as they are the only ones not members of the gamma-proteobacteria, this is unsurprising. The rest of the sequences were more than 50% identical. While most changes were genus-specific, at position 378, SS9, PE36, CNPT3, and KT99 all had serines where their comparison strain had threonines (Figure 4.19). These exchanges are at the very end and do not represent drastic changes in biochemistry. It is, therefore, doubtful that they confer significant adaptation.

Amino Acid Composition

Several studies have implicated amino acid composition as necessary for a protein's adaptation to extreme environments, revealing residues that are thought to be

involved in structural stability of proteins under extreme conditions (Chilukuri and Bartlett 1997, Fukuchi et al 2003). Fukuchi et al (2003), studied thermophiles and halophiles, finding that both showed high ratios of acidic residues. An investigation of SSB proteins in marine *Shewanella* revealed low temperature and high pressure adaptations such as increased overall hydrophilicity, decreased asparagines/glutamine and arginine/lysine ratios, and a decrease in helix-destabilizing residues (Chilukuri and Bartlett, 1997). No previous studies have investigated piezophiles across genera or protein types.

For each of the fifteen proteins, amino acid composition was calculated by combining the protein sequences of piezophiles (*piezo*) and the piezosensitive strains (*non*) in to their respective groups. This resulted in the amino acid usage of piezophiles and piezosensitive strains in terms of percentage that each amino acid represents. Comparisons were also made in terms of composition of residue types. Amino acids were classified as being tiny, small, aliphatic, aromatic, non-polar, polar, charged, basic and acidic. Compositions were then compared between piezophiles and non-piezophiles. Statistical tests (independent, two-tailed T-test and chi-squared) were performed comparing piezophiles to non-piezophiles across amino acid or amino acid classification. No statistically significant variations were found between the two groups (Table 4.4, 4.5). Similarly, no trends were discovered in protein length, molecular weight or isoelectric point (Table 4.6).

Protein volume, a function of secondary structure, is the key factor in stability and function at high pressure (Gross and Jaenicke, 1994; Somero, 1990). Since alpha helices are more compressible, they are thought to be less adaptive to high pressure environments

than beta sheets (Gross and Jaenicke, 1994; Somero, 1990). Statistical tests revealed no significant differences between secondary structure composition of piezophiles and non-piezophiles (Table 4.7). Volume adaptations may occur during the folding in to the tertiary structure.

Conclusion

The use of genomic technology has embraced by a few groups studying deep-sea biology as a tool to better understand the functioning of organsisms adapted to life at high pressure. Since global genome comparisons have indicated that it is unlikely that piezophily due to the presence or absence of key genes (Lauro, 2007), research has focused in on sequences of proteins. As protein function is a factor of its structure and structure is determined by its amino acid sequence (Dunker et al, 2002), it seems likely that a protein's ability to hold structure and perform function in the face of environmental stresses is rooted in its amino acid sequence. In this study, we investigated protein sequences from six piezophiles and their piezosensitive relatives to determine amino acid substitutions in 15 proteins from pressure-sensitive processes, in the hopes that it would reveal universal piezophilic signatures. The inability to uncover such signatures implies that piezophily is not so simply explained.

While previous studies of this kind focused on small numbers of isolates from similar phylogenetic or geographical backgrounds (Bidle and Bartlett, 1999; Kato et al, 1996; Saito et al, 2006; Saito and Nakayama, 2004; Welch and Bartlett, 1997; Welch and Bartlett, 1998), the genome sequences used in this study represent a wide range of phylogenetic piezophilic representatives from several locations in the Pacific Ocean (Figure 4.1). This diversity among sequences allowed investigation in to the universal

adaptations of proteins to high pressure. Although these 15 proteins revealed no universal piezophilic signatures, the results serve as an excellent prelude to future research on protein behavior at high hydrostatic pressure.

The proteins investigated were chosen due to previous research that indicated their involvement in piezophilic functioning. Yet, they represent only a fraction of those that could be involved in piezophily. It is possible that proteins not studied here have primary structure alterations that confer adaptation across genera. Furthermore, amino acids substitutions were analyzed in terms of their physical and biochemical classifications, where substitutions within these groupings were not included. Individual amino acids within these classifications do show differences in size and shape, and therefore, can behave differently within protein structure. A more stringent analysis of the effects of individual amino acids on the protein structure and resultant volume changes would be necessary to affirm substitution trends. It may also be that too few sequences were considered to deduce a pattern. This will be resolved through the continued sequencing and isolation of deep-sea microbes.

The results of this study indicate that it is unlikely for universal adaptations to be held in single amino acid substitutions. The adaptive qualities of the proteins may not be in the single amino acid replacements, but in the overall features of the protein that would influence secondary and tertiary structures. Some of these protein features, such as size and isoelectric point, were examined without trends, but other factors and investigation into structure and volume changes would be necessary. Additionally, there is some evidence to indicate that tertiary structure is affected by properties other than primary structure alone. Osmolytes, both produced intracellularly or taken up from the

environment, may interact with proteins and influence folding (Bartlett, 2002). Also, proteins such as H-NS have multiple structural isomers that are believed to interconvert with environmental signals (Hulton et al, 1990). It is also likely that protein adaptations are better investigated within genera. Piezophily is found in several branches of the phylogenetic tree (Lauro et al, 2007) and it is possible that is has evolved multiple times. Each time, microbes may have adopted different strategies surviving life in the deep sea. In this way, adaptations would vary with phylogeny and could, therefore, be best investigated within, as opposed to across, separately evolved groups. As technology continues to improve and environmental sequence data increases, such investigations in microbial deep-sea adaptations will become more readily available.

Table 4.1: List of piezophilic strains used in this study, including depth, location, and temperature of isolation, and comparison strains and habitats.

Strain	Depth (m)	Location	Temperature (°C)	Reference	Comparison strain	Description	Reference
Carnobacterium AT7	2500	Aleutian Trench	1.8	Lauro et al 2007	Enterococcus faecalis	Human intestinal bacterium	Paulsen et al, 2003
Photobacterium profundum SS9	2551	Sulu Sea	9.8	Delong et al 1997	Photobacterium profundum 3TCK	San Diego Bay	Campanaro et al, 2005
Moritella PE36	3584	Pacific Ocean (California coast)	4	Delong *et al* 1997; Yayanos 1986	Moritella viscosa	Atlantic salmon pathogen	Benediktsdottir *et al.* 2000
Psychromonas CNPT3	5700	Central North Pacific	2–4	Delong et al 1997; Yayanos et al 1979	Psychromonas ingrahamii 37	Arctic ice	Auman et al, 2006
Shewanella benthica KT99	9856	Tonga-Kermadec Trench	1.8	Lauro et al 2007	Shewanella frigidimarina	Antarctic sea ice	Bowman et al, 1997
Colwellia MT41	10,476	Mariana Trench	2	Yayanos et al 1981	Colwellia psyrerythraea	Arctic sea ice	Methe et al, 2005

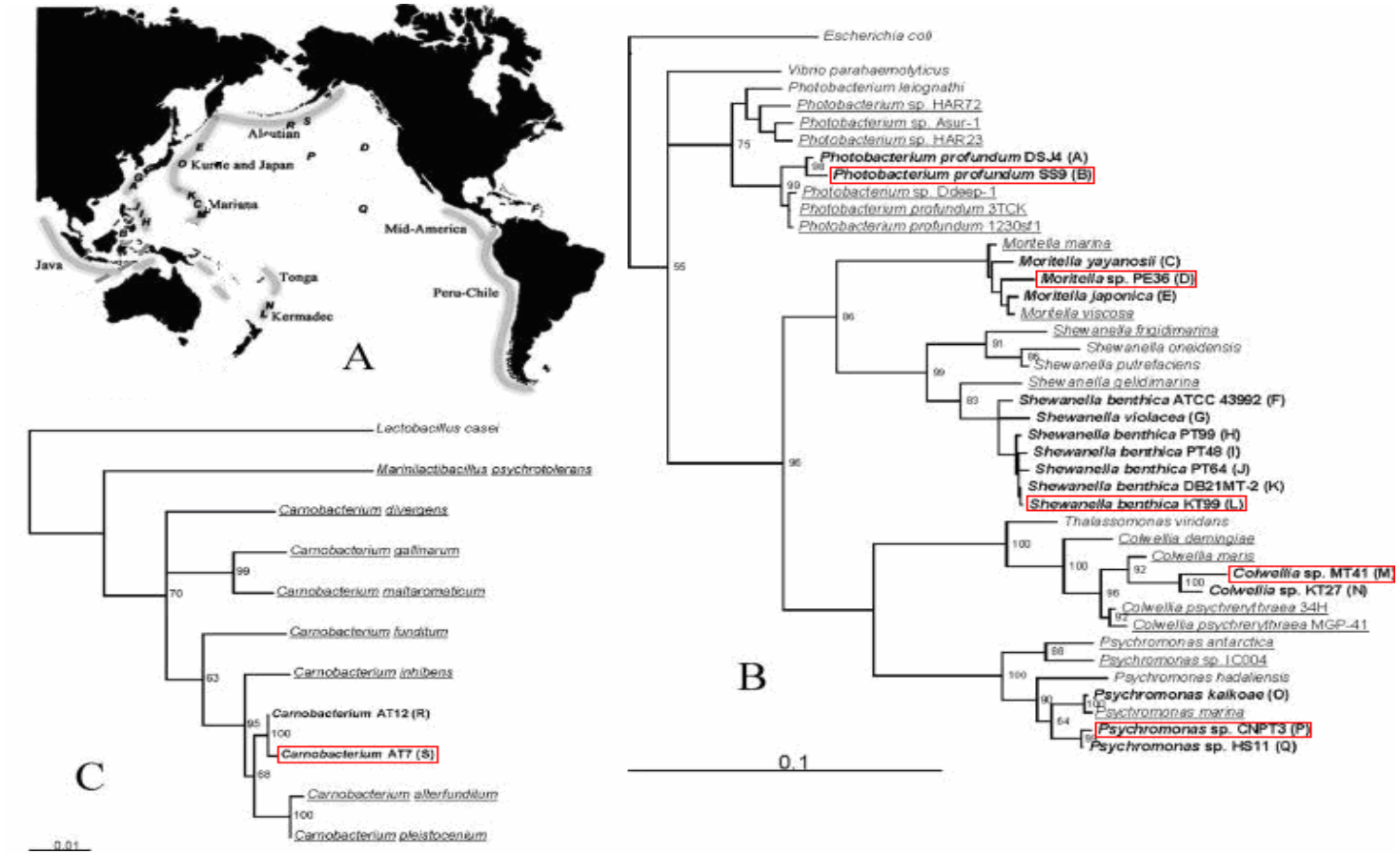

Figure 4.1: Map and phylogenetic tree of cultured deep-sea isolates and relatives. Piezophilic and piezotolerant strains highlighted in bold with letters corresponding to location of isolation on the map (psychrophilic and psychrotolerant strains underlined). Piezophilic strains of interest in this study are indicated by red squares. (A) Approximate locations of site of collection for each piezophilic strain. (B) Phylogenetic tree of the relationships of the gamma-proteobacteria. (C) Phlyogenetic tree of the relationships of the *Carnobacteraceae* strains. Adapted from Lauro et al (2007).

Table 4.2: List of proteins studied, including their function in the cell, COG category (where appropriate), and reference.

Protein	Function	COG	Reference
ToxR	piezo-sensor and transcriptional regulator		Welch and Bartlett 1998
Mdh	Malate dehydrogenase; citric acid cycle enzyme	energy production and conversion	Saito and Nakayama 2004, Saito et al. 2006, Welch and Bartlett 1997
RecD	Component of exonuclease V	DNA replication, recombination and repair	Bidle and Bartlett, 1999; Finch et al, 1986
SeqA	replication initiation regulator	DNA replication, recombination and repair	Lauro 2007
DnaK	heat shock protein	signal transduction mechanisms	Welch et al 1993
DiaA	Phosphoheptose isomerase		Lauro 2007
H-NS	Histone-like nucleoid structuring protein	general function prediction only	Dersch et al 1994, Ishii et al 2002
FabF	3-oxoacyl-[acyl-carrier-protein] synthase II; Fatty acid biosynthesis	Secondary metabolites biosynthesis/Lipid metabolism	Allen and Bartlett 2000
FabB	3-oxoacyl-[acyl-carrier-protein] synthase I; Fatty acid biosynthesis	Secondary metabolites biosynthesis/Lipid metabolism	Lauro 2007
MreB	Involved in cell morphogenesis in rod shaped bacteria	Cell division and chromosome partitioning	
FtsZ	Essential for cell division in prokaryotes	Cell division and chromosome partitioning	Ishii et al 2004
CydD	Formation of cytochrome *d* oxidase	Secondary metabolites biosynthesis, transport and catabolism	Bebbington and Williams, 1993; Kato et al., 1996
RseB	sigma-E factor regulatory protein	Signal transduction mechanisms	Lauro 2007
RpoE	RNA polymerase sigma factor	transcription	Lauro 2007
Ans	L-asparaginase		Lauro 2007

Table 4.3: List of the colors used to identify amino acid properties in the sequence alignments.

Color	Property	Amino acids
RED	Small (small, hydrophobic and aromatic)	A V F P M I L W
BLUE	Acidic	D E
MAGENTA	Basic	R K
GREEN	Hydroxyl and amine	S T Y H C N G Q

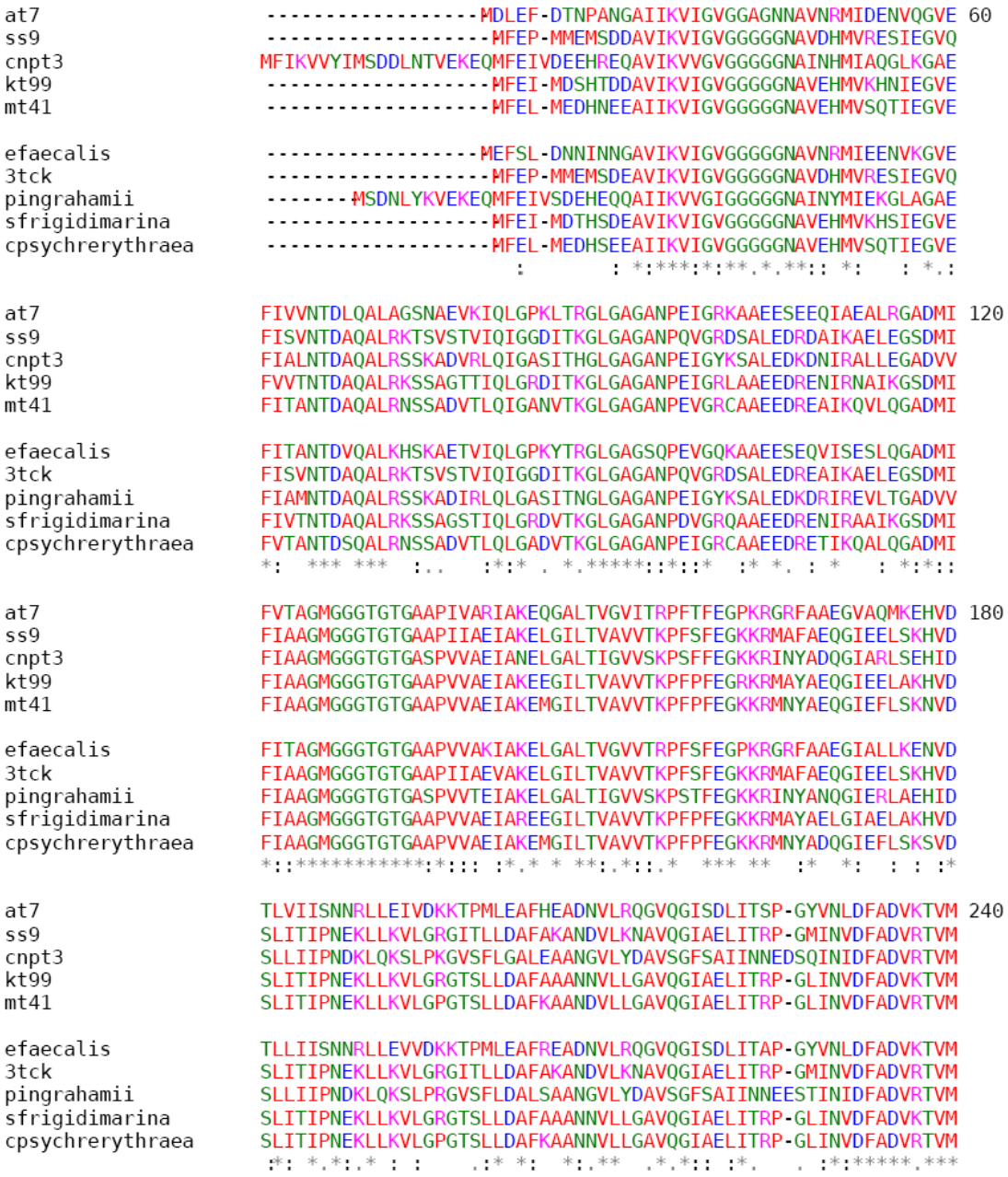

Figure 4.2: FtsZ protein alignment with colors depicting amino acid properties as described in Table 3. Proteins from deep-sea strains are grouped on top, listed with increasing piezophilicity. Comparison strains are grouped on the bottom. Symbols depict similarity between amino acids, "*" means identical amino acids, ":" are conserved substitutions, "." are semi-conserved substitutions.

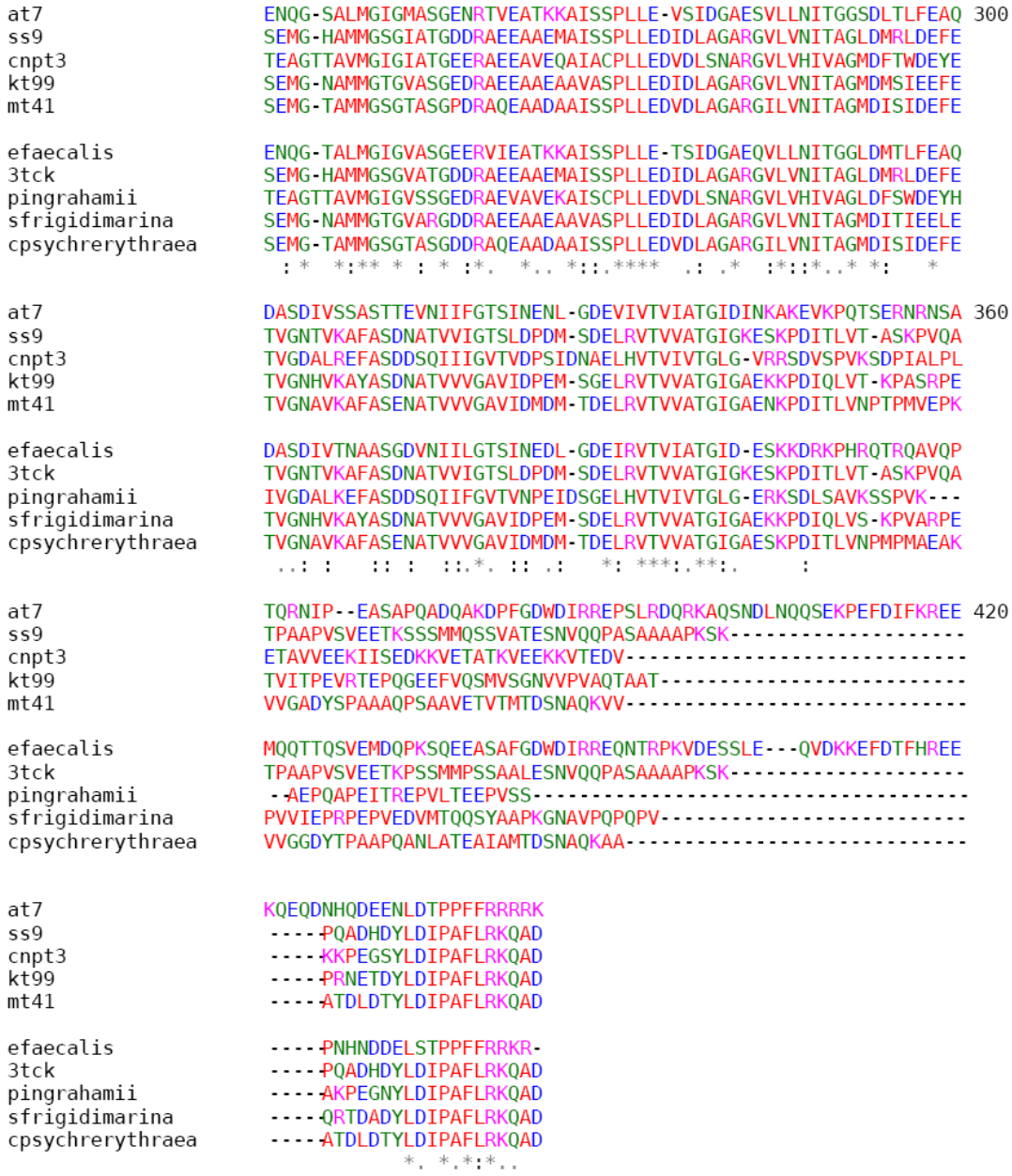

Figure 4.2, Continued: FtsZ protein alignment with colors depicting amino acid properties as described in Table 3. Proteins from deep-sea strains are grouped on top, listed with increasing piezophilicity. Comparison strains are grouped on the bottom. Symbols depict similarity between amino acids, "*" means identical amino acids, ":" are conserved substitutions, "." are semi-conserved substitutions.

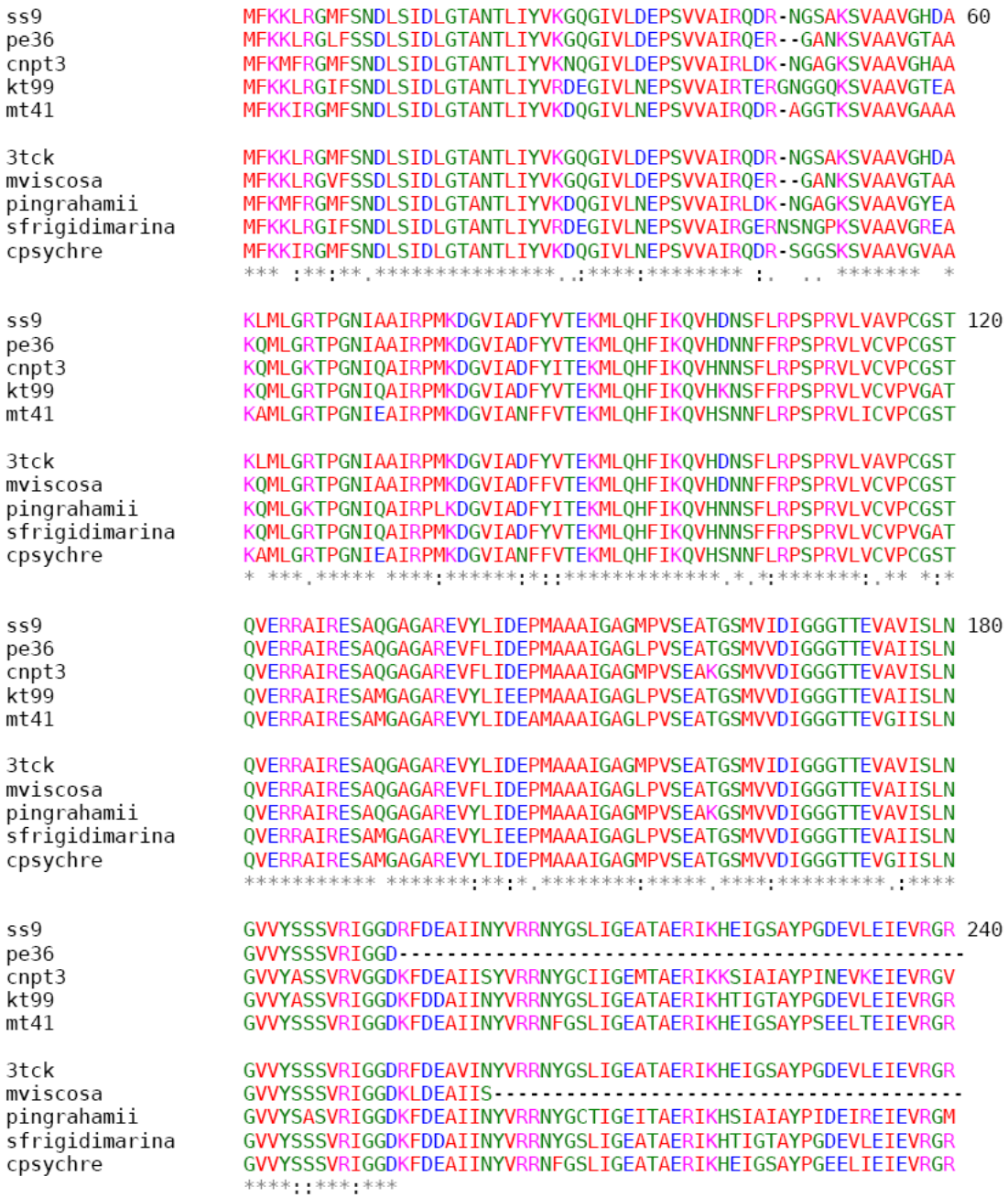

Figure 4.3: MreB protein alignment with colors depicting amino acid properties as described in Table 3. Proteins from deep-sea strains are grouped on top, listed with increasing piezophilicity. Comparison strains are grouped on the bottom. Symbols depict similarity between amino acids, “*” means identical amino acids, “:” are conserved substitutions, “.” are semi-conserved substitutions.

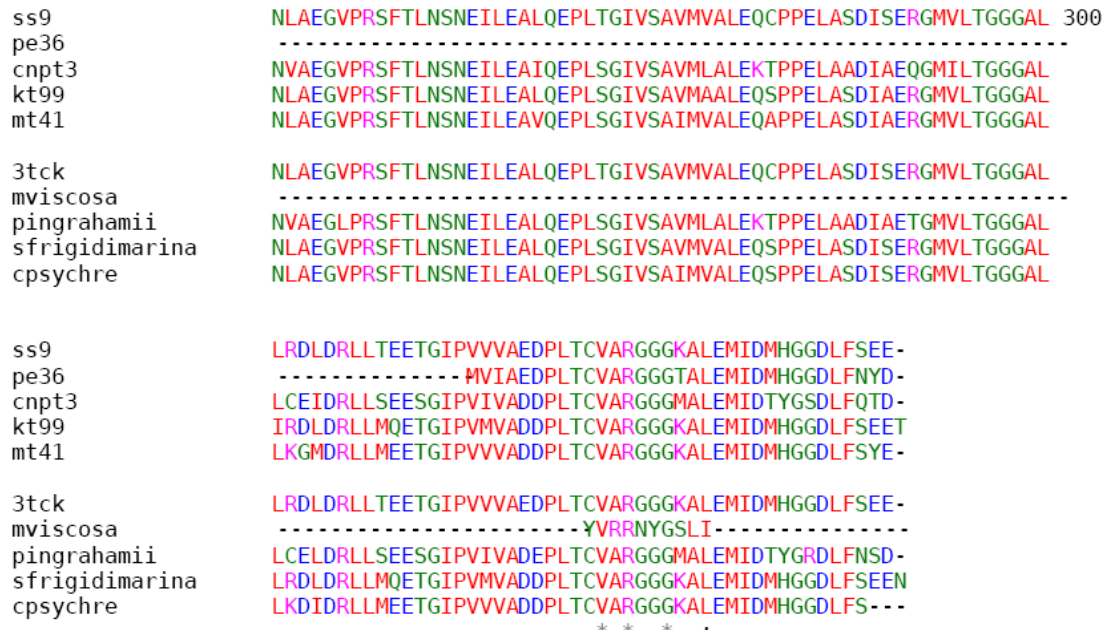

Figure 4.3, Continued: FtsZ protein alignment with colors depicting amino acid properties as described in Table 3. Proteins from deep-sea strains are grouped on top, listed with increasing piezophilicity. Comparison strains are grouped on the bottom. Symbols depict similarity between amino acids, "*" means identical amino acids, ":" are conserved substitutions, "." are semi-conserved substitutions.

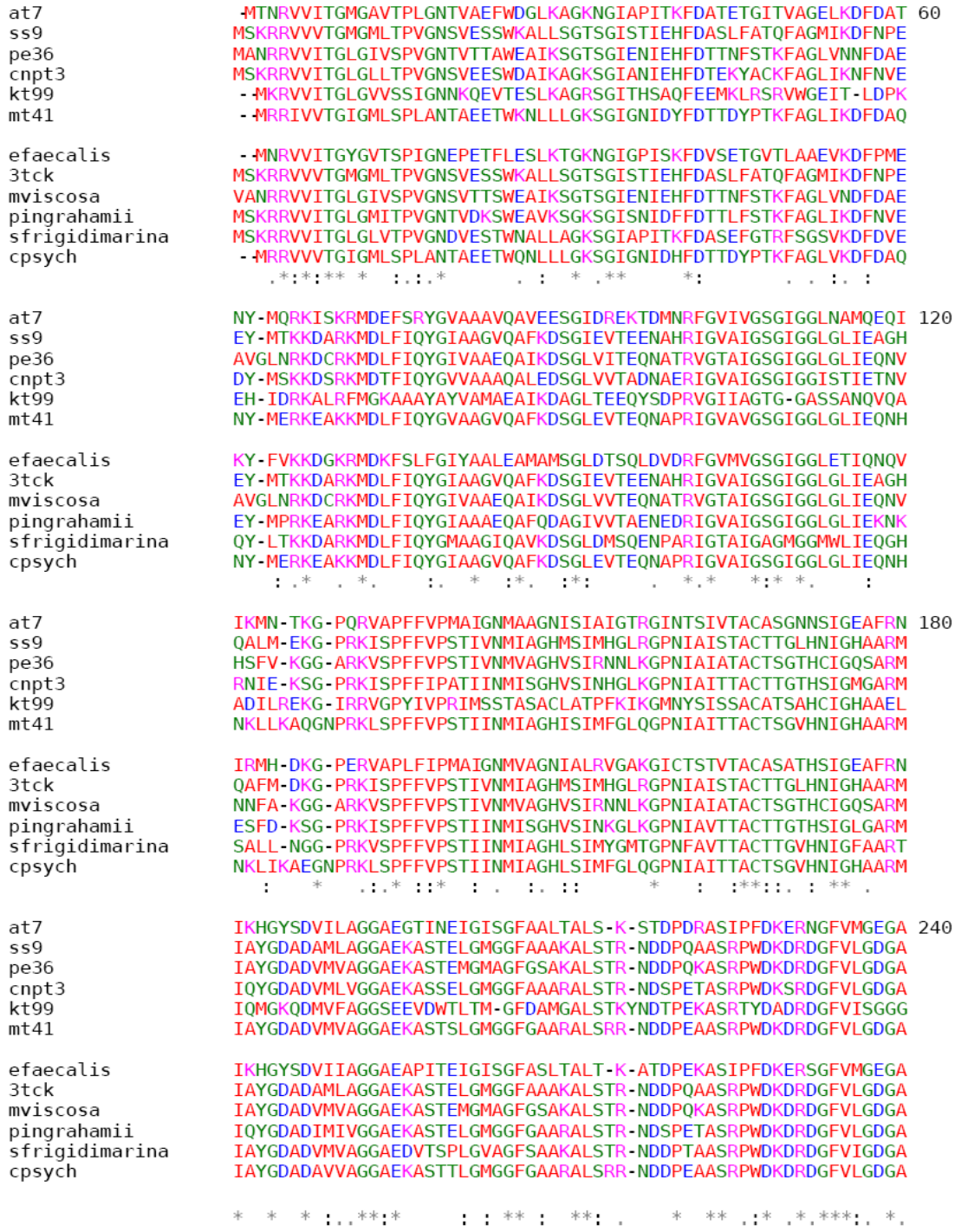

Figure 4.4: FabF protein alignment with colors depicting amino acid properties as described in Table 3. Proteins from deep-sea strains are grouped on top, listed with increasing piezophilicity. Comparison strains are grouped on the bottom. Symbols depict similarity between amino acids, "*" means identical amino acids, ":" are conserved substitutions, "." are semi-conserved substitutions. Regions with substitutions of interest are highlighted in yellow.

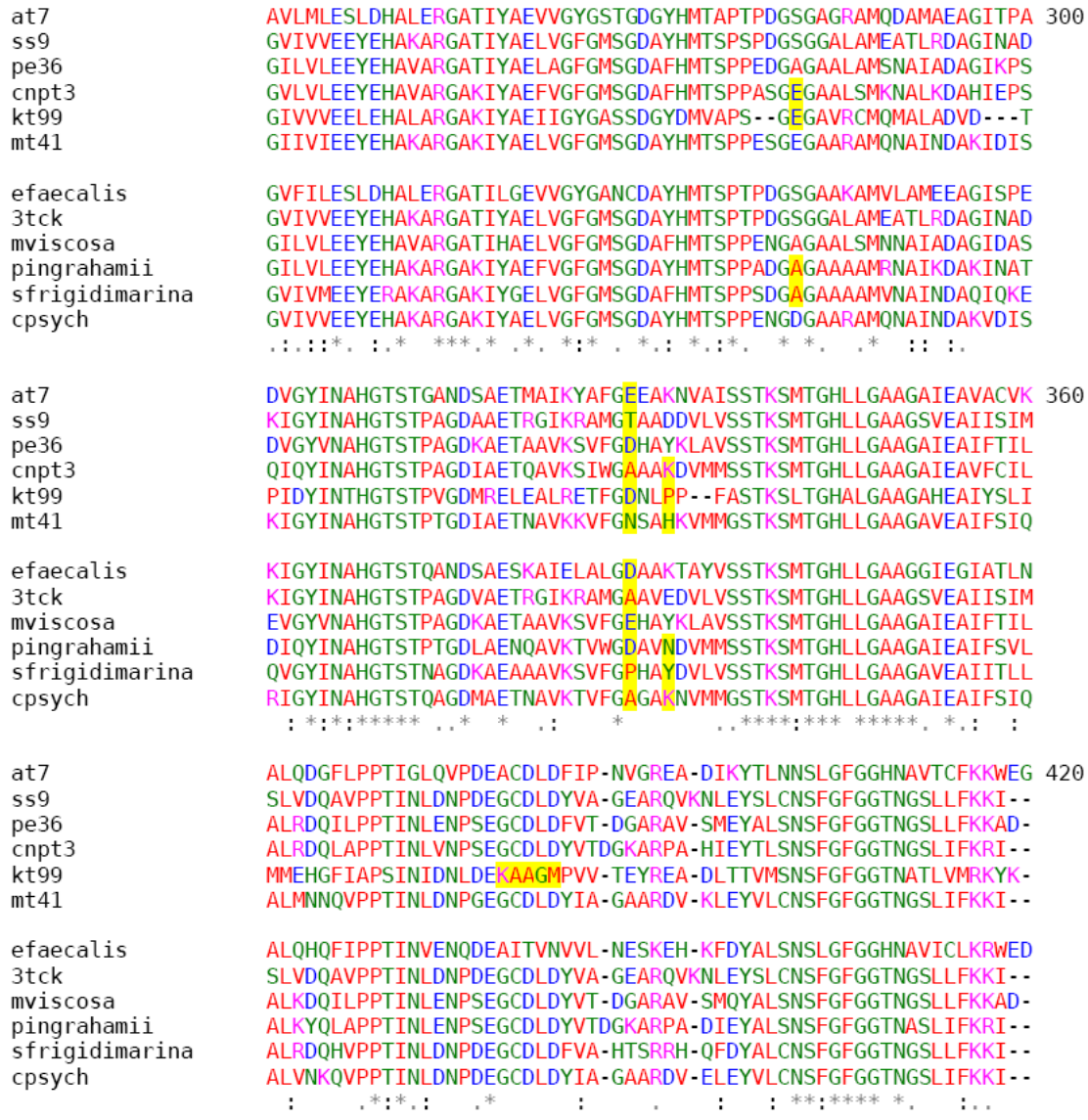

Figure 4.4, Continued: FabF protein alignment with colors depicting amino acid properties as described in Table 3. Proteins from deep-sea strains are grouped on top, listed with increasing piezophilicity. Comparison strains are grouped on the bottom. Symbols depict similarity between amino acids, "*" means identical amino acids, ":" are conserved substitutions, "." are semi-conserved substitutions. Regions with substitutions of interest are highlighted in yellow.

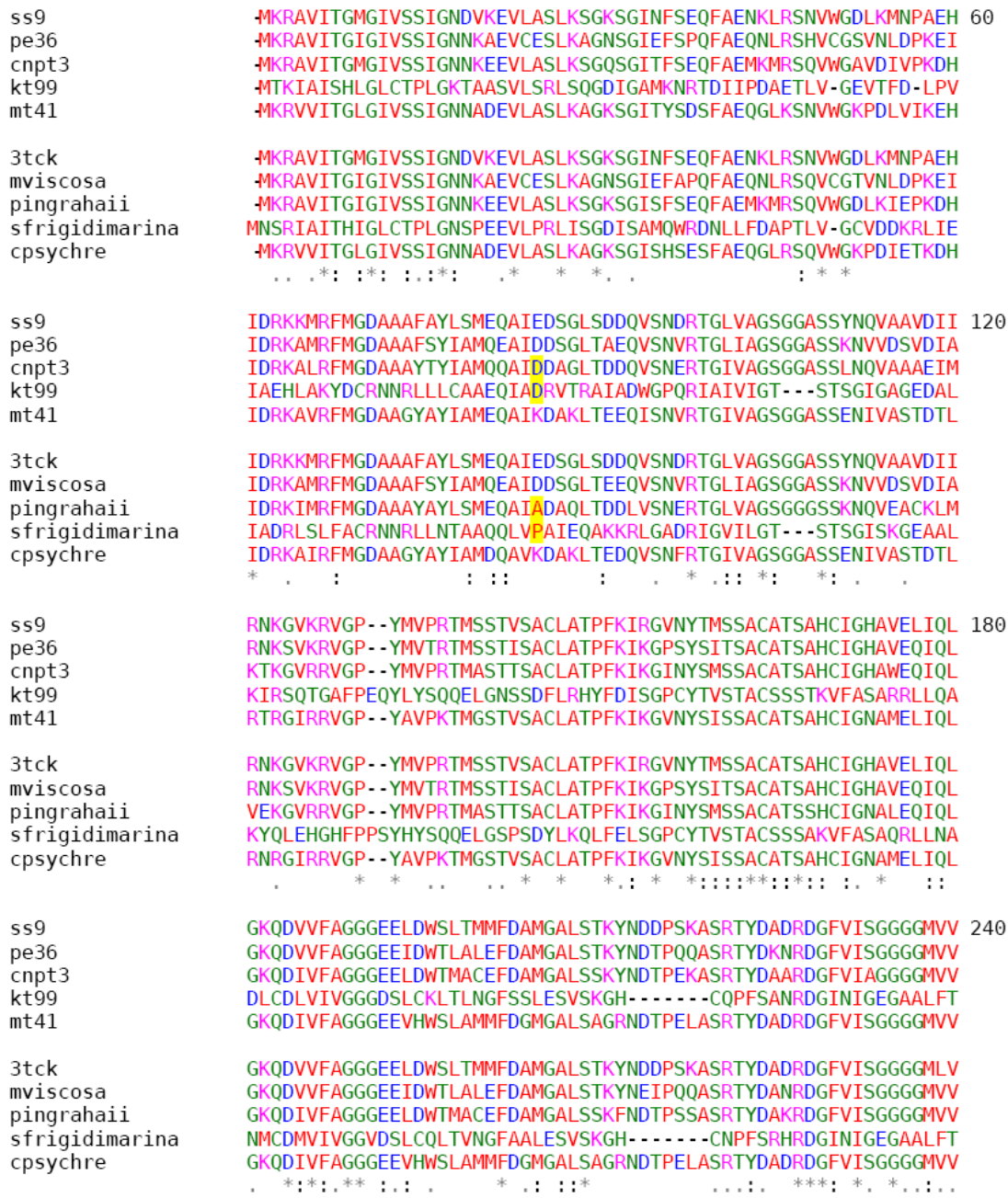

Figure 4.5: FabB protein alignment with colors depicting amino acid properties as described in Table 3. Proteins from deep-sea strains are grouped on top, listed with increasing piezophilicity. Comparison strains are grouped on the bottom. Symbols depict similarity between amino acids, "*" means identical amino acids, ":" are conserved substitutions, "." are semi-conserved substitutions. Regions with substitutions of interest are highlighted in yellow.

```
ss9                 IEELEHALARGAKIYGEVVGYGATSDGYDMVA--PSGEGAVRCMKMAMKDLD---APIDY 300
pe36                VEELEHALARGATIYAEITGYGATSDGYDMVA--PSGEGAIRCMKQALETVD---GEIDY
cnpt3               VEELEHALARGAKIYAEIVGYGATSDGYDMVA--PSGEGAVRCMRQALATVD---GKVDY
kt99                LEKLTADDSSSVI----LAGIGESSDAHHISAPHPEGRGAIAAIQSALEDANLAAEDIDY
mt41                VEELEHALARGAHIYAELVGYGATSDGYDMVA--PSGEGAVRCMKQAMQGVE---GKIDY

3tck                IEELEHALARGAKIYGEVVGYGATSDGYDMVA--PSGEGAVRCMKMAMQDLD---VPIDY
mviscosa            VEELEHALARGATIYAEITGYGATSDGYDMVA--PSGEGAVRCMQQALETVD---GEIDY
pingrahaii          VEELEHALARGAKIYAEIVGYGATSDGYDMVA--PSGEGAVRCMRQALATVD---GEVDY
sfrigidimarina      LERGQSDVM--------LAGIGESADAHHISAPHPQGLGAIAAMQQALTEANITAKQIDY
cpsychre            VEELEHALARGAHIYAELVGYGATSDGYDMVA--PSGEGAVRCMQQAMQGVE---GKLDY
                    :*               :.* * ::*.: : *  *.* **: .:. *:   :     :**

ss9                 INTHGTSTPVGDVKELGAIQELFGKNSPAISATKAMTGHALGAAGVHEAIYSTLMLN--- 360
pe36                LNTHGTSTPVGDTKELEAIQAVFGHKSPKISATKSLTGHALGAAGVHEAIYSLIMME---
cnpt3               LNTHGTSTPVGDTKELGAINEVFGLDAPKISATKSQTGHSLGAAGVHEAIFSMIMME---
kt99                INLHGTATPKNDAMESRAIKQVFGSTPPPCSSTKPITGHTLGAAGAIEAAFCYLLLSQES
mt41                LNTHGTSTPVGDVKELGAIQELFGDDSPAISATKAMTGHALGAAGVHEAIYSILMME---

3tck                INTHGTSTPVGDVKELGAIQELFGENSPAISATKAMTGHALGAAGVHEAIYSTLMLN---
mviscosa            LNTHGTSTPVGDTKELEAIQAVFGHKSPKISATKSLTGHALGAAGVHEAIYSLIMME---
pingrahaii          LNTHGTSTPVGDTKELEAINIVFPDKAPKISATKSQTGHSLGAAGVHETIYSLLMME---
sfrigidimarina      LNLHGTATPKNDAMESRAVTAVFGHSLPACSSTKPLTGHTLGAAGAIEAAFCYLLLSPLN
cpsychre            LNTHGTSTPVGDVKELGAIQELFGEDSPAISATKAMTGHALGAAGVHEAIYSILMME---
                    :* ***:** .*. *  *:  :*    *  *:**. ***:*****. *: :. :::.

ss9                 -NNFVAPSINIDNLDPAAEGLDIVTETR-ET---ELTTVMSNSFGFGGTNATLVIKKYEG 420
pe36                -NSFIAASANIEDLDEKADGLDIVSGQAIDA---ELTNVMSNSFGFGGTNATLVFSKFTK
cnpt3               -HNFISPSINIDVLDELAVNLPIVKNKSIEA---KLNIVMSNSFGFGGTNATLVFKRFTK
kt99                SRQTLPPQIWDGIQDPQDPSLPLVSQGQ-SA---EVNYLMSNSFAFGGSNASLILARNPR
mt41                -HSFVAPSINIDNLDEKAHGLDIVTEKR-DV---ELNLIMSNSFGFGGTNATLVMKKYK-

3tck                -NNFVAPSINIDNLDPAAEGLDIVTETR-ET---ELKTVMSNSFGFGGTNATLVIKKYED
mviscosa            -NSFIAASANIQELDEKANGLDIVSGQAIDA---ELTNVMSNSFGFGGTNATLVFSKFTK
pingrahaii          -NDFISPSINIDEIDEKAVNLPIVANKAVDA---KLNIVMSNSFGFGGTNATLVFKRFS-
sfrigidimarina      HDKRLPPHIWDGQQDANDPNIALVSANN-NSGLNQLNFIMSNSFAFGGSNASLIFCRKDA
cpsychre            -NNFVAPSINIDNLDEKAQGLDIVTEKR-DV---ELNLVMSNSFGFGGTNATLVMKKYK-
                      . :..  .    *    .: :*     .    ::. :*****.***:**:*:: .
```

Figure 4.5, Continued: FabB protein alignment with colors depicting amino acid properties as described in Table 3. Proteins from deep-sea strains are grouped on top, listed with increasing piezophilicity. Comparison strains are grouped on the bottom. Symbols depict similarity between amino acids, "*" means identical amino acids, ":" are conserved substitutions, "." are semi-conserved substitutions. Regions with substitutions of interest are highlighted in yellow.

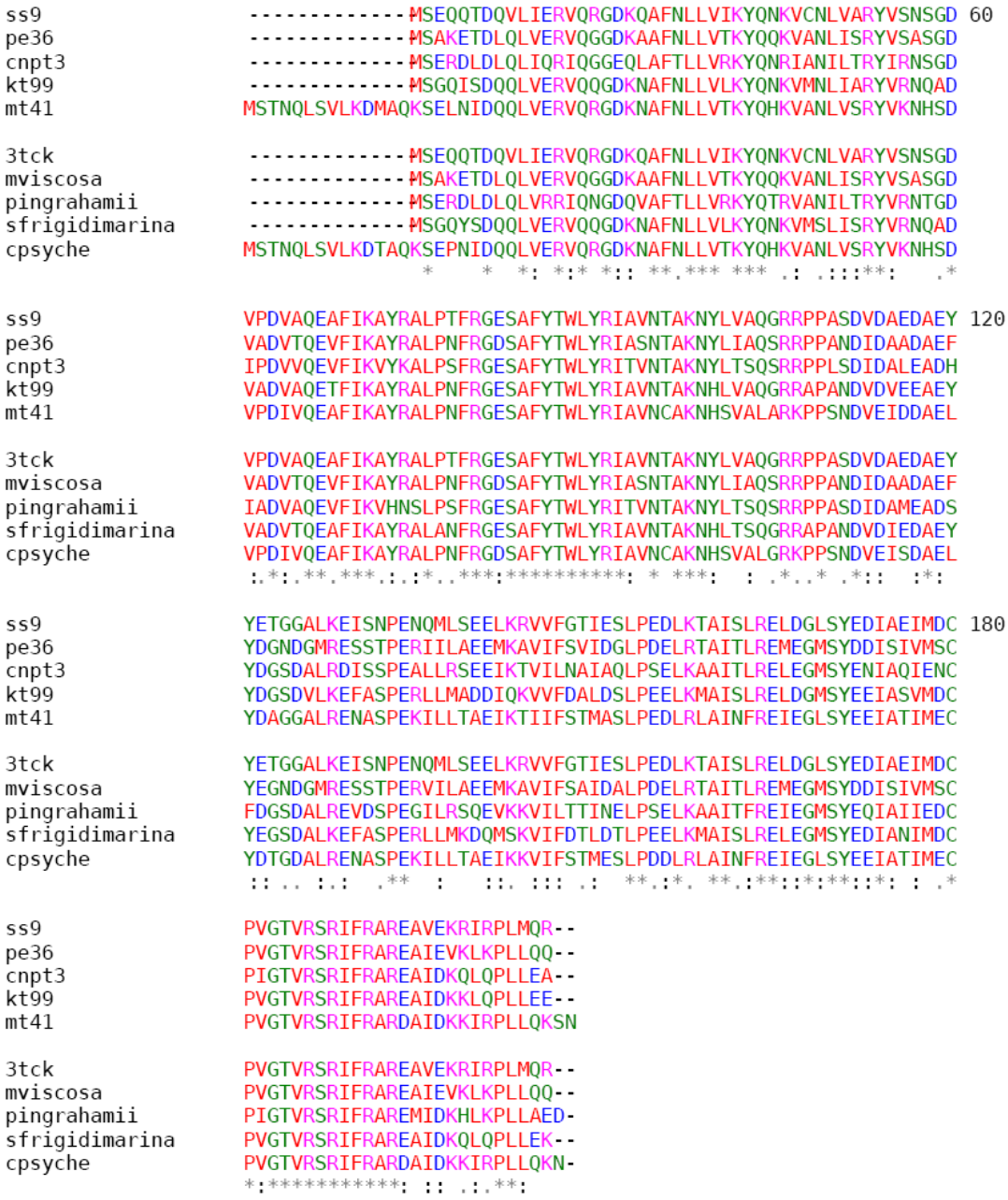

Figure 4.6: RpoE protein alignment with colors depicting amino acid properties as described in Table 3. Proteins from deep-sea strains are grouped on top, listed with increasing piezophilicity. Comparison strains are grouped on the bottom. Symbols depict similarity between amino acids, "*" means identical amino acids, ":" are conserved substitutions, "." are semi-conserved substitutions. Regions with substitutions of interest are highlighted in yellow.

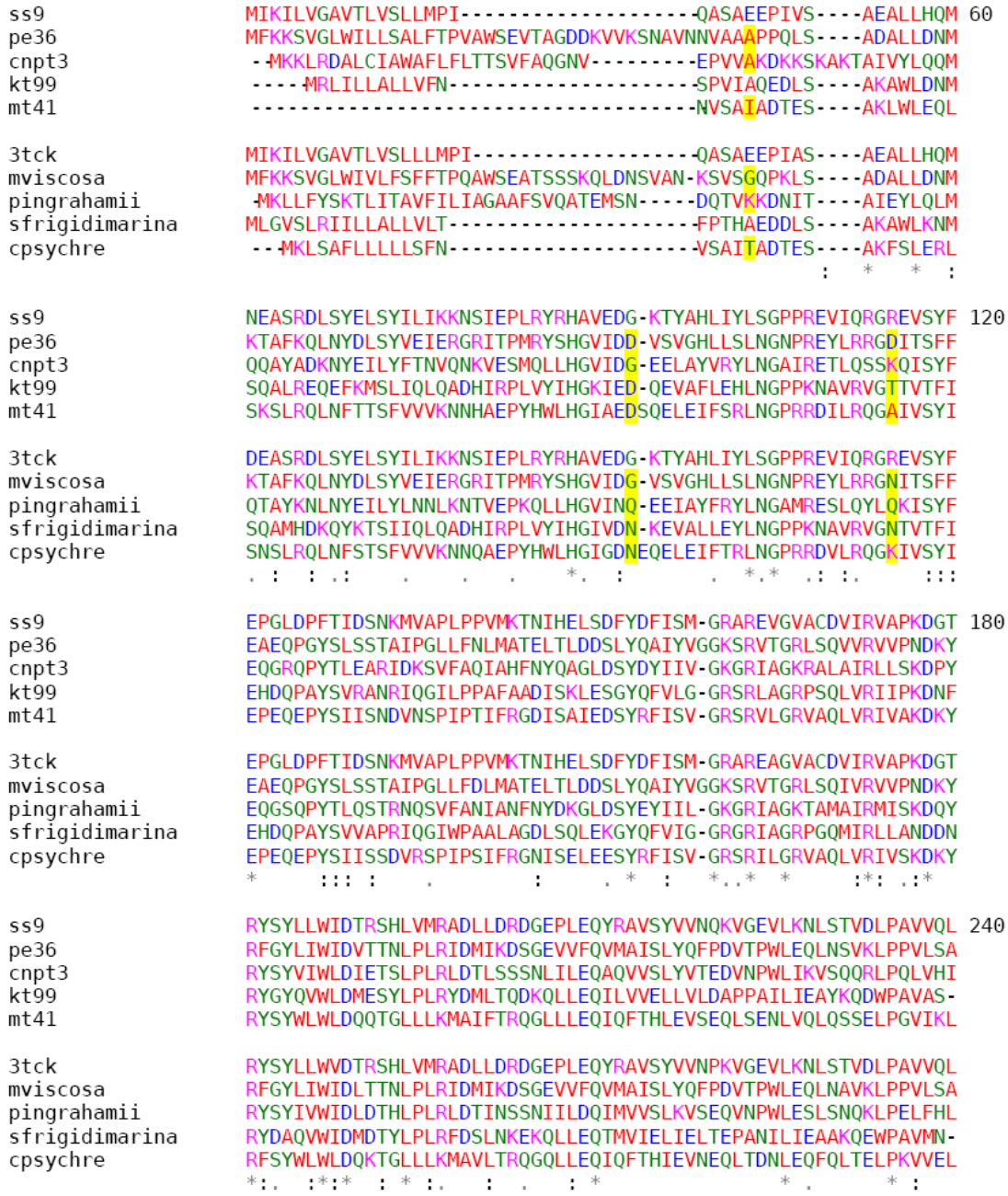

Figure 4.7: RseB protein alignment with colors depicting amino acid properties as described in Table 3. Proteins from deep-sea strains are grouped on top, listed with increasing piezophilicity. Comparison strains are grouped on the bottom. Symbols depict similarity between amino acids, "*" means identical amino acids, ":" are conserved substitutions, "." are semi-conserved substitutions. Regions with substitutions of interest are highlighted in yellow.

Figure 4.7, Continued: RseB protein alignment with colors depicting amino acid properties as described in Table 3. Proteins from deep-sea strains are grouped on top, listed with increasing piezophilicity. Comparison strains are grouped on the bottom. Symbols depict similarity between amino acids, "*" means identical amino acids, ":" are conserved substitutions, "." are semi-conserved substitutions. Regions with substitutions of interest are highlighted in yellow.

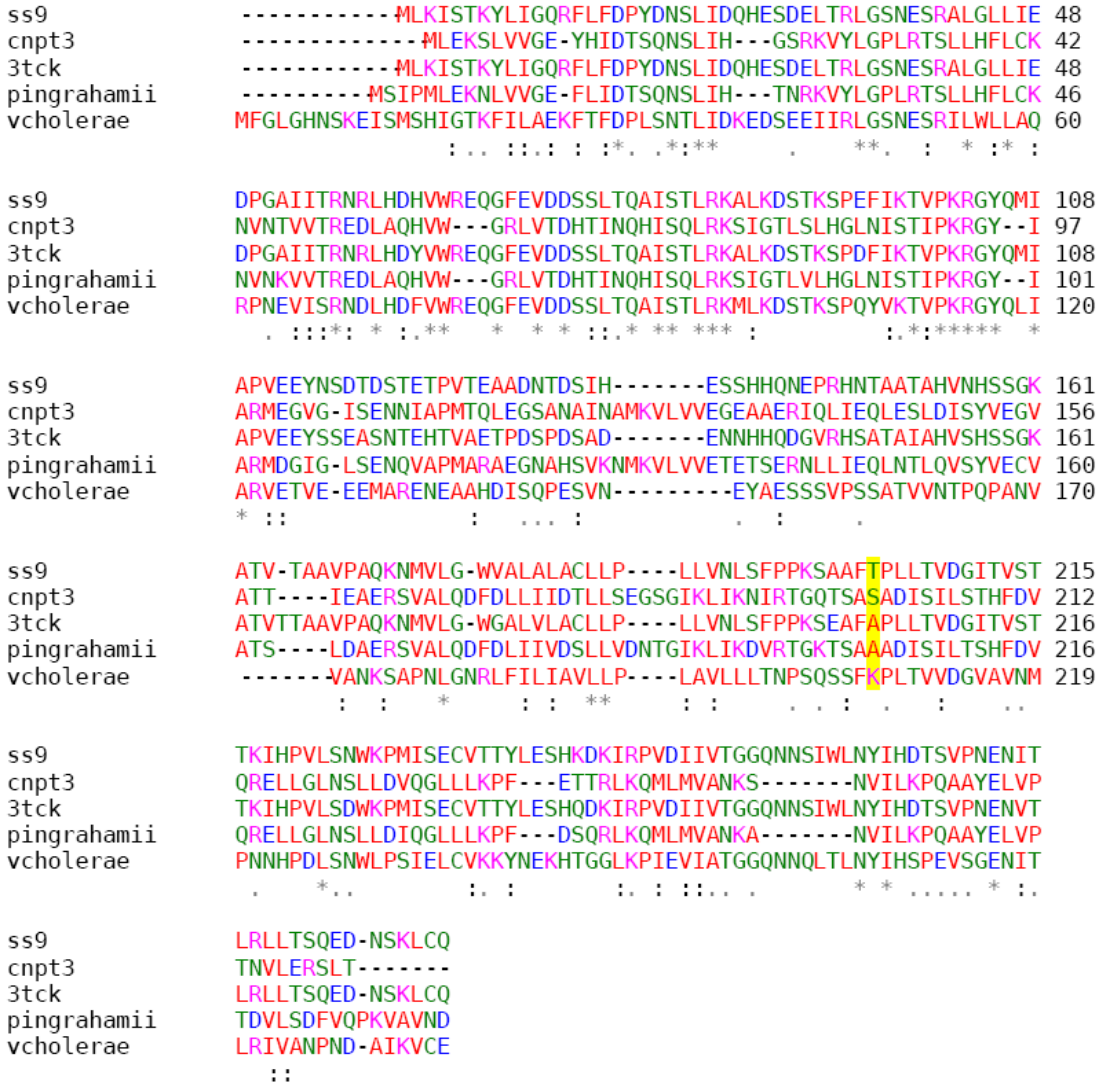

Figure 4.8: ToxR protein alignment with colors depicting amino acid properties as described in Table 3. Proteins from deep-sea strains are grouped on top, listed with increasing piezophilicity. Comparison strains are grouped on the bottom. Symbols depict similarity between amino acids, "*" means identical amino acids, ":" are conserved substitutions, "." are semi-conserved substitutions. Regions with substitutions of interest are highlighted in yellow.

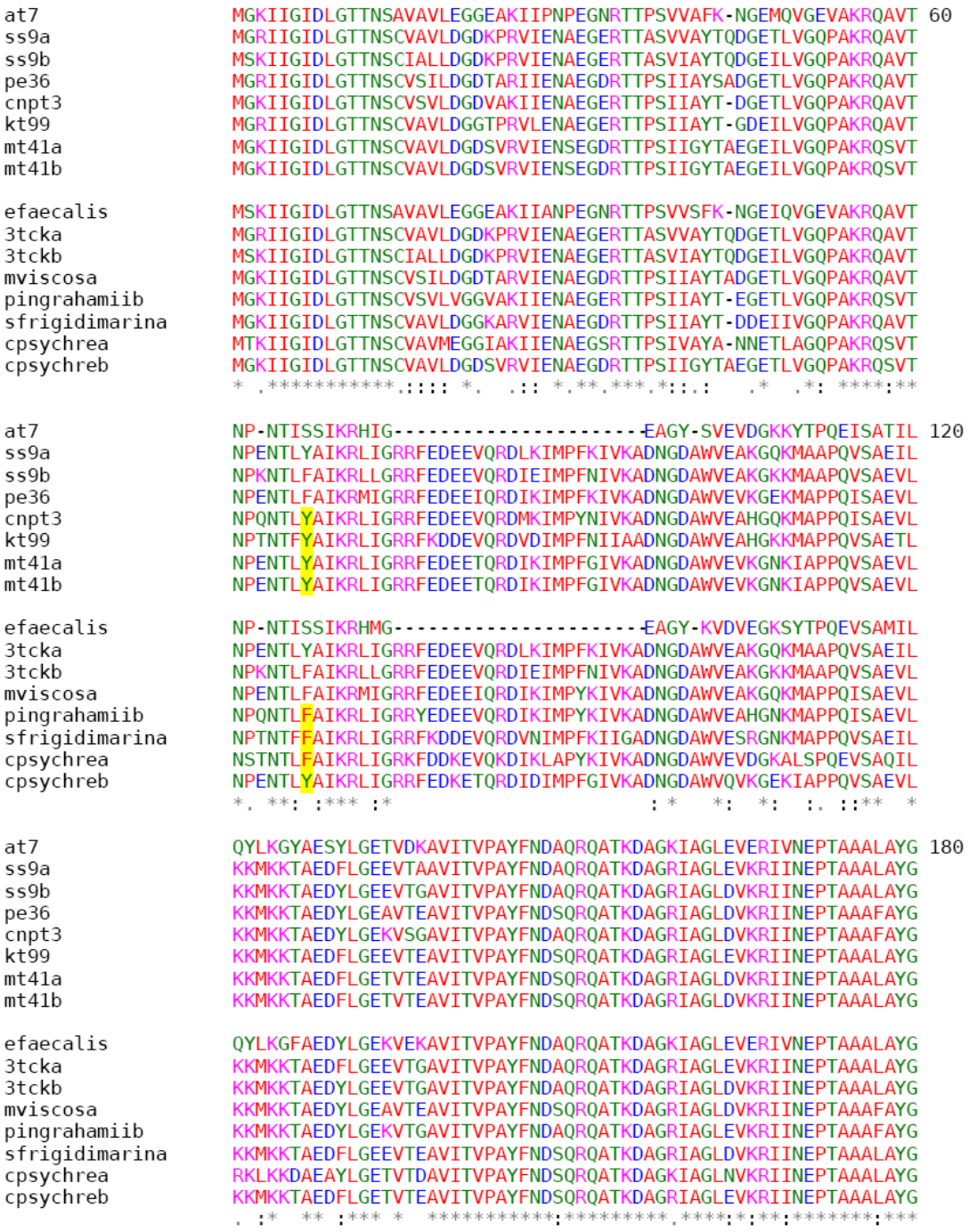

Figure 4.9: DnaK protein alignment with colors depicting amino acid properties as described in Table 3. Proteins from deep-sea strains are grouped on top, listed with increasing piezophilicity. Comparison strains are grouped on the bottom. Symbols depict similarity between amino acids, "*" means identical amino acids, ":" are conserved substitutions, "." are semi-conserved substitutions. Regions with substitutions of interest are highlighted in yellow.

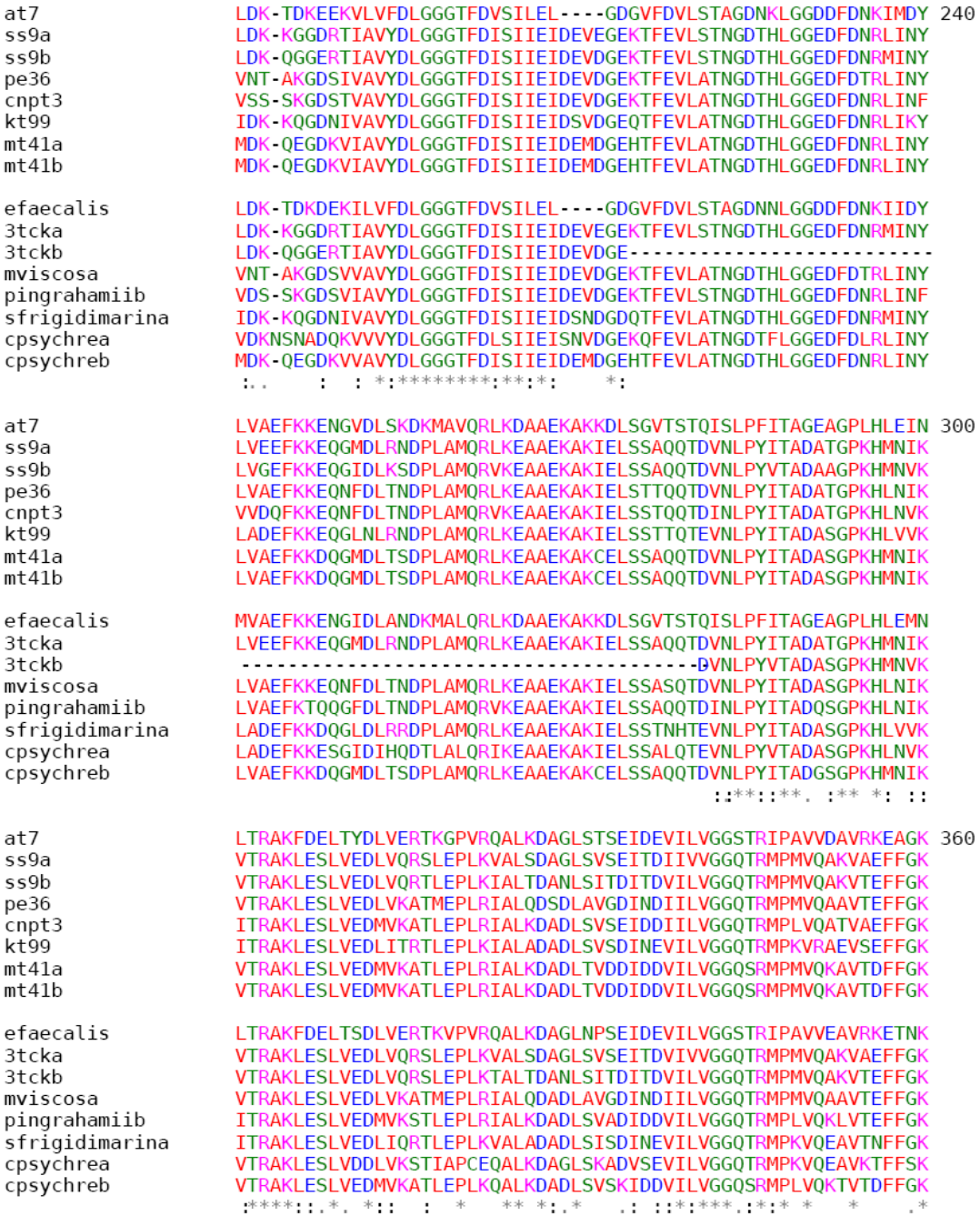

Figure 4.9, Continue: DnaK protein alignment with colors depicting amino acid properties as described in Table 3. Proteins from deep-sea strains are grouped on top, listed with increasing piezophilicity. Comparison strains are grouped on the bottom. Symbols depict similarity between amino acids, "*" means identical amino acids, ":" are conserved substitutions, "." are semi-conserved substitutions. Regions with substitutions of interest are highlighted in yellow.

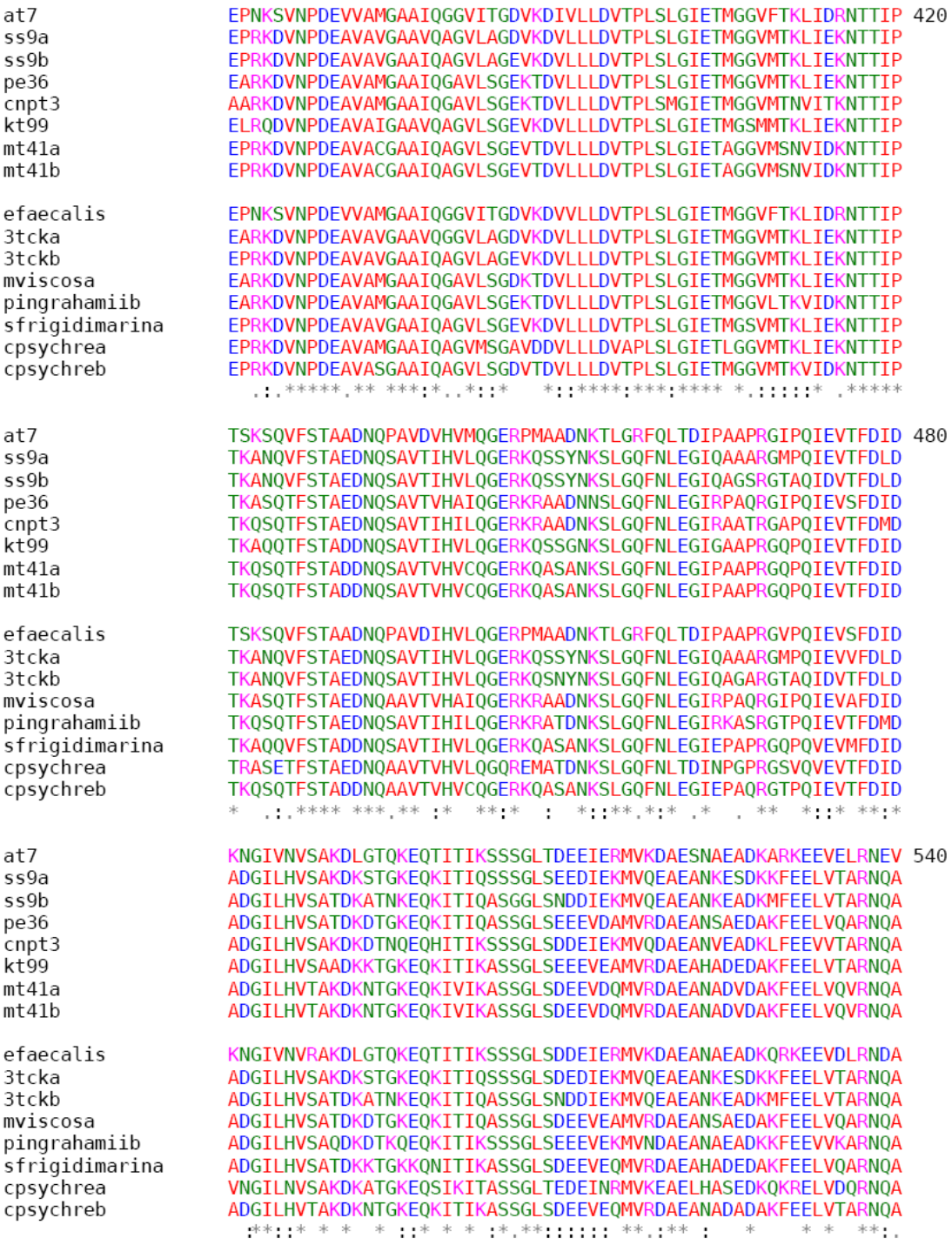

Figure 4.9, Continue: DnaK protein alignment with colors depicting amino acid properties as described in Table 3. Proteins from deep-sea strains are grouped on top, listed with increasing piezophilicity. Comparison strains are grouped on the bottom. Symbols depict similarity between amino acids, "*" means identical amino acids, ":" are conserved substitutions, "." are semi-conserved substitutions. Regions with substitutions of interest are highlighted in yellow.

Figure 4.9, Continue: DnaK protein alignment with colors depicting amino acid properties as described in Table 3. Proteins from deep-sea strains are grouped on top, listed with increasing piezophilicity. Comparison strains are grouped on the bottom. Symbols depict similarity between amino acids, "*" means identical amino acids, ":" are conserved substitutions, "." are semi-conserved substitutions. Regions with substitutions of interest are highlighted in yellow.

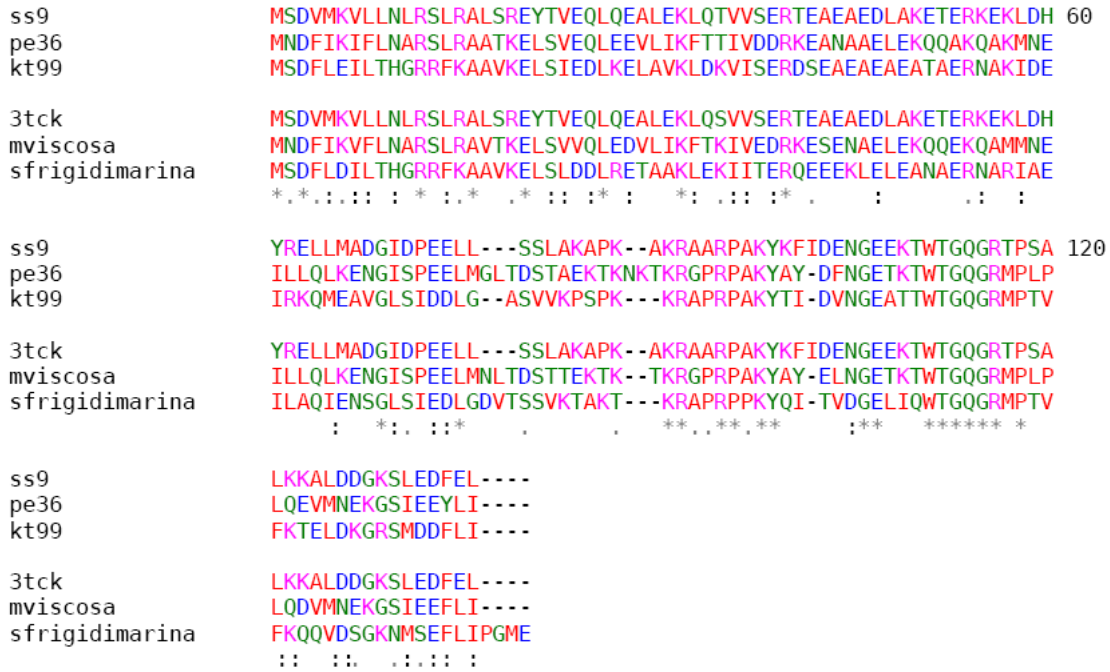

Figure 4.10: H-NS protein alignment with colors depicting amino acid properties as described in Table 3. Proteins from deep-sea strains are grouped on top, listed with increasing piezophilicity. Comparison strains are grouped on the bottom. Symbols depict similarity between amino acids, "*" means identical amino acids, ":" are conserved substitutions, "." are semi-conserved substitutions. Regions with substitutions of interest are highlighted in yellow.

Figure 4.11: Pairwise alignments of RecD, highlighting gaps between comparison strains, with *E. coli* included as a reference. Strains are paired with closely related strains underneath and listed in order of increasing piezophilicity. Symbols depict conservation among amino acids: "*" are identical, ":" are conserved, and "." are semi-conserved.

Figure 4.11, Continued: Pairwise alignments of RecD, highlighting gaps between comparison strains, with *E. coli* included as a reference. Strains are paired with closely related strains underneath and listed in order of increasing piezophilicity. Symbols depict conservation among amino acids: "*" are identical, ":" are conserved, and "." are semi-conserved.

```
at7_834              FILTGGPGTGKTTVLNGIVNLFAELNGLSLELDDYTDKIFPILLAAPTGRAAKRMNESTE
efaecalis_865        FILTGGPGTGKTTVINGIVSLFAELNGLSLDLKDYTQEMFPILLAAPTGRAAKRMNETTG
                     **************:****.**********:*.***:::********************:*
ss9_701              VDLPMMARLLDALPPHAKLILLGDRDQLASVEAGAVLGDICAFARQGYSPQQDQILSQLT
3tck_701             VDLPMMARLLDALPSHAKLILLGDRDQLASVEAGAVLGDICAFSSQGYSPQQGQVLSQLT
                     **************.*****************************: *******.*:******
pe36_675             LGDKDQLASVEAGSVLADICAYAESGYSAEQVNWLSKITDYELSQYQASDATAIYDSLCL
mviscosa_675         LGDKDQLASVEAGSVLADICAYAELGYSGGQANWLSKITDYELTQYQASDATAIYDSLCL
                     ********************** ***. *.***********:***************
cnpt3_687            LLGDRDQLASVEAGAVLGDICAFIDAGYSQTKSHQLADLTGFEHLRQSGNTRSGMADNLC
pingrahamii_698      LLGDKDQLASVEAGAVLGDICHFIESGYSAKKAQLLADLTGYTSLTKTAPRSPNMADNLC
                     ****.*************** **::*** .*:: ******:  * ::.   ..******
kt99_673             GAVLADICAGLK-ADKNQSAKWQMRYSAAYATRLSQLTGNDLSGFISASPGIGDSLCMLM
sfrigidimarina_687   GAVLADICVGLKQANQHEASPWSMRYSAAQANLIEQLTQQDVSQFIHQQPKIGDSLCMLM
                     ********.*** *:::::: *.****** *. :.*** :*:* **  .* *********
mt41_682             IDEVSMVDLPLMTRIFRALKDTAKVILLGDADQLPSVAAGSVLADIAPRPHGGFSSDNVQ
cpsychre_769         IDEVSMVDLPLMTRVFRALKDTAKVILLGDADQLPSVAAGSVLADIAPRPHGGFSLENEQ
                     **************:****************************************** :* *
ecoli_608            DEASMIDLPMMSRLIDALPDHARVIFLGDRDQLASVEAGAVLGDICAYANAGFTAERAGQ


at7_834              LPSSTIHRLLGLNGQEKPTAELSDRELEGGLLIVDEMSMVDTWLANQLLRAVPQNMQVIF
efaecalis_865        LPASTIHRLLGLTGREK-NPSLTAKELEGGLLIVDEMSMVDTWLANTLLKAIPTNMQVIF
                     **:*********.*.** ...*: .********************** **.*:* ******
ss9_701              GFNVSGIRLGTHDDNSVSPIADSLCMLQKSYRFHAESGIGKLAKAINDGKPYDVDRICQH
3tck_701             GFNVSGIHLGMHEGNSVSPIADSLCMLQKSYRFHAESGIGKLAKAINDGKPYDVEKVCQH
                     *******.** *:.*****************************************:.:***
pe36_675             LRKSYRFDAESGIGCLAGAVNRGDLNEFDRVWENQHDDIHLHPLSTETYDALITMAKQGY
mviscosa_675         LRKSYRFDAESGIGCLAGAVNRGDLNEFDRVWENQHDDIHLHPLSTEAYDALITMAKQGY
                     ************************************************:***********
cnpt3_687            LLKKSYRFDQYSGIGFLGKAINQGQASPEQVLALCEKYDDLAHYENNESSLPILNKMILD
pingrahamii_698      LLRKSYRFDQYSGIGYLAKAINSGRANSKKVMALCDDYNDLQHYPNTELGKKELAALVLE
                     **.************:*.****.*.*..::*:***:.*:** ** *.* .   *  ::*:
kt99_673             HSHRFKGDAGIGQLATAVNNSDKSRIMEVWQSGYDELHWIEHQKTPTG-NAGLEQLLTQS
sfrigidimarina_687   HSHRFRGDAGIGLLANAVNQSNVHAIRQVWQHDHAELNWFEHQHNAQGDNTGKQALLALS
                     *****.***** **.***:*:   * :*** .: **:*:***:.. * *:* : **: *
mt41_682             YLRQVCQLSAEQASEFLSGQ----LTGHTSNVDYLSFLLKSRRFDGKGGIGLLANSVIQG
cpsychre_769         YLAQVCQLLPAQVSDYFSAHRLPSQSAEQSSFDYLTFLVKSRRFDGKGGIGLIANSVIKG
                     ** ***** . *.*:::*.:     :.  *..***:**:************:*****:*
ecoli_608            LSRLTGTHVPA---GTGTE--AASLRDSLCLLQKSYRFGSDSGIGQLAAAINRGDKTAVK


at7_834              VGDKDQLPSVGPGQVLHDLIRAKQIPSRELTEIYRQDDGSSIIPLAHAIKEGQLPADFTK
efaecalis_865        VGDKDQLPSVGPGQVLHDLLQINEIPKCELNEIYRQGDGSSIIPLAHEIKEGKLPADFQK
                     ********************:. ::**. **.*****.********** ****:***** *
ss9_701              DFKDINHYPLSAENYQKMINLTVTFYRDYLDAIEHKQEPEVVLKAFSSVRLLCALREGDF
3tck_701             DFKDINHYPLSAENYQKMINRTVTFYRDYLDAIEHKQEPEVVLKAFSSVRLLCALREGDF
                     ******************** ************************************
pe36_675             QDYLAKVVVSPSDDEAKSILRSFNDFQILTAIREGDFGVHGLNQRIEKALQSARKIRKNK
mviscosa_675         QDYLAKVVVSPSEDEAKSILRSFNDFQILTAIREGDFGVHGLNQRIEKALQSARKIRKNQ
                     ************:**************************************************:
cnpt3_687            GYQPYLTLLHHTSKQQTLSPKALLKAFNKFQILCAVREGDWGVQGLNQRAEKVLASAGLI
pingrahamii_698      GYRPYLQNLSTITVDNRNLAGQLLKQFNQFKILCATREGEWGVKCLNKLCETVLEKAGLL
                     **.***  *   : ::   .  *** **:*:****.***:***: **: .*.** .***:
kt99_673             VSHYRAYLEMAKD-----------PQSDPLD-----------IIECYNEFRILCAMRAGE
sfrigidimarina_687   VSAYGEYLSLVQENQQRINQQTIIDQADRQDSVITLPHSEQTIIDSFNQYRVLCAMRAGE
                     ** *  **.:.::            *:*  *          **:.:*::*:********
mt41_682             DVKRSWQLLCD-------------------------AHSQQAGQLSLAQGELANWLAPLV
cpsychre_769         DVKRSWRLLNDSLSQSTQEGFIQGDYLQAGYLQGDKSQKQQPSQLTLAQGDLTTWLAPLV
                     ******.** *                        ::.**..**:****:*:.******
ecoli_608            TVFQQDFTDIEKRLLQSGEDYIAMLEEALAGYGRYLDLLQARAEPDLIIQAFNEYQLLCA
```

Figure 4.11, Continued: Pairwise alignments of RecD, highlighting gaps between comparison strains, with *E. coli* included as a reference. Strains are paired with closely related strains underneath and listed in order of increasing piezophilicity. Symbols depict conservation among amino acids: "*" are identical, ":" are conserved, and "." are semi-conserved.

```
at7_834              NKKDRSFFQCNTYQIEPVIRQVVEKAKEKGFTAQDIQVLAPMYRGPAGIDALNKMMQEIF
efaecalis_865        NQKDRSFFASDIGHIEEYIRQIVTKAKAKGFTPQDIQVLAPMYRGAAGIDALNKMMQEIF
                     *:****** .:  :**  ***:* *** ****.***********.**************
ss9_701              GVVGLNQRIERELVRIGKISQTDDTWYEGRPVMITRNDHSLGLYNGDIGIAMIDEKVDSN
3tck_701             GVVGLNQRIERELVRIGKISQTDDTWYEGRPVMITRNDHGLGLYNGDIGIAMIDEKVDPN
                     ************************************.******************.*
pe36_675             SEWYAGRPIMITNNDHGLGLYNGDIGICMQDDDERMRVYFEMADGNVQSFLPSRLPPHQT
mviscosa_675         SEWYAGRPIMIINNDHGLGLYNGDIGICMQDEDDRMRVYFEMADGNVQSFLPSRLPPHQT
                     ********** ******************:*:**************************
cnpt3_687            DKSV---EAQQLAQTWYLGRPIMITQNNYHLGLYNGDIGLCLKDEKGQLRVYFEMPDGDV
pingrahamii_698      TTGQVHKDPQHAYVYWYVGRPVMITQNNYHLGLYNGDIGLCLLDESLQLRVYFQMPDASI
                      ..    :.*:    **:***:******************** **. ******:***..:
kt99_673             YGVDGINQGVRDALKQAKLISPEHEFYPGRPVIIQSNDYNLGLFNGDIGLILPDSAHQQA
sfrigidimarina_687   FGVEGMNIAITDSLKHAKLINPQNEFYLGRPIIIQSNDYNLGLFNGDIGLILQDSTK---
                     :**:*:* .: *:**:****.*::*** ***:******************** **::
mt41_682             KQYYQPIEFCANVSDAFALLAQFRVLCATRQGDYGVERLNEVIKHYLGKNKALYQQNEPA
cpsychre_769         KQYYQPIEHCSTVSDAFVLLSQFRVLCATRQGDYGVERLNEVIKSYLGKNNAPYQQHQQT
                     ********.*:.*****.**:************************ *****:* ***:: :
ecoli_608            LREGPFGVAGLNERIEQFMQQKRKIHRNPHSRWYEGRPVMIARNDSALGLFNGDIGIAL-


at7_834              NPNPSGRRKEVKFNEKSYRIGDKVLQLVNYPEMNVFNGDMGEIVGINLAKETEDKVDEIV
efaecalis_865        NPN-DGKKKEVKWNDTVYRIGDKVLQLVNTPELNVFNGDMGEIVGITLAKDSEDKVDELV
                     *** .*..****:*:. ************ **:*************.***::******:*
ss9_701              KPHLRVFFEMPDGSIRSVLPSRLPEHETVYAMTIHKSQGSEFADTLMVLPADFTPILTRE
3tck_701             KPHLRVFFEMPDGAIRSVLPSRLPEHETVYAMTIHKSQGSEFEDTLMVLPADFTPILTRE
                     *************:*************************** ****************
pe36_675             VFAMTIHKSQGSEFRHTVMILPNRFNPVLTRELVYTGITRAKEKLDIFTNVTVLKKAIKL
mviscosa_675         VFAMTIHKSQGSEFRHTVMILPNRFNPVLTRELVYTGITRAKEKLDIFTNVTVMKKAIKL
                     ************************************************:******
cnpt3_687            ADFQPSRLPSHETVFAMTVHKSQGSEFEHTLLALPQKSLPVVTRELIYTGITRAKLRLSL
pingrahamii_698      ADFQPSRLPNHETVFAMTVHKSQGSEFEHTVLALPENHVSVVSRELIYTAITRAKKQLTL
                     *********.*********************:****:: :.**:******.***** .*:*
kt99_673             TGAAPRLMAHFIQADGSILKVLPARLPSHDTCFAMTVHKSQGSEFSKVAMVLPLWPSQAQ
sfrigidimarina_687   ---PSRLMAHFVMADGSLLKVLPARLPSHDTCYAMTVHKSQGSEFDKVSFVLPHKPSSAQ
                        ..******: ****:**************:************.**::*** .**.**
mt41_682             LYHGQPIMINENDYRLGLYNGDIGIIWRVIDQAGKTHLMACFEDTSLESNTA-----EDT
cpsychre_769         LYHGQPIMINENDYRLGLYNGDIGIIWKVIDQAGKTHLMACFEDTTLESKSAESKVDEST
                     ***************************.*****************:***::*     *.*
ecoli_608            -----DRGQGTRVWFAMPDGNIKSVQPSRLPEHETTWAMTVHKSQGSEFDHAALILPSQR


at7_834              IQFDANEVVYKRNEWIKITLAYCCSIHKSQGSEFKMVILPMVQNYHRMLRRDLLYTAITR
efaecalis_865        LQFDNNEVTYKRNEWNKITLSYCCSIHKAQGSEFRMVLLPMVHQYSRMLQRNLLYTAVTR
                     :*** ***.****** ****:*******:*****.**:****::* ***.*:*****:**
ss9_701              LIYTGVTRAKSRLYLFAQPDVIARAVRLRTERASGLARLLG
3tck_701             LIYTGVTRAKTRLYLFAQPDVIARAVRLRTERASGLACLLG
                     **********:************************** ***
pe36_675             RTERVSGLMALLNQD
mviscosa_675         RTERVSGLMTLLNKD
                     *********:***:*
cnpt3_687            FADLTLMARAIKQKTLRYSRLVERLKGDSA------------
pingrahamii_698      FADLSLMASAMRHKTKRFSRLVERLKSPSVTLAEKVKQKIKE
                     ****:*** *:.:** *:********. *.
kt99_673             KQLLTKELVYTAITRAKHQFTCLGSQTVFEQASLQATKRSSGLARRLWQ
sfrigidimarina_687   WQLLSKELLYTAITRAKNHFYCLGTSMIFEKASKQATSRASGLAWRLWG
                     .***:***:********::* ***:. :**:** ***.*:****.***
mt41_682             DTGQQAIRQILPSRLPKFESVYAMTIHKTQGSEFAHVALVISATQSAAHH----KQQRAS
cpsychre_769         ETSQLTIRQILPSRLPQFESVYAMTIHKTQGSEFSHVAMVISATQSEQQTLQQGKQQGGS
                     :*.* :**********:*****************:***:*******  :     *** .*

ecoli_608            TPVVTRELVYTAVTRARRRLSLYADERILSAAIATRTERRSGLAALFSSRE
```

Figure 4.11, Continued: Pairwise alignments of RecD, highlighting gaps between comparison strains, with *E. coli* included as a reference. Strains are paired with closely related strains underneath and listed in order of increasing piezophilicity. Symbols depict conservation among amino acids: "*" are identical, ":" are conserved, and "." are semi-conserved.

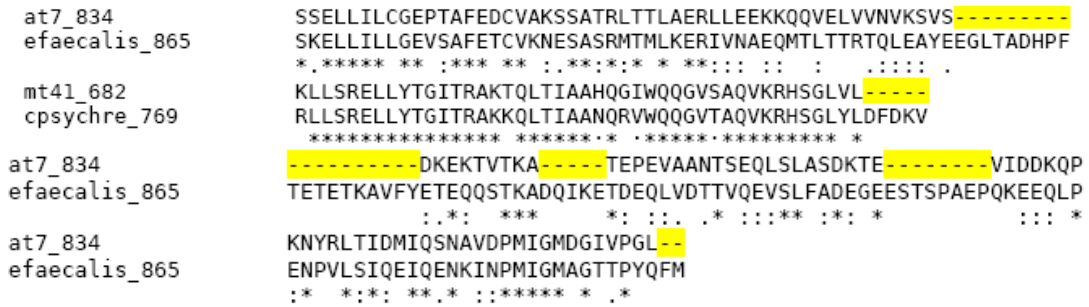

Figure 4.11, Continued: Pairwise alignments of RecD, highlighting gaps between comparison strains, with *E. coli* included as a reference. Strains are paired with closely related strains underneath and listed in order of increasing piezophilicity. Symbols depict conservation among amino acids: "*" are identical, ":" are conserved, and "." are semi-conserved.

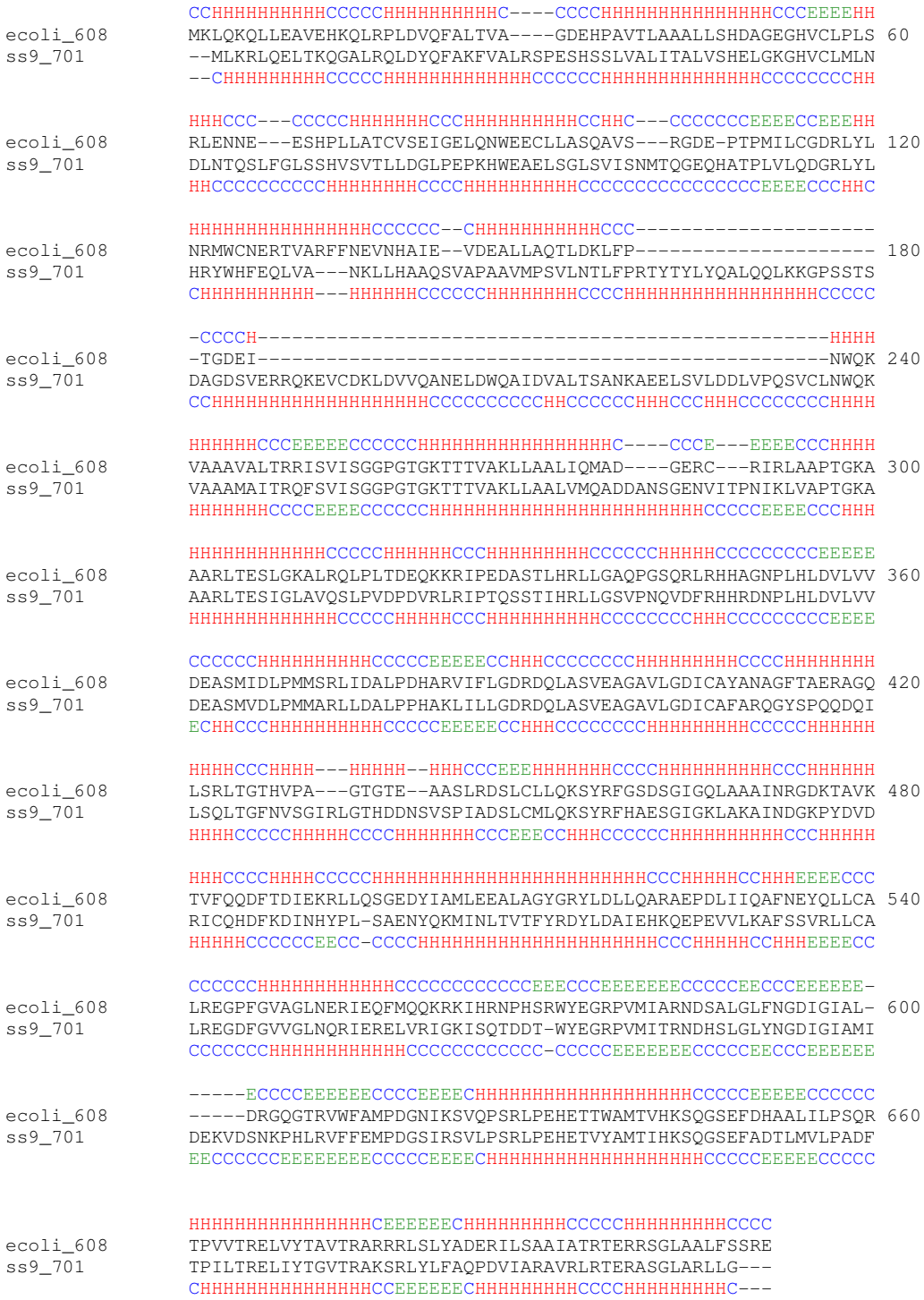

Figure 4.12: RecD secondary structure prediction comparison between *E. coli* and *P. profundum* SS9, showing the secondary structure in the *E. coli* gaps. Numbers following the name are the length of the protein in numbers of amino acids. Symbols for structures are as follows: green E is strand, red H is helix, and blue C is coil.

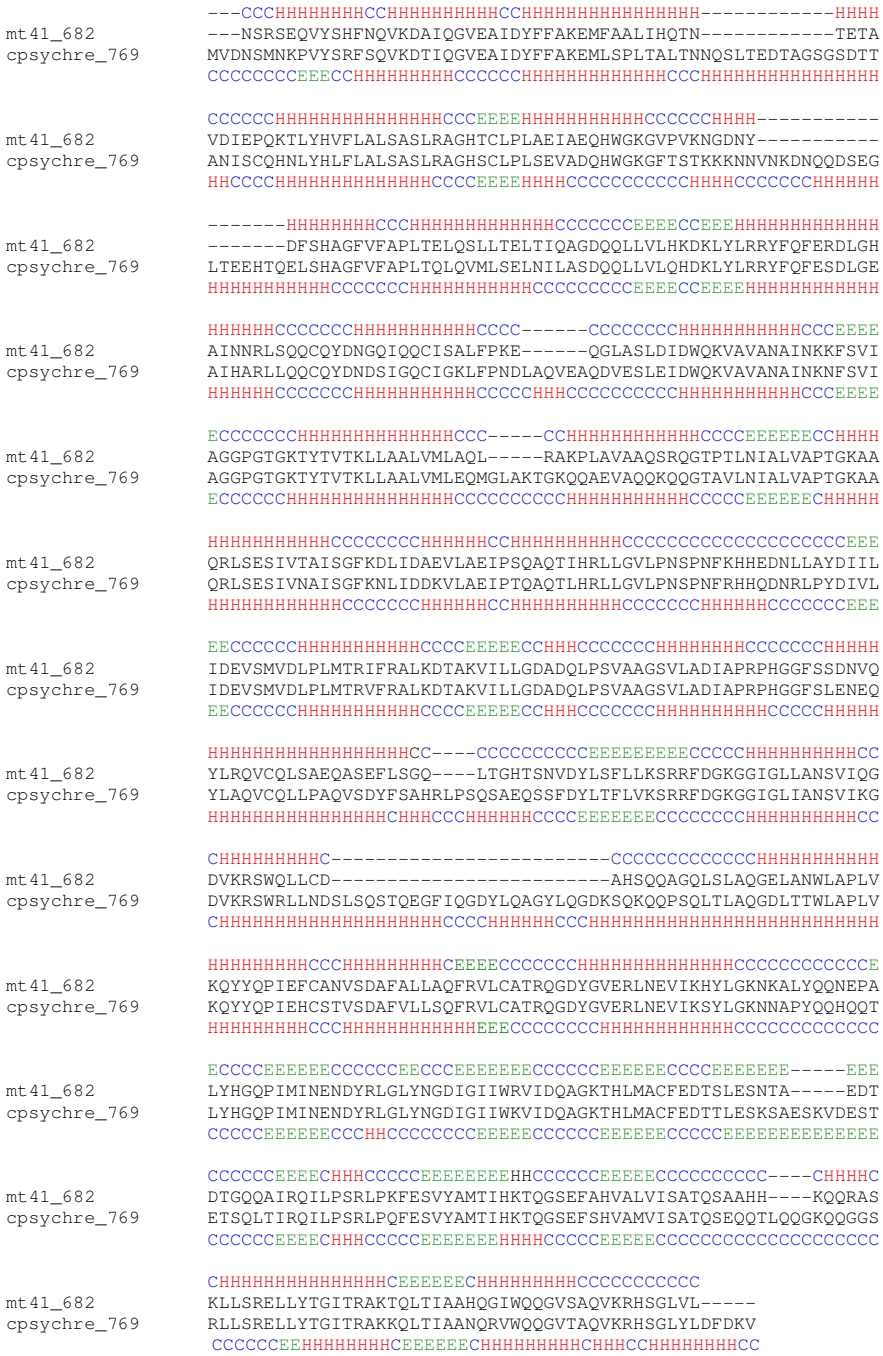

Figure 4.13: RecD secondary structure prediction comparison between MT41 and *C. psychrerthrea*, showing the secondary structure in the MT41 gaps. Numbers following the name are the length of the protein in numbers of amino acids. Symbols for structures are as follows: green E is strand, red H is helix, and blue C is coil.

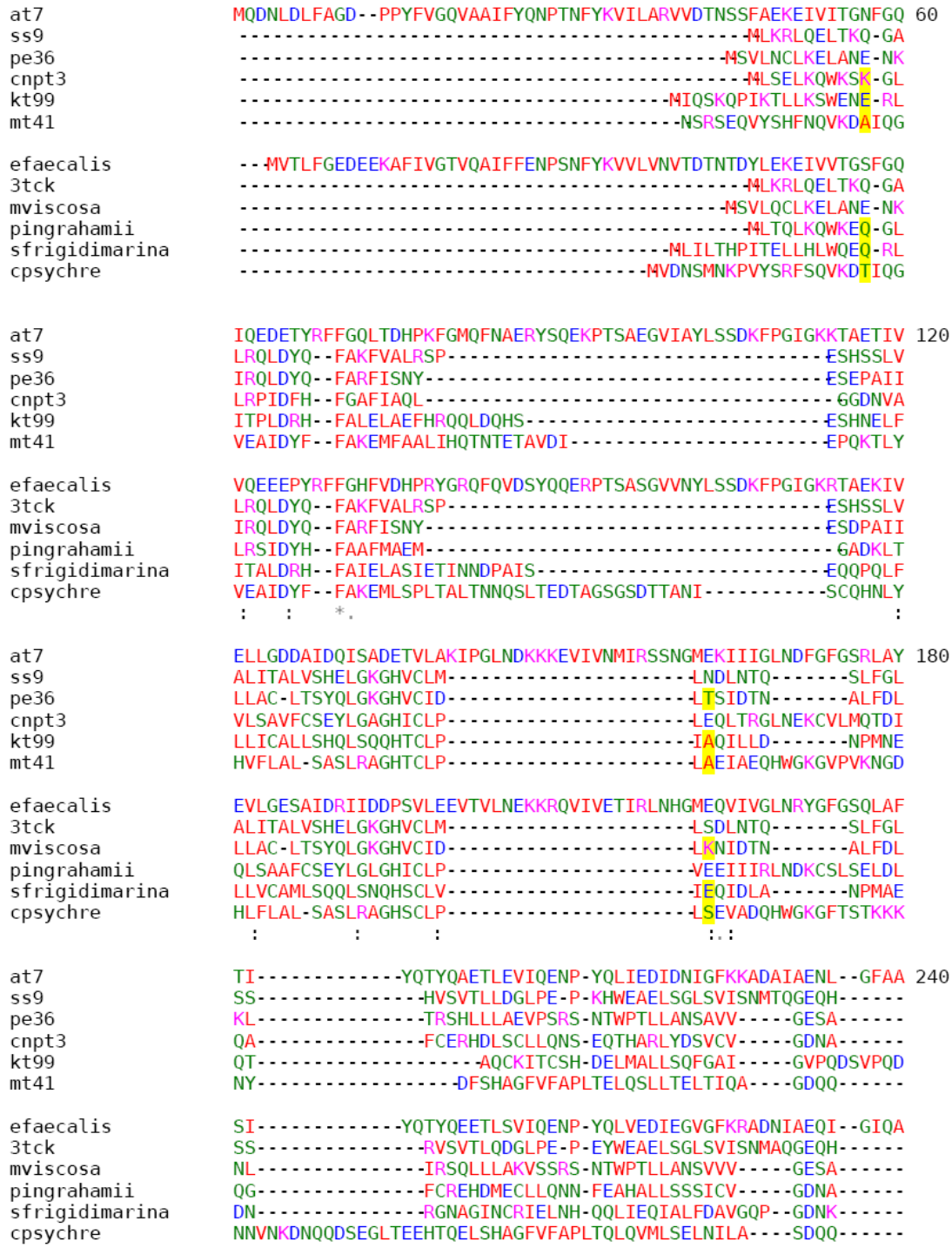

Figure 4.14: RecD protein alignment with colors depicting amino acid properties as described in Table 3. Proteins from deep-sea strains are grouped on top, listed with increasing piezophilicity. Comparison strains are grouped on the bottom. Symbols depict similarity between amino acids, "*" means identical amino acids, ":" are conserved substitutions, "." are semi-conserved substitutions. Regions with substitutions of interest are highlighted in yellow.

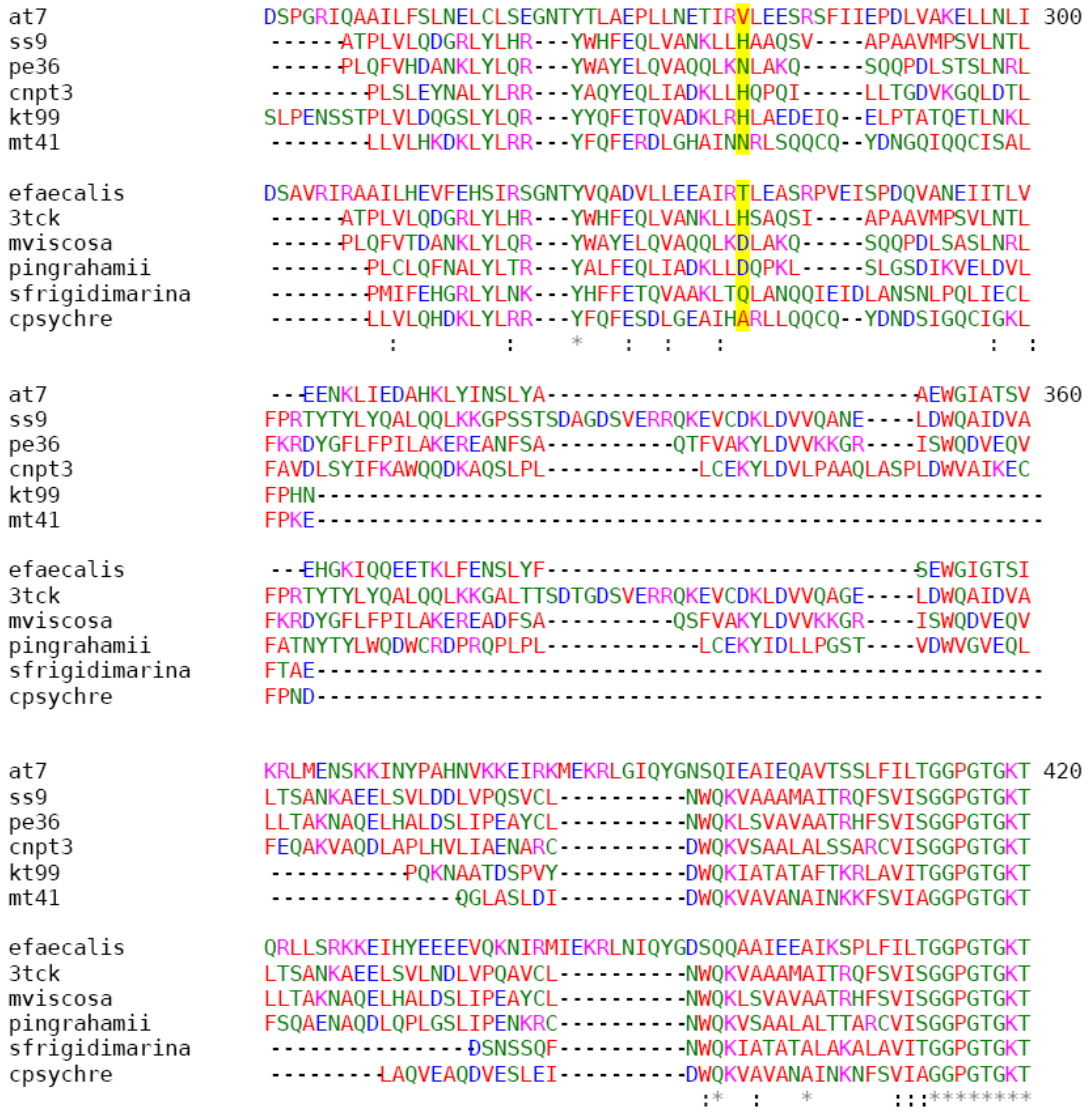

Figure 4.14, Continued: RecD protein alignment with colors depicting amino acid properties as described in Table 3. Proteins from deep-sea strains are grouped on top, listed with increasing piezophilicity. Comparison strains are grouped on the bottom. Symbols depict similarity between amino acids, "*" means identical amino acids, ":" are conserved substitutions, "." are semi-conserved substitutions. Regions with substitutions of interest are highlighted in yellow.

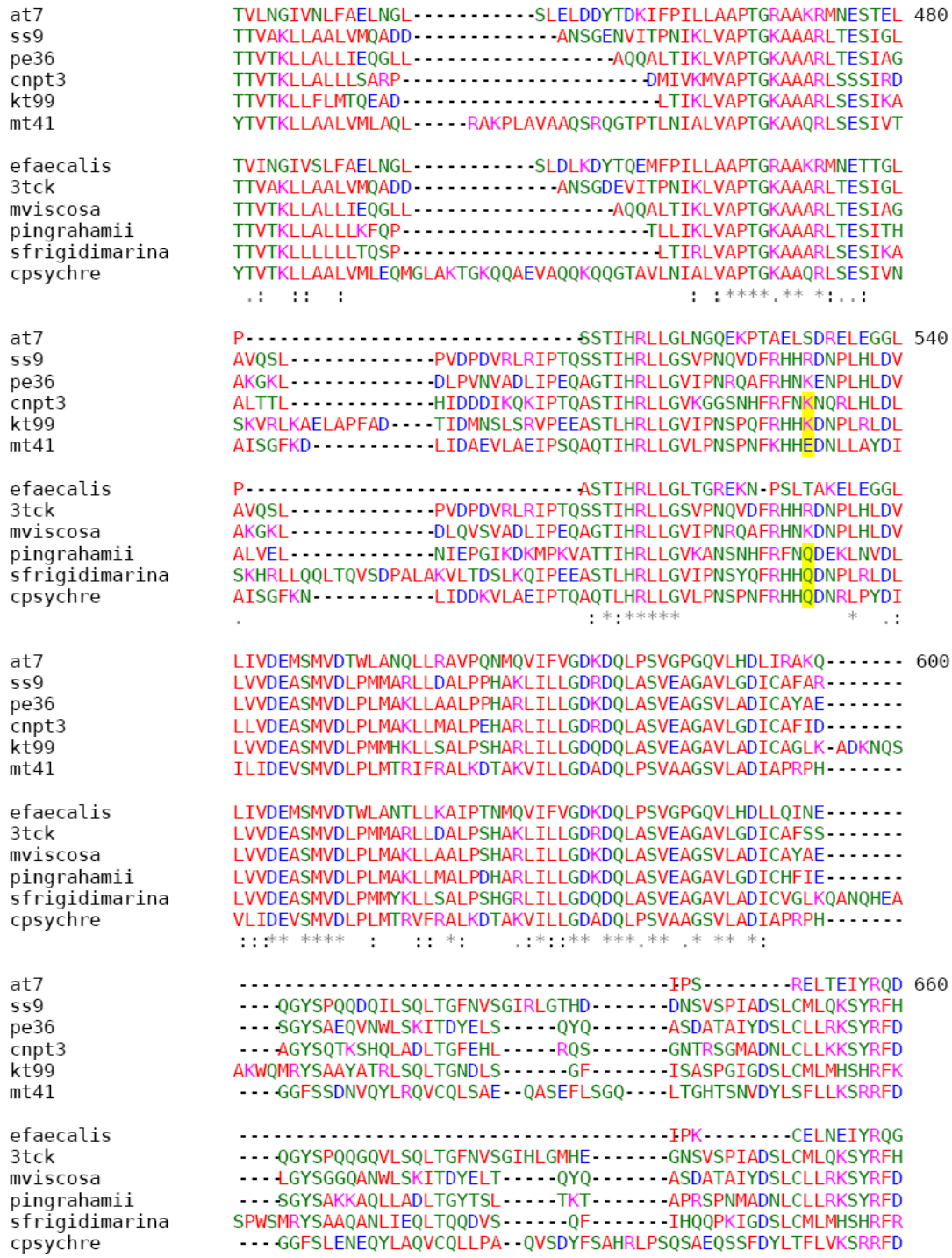

Figure 4.14, Continued: RecD protein alignment with colors depicting amino acid properties as described in Table 3. Proteins from deep-sea strains are grouped on top, listed with increasing piezophilicity. Comparison strains are grouped on the bottom. Symbols depict similarity between amino acids, "*" means identical amino acids, ":" are conserved substitutions, "." are semi-conserved substitutions. Regions with substitutions of interest are highlighted in yellow.

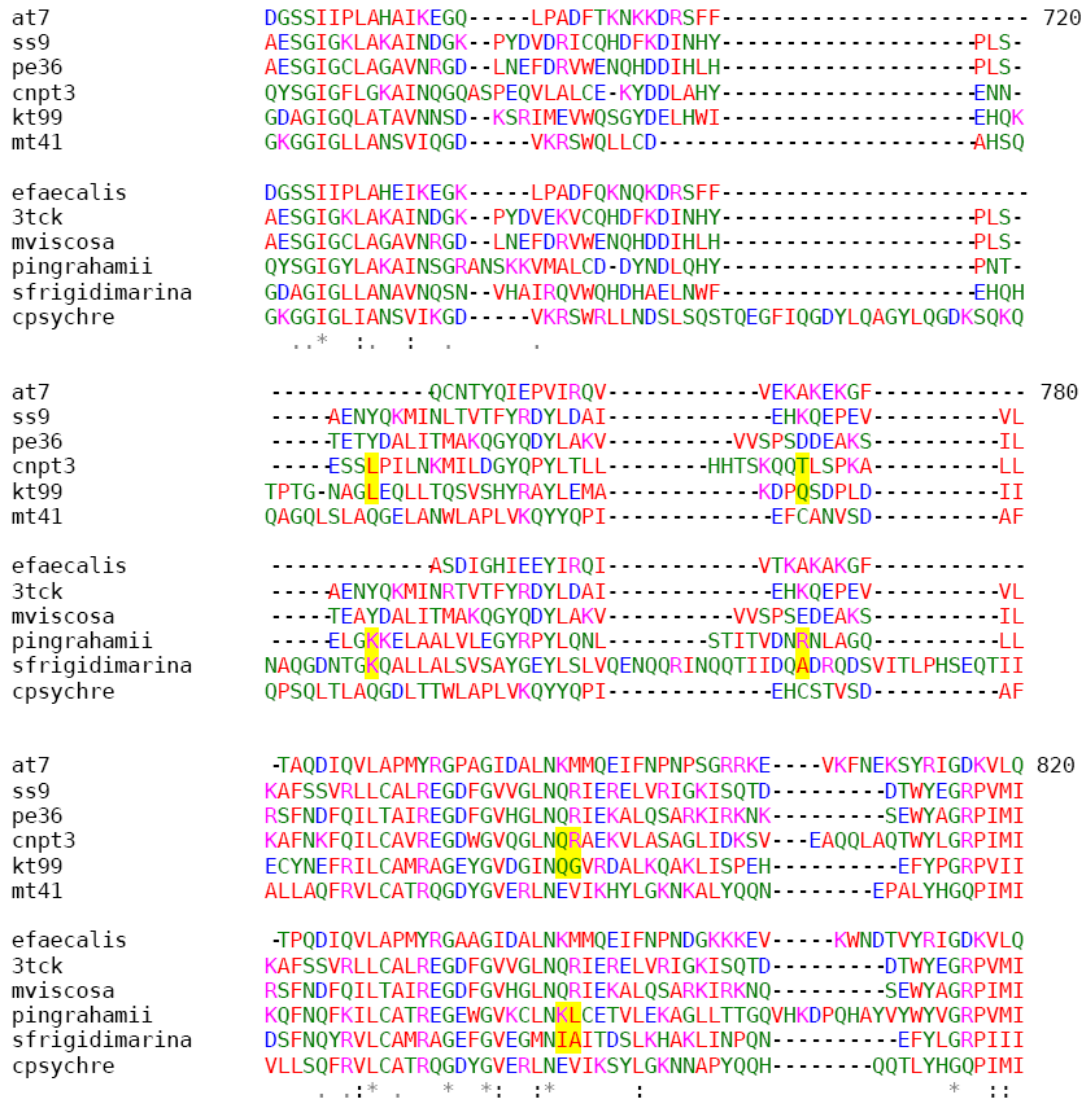

Figure 4.14, Continued: RecD protein alignment with colors depicting amino acid properties as described in Table 3. Proteins from deep-sea strains are grouped on top, listed with increasing piezophilicity. Comparison strains are grouped on the bottom. Symbols depict similarity between amino acids, "*" means identical amino acids, ":" are conserved substitutions, "." are semi-conserved substitutions. Regions with substitutions of interest are highlighted in yellow.

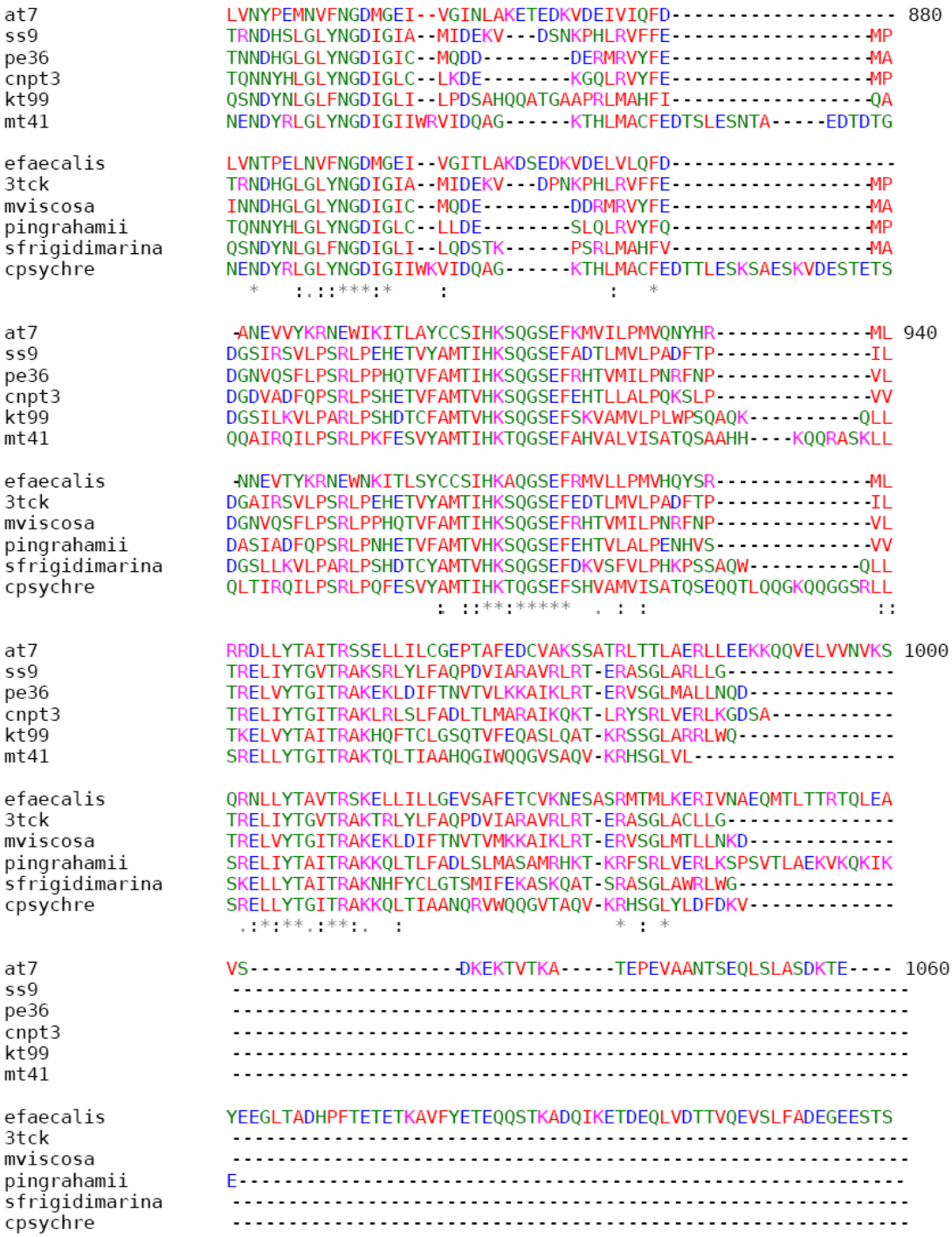

Figure 4.14, Continued: RecD protein alignment with colors depicting amino acid properties as described in Table 3. Proteins from deep-sea strains are grouped on top, listed with increasing piezophilicity. Comparison strains are grouped on the bottom. Symbols depict similarity between amino acids, "*" means identical amino acids, ":" are conserved substitutions, "." are semi-conserved substitutions. Regions with substitutions of interest are highlighted in yellow.

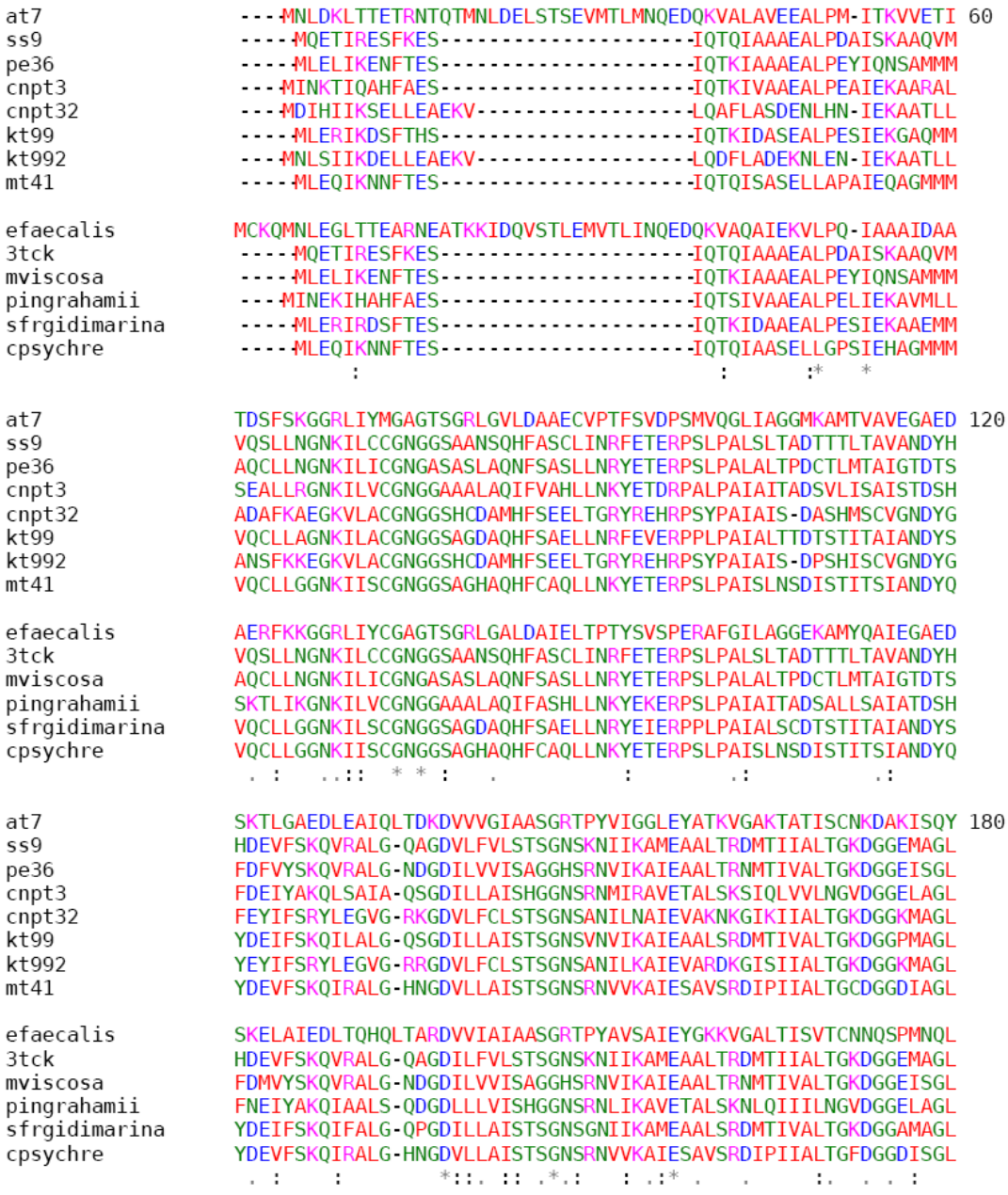

Figure 4.15: DiaA protein alignment with colors depicting amino acid properties as described in Table 3. Proteins from deep-sea strains are grouped on top, listed with increasing piezophilicity. Comparison strains are grouped on the bottom. Symbols depict similarity between amino acids, "*" means identical amino acids, ":" are conserved substitutions, "." are semi-conserved substitutions. Regions with substitutions of interest are highlighted in yellow.

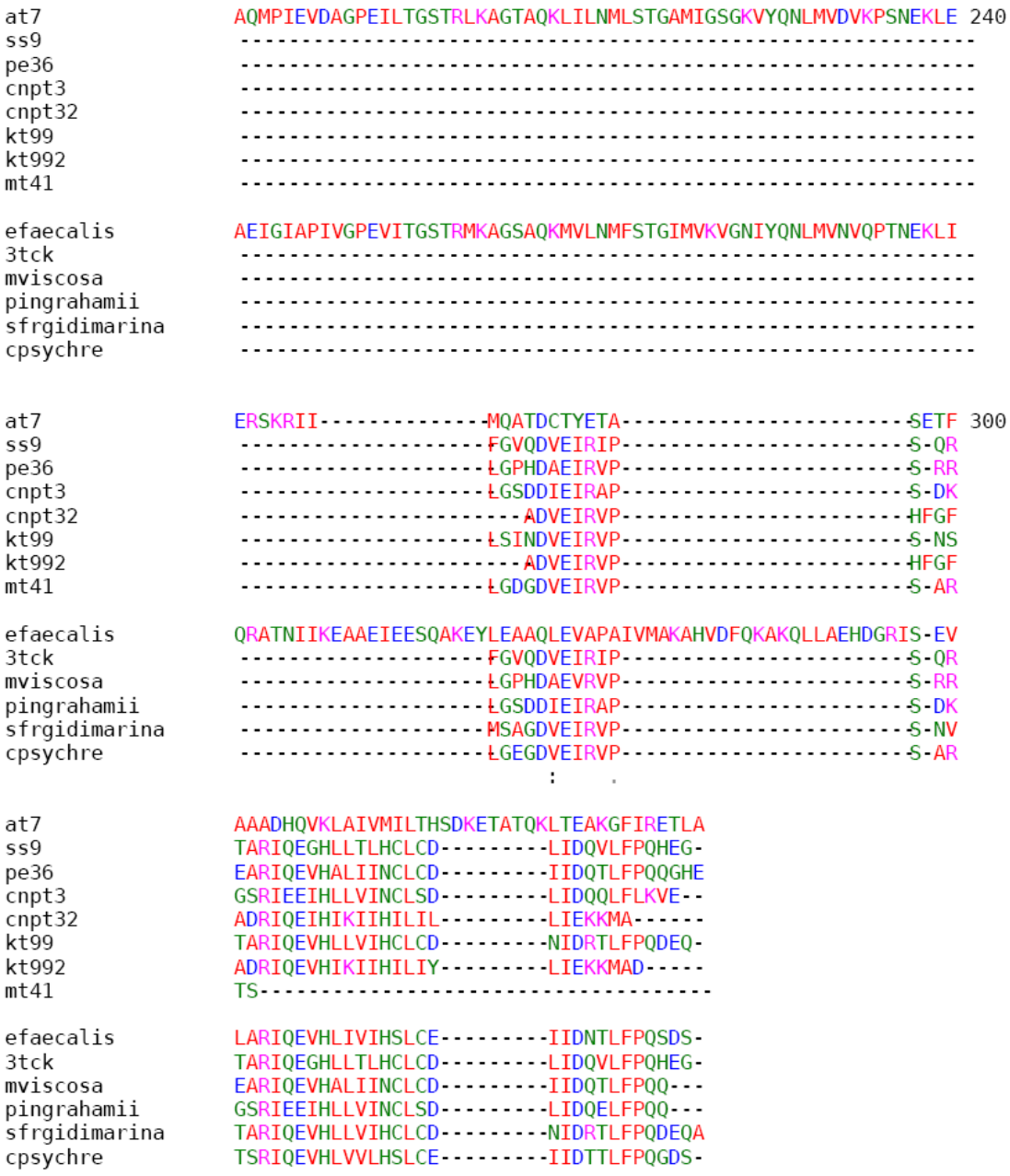

Figure 4.15, Continued: DiaA protein alignment with colors depicting amino acid properties as described in Table 3. Proteins from deep-sea strains are grouped on top, listed with increasing piezophilicity. Comparison strains are grouped on the bottom. Symbols depict similarity between amino acids, "*" means identical amino acids, ":" are conserved substitutions, "." are semi-conserved substitutions. Regions with substitutions of interest are highlighted in yellow.

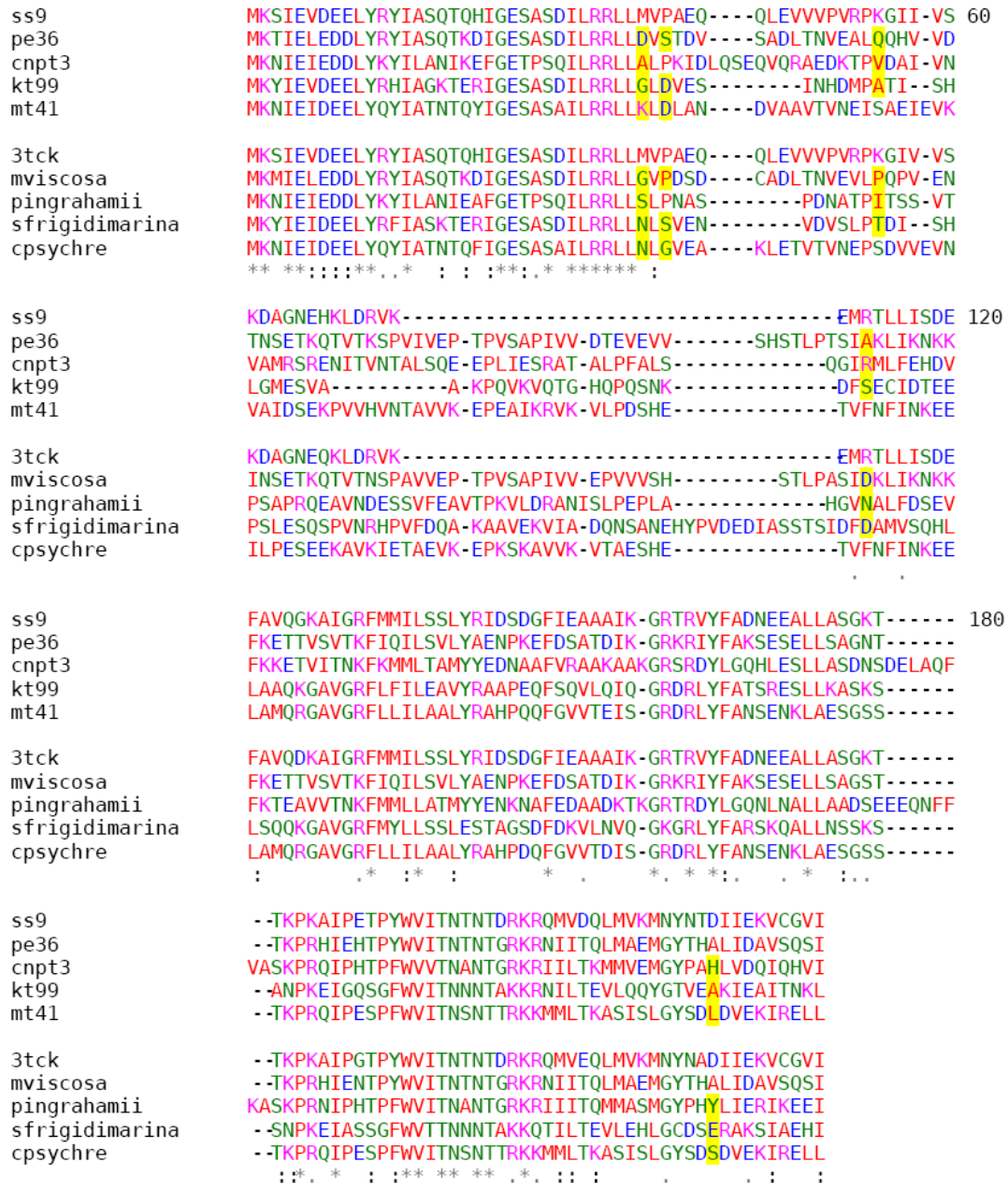

Figure 4.16: SeqA protein alignment with colors depicting amino acid properties as described in Table 3. Proteins from deep-sea strains are grouped on top, listed with increasing piezophilicity. Comparison strains are grouped on the bottom. Symbols depict similarity between amino acids, "*" means identical amino acids, ":" are conserved substitutions, "." are semi-conserved substitutions. Regions with substitutions of interest are highlighted in yellow.

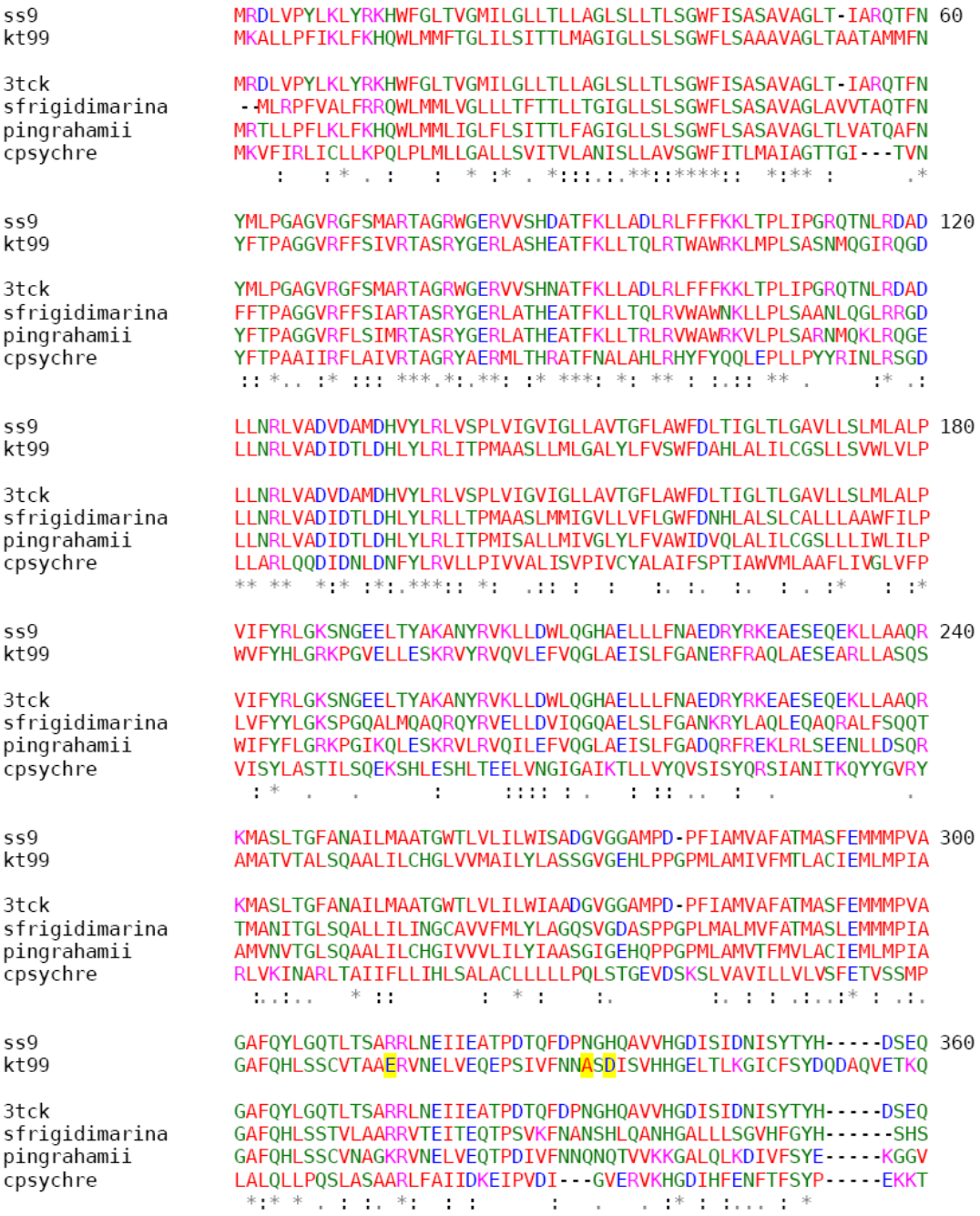

Figure 4.17: CydD protein alignment with colors depicting amino acid properties as described in Table 3. Proteins from deep-sea strains are grouped on top, listed with increasing piezophilicity. Comparison strains are grouped on the bottom. Symbols depict similarity between amino acids, “*” means identical amino acids, “:” are conserved substitutions, “.” are semi-conserved substitutions. Regions with substitutions of interest are highlighted in yellow.

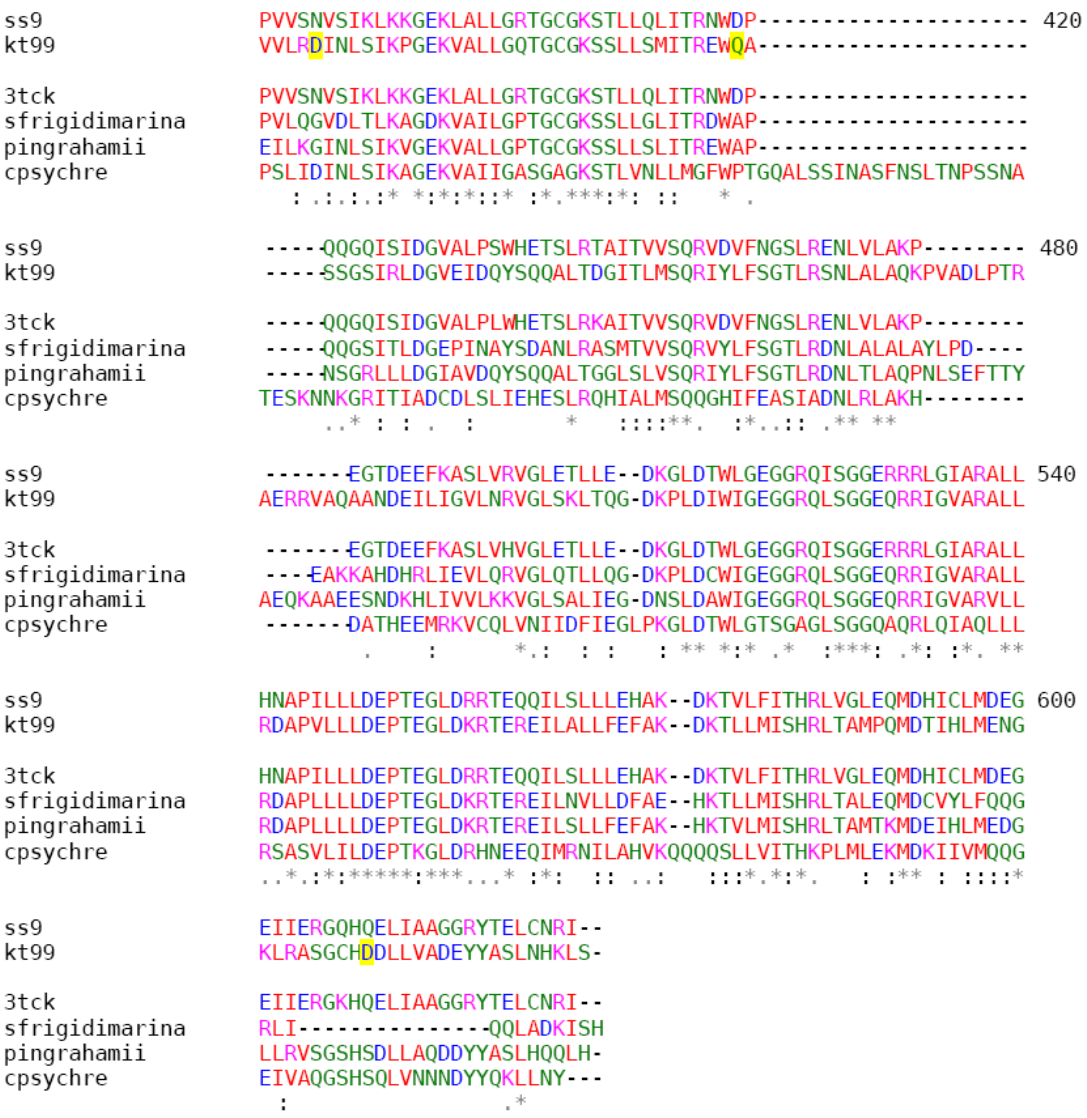

Figure 4.17, Continued: CydD protein alignment with colors depicting amino acid properties as described in Table 3. Proteins from deep-sea strains are grouped on top, listed with increasing piezophilicity. Comparison strains are grouped on the bottom. Symbols depict similarity between amino acids, "*" means identical amino acids, ":" are conserved substitutions, "." are semi-conserved substitutions. Regions with substitutions of interest are highlighted in yellow.

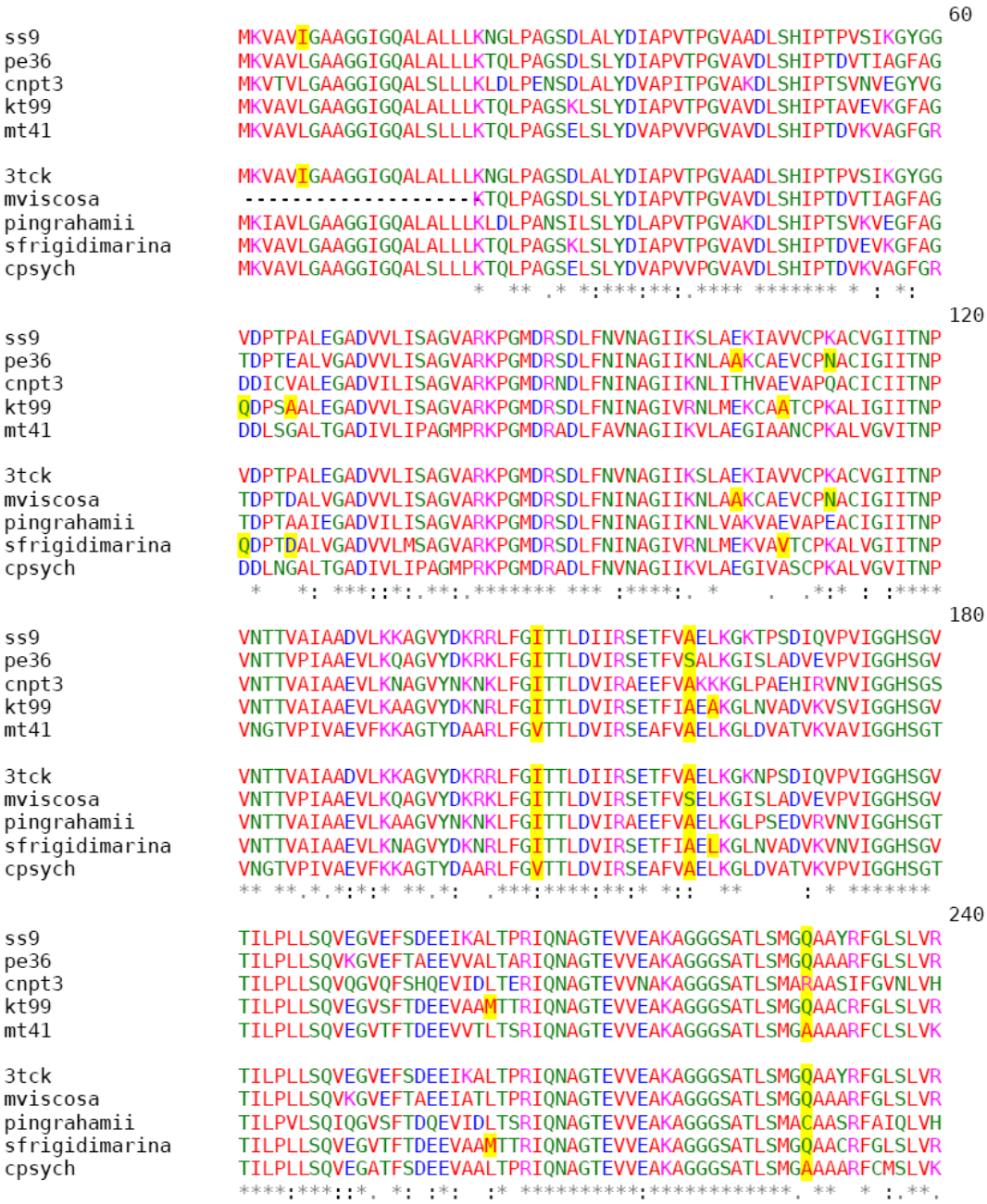

Figure 4.18: MDH protein alignment with colors depicting amino acid properties as described in Table 3. Proteins from deep-sea strains are grouped on top, listed with increasing piezophilicity. Comparison strains are grouped on the bottom. Symbols depict similarity between amino acids, "*" means identical amino acids, ":" are conserved substitutions, "." are semi-conserved substitutions. Regions with substitutions of interest are highlighted in yellow.

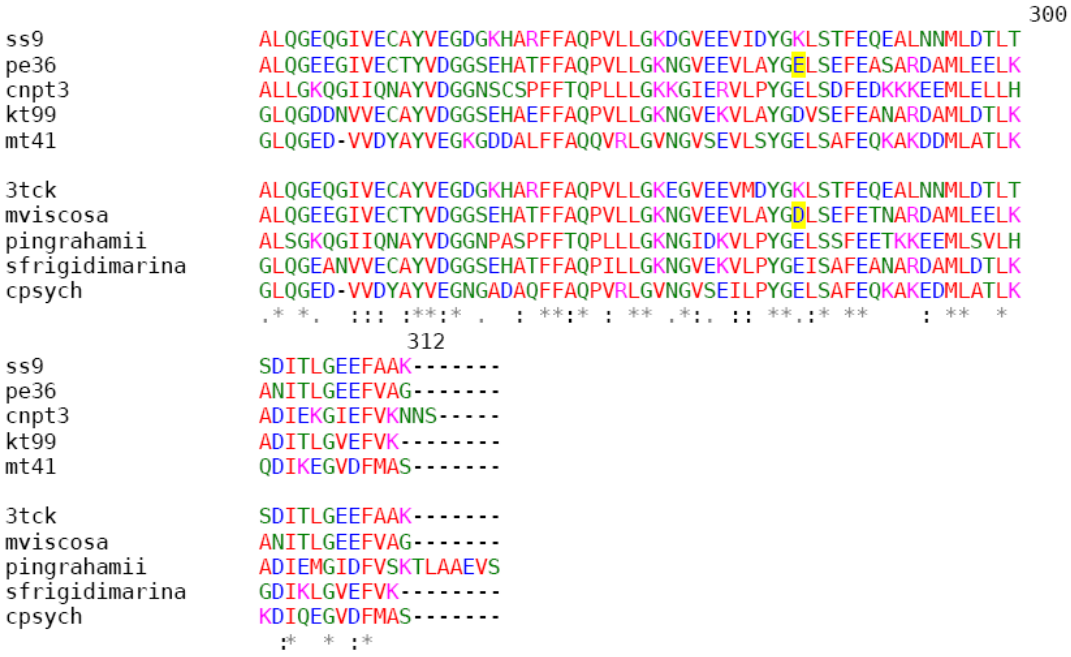

Figure 4.18, Continued: MDH protein alignment with colors depicting amino acid properties as described in Table 3. Proteins from deep-sea strains are grouped on top, listed with increasing piezophilicity. Comparison strains are grouped on the bottom. Symbols depict similarity between amino acids, "*" means identical amino acids, ":" are conserved substitutions, "." are semi-conserved substitutions. Regions with substitutions of interest are highlighted in yellow.

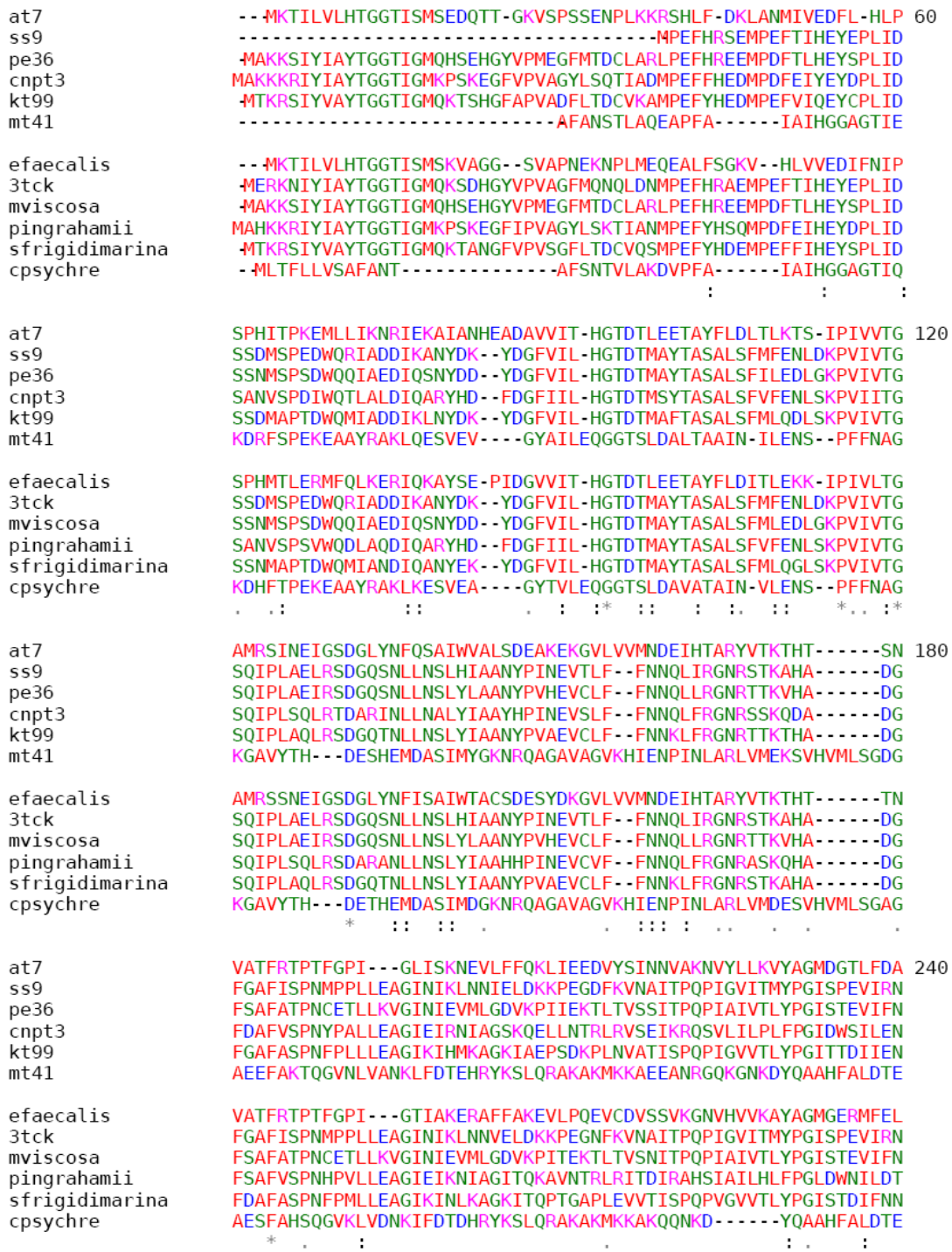

Figure 4.19: AsnA protein alignment with colors depicting amino acid properties as described in Table 3. Proteins from deep-sea strains are grouped on top, listed with increasing piezophilicity. Comparison strains are grouped on the bottom. Symbols depict similarity between amino acids, "*" means identical amino acids, ":" are conserved substitutions, "." are semi-conserved substitutions. Regions with substitutions of interest are highlighted in yellow.

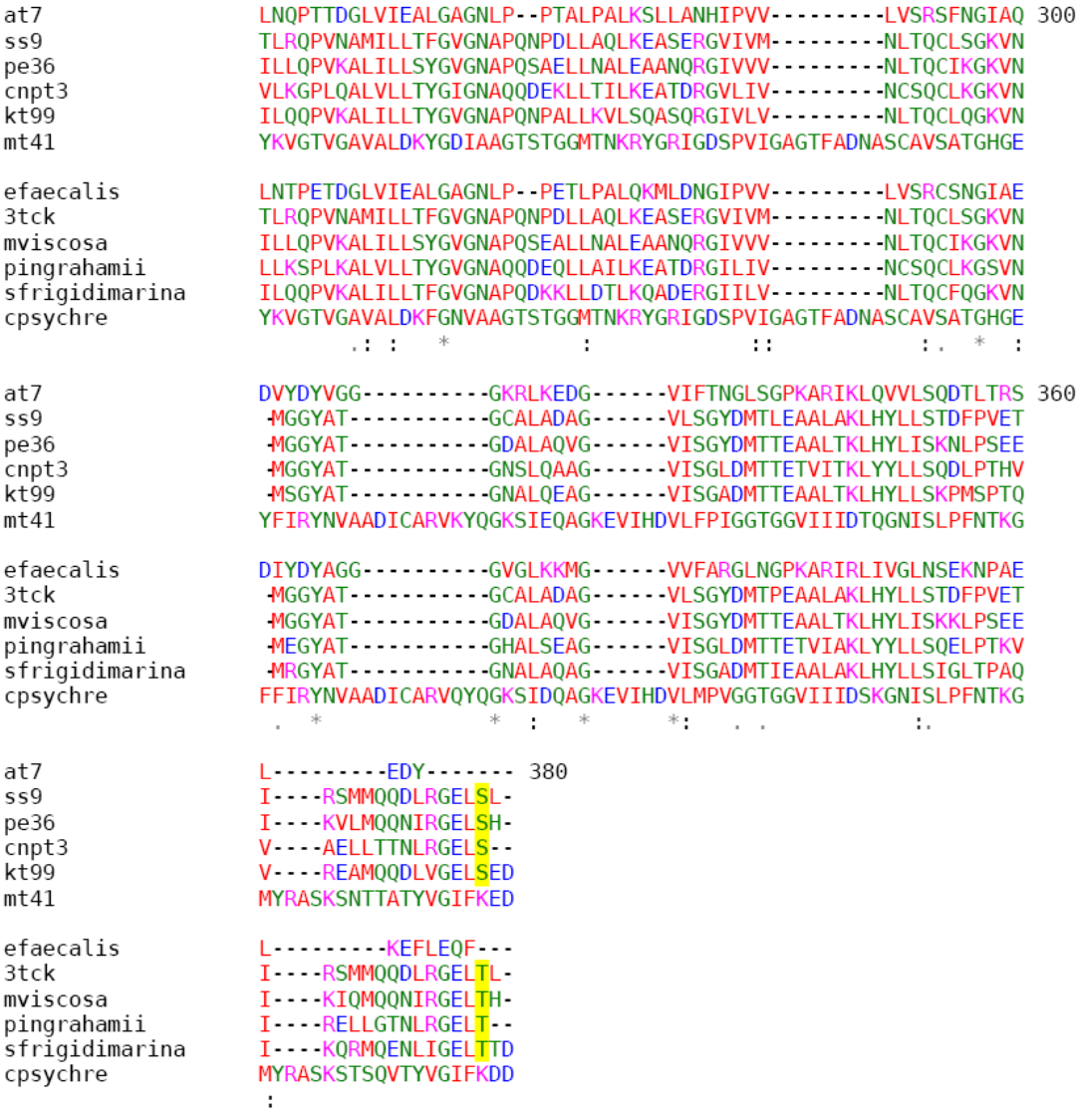

Figure 4.19, Continued: AsnA protein alignment with colors depicting amino acid properties as described in Table 3. Proteins from deep-sea strains are grouped on top, listed with increasing piezophilicity. Comparison strains are grouped on the bottom. Symbols depict similarity between amino acids, "*" means identical amino acids, ":" are conserved substitutions, "." are semi-conserved substitutions. Regions with substitutions of interest are highlighted in yellow.

Table 4.4: Amino acid composition (in percent) of each protein as found in the piezophilic (*piezo*) and non-piezophilic (*non*) strains.

	Asn		CydD		DiaA		DnaK		FabB		FabF		FtsZ		HNS	
	piezo	*non*	*piezo*	*non*	*piezo*	*non*	*piezo*	*non*	*piezo*	*non*	*piezo*	*non*	*piezo*	*non*	*piezo*	*non*
A = Ala	8.813	8.96	8.624	8.443	10.9	11.341	11.316	11.528	11.491	11.26	12.642	12.444	12.08	12.195	9.848	7.75
C = Cys	0.713	0.846	0.634	0.713	1.898	1.916	0.257	0.199	1.436	1.538	0.891	0.768	0.1	0.087	0	0
D = Asp	5.349	4.978	4.597	4.136	5.511	4.828	8.526	8.436	5.844	5.655	5.551	5.899	6.667	6.707	6.313	5.5
E = Glu	5.502	5.326	5.358	4.991	6.92	7.126	9.298	9.055	5.498	5.655	5.348	5.051	8.17	7.84	12.626	13.25
F = Phe	3.719	4.032	4.407	4.164	2.633	2.299	2.453	2.433	3.071	3.224	3.647	4.081	2.957	2.787	2.525	2.5
G = Gly	7.794	8.213	7.039	6.988	7.777	7.203	7.28	7.18	10.649	10.417	12.277	12.364	9.825	10.279	4.798	4.75
H = His	2.241	2.539	2.029	2.282	2.449	2.222	0.89	0.878	2.08	2.133	2.188	2.222	1.003	0.915	0.505	0.5
I = Ile	7.387	7.267	7.609	6.845	9.369	9.885	6.845	6.741	6.885	6.399	7.455	7.111	7.519	7.274	4.545	4.75
K = Lys	5.196	5.177	4.597	3.993	5.083	4.215	7.854	8.117	5.052	4.861	4.943	5.091	4.912	4.704	10.606	9.75
L = Leu	10.29	9.457	13.76	16.515	10.165	10.345	6.963	7.08	6.637	7.093	5.835	6.141	6.617	7.143	10.859	11.75
M = Met	2.955	3.335	2.79	2.653	2.817	2.759	2.255	2.134	3.616	3.77	3.89	3.556	3.258	3.223	3.03	3.25
N = Asn	5.298	5.426	3.932	3.423	4.287	4.598	3.66	3.929	3.863	4.563	4.335	4.485	4.311	4.007	3.03	3.5
P = Pro	4.636	4.579	2.949	3.423	2.572	3.065	3.205	3.051	2.922	2.927	3.404	3.475	3.86	3.963	3.535	3.5
Q = Gln	3.974	4.082	4.375	4.621	3.919	5.287	5.302	5.445	2.823	3.075	1.945	2.02	3.208	3.092	3.283	4.5
R = Arg	2.853	2.887	4.756	5.248	3.735	3.602	3.442	3.391	3.517	3.522	3.687	3.354	3.759	3.702	6.061	6
S = Ser	6.368	5.923	6.595	5.876	7.41	7.433	4.372	4.448	8.866	8.73	6.888	6.707	5.313	5.575	5.303	5.75
T = Thr	6.673	6.62	5.58	5.619	5.634	5.364	6.924	6.741	5.993	5.506	6.24	6.182	6.266	6.185	6.313	6.25
V = Val	6.266	6.62	6.848	6.218	5.144	5.057	7.794	7.858	6.686	6.746	5.916	6.343	9.373	9.451	4.293	4.5
W = Trp	0.306	0.299	0.919	1.312	0	0	0.138	0.14	0.495	0.446	0.527	0.525	0.1	0.087	0.758	0.75
Y = Tyr	3.668	3.435	2.6	2.539	1.776	1.456	1.227	1.217	2.576	2.48	2.391	2.182	0.702	0.784	1.768	1.5

Table 4.4, Continued: Amino acid composition (in percent) of each protein as found in the piezophilic (*piezo*) and non-piezophilic (*non*) strains. Results (p-values) of statistical analyses shown on the right.

	Mdh		**MreB**		**RecD**		**RpoE**		**RseB**		**SeqA**		**ToxR**			
	piezo	*non*	*piezo*	*non*	*piezo*	*non*	*piezo*	*non*	*piezo*	*non*	*piezo*	*non*	*piezo*	*non*	*T-test*	*Chi*
A = Ala	12.5	12.427	9.975	9.381	9.219	8.237	10.041	9.019	7.06	6.547	8.359	8.124	6.239	6.283	0.882	0.999
C = Cys	1.09	0.906	0.991	0.938	1.27	1.251	0.717	0.613	0.183	0.06	0.204	0.301	0.713	0.873	0.939	0.999
D = Asp	5.256	5.178	5.019	4.941	5.833	5.461	6.455	6.743	5.234	5.165	4.995	5.216	4.991	6.457	0.771	0.999
E = Glu	6.09	6.149	7.435	7.567	5.503	6.075	7.889	8.056	5.782	6.006	8.257	8.626	6.417	5.236	0.873	1
F = Phe	3.141	3.236	2.788	2.752	3.246	3.276	3.279	3.853	3.043	3.183	2.854	3.109	1.783	2.094	0.46	1
G = Gly	11.026	10.939	10.967	10.757	5.88	5.87	3.893	3.503	5.478	5.465	3.874	3.711	4.991	4.538	0.911	1
H = His	1.154	0.971	1.177	1.126	2.728	2.662	0.512	0.788	1.704	1.742	1.937	1.505	3.565	3.316	0.648	0.999
I = Ile	6.474	6.537	8.178	8.193	5.786	5.666	7.07	7.18	7.425	7.327	8.257	7.623	7.487	6.457	0.704	1
K = Lys	5.385	5.243	3.47	3.377	5.503	5.324	5.02	6.13	5.234	5.105	7.136	6.82	4.813	5.236	0.889	0.999
L = Leu	10.385	10.032	7.745	8.005	12.747	12.537	9.529	9.282	10.469	10.811	8.461	8.225	12.478	12.391	0.941	0.999
M = Met	1.538	1.748	3.841	3.627	1.999	2.116	2.254	2.364	2.191	2.342	2.752	2.608	1.783	1.920	0.73	1
N = Asn	3.397	3.43	3.47	3.752	3.975	4.3	4.611	4.729	3.956	5.105	4.383	5.416	5.526	5.410	0.441	0.999
P = Pro	3.846	4.401	4.213	4.315	4.045	3.709	3.791	3.415	4.626	4.505	3.568	4.413	3.922	4.188	0.672	0.999
Q = Gln	2.821	2.718	2.54	2.376	6.256	6.849	4.918	4.466	5.66	5.646	4.383	3.31	4.1	4.363	0.778	0.999
R = Arg	2.628	2.589	6.01	6.316	4.563	4.323	7.992	7.793	5.113	4.625	5.505	4.915	4.278	4.188	0.885	1
S = Ser	5.128	5.178	6.32	6.817	6.632	6.644	6.455	6.305	8.278	7.988	6.932	8.726	8.913	8.551	0.881	1
T = Thr	5.641	6.019	4.523	4.19	5.198	5.575	3.893	4.553	3.956	4.565	6.83	6.419	8.2	7.155	0.673	0.999
V = Val	10.962	10.809	9.418	9.506	5.95	6.394	7.07	6.655	9.191	8.589	7.849	7.523	6.952	8.551	0.989	0.999
W = Trp	0	0	0	0	0.8	0.842	0.512	0.701	1.522	1.441	0.51	0.502	0.891	0.873	0.978	1
Y = Tyr	1.538	1.489	1.921	2.064	2.869	2.89	4.098	3.853	3.895	3.784	2.956	2.909	1.961	1.920	0.851	1

Table 4.5: Comparison of amino acid residue usages by percentage between piezophiles (*piezo*) and the non-piezophile, shallow water organisms (*non*). Residues are divided in to categories according to EMBOSS Pepstats (Rice et al, 2000) as follows: Tiny (A+C+G+S+T), Small (A+B+C+D+G+N+P+S+T+V), Aliphatic (I+L+V), Aromatic (F+H+W+Y), Non-polar (A+C+F+G+I+L+M+P+V+W+Y), Polar (D+E+H+K+N+Q+R+S+T+Z), Charged (B+D+E+H+K+R+Z), Basic (H+K+R), Acidic (B+D+E+Z). P values for the T-test and Chi-squared tests also shown

	AsnA		**CydD**		**DiaA**		**DnaK**		**FabB**		**FabF**		**FtsZ**		**HNS**	
	piezo	*non*	*piezo*	*non*	*piezo*	*non*	*piezo*	*non*	*piezo*	*non*	*piezo*	*non*	*piezo*	*non*	*piezo*	*non*
Tiny	30.362	30.562	28.472	27.638	33.619	33.257	30.148	30.096	38.435	37.45	38.938	38.465	33.584	34.321	26.263	24.5
Small	51.91	52.165	46.798	44.837	51.133	50.805	53.333	53.371	57.751	57.341	58.144	58.667	57.794	58.449	43.434	41.5
Aliphatic	23.943	23.345	28.218	29.578	24.679	25.287	21.602	21.679	20.208	20.238	19.206	19.596	23.509	23.868	19.697	21
Aromatic	9.934	10.304	9.956	10.297	6.859	5.977	4.708	4.667	8.222	8.284	8.752	9.01	4.762	4.573	5.556	5.25
Non-polar	56.546	57.043	58.18	59.812	55.052	55.326	49.733	49.561	56.464	56.3	58.874	58.99	56.391	57.274	45.96	45
Polar	43.454	42.957	41.82	40.188	44.948	44.674	50.267	50.439	43.536	43.7	41.126	41.01	43.609	42.726	54.04	**55**
Charged	21.141	20.906	21.338	20.65	23.699	21.992	30.01	29.876	21.991	21.825	21.718	21.616	24.511	23.868	36.111	35
Basic	10.29	10.602	11.382	11.523	11.268	10.038	12.186	12.385	10.649	10.516	10.818	10.667	9.674	9.321	17.172	16.25
Acidic	10.851	10.304	9.956	9.127	12.431	11.954	17.824	17.491	11.342	11.31	10.9	10.949	14.837	14.547	18.939	18.75

	Mdh		**MreB**		**RecD**		**RpoE**		**RseB**		**SeqA**		**ToxR**			
	piezo	*non*	*piezo*	*non*	*piezo*	*non*	*piezo*	*non*	*piezo*	*non*	*piezo*	*non*	*piezo*	*non*	Ttest	Chi
Tiny	35.385	35.469	32.776	32.083	28.198	27.577	25	23.993	24.954	24.625	26.198	27.282	29.055	27.4	0.795	0.999
Small	58.846	59.288	54.895	54.597	48.001	47.44	46.926	45.534	47.961	47.988	46.993	49.85	50.446	52.007	0.985	0.999
Aliphatic	27.821	27.379	25.341	25.704	24.483	24.596	23.668	23.117	27.085	26.727	24.567	23.37	26.916	27.4	0.903	1
Aromatic	5.833	5.696	5.886	5.941	9.643	9.67	8.402	9.194	10.164	10.15	8.257	8.024	8.2	8.202	0.993	1
Non-polar	62.5	62.524	60.037	59.537	53.81	52.787	52.254	50.438	55.082	54.054	49.643	49.047	49.198	50.087	0.941	1
Polar	37.5	37.476	39.963	40.463	46.19	47.213	47.746	49.562	44.918	45.946	50.357	50.953	50.802	49.913	0.941	1
Charged	20.513	20.129	23.11	23.327	24.13	23.845	27.869	29.51	23.068	22.643	27.829	27.081	24.064	24.433	0.849	0.999
Basic	9.167	8.803	10.657	10.819	12.794	12.309	13.525	14.711	12.051	11.471	14.577	13.24	12.656	12.74	0.756	0.999
Acidic	11.346	11.327	12.454	12.508	11.336	11.536	14.344	14.799	11.016	11.171	13.252	13.842	11.408	11.693	0.949	1

Table 4.6: Lists the length (in number of amino acids), molecular weight, and isoelectric point (pI) for each of the amino acid sequences investigated.

		length	*molecular weight*	*pI*
AsnA	AT7	320	34977.1	6.24
	E faecalis	322	35045.3	5.43
	SS9	299	32865.4	4.49
	3TCK	336	36963.1	4.65
	PE36	336	36701.9	4.7
	M viscosa	336	36792	4.79
	CNPT3	336	37133.5	4.9
	P ingrahamii	336	36994.4	6.65
	KT99	337	36573	4.96
	S frigidimarina	337	36612.1	5.65
	MT41	335	35716.1	7.32
	C psychrerithreae	342	36350.98	8.23
CydD	AT7	571	64380.2	5.2
	E faecalis	588	66532.9	5.88
	SS9	573	63314.3	6.71
	3TCK	573	63331.5	7.05
	PE36	587	65370.5	7.43
	M viscosa	587	65393.7	7.35
	CNPT3	583	63541.7	9.23
	P ingrahamii	592	65779.2	9.04
	KT99	597	65697.7	7.4
	S frigidimarina	567	62297.7	8.88
	MT41	243	26397.9	6.63
	C psychrerithreae	599	66329.8	8.97
DiaA	AT7	296	31569.1	4.62
	E faecalis	300	32157.9	5.09
	SS9	196	21065.9	5.23
	3TCK	196	21079.9	5.23
	PE36	197	21212.3	4.82
	M viscosa	194	20859	4.8
	CNPT3 (a)	196	20759.8	4.85
	CNPT3 (b)	191	20868	6.6
	P ingrahamii	195	20793.9	4.86
	KT99 (a)	196	20981.8	4.42
	KT99 (b)	192	21199.2	6.16
	S frigidimarina	197	21030.9	4.25
	MT41	196	20909.6	4.54
	C psychrerithreae	196	20921.6	4.67
DnaK	AT7	609	65499.3	4.3
	E faecalis	609	65583.5	4.33
	SS9 (a)	639	69358.9	4.45
	SS9 (b)	641	69250.6	4.41
	3TCK (a)	639	69322.9	4.45
	3TCK (b)	575	62030.5	4.38
	PE36	641	69359.8	4.29
	M viscosa	642	69112.5	4.3
	CNPT3	635	68683.9	4.4
	P ingrahamii	640	69615.2	4.48
	KT99	640	69106.5	4.39
	S frigidimarina	638	68878.2	4.45
	MT41 (a)	638	68776.9	4.23
	MT41 (b)	612	66102.2	4.32
	C psychrerithreae (a)	633	68178.5	4.56
	C psychrerithreae (b)	638	68765.9	4.27

Table 4.6, Continued: Lists the length (in number of amino acids), molecular weight, and isoelectric point (pI) for each of the amino acid sequences investigated.

FabB	SS9	404	42670.2	4.92
	3TCK	404	42798.3	4.72
	PE36	405	42488.7	4.77
	M viscosa	405	42502.7	4.65
	CNPT3	405	42870.7	5.89
	P ingrahamii	404	42901.6	5.15
	KT99	402	42378.6	5.21
	S frigidimarina	400	42563.1	6.62
	MT41	403	42343	5.42
	C psychrerthreae	403	42376.8	5.02
FabF	AT7	413	42941.3	4.76
	E faecalis	411	43069.8	4.94
	SS9	413	42942.5	5.4
	3TCK	413	43016.7	5.4
	PE36	414	42774.1	5.4
	M viscosa	414	42726.9	5.23
	CNPT3	413	43310.1	6.35
	P ingrahamii	413	43546.2	5.48
	KT99	403	42894.6	5.29
	S frigidimarina	412	43035.6	6
	MT41	412	43412.3	6.61
	C psychrerthreae	412	43290.9	5.62
FtsZ	AT7	418	44979.2	4.47
	E faecalis	410	44147.4	4.49
	SS9	394	40983.6	4.33
	3TCK	394	40946.6	4.34
	M viscosa	330	33741.2	4.01
	CNPT3	409	43359	4.26
	P ingrahamii	388	40942.1	4.3
	KT99	388	40469.9	4.36
	S frigidimarina	388	40457	4.42
	MT41	386	39863.2	4.04
	C psychrerthreae	386	39749	4.02
H-NS	SS9	131	14893.8	4.83
	3TCK	131	14879.8	4.83
	PE36	135	15377.6	5.39
	M viscosa	133	15318.6	4.89
	KT99	130	14463.4	5.08
	S frigidimarina	136	15301.4	4.89
Mdh	SS9	312	32293.1	4.71
	3TCK	312	32338.2	4.72
	PE36	312	31926.6	4.34
	M viscosa	292	30207.4	4.25
	CNPT3	314	33164.2	5.63
	P ingrahamii	319	32997	4.69
	KT99	311	31951.7	4.62
	S frigidimarina	311	32143.1	4.87
	MT41	311	31887.7	4.48
	C psychrerthreae	311	31876.7	4.47

Table 4.6, Continued: Lists the length (in number of amino acids), molecular weight, and isoelectric point (pI) for each of the amino acid sequences investigated.

MreB	SS9	347	36961.2	4.65
	3TCK	347	369471	4.65
	PE36	224	23642.1	5.73
	M viscosa	209	22209.6	9.14
	CNPT3	347	36916.6	4.73
	P ingrahamii	347	37121.7	4.62
	KT99	349	37210.5	4.77
	S frigidimarina	349	37336.6	4.77
	MT41	347	36776.2	4.83
	C psychrerithreae	347	36904.3	4.76
RecD	A17	834	93222.2	4.7
	E faecalis	865	97476.2	4.52
	SS9	701	77416.4	6.62
	3TCK	701	77251.3	6.44
	PE36	675	75266.1	7.11
	M viscosa	675	75198.1	7.21
	CNPT3	683	76153.7	7.55
	P ingrahamii	698	78023.2	7.48
	KT99	673	74493.9	6.74
	S frigidimarina	687	76412.2	6.63
	MT41	682	75127.5	7.25
	C psychrerithreae	769	85229.2	6.52
RpoE	E faecalis	165	19696.8	8.79
	SS9	192	21857.8	4.98
	3TCK	192	21847.8	4.98
	PE36	192	21524.5	4.92
	M viscosa	192	21510.4	4.92
	CNPT3	192	21884.9	5.4
	P ingrahamii	193	22033	5.4
	KT99	192	21867.8	4.64
	S frigidimarina	192	22030	4.95
	MT41	208	23584	8.64
	C psychrerithreae	208	23653.9	8.64
RseB	SS9	331	37068.4	6.41
	3TCK	332	37079.3	5.89
	PE36	347	38189.7	5.01
	M viscosa	346	38321.9	6.82
	CNPT3	332	37903.9	9.71
	P ingrahamii	335	38361.9	6.54
	KT99	310	34795	5.22
	S frigidimarina	315	35208.6	5.62
	MT41	323	36514.4	6.64
	C psychrerithreae	337	38176.2	5.71
SeqA	SS9	178	20178.3	7.51
	3TCK	178	20125.3	7.35
	PE36	206	22964	6.37
	M viscosa	204	22621.8	5.77
	CNPT3	213	24276	7.87
	P ingrahamii	207	23299.4	4.95
	KT99	184	20577.3	6.96
	S frigidimarina	208	23016.7	5.33
	MT41	200	22434.7	6.29
	C psychrerithreae	200	22379.4	5.01
ToxR	SS9	290	32175.2	6.08
	3TCK	291	32159.1	5.33
	CNPT3	271	29856.4	6.52
	P ingrahamii	282	31263.2	7.27
	V cholerae	294	32522.8	5.2

Table 4.7: Percentage of predicted secondary structure conformations for each protein. Piezophiles are shown on top and non-piezophiles on the bottom. Statistical test results (p-values) of the differences between the two groups are shown on the right.

Asn	*At7*	*SS9*	*PE36*	*CNPT3*	*KT99*	*MT41*	TTEST	CHI
Coils	48.13%	48.83	49.7	47.32	50.45	46.27	0.398	0.998
Sheets	21.88%	21.07	20.54	22.02	20.47	20.3	0.903	0.999
Helices	30.00%	30.1	29.76	30.65	29.08	33.43	0.401	0.993

	E. faecalis	*3TCK*	*M. viscosa*	*P. ingrahamii*	*S. frigid*	*C. psychre*
Coils	46.27	47.62	49.4	47.62	49.26	43.57
Sheets	22.05	22.32	20.54	22.02	20.18	19.59
Helices	31.68	30.06	30.06	30.36	30.56	36.84

CydD	*At7*	*SS9*	*PE36*	*CNPT3*	*KT99*	*MT41*	TTEST	CHI
Coils	23.47	23.73	24.53	24.7	23.28	25	0.906	0.998
Sheets	9.81	8.9	9.03	9.26	9.21	11.33	0.292	0.97
Helices	66.73	67.36	66.44	66.04	67.5	63.67	0.599	0.999

	E. faecalis	3TCK	M. viscosa	P. ingrahamii	S. frigid	C. psychre
Coils	23.98	24.08	24.7	22.47	23.99	25.88
Sheets	8.67	8.9	9.37	9.12	9.88	8.85
Helices	67.35	67.02	65.93	68.41	66.14	65.28

DiaA	*At7*	*SS9*	*PE36*	*CNPT3a*	*CNPT3b*	*KT99a*	*KT99b*	*MT41*	TTEST	CHI
Coils	30.74	31.12	31.47	29.59	27.75	31.12	27.6	33.16	0.253	0.999
Sheets	6.76	14.29	13.71	14.29	14.66	14.29	14.58	14.29	0.838	0.999
Helices	62.5	54.59	54.82	56.12	57.59	54.59	57.81	52.55	0.594	0.999

	E. faecalis	*3TCK*	*M. viscosa*	*P. ingrahamii*	*S. frigid*	*C. psychre*
Coils	30	31.12	30.41	29.74	33.5	33.16
Sheets	8.33	14.29	13.92	13.33	13.71	14.29
Helices	61.67	54.59	55.67	56.92	52.79	52.55

DnaK	*At7*	*SS9a*	*SS9b*	*PE36*	*CNPT3*	*KT99*	*MT41a*	*MT41b*	TTEST	CHI
Coils	38.42	38.97	39.16	39.31	38.58	39.22	38.71	36.44	0.242	0.999
Sheets	22.17	22.69	22.62	22.46	22.99	22.81	22.88	23.04	0.687	1
Helices	39.41	38.34	38.22	38.22	38.43	37.97	38.4	40.52	0.223	0.999

	E. faecalis	*3TCKa*	*3TCKb*	*M. viscosa*	*S. frigid*	*S. frigid*	*C. psychrea*	*C. psychreb*
Coils	38.59	38.97	39.13	39.1	39.06	39.18	39.02	39.03
Sheets	21.67	22.69	23.3	23.21	23.13	22.73	23.06	22.57
Helices	39.74	38.34	37.57	37.69	37.81	38.09	37.91	38.4

FabB	*At7*	*SS9*	*PE36*	*CNPT3*	*KT99*	*MT41*	TTEST	CHI
Coils	x	44.8	45.19	46.17	48.76	45.16	0.762	1
Sheets	x	16.58	16.3	16.05	14.68	16.87	0.891	1
Helices	x	38.61	38.52	37.78	36.57	37.97	0.331	0.999

	E. faecalis	*3TCK*	*M. viscosa*	*P. ingrahamii*	*S. frigid*	*C. psychre*
Coils	x	43.56	45.68	45.79	48.25	45.16
Sheets	x	17.57	16.3	16.34	12.75	16.87
Helices	x	38.86	38.02	37.87	39	37.97

Table 4.7, Continued: Percentage of predicted secondary structure conformations for each protein. Piezophiles are shown on top and non-piezophiles on the bottom. Statistical test results (p-values) of the differences between the two groups are shown on the right.

FabF	*At7*	*SS9*	*PE36*	*CNPT3*	*KT99*	*MT41*	TTEST	CHI
Coils	47.94	46.36	46.86	46.25	44.91	46.12	0.451	0.999
Sheets	15.25	15.78	15.46	16.22	16.63	16.02	0.125	1
Helices	36.8	37.86	37.68	37.53	38.46	37.86	0.996	1

	E. faecalis	*3TCK*	*M. viscosa*	*P. ingrahamii*	*frigidimari*	*C. psychre*
Coils	47.45	46.49	46.86	46.97	47.33	45.63
Sheets	15.33	15.5	15.22	15.74	15.53	15.78
Helices	37.23	38.01	37.92	37.29	37.14	38.59

FtsZ	*At7*	*SS9*	*PE36*	*CNPT3*	*KT99*	*MT41*	TTEST	CHI
Coils	51.91	48.48	x	45.23	47.94	47.15	0.426	0.999
Sheets	16.99	18.02	x	17.11	18.04	18.13	0.476	1
Helices	31.1	33.5	x	37.65	34.02	34.72	0.487	1

	E. faecalis	*3TCK*	*M. viscosa*	*P. ingrahamii*	*S. frigid*	*C. psychre*
Coils	49.76	49.75	40	44.07	48.2	47.15
Sheets	17.8	17.26	20.91	17.27	17.78	17.88
Helices	32.44	32.99	39.09	38.66	34.02	34.97

H-NS	*At7*	*SS9*	*PE36*	*CNPT3*	*KT99*	*MT41*	TTEST	CHI
Coils	x	35.88	37.04	x	36.15	x	0.407	0.999
Sheets	x	0	2.22	x	0	x	0.445	0.998
Helices	x	64.12	60.74	x	63.85	x	0.919	0.999

	E. faecalis	*3TCK*	*M. viscosa*	*P. ingrahamii*	*S. frigid*	*C. psychre*
Coils	x	35.88	33.83	x	36.76	x
Sheets	x	0	3.76	x	1.47	x
Helices	x	64.12	62.41	x	61.76	x

Mdh	*At7*	*SS9*	*PE36*	*CNPT3*	*KT99*	*MT41*	TTEST	CHI
Coils	x	34.94	34.94	34.08	33.76	32.8	0.672	1
Sheets	x	19.87	19.87	20.06	20.26	21.54	0.127	0.999
Helices	x	45.19	45.19	45.86	45.98	45.66	0.258	1

	E. faecalis	*3TCK*	*M. viscosa*	*P. ingrahamii*	*S. frigid*	*C. psychre*
Coils	x	33.97	36.3	34.17	34.41	33.12
Sheets	x	19.55	18.84	19.44	19.61	20.58
Helices	x	46.47	44.86	46.39	45.98	46.3

MreB	*At7*	*SS9*	*PE36*	*CNPT3*	*KT99*	*MT41*	TTEST	CHI
Coils	x	38.62	38.39	37.75	37.82	40.42	0.124	0.999
Sheets	x	20.17	27.68	20.46	21.2	20.21	0.936	0.999
Helices	x	41.21	33.93	41.79	40.97	39.37	0.557	0.999

	E. faecalis	*3TCK*	*M. viscosa*	*P. ingrahamii*	*S. frigid*	*C. psychre*
Coils	x	38.62	37.8	35.45	37.54	37.46
Sheets	x	20.17	28.23	20.46	19.2	20.75
Helices	x	41.21	33.97	44.09	43.27	41.79

Table 4.7, Continued: Percentage of predicted secondary structure conformations for each protein. Piezophiles are shown on top and non-piezophiles on the bottom. Statistical test results (p-values) of the differences between the two groups are shown on the right.

RecD	*At7*	*SS9*	*PE36*	*CNPT3*	*KT99*	*MT41*	TTEST	CHI
Coils	40.17	38.66	33.04	38.28	37.59	37.24	0.658	0.999
Sheets	15.47	10.13	10.67	11.21	11.14	14.81	0.79	0.999
Helices	44.36	51.21	56.3	50.51	51.26	47.95	0.829	0.999

	E. faecalis	*3TCK*	*M. viscosa*	*P. ingrahamii*	*S. frigid*	*C. psychre*
Coils	43.47	37.09	32.15	35.53	34.79	36.93
Sheets	14.34	11.13	12.59	12.46	11.5	13.13
Helices	42.2	51.78	55.26	52.01	53.71	49.93

RpoE	*At7*	*SS9*	*PE36*	*CNPT3*	*KT99*	*MT41*	TTEST	CHI
Coils	x	28.65	27.6	29.69	29.17	32.21	0.999	0.998
Sheets	x	0	0	0	0	0		
Helices	x	71.35	72.4	70.31	70.83	67.79	0.999	9.99E-01

	E. faecalis	*3TCK*	*M. viscosa*	*P. ingrahamii*	*S. frigid*	*C. psychre*
Coils	28.48	28.8	28.65	30.05	27.6	32.21
Sheets	0	0	0	0	0	0
Helices	71.52	71.2	71.35	69.95	72.4	67.79

RseB	*At7*	*SS9*	*PE36*	*CNPT3*	*KT99*	*MT41*	TTEST	CHI
Coils	x	44.41	44.09	39.16	40.32	47.06	0.45	0.999
Sheets	x	38.37	38.9	39.46	40.32	40.56	0.133	0.999
Helices	x	17.22	17	21.39	19.35	12.38	0.182	0.994

	E. faecalis	*3TCK*	*M. viscosa*	*P. ingrahamii*	*S. frigid*	*C. psychre*
Coils	x	43.67	42.77	38.81	38.41	43.92
Sheets	x	38.25	37.57	39.4	39.68	37.69
Helices	x	18.07	19.65	21.79	21.9	18.4

SeqA	*At7*	*SS9*	*PE36*	*CNPT3*	*KT99*	*MT41*	TTEST	CHI
Coils	x	44.97	50	50.7	44.57	46	0.645	0.999
Sheets	x	0	5.34	5.16	5.98	7.5	0.779	0.927
Helices	x	55.03	44.66	44.13	49.46	46.5	0.843	1

	E. faecalis	*3TCK*	*M. viscosa*	*P. ingrahamii*	*S. frigid*	*C. psychre*
Coils	x	41.01	48.53	49.28	45.19	47.5
Sheets	x	3.93	6.37	5.31	4.81	5.5
Helices	x	55.06	45.1	45.41	50	47

ToxR	*At7*	*SS9*	*PE36*	*CNPT3*	*KT99*	*MT41*	TTEST	CHI
Coils	x	52.41	x	50.18	x	x	0.803	0.999
Sheets	x	22.76	x	17.71	x	x	0.999	0.999
Helices	x	24.83	x	32.1	x	x	0.878	0.999

	E. faecalis	*3TCK*	*M. viscosa*	*P. ingrahamii*	*S. frigid*	*C. psychre*
Coils	x	55.67	x	48.94	x	x
Sheets	x	21.31	x	19.15	x	x
Helices	x	23.02	x	31.91	x	x

Appendix A

MT41 RNA extractions

RNA was extracted from the *Colwellia* sp. MT41 for the purpose of microarray analysis to examine the differential gene expression at optimal and suboptimal pressures. MT41 is the most obligate piezophile in isolation. It was isolated from a sample taken at 10,476m depth in the Mariana Trench in December of 1978 (Yayanos *et al*, 1981). The unique growth characteristics of MT41 make it a difficult strain to grow and work with. As an obligate piezophile from the hadal zone, MT41 exhibits optimal growth at 14,500 psi and demonstrates no growth at pressures less than 7,500 psi. Previous work with MT41 has shown that the entire culture dies within 6 hours of decompression. MT41 grows to low cell densities (growth ceases at $2x10^7$ cells/ml) (Yayanos et al, 1981) and, therefore, must be grown in large batches. Similarly, special pressure vessels must be used that are capable of maintaining large volumes at high pressure. Lastly, as is characteristic of deep sea bacteria, MT41 grows slowly, with a generation time between 25 and 33 hours at temperatures and pressures found in its natural environment (Yayanos et al, 1981), meaning that cultures took several weeks to reach optimum cell densities and exponential growth phase.

Culturing of MT41 followed protocols previously established (Yayanos *et al*, 1981). The bacteria were grown in marine broth 2216 (Difco) at 2ºC. Initial batch cultures were grown for approximately four weeks at 13,500 psi before being transferred in to fresh media and housed in 4 two-liter sealed bags. A total of four liters of media were grown for approximately three weeks at the two tested pressures: high pressure at 13,500 psi) and low pressure at 9,300 psi). At the end of three weeks, the vessels were

decompressed and kept on ice until the cells were harvested. Since the organism is pressure sensitive and RNA is easily degraded, all procedures were done as quickly as possible to ensure the quality of the results and precautions were taken to reduce exposure to RNases.

At higher cell densities, MT41 forms flaky aggregations, that are visible floating in the media column. These aggregations are not conducive to centrifugation; they remain suspended in the media rather than being "spun down". Cell harvesting, then, had to be done via filtration. The protocol for cell harvesting went through several trials to get adequate quantity and quality of RNA. In total, three biological replicates were produced at both high and low pressures.

In the initial trial, the liquid cultures were filtered through pre-sterilized Millipore nitrocellulose .22 µm filters. Each filter corresponded to approximately 250-300 ml of the culture. Filtration of four liters took approximately 5.5 hours. Later trials followed the same steps for RNA isolation and purification, differing only in filtration methods. Alteration was required because the resulting RNA in the initial trial was highly degraded due to excessively long extraction times. The later trials involved changes in the filters used (to reduce filtration times) and the addition of RNAlater (Ambion) to the pre-filtered cultures to preserve the RNA until extraction. Filtering times were improved dramatically by using pre-sterilized Pall Supor membrane .2 µm. These filters broke apart upon exposure to the RNA isolation solution, but did not seem to effect the quality of the RNA.

Immediately after filtering through the liquid, the filters were placed in a polypropylene culture tube containing 4 ml RNABee RNA isolation solution (Iso-Tex

Diagnostics) and stored on ice. The filtrand was removed from the filter and in to the isolation solution by pipetting and light vortexing, and the filter was removed. After adding 800 µl of chloroform, the solution was mixed by inversion for 20 seconds and then allowed to sit on ice for 5 minutes. The solution was then centrifuged at 9,000 rpm for 15 minutes at 4°C. Two milliliters of the resulting aqueous phase was transfered to another tube, mixed with 2 ml isopropanol, and allowed to sit at room temperature for 10 minutes. The samples were centrifuged for 10 minutes at 10,000rpm (4°C). The supernatant was removed by decanting and the remaining pellet was washed in 4 ml of 75% ethanol (prepared with DEPC-treated water). The RNA was repelleted by centrifugation (8,000 rpm for 5 minutes). The supernatant was removed and the pellet dried under a laminar flow hood for 10 minutes. A second RNA cleanup was performed using the RNeasy column minikit (Qiagen) following the manufacturer's instructions with the optional in-column DNase digestion. The quality (degree of degradation) of the RNA was visualized on an electrophoresis gel, while the purity and quantity was determined using a NanoDrop spectrophotometer set for RNA analysis. As the 23S ribosomal subunit is the first to show degradation, a high quality RNA extraction was one in which the gel results showed the quantity of 23S at least as large as the 16S.

Cell density was measured through optical density at 600 nm. Optical densities were very low, presumably due to the clumped growth that MT41 exhibits, therefore the density was reassessed using DAPI stain and microscopy. The cultures were mixed by vortexing to break up the aggregations and allow for more even distribution of cells. Three aliquots, 100 µl, 500 µl, and 1 ml, of the culture were placed in 1.5 ml microcentrifuge tubes and fixed with a 1:10 dilution of fixation solution (4%

formaldehyde in PBS) overnight. The sample volumes were brought up to 5 ml using 1X PBS (phosphate buffered saline) solution and mixed by vortexing. The solutions were then filtered on to polycarbonate, black, .22µm, 25mm diameter filters. The filters were washed thoroughly with MilliQ water before being allowed to dry. To prepare the slides, 10 µl of DAPI stain plus Vectashield mounting solution was put on the slide and the dry filter placed on top. Another 10 µl of the stain plus mounting solution was added to the top of the filter and the coversheet placed on top. The slides were stored in the dark at -20°C. The cells were then counted under a microscope and used to calculate the average concentration of the cell cultures.

Appendix B

SS9 RNA extractions

RNA was extracted from two strains of *Photobacterium profundum* SS9, called TW30 and DB110, for the purpose of microarray analysis. The two strains differed in their ribotypes with TW30 demonstrating a short ribotype and DB110 a long ribotype. The long ribotypes, more common in the deep sea, are believed to be adaptations to deep sea conditions. The analysis sought to explore the differential gene expression of each strain under conditions of high and low pressure and high and low nutrient availability. Both strains TW30 and DB110 were grown under four different conditions, with three biological replicates. The four conditions were as follows: complete medium at high pressure, complete medium at low pressure, minimal medium at high pressure and minimal medium at low pressure. Complete media was 75% 2216 Marine Broth (Difco), with 100mM HEPES, and 20mM glucose. Minimal media was MOPS M4 made of 1X MOPS modified buffer (Teknova), 1X K_2HPO_4, 954mM NH_4Cl, 20mM glucose, and 32g/L sea salts (Sigma).

Starter cultures were grown in 4 ml of complete media with 40 µg rifampicin overnight in a 15ºC shaking incubator. The cultures were then used to inoculate 50 ml of complete or minimal media in a 1:150 dilution. While kept on ice, the inoculated media was transfered in to heat-sealable plastic bulbs with no gas space. The heat-sealed bulbs were placed in vessels for their appropriate pressure. The high pressure cultures were placed in pressure vessels and pressurized to 28MPa as previously described (Yayanos). Those grown at low pressure were placed in a water-filled flask and wrapped in foil to keep dark. All cultures were grown at 15ºC.

Cultures were grown to log phase, as determined by an optical density between .1 and .3 (at 600nm), which typically took between 24 and 48 hours. The growth times varied depending on the condition and was decided upon by visual scanning of the cultures and trial and error. When the cultures reached an adequate density, they were decompressed and the RNA harvested. Due to the unstable nature of RNA and the potential effects of pressure changes on RNA expression, all procedures were done quickly, at low temperature, and with reduced exposure to RNases. The total time from decompression to the addition of RNABee was less than ten minutes.

Approximately 35 ml of the liquid culture was transferred to an Rnase-free centrifuge tube. The tubes were centrifuged at 6,000 rpm for 4 minutes at 4ºC. The supernatant was discarded via decanting and the pellet resuspended in 4 ml RNABee (Iso-Tex Diagnostics) by pipetting. The solution was then transfered to a sterile polypropylene culture tube. After adding 800 µl of chloroform, the solution was mixed by inversion for 20 seconds and then allowed to sit on ice for 5 minutes. The solution was then centrifuged at 9,000 rpm for 15 minutes at 4°C. Two milliliters of the resulting aqueous phase was transfered to another tube, mixed with 2 ml isopropanol, and allowed to sit at room temperature for 10 minutes. The samples were centrifuged for 10 minutes at 10,000rpm (4°C). The supernatant was removed by decanting and the remaining pellet was washed in 4 ml of 75% ethanol (prepared with DEPC-treated water). The RNA was repelleted by centrifugation (8,000 rpm for 5 minutes). The supernatant was removed and the pellet dried under a laminar flow hood for 10 minutes. A second RNA cleanup was performed using the RNeasy column minikit (Qiagen) following the manufacturer's

instructions with the optional in-column Dnase digestion. The quality (degree of degradation) of the RNA was visualized on an electrophoresis gel. As the 23S ribosomal subunit is the first to show degradation, a high quality RNA extraction was one in which the gel results showed the quantity of 23S at least as large as the 16S.

References

Allen, E. E. and D. H. Bartlett. 2002. Structure and regulation of the omega-3 polyunsaturated fatty acid synthase genes from the deep-sea bacterium *Photobacterium profundum* strain SS9. Microbiology. **148**: 1903-1913.

Allen, E. E. and D. H. Bartlett. 2000. FabfF is required for piezoregulation of *cis*-vaccenic acit dlevels and piezophilic growth of the deep-sea bacterium *Photobacterium profundum* strain SS9. J Bacteriol. **182**(5): 1264-1271.

Auman, A. J., J. L. Breezee, J. J. Gosink, P. Kampfer, and J. Staley. 2006. *Psycrhomonas ingrahamii* sp. nov., a novel gas vacuolate, psychrophilic bacterium isolates from Arctic polar sea ice. Int J Syst Evol Micr. **56**: 1001-1007.

Bartlett, D. H. 2002. Pressure effects on in vivo microbial processes. Biochim Biophys Acta. **1595**: 367-381.

Bebbington, K. J. and H. D. Williams. 1993. Investigation of the role of *cydD* gene product in the production of a functional cytochrome *d* oxidase in *Echerichia coli*. FEMS Microbiol Lett. **112**: 19-24.

Benediktsdottir, E., L. Verdonck, C. Sproer, S. Helgason, and J. Swings. 2000. Characterization of *Vibrio viscosus* and *Vibrio wodanis* isolated at different geographical locations: a proposal for reclassification of *Vibrio viscosus* as *Moritella viscosa* comb. nov. Int J Syst Evol Micr. **50**: 479-488.

Bi, E. F. and J. Lutkenhaus. 1991. FtsZ ring structure associated with division in *Escherichia coli*. Nature. **354**: 161-164.

Bidle, K. A. and D. H. Bartlett. 1999. RecD function is required for high-pressure growth of a deep-sea bacterium. J Bacteriol. **181**(8): 2330-2337.

Bourns, B., S. Franklin, L. Cassimeris, and E. D. Salmon. 1996. Hydrostatic pressure has different effects on the assembly of tubulin, actin, myosin II, vinculin, talin, vimentin, and cytokeratin in mammalian tissue cells. Exp Cell Res. **227**: 254-256.

Bowman, J.P., S.A. McCammon, D.S. Nichols, J.H. Skerratt, S.M. Rea, P.D. Nichols, and T.A. McMeekin. 1997. *Shewanella gelidimarina* sp. nov. and *Shewanella frigidimarina* sp. nov., novel Antarctic species with the ability to produce eicosapentaenoic acid (20:5 omega 3) and grow anaerobically by dissimalatory Fe(III) reduction. Int J Syst Bacteriol. **47**:1040-1047.

Braganza, L. F. and D. L. Worcester. 1986. Structural changes in lipid bilayers and biological membranes caused by hydrostatic pressure. Biochemistry-US. **25**: 144-150.

Breezee, J., N. Cady, and J.T. Staley. 2004. Subfreezing growth of the sea ice bacterium *Psychromonas ingrahamii*. Microb Ecol. **47**(3):300-304.

Chi, E. and D. H. Bartlett. 1995. An *rpoE*-like locus controls outer membrane protein synthesis and growth at cold temperatures and high pressures in the deep-sea bacterium *Photobacterium* sp. strain SS9. Mol Microbiol. **17**(4): 713-726.

Chilikuiri, L. N. and D. H. Bartlett. 1997. Isolation and characterization of the gene encoding single-stranded-DNA-binding protein (SSB) from four marine *Shewanella* strains that differ in their temperature and pressure optima for growth. Microbiology. **143**: 1163-1174.

Campanaro, S., A. Vezzi, N. Vitulo, F. M. Lauro, M. D'Angelo, F. Simonato, A. Cestaro, G. Malacrida, G. Bertoloni, G. Valle, and D. H. Bartlett. 2005. Laterally transferred elements and high pressure adaptation in *Photobacterium profundum* strains. BMC Genomics. **6**:122.

De Las Penas, A., L. Connolly, and C. A. Gross. 1997. The σ^E –mediated response to extracytoplasmic stress in *Escherichia coli* is transduced by RseA and RseB, two negative regulators of σ^E . Mol Microbiol. **24**(2): 373-385.

Del Casale, T., P. Sollitti, and R. H. Chesney. 1983. Cytoplasmic L-asparaginase: isolation of a defective strain and mapping of *ansA*. J Bacteriol. **154**(1): 513-515.

Delong, E. and A. Yayanos. 1986. Biochemical function and ecological significance of novel bacterial lipids in deep-sea procaryotes. Appl Environ Microbiol. **51**(4): 730-737.

Delong, E. F. and A. A. Yayanos. 1985. Adaptation of the membrane lipids of a deep-sea bacterium to changes in hydrostatic pressure. Science. **228**(4703): 1101-1103.

Delong, E., D. Franks, and A. Yayanos. 1997. Evolutionary relationships of cultivated psychrophilic and barophilic deep-sea bacteria. Appl Environ Microbiol. **63**(5): 2105-2108.

Dersch, P., S. Kneip, and E. Bremer. 1994. The nucleoid-associated DNA-binding protein H-NS is required for the efficient adaptation of *Escherichia coli* K-12 to a cold environment. Mol Genet Genomics. **245**: 255-259.

Dunker, A. K., C. J. Brown, and Z. Obradovic. 2002. Identification and functions of usefully disordered proteins. Adv Protein Chem. **62**: 25-49.

Edgar, R.C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. **32**(5): 1792-1797.

Ewing, B. and P. Green. 1998. Base-calling of automated sequencer tracese using phred. II. Error probabilities. Genome Res. **8**: 186-194.

Ewing, B., L. Hillier, M.C. Wendl, and P. Green. 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res. **8**: 175-185.

Farmer, J. 2006. The family Vibrionaceae. In *The Prokaryotes* M. Dworkin, S. Falkow, E. Rosenberg, K. H. Schleifer and E. Stackebrandt (eds). **6**(3.3): 495-507.

Farmer, J. and F. Hickman-Brenner. 2006. The genera *Vibrio* and *Photobacterium*. In *The Prokaryotes* M. Dworkin, S. Falkow, E. Rosenberg, K. H. Schleifer and E. Stackebrandt (eds). **6**(3.3): 508-563.

Finch, P.W., A. Storey, K. Brown, I. D. Hickson, and P. T. Emmerson. 1986. Complete nucleotide sequence of *recD*, the structural gene for the α subunit of Exonuclease V of *Escherichia coli*. Nucleic Acids Res. **14**(21): 8583-8594.

Fruton, J. 1979. Early theories of protein structure. Ann NY Acad Sci. **325**(1): 1-20.

Fukuchi, S., K. Yoshimune, M. Wakayama, M. Moriguchi, and K. Nishikawa. 2003. Unique amino acid composition of proteins in halophilic bacteria. J Mol Bacteriol. **327**: 347-357.

Garwin, J. L., A. L. Klages, and J. E. Cronan. 1980. Structural, enzymatic, and genetic studies of beta-ketoacyl-acyl carrier protein synthases I and II of *Escherichia coli*. J Biol Chem. **225**: 438-463.

Goldberg, S.M.D., J. Johnson, D. Busam, T. Feldblyum, S. Ferriera, R. Friedman, A. Halpuern, H. Khouri, S.A. Kravitz, F.M. Lauro, K. Li, Y.H. Rogers, R. Strausberg, G. Sutton, L. Tallon, T. Thomas, E. Venter, M. Frazier, and J.C. Venter. 2006. A Sanger/pyrosequencing hybrid approach for the generation of high-quality draft assemblies of marine microbial genomes. P Natl Acad Sci USA. **103**: 11240-11245.

Gordon, D., C. Abajian, and P. Green. 1998. Consed: a graphical tool for sequence finishing. Genome Res. **8**: 195-202.

Gross, M. and R. Jaenicke. 1994. Proteins under pressure: the influence of high hydrostatic pressure on structure, function and assembly of proteins and protein complexes. Eur J Biochem. **221**: 617-630.

Hammes, W. and C. Hertel. 2006.The Genera Lactobacillus and Carnobacterium. In *The Prokaryotes* M. Dworkin, S. Falkow, E. Rosenberg, K. H. Schleifer and E. Stackebrandt (eds). **4**: 320-403.

Heath, R. J. and C. O. Rock. 1996. Roles of FabA and FabZ beta-hydoxyacyl-acyl carrier protein dehydratases in *Escherichia coli* fatty acid biosynthesis. The Journal of Biological Chemistry. **271**(44): 27795-27801.

Hulton, C. S. J., A. Seirafi, J. C. D. Hinton, J. M. Sidebotham, L. Waddell, G. D. Pavitt, T. Owen-Hughes, A. Spassy, H. Buc, and C. F. Higgins. 1990. Histone-like protein H1 (H-NS), DNA supercoiling, and gene expression in bacteria. Cell. **63**(3): 631-642.

Ishida, T., N. Akimitsu, T. Kashioka, M. Hatano, T. Kubota, Y. Ogata, K. Sekimizu, and T. Katayama. 2004. DiaA, a novel DnaA-binding protein, ensures the timely initiation of *Escherichia coli* chromosome replication. J Biol Chem. **279**(44): 45546-45555.

Ishii, A., T. Oshima, T. Sato, K. Nakasome, H. Mori and C. Kato. 2005. Analysis of hydrostatic pressure effects on transcription in *Escherichia coli* by DNA microarray procedure. Extremophiles. **9**: 65-73.

Ishii, A., T. Sato, M. Wachi, K. Nagai, and Chiaki Kato. 2004. Effects of high hydrostatic pressure on bacterial cytoskeleton FtsZ polymers *in vivo* and *in vitro*. Microbiology. **150**: 1965-1972.

Jack, R. W., J. R. Tagg, and B. Ray. 1995. Bacteriocins of gram-positive bacteria. Microbiol Rev. **59**(2): 171-200.

Jones, D.T. (1999) Protein secondary structure prediction based on position-specific scoring matrices. J Mol Biol. **292**: 195-202.

Jones, L. J. F., R. Carballido-Lopez, and J. Errington. 2001. Control of cell shape in bacteria: helical, actin-like filaments in *Bacillus subtilis*. Cell. **104**: 913-922.

Kamimura, K., H, Fuse, O. Takimura, and Y. Yamaoka. 1993. Effects of growth pressure and temperature on fatty acid composition of a barotolerant deep-sea bacterium. Appl Environ Microbiol. **59**: 924-926.

Kanehisa, M. and S. Goto. 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. **28**: 27-30.

Katayama, T., M. Tanaka, J. Moriizumi, T. Nakamura, A. Brouchkov, T. A. Douglas, M. Fukuda, F. Tomita, and K. Asano. 2007. Phylogenetic analysis of bacteria preserved in a permafrost ice wedge for 25,000 years. Appl Environ Microbiol. **73**(7): 2360-2363.

Kato, C., H. Tamegai, A. Ikegami, R. Usami, and K. Horikoshi. 1996. Open reading frame 3 of the barotolerant bacterium strain DSS12 is complementary with *cydD* in *Escherichia coli*: *cydD* functions are required for cell stability at high pressure. J Biochem. **120**: 301-305.

Lauro, F. M. 2007. Evolutionary and functional genomics of bacteria from the cold deep sea. PhD Dissertation, University of California, San Diego.

Lauro, F.M., K. Tran, A. Vezzi, N. Vitulo, G. Valle, and D.H. Bartlett. 2007. Large-scale transposon mutagenesis of *Photobacterium profundum* SS9 reveals new genetic loci important for growth at low temperature and high pressure. J Bacteriol. doi:10.1128/JB.01176-07.

Lauro, F.M., R.A. Chastain, L.E. Blakenship, A.A. Yayanos, and D.H. Bartlett. 2007. The unique 16S rRNA genes of piexophiles reflect both phylogeny and adaptation. Appl Environ Microbiol. **73**(3): 838-845.

Leisner, J. J., B. G. Laursen, H. Prevost, D. Drider, and P. Dalgaard. 2007. *Carnobacterium*: positive and negative effects in the environment and in foods. FEMS Microbiol Rev. **31**(5): 592-613.

Lu, M., J. L. Campbell, E. Boye, and N. Kleckner. 1994. SeqA: a negative modulator of replication initiation in E. coli. Cell. **77**: 413-426.

Lu, Y. J., Y. M. Zhang, K. D. Grimes, K. Qi, R. E. Lee, and C. O. Rock. 2006. Acyl-phosphates initiate membrane phospholipids synthesis in gram-positive pathogens. Mol Cell. **23**: 723-772.

Markowitz, V. M., E. Szeto, K. Palaniappan, Y. Grechkin, K. Chu, I. A. Chen, I. Dubchak, I. Anderson, A. Lykidis, K. Mavromatis, N. N. Ivanova, and N. C. Kyrpides. 2007. The integrated microbial genomes (IMG) system in 2007: data content and analysis tool extensions. Nucleic Acids Res. doi:10.1093/nar/gkm846

Masson, P., C. Tonello, and C. Balny. 2001. High pressure biotechnology in medicine and pharmaceutical science. J Biomed Biotechnol. **1**(2): 85-88.

McCarty, J. S., and G. C. Walker. 1991. DnaK as a thermommeter: threonine-199 is site of autophosphorylation and is critical for ATPase activity. P Natl Acad Sci USA. **88**: 9513-9517.

Methé, B.A., K. E. Nelson, J. W. Deming, B. Momen, E. Melamud, X. Zhang, J. Moult, R. Madupu, W. C. Nelson, R. J. Dodson, L. M. Brinkac, S. C. Daugherty, A. S. Durkin, R. T. DeBoy, J. F. Kolonay, S. A. Sullivan, L. Zhou, T. M. Davidsen, M. Wu, A. L. Huston, M. Lewis, B. Weaver, J. F. Weidman, H. Khouri, T. R. Utterback, T. V. Feldblyum, and C. M. Fraser. 2005. The psychrophilic lifestyle as revealed by the genome sequence of *Colwellia psychrerythraea* 34H through genomic and proteomic analyses. P Natl Acad Sci USA. **102**(31): 10913-10918.

Miller, V.L., R.K. Taylor, and J.J. Mekalanos. 1987. Cholera toxin transcriptional activator ToxR is a transmembrane DNA binding protein. Cell. **48**: 271-279.

Missiakas, D., M. P. Mayer, M. Lemaire, C. Georgopoulos, and S. Raina. 1997. Modulation of the *Escherichia coli* σ^E (RpoE) heat-shock transcription-factor activity by the RseA, RseB, and RseC proteins. Mol Microbiol. **24**(2): 355-371.

Mountfort, D. O., F. A. Rainey, J. Burghardt, H. F. Haspar, E. Stackebrandt. 1998. *Psychromonas antarcticus* gen. nov., sp. nov., a new aerotolerant anaerobic, halophilic psychrophile isolated from pond sediment of the McMurdo Ice Shelf, Antarctica. Arch. Microbiology. **169**: 231-238.

Mozhaev, V. V., K. Heremans, J. Frank, P. Masson, and C. Balny. 1994. Exploiting the effects of high hydrostatic pressure in biotechnological applications. Trends Biotechnol. **12**: 493-501.

Mozhaev, V. V., K. Heremans, J. Frank, P. Masson, and C. Balny. 1996. High pressure effects on protein structure and function. Proteins. **24**: 81-91.

Myers, E.W., G.G. Sutton, A.L. Delcher, I.M. Dew, D.P. Fasulo, M.J. Flanigan, S.A. Kravitz, C.M. Mobarry, K.H. Reinert, K.A. Remington, E.L. Anson, R.A. Bolanos, H.H. Chou, C.M. Jordan, A.L. Halpern, S. Lonardi, E.M. Beasley, R.C. Brandon, L. Chen, P.J. Dunn, Z. Lai, Y. Liang, D.R. Nusskern, M. Zhan, Q. Zhang, X. Zheng, G.M. Rubin, M.D. Adams, and J.C. Venter. 2000. A whole-genome assembly of *Drosophila*. Science. **287**: 2196-2204.

Nogi, Y. C. Kato, and K. Horikoshi. 2002. Psychromonas kaikoae sp. nov., a novel piezophilic bacterium from the deepest cold-seep sediments in the Japan Trench. Int J Syst Evol Micr. **52**(5): 1527-1532.

Nogi, Y., N. Masui, and C. Kato. 1998. *Photobacterium profundum* sp. nov., a new, moderately barophilic bacterial species isolated from a deep-sea sediment.

Extremeophiles. **2**: 1-7.

O'Conner, L., A. Coffey, C. Daly, and G. F. Fitzgerald. 1996. AbiG, a genomtypically novel abortive infection mechanism encoded by plasmid pCI750 of *Lactococcus lactis* subsp. *cremoris* UC653. App Environ Microbiol. **62**(9): 3075-3082

Paulson, I.T., L. Banderjei, G.S. Meyers, K.E. Nelson, R. Seshadri, T.D. Read, D.E. Fouts, J.A. Eisen, S.R. Gill, J.F. Heidelberg, H. Tettelin, R.J. Dodson, L. Umayam, L. Brinkac, M. Beanan, S. Daugherty, R.T. DeBoy, S. Durkin, J. Kolonay, R. Madupu, W. Nelson, J. Vamatheyan, B. Tran, J. Upton, T. Hansen, J. Shetty, H. Khouri, T. Utterback, D. Radune, K.A. Ketchum, B.A. Dougherty, and C.M. Fraser. 2003. Role of mobile DNA in the evolution of vancomycin-resistant Enterococcus faecalis. Science. **299**(5615): 1999-2002.

Pikuta, E. V., D. Marsic, A. Bej, J. Tang, P. Krader, and R. B. Hoover. 2005. *Carnobacterium pleistocenium* sp. nov., a novel psychrotolerant, facultative anaerobe isolated from permafrost of the Fox Tunnel in Alaska. Int J Syst Evol Micr. **55**: 473-478.

Rice, P. I. Longden, and A. Bleasby. 2000. EMBOSS: the European molecular biology open software suite. Trends Genet. **16**(6): 276-277.

Rothman, J. E. 1989. Polypeptide chain binding proteins: catalysts of protein folding and related processes in cells. Cell. **59**: 591-601.

Rouviere, P.E., A. De Las Penas, J. Mecsas, C. Z. Lu, K. E. Rudd, and C.A. Gross. 1995. *rpoE*, the gene encoding the second heat-shock sigma factor, σ^E, in *Escherichia coli*. EMBO J. **14**(5): 1032-1042.

Saito, R., and A. Nakayama. 2004. Differences in malate dehydrogenases from the obligately piezophilic deep-sea bacterium *Moritella* sp. strain 2D2 and the psychrophilic bacterium *Moritella* sp. strain 5710. FEMS Microbiol Lett. **233**: 165-172.

Saito, R., C. Kato, and A. Nakayama. 2006. Amino acid substitutions in malate dehydrogenases of piezophilic bacteria isolated from intestinal contents of deep-sea fishes retrieved from the abyssal zone. J Gen Appl Microbiol. **52**: 9-19.

Schatz, M. C., A. M. Phillippy, B. Shneiderman, S. L. Salzberg. 2007. Hawkeye: a visual analytics tool for genome assemblies. Genome Biol. **8**: R34.

Simonato, F., S. Campanaro, F. M. Lauro, A. Vezzi, M. D'Angelo,, N. Vitulo, G. Valle, D. H. Bartlett. 2006. Piezophilic adaptation: a genomic point of view. J Biotechnol. **126**: 1-25.

Sogin, M. L., H. G. Morrison, J. A. Huber, D. M. Welch, S. M. Huse, P. R. Neal, J.

M. Arrieta, and G. J. Herndl. 2006. Microbial diversity in the deep sea and the underexplored "rare biosphere". P Natl Acad Sci USA. **103**(32): 12115-12120.

Somero, G.N. 1990. Life at low volume change: hydrostatic pressure as a selective factor in the aquatic environment. Am Zool. **30**: 123-135.

Staley, J. T. and A. Konopka. 2003. Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. Annu Rev Microbiol. **39**: 321-346.

Tatusov, R. L., E. V. Koonin, and D. J. Lipman. 1997. A genomic perspective on protein families. Science. **278**(5338): 632-637.

Teplyakov, A., G. Obmolova, M. A. Badet-Denisot, B. Badet, and I. Polikarpov. 1998. Involvement of the C terminus in intramolecular nitrogen channeling in glucosamine 6-phophate synthase: evidence from a 1.6A crystal structure of the isomerase domain. Structure. **6**(8): 1047-1055.

Wang, F., J. Wang, H. Jian, B. Zhang, S. Li, F. Wang, X. Zeng, L. Gao, D. H. Bartlett, J. Yu, S. Hu, X. Xiao. 2008. Environmental adaptation: genomic analysis of the piezotolerant and psychrotolerant deep-sea iron reduction bacterium *Shewanella piezotolerans* WP3. PLoS ONE. **3**(4): e1937.

Wegrzyn, A., B. Wrobel, and G. Wegrzyn. 1999. Altered biological properties of cell membranes in *Escherichia coli* dnaA and seqA mutants. Mol Gen Genet. **261**(4/5): 762-769.

Welch, T. J., A. Farewell, F. C. Neidhardt, and D. H. Bartlett. 1993. Stress response of *Escherichia coli* to elevated hydrostatic pressure. J Bacteriol. **175**(22): 7170-7177.

Welch, T.J. and D.H. Bartlett. 1997. Cloning, sequencing and overexpression of the gene encoding malate dehydrogenase from the deep-sea bacterium *Photobacterium* species strain SS9. Biochim Biophys Acta. **1350**: 41-46.

Welch, T.J. and D.H. Bartlett. 1998. Identification of a regulatory protein required for pressure-responsive gene expression in the deep-sea bacterium *Photobacterium* species strain SS9. Mol Microbiol. **27**(5): 977-985.

Willis, R. C. and C. A. Woolfolk. 1974. Asparagine utilization in *Escherichia coli*. J Bacteriol. **118**: 231-241.

Worden, A., M. Culvelier, and D. Bartlett. 2006. In-depth analysis of marine microbial community genomics. Trends Microbiol. **14**: 331-336.

Wren, B. W. 2006. Prokaryotic genomics. In *The Prokaryotes* M. Dworkin, S. Falkow,

E. Rosenberg, K. H. Schleifer and E. Stackebrandt (eds). **1**: 246-260.

Yayanos, A. A. 1995. Microbiology to 10,500 Meters in the Deep-Sea. Annu Rev Microbiol. **49**: 777-805.

Yayanos, A. A. and A. S. Dietz. 1982. Thermal inactivation of a deep-sea barophilic bacterium, isolate CNPT-3. Appl Environ Microbiol. **43**(6): 1481-1489.

Yayanos, A. A. and E. C. Pollard. 1969. A study of the effects of hydrostatic pressure on macromolecular synthesis in *Escherichia coli*. Biophysics. **9**: 1464-1482.

Yayanos, A.A. 1986. Evolutional and ecological implications of the properties of deep-sea barophilic bacteria. P Natl Acad Sci USA. **83**: 9542-9546.

Yayanos, A.A. 1998. Empirical and theoretical aspects of life at high pressure in the deep sea, p. 47-92. In K. Horikoshi and W.D. Grant (ed.), Extremophiles. Microbial life in extreme environments. John Wiley and Sons, Inc., New York, N.Y.

Yayanos, A.A., A.S. Dietz, and R. Van Boxtel. 1981. Obligately barophilic bacterium from the Mariana Trench. P Natl Acad Sci USA. **78**: 5212-5215.

Zobell, C. and R. Morita. 1957. Barophilic bacteria in some deep sea sediments. J Bacteriol. **73**(4): 563-568

ZoBell, C. E. and A. B. Cobet. 1962. Growth, reproduction, and death rates of *Escherichia coli* at increased hydrostatic pressures. J Bacteriol. **84**: 1228-1236.

ZoBell, C. E. and A. B. Cobet. 1964. Filament formation by *Escherichia coli* at increased hydrostatic pressures. J Bacteriol. **87**: 710-719.