WELLS, MICHAEL BANNER, M.S. Se-ing Bacteria in a New Light: Investigating Selenium Metabolism in *Bacillus selenitireducens* MLS10. (2015)
Directed by Dr. Malcolm Schug 84 pp.

Prokaryotes metabolize selenium through incorporation into the 21st amino acid selenocysteine, and the respiration of the selenium oxyanions, selenite and selenite.I investigated the physiological function and evolution of selenoproteins in*Bacillus selenitireducens* MLS10 by annotating the selenoproteome of MLS10 and constructing phylogenies of selenoproteins. I investigated the physiology of selenite respiration in MLS10 by obtaining protein profiles using SDS-PAGE, determining the cytochrome content using the pyridine hemochrome assay, and testing for enzyme activity in native gels using selenite-grown MLS10 cells. My research demonstrates that the Bacilli exploit Sec residues far more than has heretofore been appreciated, that the use of Sec residues in MLS10 is ancestral, and suggests that extensive horizontal gene transfer characterizes the evolution of selenoproteins in Gram-positive bacteria and the δ-Proteobacteria. Finally, my research provides evidence that selenite respiration is an inducible respiratory pathway in MLS10, and suggests future directions for further testing of this hypothesis.

SE-ING BACTERIA IN A NEW LIGHT: INVESTIGATING SELENIUM METABOLISM IN *BACILLUS SELENITIREDUCENS* MLS10

by

Michael Banner Wells

A Thesis Submitted to
the Faculty of The Graduate School at
The University of North Carolina at Greensboro
in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Greensboro
2015

Approved by

_____
Committee Chair

Tezimi Merve Seven'e ithaf ederim. Çalışmalarım boyunca yoluma ışık olduğunuz ve gerçek bilim insanı olmayı bana öğrettiğiniz için size minnettarım.

APPROVAL PAGE

This thesis written by Michael Banner Wells has been approved by the following committee of the Faculty of The Graduate School at The University of North Carolina at Greensboro.

Committee Chair _____

Committee Members _____

_____

_____

_06/26/15_____
Date of Acceptance by Committee

_06/26/15_____
Date of Final Oral Examination

ACKNOWLEDGEMENTS

I would like to thank Dr. Malcolm Schug for his invaluable assistance and encouragement in helping me carry out this research. Without his unflagging support, this research project would not have been possible. I would also like to thank Dr. Ronald Oremland of the United States Geological Survey and Dr. John Stolz of Duquesne University for letting me do a great deal of this research intheir labs. I am grateful to Dr. Robert Cannon, Dr. David Remington, and Dr. John Stolz for serving on my committee. Their insights and critiques have immeasurably improved my science. Finally, I am grateful to the UNCG Biology Department and Sigma Xi for funding my research, to Dr. Karen Katula for giving me access to her lab for the SDS-PAGE gels, and finally to Mr. Carl Manner for his time spent teaching me how to prepare and run protein gels.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER I

INTRODUCTION AND PROJECT AIMS

The long-term goal of my research is to understand the role of selenium metabolism during the evolution of prokaryotes. My thesis research will focus on two specific studies that address this question as follows: 1) What is the evolutionary history of selenoproteins in the Gram-positive Bacilli? and 2) Is the ability of some Bacilli to use the selenium oxyanion, selenite, as a terminal electron acceptor during anaerobic respiration an inducible respiratory pathway distinct from other respiratory pathways?

It is Unknown Why Some Organisms Use Sec in Certain Oxidoreductases, Rather Than Cys

In many organisms the trace element selenium is exploited metabolically in a way that has no known analogue in other trace elements - it is incorporated into the amino acid selenocysteine (Sec) (Stolz et al. 2006) co-translationally during protein synthesis using the UGA opal codon (Chambers et al. 1986; Zinoni et al. 1987; Stolz et al. 2006). In contrast, other trace elements such as nickel, iron, and molybdenum serve mainly as co-factors in several enzymes mediating metabolic reactions. The components of the Sec synthesis and insertion architecture were first elucidated in *Escherichia coli*.

Selenocysteine is not ligated onto a $tRNA^{Sec}$ in the manner of the twenty most common amino acids, but is synthesized from a seryl-$tRNA^{Ser}$ (SelC) (Leinfelder et al. 1988)by a selenocysteine synthase (SelA), using selenophosphate as the selenium donor

Forchhammer and Böck 1991). Leinfelder et. al (1990) discovered that the selenophosphate selenium donor is supplied by SelD, which catalyzes the formation of selenophosphate from selenide and ATP. Thus, SelA and SelD mediate the synthesis of Sec. A stem and loop mRNA structure, known as the Sec insertion sequence (SECIS) element, allows the cell's translation machinery to differentiate between UGA's dual roles as a Sec and a stop codon (Zinoni et al. 1990). The final element mediating the insertion of Sec is the Sec specific elongation factor (SelB) (Heider et al. 1992) that promotes the translation of Sec by forming a complex with GTP, SECIS, and the $tRNA^{Sec}$ during translation.

It has been experimentally demonstrated in several Gram-positive bacteria and several methanogenic archaeans that the cellular machinery required to synthesize Sec is strongly conserved across prokaryotes (Tormay et al. 1994; Gursinsky et al. 2000; Rother et al. 2001), and the use of Sec residues is widespread in prokaryotes (Zhang et al. 2006). This machinery is energetically expensive to maintain, yet most organisms encode one or twoselenoproteins (the full complement of selenoproteins in an organism is referred to as the selenoproteome), with the exception of the Actinobacteria, the Clostridia, and the δ-proteobacteria, with some representatives from these taxa incorporating Sec into six or more selenoproteins (Zhang et al. 2006). The Clostridia and the δ-proteobacteria are distinct among prokaryotes additionally for being the only taxa that use all bacterial selenoproteins discovered to date, with most of the selenoproteins reported in the primary literature being restricted to these taxa (Zhang et al. 2006 and Stock and Rother 2009).

All selenoproteins discovered to date are oxidoreductases (Zhang et al. 2006), suggesting that Sec residues confer a functional advantage to organisms by facilitating the electron transfer. However, there appears to be no fitness advantage to incorporating Sec into proteins, as there has been no single demonstration that oxidoreductases that incorporate Sec are physiologically more effective in catalyzing electron transfer. Furthermore, homologous proteins that incorporate cysteine (Cys) in lieu of Sec exist in most organisms (Zhang et al. 2006). Thus, it is unknown what advantage is conferred to organisms that have evolved the capacity to synthesize and incorporate Sec.

It is possible that Sec confers a fitness advantage in challenging environments in nature. Zhang et al. (2006) found that a higher percentage of prokaryotes use Sec in thermophilic and anaerobic environments relative to mesophilic and aerobic environments. Additionally, studies focused on Sec use in the Gram-positive Clostridia, a class of organisms in the Firmicutes phylum, demonstrate that Sec use is closely associated with energy generation, including fermentative pathways and especially acetogenesis (Stock and Rother 2009). Both of these observations suggests that Sec use in bacteria may be associated with physiological processes involved with energy generation, particularly energy yielding pathways associated with low free energies.

The Bacilli, a sister class to the Clostridia in the Firmicutes phylum that is characterized by different energy-generating physiologies, have previously been found to be a selenoprotein poor class of organisms (Zhang et al. 2006). Nonetheless, previous research has found that the genome of the Gram-positive bacteria *Bacillus selenitireducens* MLS10 contains at least 4 selenoproteins (Lucas et al. 2013), though the

genome of MLS10 has not been systematically annotated for selenoproteins. These findings demonstrate that MLS10 utilizes selenoproteins to a significant extent, as most organisms only encode one or two selenoproteins (Zhang et al. 2006). MLS10 is thus an ideal system to use in investigating the evolution of selenoproteins in the Bacilli, to determine to what extent selenoproteins in the Bacilli are ancestral, conserved traits, relatively recent derived adaptations, and to determine to what extent instances of horizontal gene transfer between the bacili and other bacterial taxa have shaped the evolution of selenoproteins in the Bacilli.

The Ability to Use the Selenium Oxyanion Selenite During Respiration is Unusual in Prokaryotes

Although the ability of bacteria to generate energy from selenite remains poorly understood, there are several prokaryotes known to use the selenium oxyanions, selenate and selenite, as terminal electron acceptors during anaerobic respiration (Schröder et al. 1997; Switzer Blum et al. 1998; Stolz et al. 1999; Huber et al. 2000; Blum et al. 2001; Takai et al. 2002; Pearce et al. 2009). Prokaryotes known to use selenate and selenite in respiration are phylogenetically diverse (Stolz et al. 2006). It is unclear if the ability to use selenite in respiration is an independent respiratory pathway, or if selenite respriation is mediated by a previously characterized respiratory pathway. Selenate respiration appears to be an ancient metabolic process, mediated by a distinct, inducible selenate reductase, indicating that using selenate as a source of energy is an ancient evolutionary adaptation in prokaryotes (Schröder et al. 1997; Krafft et al. 2000; McEwan et al. 2002).

MLS10 is involved in the biogeochemical cycling of many elements including selenium through respiration of the selenium oxyanion selenite (Switzer Blum et al. 1998). This ability to respire selenite may represent a distinct respiratory pathway that evolved early in the history of prokaryotes as a form of energy generation, and is distinct from respiratory pathways that prokaryotes use to respire other electron acceptors. This unique feature of selenium metabolism may also be important to environmental remediation approaches in natural landscapes polluted with selenite because an enzyme mediating the dissimilatory reduction of selenite may be used to remediate selenite contaminated environments that are highly toxic to humans and other organisms (Switzer Blum et al. 1998).

MLS10 is an ideal organism to use to investigate the physiology of selenite respiration in prokaryotes. Previous research has demonstrated that MLS10 grows readily on selenite (Switzer Blum et al. 1998), allowing for sufficient cell yields to be obtained for proteomic analysis. Proteomic approaches can be used to compare to the respiratory pathways of other terminal electron acceptors in MLS10, as the respiratory pathways used by MLS10 to respire other terminal electron acceptors have already been well-characterized in multiple bacterial taxa. The arsenate reductase of MLS10 has previously been characterized (Afkar et al. 2003) and methods for assaying enzyme activity of the arsenate reductase were established from studies of the arsenate reductase from other prokaryotes (Richey et al. 2009). The respiratory pathways of nitrate (Sparacino-Watkins et al. 2014) and fumarate respiration (Kröger et al. 1992) are well characterized, including the cytochromes involved in nitrate and fumarate respiration.

The specific cytochrome signatures associated with both respiratory processes allows for cytochrome content assays to identify differences between cytochromes associated with selenite respiration and the respiration of nitrate and fumarate.

Significance

Annotating the selenoproteome of MLS10, and constructing phylogenies of the selenoprotein genes will provide insights into the functional role of Sec in the Bacilli, thus making a fundamental contribution to our understanding of the inheritance of Sec traits in this taxon, revealing novel insights into the evolutionary history and potential physiological roles of selenoproteins in the Bacilli. A detailed characterization of Sec use in the Bacilli may also suggest possible functions associated with Sec uses that are broadly conserved across bacterial taxa.

In addition to the ability of many prokaryotes to incorporate selenium into the 21st amino acid selenocysteine, the ability of some prokaryotes to use selenite, an oxyanion of selenium, as a terminal electron acceptor in anaerobic respiration remains poorly understood, and the physiology of selenite respiration remains unknown. Ancient respiratory pathways associated with particular electron acceptors are characterized by distinct respiratory pathways, including a particular cytochrome content, differential expression of proteins involved in the electron transport pathway, and an enzyme to mediate the dissimilatory reduction of the terminal electron acceptor that is inducible only when cells are grown in the presence of that particular electron acceptor. Investigating the biochemistry and physiology of selenite respiration in MLS10 may shed light on whether selenite respiration is an ancient adaptation in prokaryotes, or an

adaptation that some bacteria have evolved in response to more recent selective pressures.

My studies of selenoprotein genes in the genome of MLS10 and use of selenite as a terminal electron acceptor in respiration in MLS10 will answer questions about the basic biology and evolution of how and when prokaryotic organisms acquired the ability to use selenium in basic biological functions. These studies will additionally inform us about how to approach the studies of how prokaryotes may use and metabolize toxic substances. Such information will also provide important basic knowledge for the application of selenite respiring bacteria for remediating environments contaminated with selenium oxyanions, which remains a significant environmental problem in seleniferous environments (Stolz and Oremland 1999).

Project Goals and Specific Aims

The goal of my project is to determine to what extent Sec use in the Bacilli represents ancestral adaptations, derived traits unique to the Bacilli, or instances of horizontal gene transfer, and thus to what extent selenoproteins in the Bacilli and Clostridia represent phylogenetically distinct lineages, and to determine if selenite respiration is an inducible pathway distinct from other pathways involved in anaerobic respiration.

My project has two specific aims that focus on assessing the role of the metabolism of selenium in the evolution of prokaryotes through investigating two different aspects of selenium metabolism. One aim will investigate whether Sec use in MLS10 is closely linked with energy generation, as in the Clostridia, with selenoproteins

involved in physiological processes not exploited by the Clostridia representing either ancestral traits lost in the Clostridia, recent adaptations unique to the Bacilli forming a distinct phylogenetic lineage from Cys-containing homologs in the Clostridia, or horizontal gene transfer events with more distantly related bacterial taxa, while shared physiological processes between the Bacilli and the Clostridia would be phylogenetically closely related, thus offering an assessment of the likelihood of whether natural selection acts to conserve Sec residues in selenoproteins because of an adaptive advantage associated with Sec use in facilitating crucial metabolic processes in bacterial taxa.

My second aim will investigate whether selenite respiration in MLS10 is a unique respiratory pathway mediated by a distinct, inducible selenite reductase (Sir), and thus represents an ancient metabolic pathway in prokaryotes.

General Hypothesis

My general hypothesis is that proteins involved in the metabolism of selenium represent ancient adaptations in prokaryotic lineages.

Aim 1: I will determineif the content of the selenoproteome of MLS10 is similar to the content of the selenoproteomes of the Clostridia, or if there are any differences in the selenoproteome of MLS10 that are either unique to the Bacilli, or shared between the Bacilli and other distantly related taxa.

I will annotate the selenoproteome of MLS10 to determine if the proteins that incorporate Sec residues in MLS10differ from other published bacterial and archaeal species, particularly the Gram-positive Clostridia.

1. I predict that the annotation of the selenoproteome of MLS10 will reveal that the content of the selenoproteome of MLS10 is similar to the content of the selenoproteomes of the Clostridiawith some possibly novel selenoproteins that are either unique to the Bacilli or shared with the Bacilli and more distant taxa involved in anaerobic respiration.

Support for this prediction will consist of the identification of selenoproteins in the genome of MLS10, with probably 6-12 selenoproteins that are mostly also exploited by the Clostridia, and with some potentially novel selenoprotein candidates identified in cytochrome and molybdopterin oxidoreductase genes. This result will suggest that Sec residues offer an inherent catalytic advantage over Cys residues in energetically challenging environments, and therefore that natural selection favors the use of Sec residues in prokaryotes in physiological processes that a particular prokaryote frequently exploits to survive in an energetically challenging environment. If no potential selenoproteins are identified among genes involved in anaerobic respiration in MLS10, this will suggest that patterns of Sec use may not vary significantly across bacterial taxa becausethe expansion of Sec incorporation into novel oxidoreductases is not advantageous for Sec utilizing organisms, but confers a fitness advantage only in very specific physiological processes. Bacteria in which these physiological pathways do not operate lose Sec synthesis and incorporation traits.

2. I predict that a phylogenetic analysis of the selenoproteins I identify in the selenoproteome of MLS10 will reveal a subset of selenoproteins that are derived in MLS10 and other Bacilli, and a subset of selenoproteins that are conserved across both the Bacilli and Clostridia lineages.

Two phylogenies would support my prediction. A phylogeny of MLS10 selenoproteins that reveal phylogenetic trees featuring the Bacilli and the Clostridiahomologs in a monophyletic clade containing either all Sec residues, or Sec and Cys residues with the most deeply branched nodes containing Sec residues represents a scenario where selenoprotein genes are ancestral to the Bacilli. Phylogenies that contain Bacilli and Clostridia homologs in a monophyletic clade, or the Bacillihomologs in a monophyletic clade without any representatives from the Clostridia, with the Bacilli clade containing Sec residues against a phylogenetic background of Cys-containing homologs represents a scenario where selenoprotein genes are derived in the Bacilli.

Instances of Sec loss and horizontal gene transfer events may complicate the interpretation of the phylogenies. For example, a phylogeny that containsBacilli and Clostridiahomologs in a monophyletic clade, with the Bacilli clade containing Sec residues against a phylogenetic background of Cys-containing homologs in the Clostridia and of Sec-containing homologs in more distantly related bacterial taxa represents a Sec loss event in the Clostridia, while the Sec trait in the Bacilli are ancestral. Another scenario would involve a phylogeny with MLS10 selenoproteins, and perhaps other Bacilli, forming a clade with Sec-containing homologs from more distantly related bacterial taxa. This scenario would suggest that a particular selenoprotein gene was inherited via horizontal transfer.

Support for these predictions will indicate that natural selection acts on proteins to incorporate Sec because of a fitness advantage associated with physiological processes that bacteria such as MLS10 exploit to adapt to marginal environments. This would

suggest that bacterial use of Sec varies across bacterial taxa, with natural selection actively favoring Sec incorporation into proteins due to a property inherent to Sec residues that allows for more efficient catalysis of metabolic reactions. If all of the selenoproteins are not phylogenetically distinguishable from Clostridia selenoproteins, this will indicate that the selenoproteins are ancestral to both lineages and conserved, suggesting that Sec use varies across bacterial taxa, not because Sec residues inherently offer an adaptive advantage in challenging environments, but rather that the advantage conferred by incorporating Sec residues into proteins is more nuanced than an inherent property of Sec, and that the nature of the physiological process the selenoprotein mediates is crucial for determining whether an adaptive advantage exists for Sec incorporation.

Aim 2: I will investigate the physiology of selenite respiration in MLS10 to assess the importance of selenite as an electron acceptor in anaerobic respiration.

I will perform experiments to test three predictions to provide a strong indication whether or not selenite respiration is a distinct, inducible respiratory pathway.

1. I predict that the protein profile of the membrane (particulate) fractions of MLS10 cells grown with selenite as a terminal electron acceptor will differ from MLS10 cells grown with nitrate, arsenate, and fumarate as electron acceptors.

If the SDS-PAGE gels show band patterns of selenite-grown MLS10 cells differ from MLS10 cells grown with nitrate, arsenate, and fumarate cells, then my hypothesis will be supported. In contrast, if selenite-grown particulate fractions demonstrate a protein profile similar to fractions grown with nitrate, arsenate, or fumarate, then this may

indicate either that my hypothesis is unsupported, or that the molecular mass of the proteins in all pathways are simply similar to one another. Regardless, further testing of the hypothesis is prudent.

2. I predict that the cytochrome content of particulate fractions of MLS10 selenite-grown cells will differ from the cytochrome content of MLS10 cell grown with arsenate, nitrate, and fumarate, as determined by a spectrophotometer using the pyridine hemochrome assay.

Differences in cytochrome content between selenite-grown particulate fractions versus arsenate, nitrate, and fumarate particulate fractions will support the hypothesis, while identical cytochrome content between selenite grown cells and cells grown on other membrane fractions would suggest that selenite respiration may be mediated by a previously identified respiratory pathway in MLS10. It is also possible that MLS10 uses cytochromes exploited in other respiratory pathways along with proteins whose expression is induced only when MLS10 is grown on selenite, and thus a further test of the hypothesis is prudent.

3. I predict that native in-gel enzyme assays will demonstrate that MLS10 selenite-grown particulate fractions show different patterns of selenite reductase activity than particulate fractions grown with nitrate, arsenate, and fumarate. Native polyacrylamide gels stained with a reduced methyl viologen solution imbue the gel with a deep blue color. Then, a solution containing 5mM of a terminal electron acceptor added to the stained gel will reduce the substrate, oxidizing the methyl viologen and showing a clear band at the location of reductase activity. Different patterns of reductase activity between selenite-

grown versus nitrate, arsenate, and fumarate grown cells will support the hypothesis. If the selenite-grown particulate fractions display a pattern of selenite reductase activity that is identical to either the nitrate, arsenate, or fumarate particulate fractions, then this would indicate that selenite respiration is mediated by a previously identified respiratory pathway in MLS10.

CHAPTER II

METHODS

Specific Aim 1 Methods

Annotation of the MLS10 selenoproteome.

I annotated the selenoproteome of MLS10 using the complete genome of MLS10, which is publicly available from NCBI (Lucas et al. 2013), and the annotation tools available from the Joint Genome Institutes's Integrated Microbial Genomes system (IMG); (Markowitz et al. 2006). I annotated the selenoproteome of MLS10 manually because the only computational tool developed for detecting bacterial selenoproteins (Zhang and Gladyshev 2005) is no longer operational.

Previous bioinformatic research has documented a large number of bacterial oxidoreductases that function as selenoproteins (Zhang et al. 2006; Stock and Rother 2009). Additional research has documented putative selenoproteins in the clostridium *Alkaliphilus oremlandii* that are involved in anaerobic respiration, including one molydopterin-containing anaerobic dehydrogenase and two split soret cytochrome c precursors (Kim et al. 2009). Given that MLS10 cells are capable of exploiting many electron acceptors in anaerobic respiration, MLS10 was a prime candidate to search for similarly novel selenoproteins involved in electron transport. Therefore searched the MLS10 genome for potential selenoproteins involved in anaerobic respiration, including molydopterin-containing oxidoreductases and cytochromes.

For selenoproteins that have been previously annotated in other bacteria, I obtained the protein sequences of these selenoproteins from the primary literature, and used the basic local alignment search tool for protein sequences (Blastp); (Altschul et al. 1990) to search the MLS10 genome for homologs. I aligned the MLS10 homolog with the selenoprotein sequence from the primary literature using the Blastp tool to search for either a Cys or Sec residue that aligned with the Sec residue from the sequence obtained from the primary literature or a truncated sequence that otherwise showed strong alignment. Because current annotation software cannot recognize TGA as a Sec codon, selenoproteins are often misannotated as truncated when they are correctly identified or marked as hypothetical proteins of unknown function, depending on where the Sec residue is located in the protein (Kryukov and Gladyshev 2004; Stolz et al. 2006). Thus, truncated proteins are a crucial indicator that an oxidoreductase contains Sec.

If the homolog identified by the Blast search was truncated, I browsed the raw scaffolding data provided by the IMG system to determine if an in-frame TGA codon was present in the upstream or downstream nucleotide sequence. If this criterion was met, I searched for an in-frame start codon in the region of the nucleotide sequence that would make the length of the oxidoreductase comparable to other known homologs. I then integrated this sequence with the gene sequence coding for the oxidoreductase, and inputed it into Blastx (Altschul et al. 1997). I identified the sequence as a selenoprotein only if the Blastx results showed that the nucleotide sequence I integrated with the annotated gene sequence aligned with homologous oxidoreductases in other bacteria, and that the TGA codon aligned with the Sec residue in the homologs from other taxa.

I searched for putative selenoprotein candidates among the molydopterin-containing anaerobic oxidoreductases and the cytochromes in the MLS10 using the IMG system. The IMG system allows for the proteins of a particular genome to be listed in the form of a database, and filters can be applied to the database to supply proteins that fit a certain criteria. I used this feature to obtain a database of molydopterin-containing oxidoreductases and cytochromes in the MLS10 genome. I looked at the raw scaffolding data provided by IMG for the MLS10 genome to search for the presence of an in-frame Sec codon above the 5' end of the protein or below the 3' end of a protein. If an in-frame Sec codon was detected, I integrated this region into the protein sequence and used the Blastx tool to see if the Sec codon aligned with a Cys codon in other homologs. If this criterion was not met, I rejected the protein as a selenoprotein candidate.

Phylogenetic analysis of the MLS10 selenoproteins.

I obtained the amino acid sequences of homologous oxidoreductases in other prokaryotes using Blastx. I then obtained the nucleotide sequences for these proteins using the databases at IMG. I aligned these sequences using MUSCLE (Edgar 2004) available in MEGA (Tamura et al. 2011). I used MUSCLE because it is a superior sequence alignment method for proteins compared to other sequence alignment tools such as Clustal W (Edgar 2004). I then constructed Bayesian phylogenies of these selenoproteins using Mr. Bayes (Ronquist et al. 2012),and I constructed maximum likelihood, maximum parsimony, and neighbor joining phylogenies using MEGA (Tamura et al. 2011).

Specific Aim 2 Methods

Cultivation of MLS10 and fractionation into soluble and particulate fractions for experiments focusing on selenite respiration.

I cultivated cells of MLS10 under an N^2 gas atmosphere in liquid medium containing (in g per L^{-1}) $(NH_4)SO_4$ (0.1), $MgSO_4$ (0.025), K_2HPO_4 (0.15), KH_2PO_4 (0.08), NaCl (40), Na_2CO_3 (10.6), $NaHCO_3$ (4.2), yeast extract (0.2), and cysteine-HCl (0.125) as described by Switzer-Blum et al. (1998), and I amended this medium with 5 mL (per L^{-1}) of an SL10 trace elements solution containing HCl (10mL per L^{-1}) and (in mg per L^{-1}) $FeCl_2$ (1500), $ZnCl_2$ (70), $MnCl_2$ (100), H_3BO_3 (6), $CoCl_2$ (190), $CuCl_2$ (2), $NiCl_2$ (24), Na_2MoO_4 (36), and Na_2WO_4 (3) as described by Widdel (1983). I adjusted the pH of the medium to 9.8. I cultivated cells with lactate as an electron donor and with a variety of electron acceptors: nitrate (40mM), selenite (10mM), arsenate (20mM) or fumarate (20mM).

I harvested cells during the early exponential phase via centrifugation two times at 6500 rpm for 30 minutes, and re-suspended the harvested cells in a 10mM Tris-HCL buffer (pH 8.0), with 1mM EDTA (Buffer A). I sonicated whole cells of MLS10 on ice using a probe sonicator at 8 watts three times at 30 seconds each. I obtained soluble and particulate fractions via ultracentrifugation at 100,000 g for 30 minutes and quantified the protein content of these fractions using either a QuBit Fluorometer or the method described in Bradford (1976). The Bradford assay is a colorimetric assay that quantifies the amount of protein in a sample based on a shift of the color of the Coomassie Brilliant Blue G-250 as the reagent. The dye is red under acidic conditions, but will shift towards

blue when proteins are added to the dye. The degree to which the dye shifts blue indicates the concentration of proteins in a sample- darker shades of blue corresponds to greater concentrations of protein. I determined the concentration of the particulate and soluble protein fractions by adding 1 μg of each protein fraction to 1 mL of the Bradford reagent, and quantified the concentration of protein using a spectrophotometer. Particulate fractions (containing membrane and ribosome proteins) were re-suspended in Buffer A, and stored at -20ºC.

SDS-PAGE electrophoresis.

I determined the protein profile of the soluble and particulate fractions of MLS10 cells grown under each electron acceptor by SDS-PAGE electrophoresis using a 4% stacking gel containing SDS (50mg), 30% acrylamide solution (1.3mL), 0.5M Tris solution (6.8 pH) (2.6mL), 10% ammonium persulfate solution (70μL), TEMED (7μL), and deionized H_20 (6mL), and a 7% resolving gel containing SDS (125mg), 30% acrylamide solution (5.83mL), 0.5M Tris solution (8.8 pH) (6.25mL), 10% ammonium persulfate solution (125μL), TEMED (12.5μL), and deionized H_20 (12.9mL). Each lane contained approximately 35μg total protein, and I ran the gels for 2 hours at 40 milliamps. Prior to electrophoresis, I placed the proteins into a 10% SDS sample buffer and incubated at 90ºC as described by Stolz et al. (Stolz et al. 1997).

Pyridine-hemochrome cytochrome assay.

I obtained oxidized absorbance spectra for the visible light wavelengths (350-750nm) of the active fractions of each growth condition using a dual beam spectrophotometer using the pyridine hemochrome assay as developed by Liu and Peck

(1981). I placed proteins from each soluble and insoluble protein fraction in a 1mL solution containing 0.075M NaOH and 25% pyridine in a cuvette, and a reference cuvette containing 1 mL Buffer A was used to control for background correction. I added 10μM of sodium dithionite to the cuvettes to obtain the reduced absorbance spectra for each fraction. Horse heart cytochrome-*c* was used as a standard.

Native in-gel enzyme assays.

I prepared native gels using a 4% stacking and a 7% resolving gel as described abovewith the exception that I used the CHAPS detergent in preparing the gels, rather than SDS. Each lane contained approximately 35 μg total protein for the soluble and particulate fractions grown under each of the four conditions. The gels were run for approximately 4 hours at 150V. Prior to electrophoresis, I placed the protein samples in a 10% CHAPs buffer without the reducing agent β-mercaptoethanol (Richey et al. 2009), and incubated the samples on ice to prevent denaturation.

After electrophoresis, I soaked the gels in Buffer A amended with 10μM reduced methyl viologen and 10μM sodium dithionite, which imbues the gel with a deep blue color. I prepared a 5mM Buffer A for each electron acceptor(selenite, nitrate, arsenate, and fumarate), and I added these solutions to gels soaked in the reduced methyl viologen to assess the substrate specificity for each fraction. I determined the pattern of selenite reductase activity in each gel by looking for the formation of a clear band in each fraction. A clear band indicates an enzyme is coupling the reduction of selenite to the oxidation of methyl viologen. While the in-gel enzyme assay yields information

concerning enzyme activity, the in-gel enzyme assay cannot yield information concerning the kinetic parameters of the enzymes that mediate the reduction of the substrates.

CHAPTER III

RESULTS

Annotation of the Selenoproteome of *Bacillus selenitireducens*

My search for selenoprotein homologs of selenoproteins previously identified in the primary literature in the genome of MLS10 found seven selenoproteins (Table 1). Four of the selenoproteins I identified were previously annotated by Lucas et al (2013), including two formate dehydrogenase α subunits, a HesB-like protein, and methionine sulfoxide reductase A. I identified three more selenoproteins: a D-proline reductase, a selenophosphate synthase (SelD), and a SelW-like protein. No novel selenoprotein candidates were found among the molybdopterin oxidoreductase or cytochrome genes in the genome of MLS10. The protein sequences of each of the seven selenoproteins is shown in Appendix A. The physiological function *in vivo* has been studied in five of these selenoproteins in other organisms.

Table 1.Previously Identified Selenoproteins from the Primary Literature (Zhang et al. 2006, Kim et al. 2009, and Stock and Rother 2009), and Whether or not Homologs are Present in the Genome of MLS10.

Selenoproteins Previously Identified in the Primary Literature Present in the MLS10	Present in the MLS10 genome as a selenoprotein?
AhpD-like selenoprotein	No
AhpF selenoprotein homolog (N-terminal domain-containing)	No
ArsC-like selenoprotein	No
Betaine reductase B	No
BFD-like (2Fe-2S)-binding domain protein	No
Coenzyme F420-reducing hydrogenase α subunit	No
Coenzyme F420-reducing hydrogenase δ subunit	No
Distant AhpD homolog	No
D-proline reductase	Yes
DsbA-like	No
DsbG-like selenoprotein	No
DsrE-like selenoprotein	No
FAD+ dependent oxidoreductase (GlpC-like domain containing)	No
Formate dehydrogenase N α subunit	Yes
Glutaredoxin	No
Glutathione peroxidase	No
Glycine reductase selenoprotein A	No
Glycine reductase selenoprotein B	No
HesB-like selenoprotein	Yes
Heterodisulfide reductase, subunit A	No
Methionine sulfoxide reductase A	Yes
Methionine sulfoxide reductase B	No
NAD+ dependent formate dehydrogenase α subunit	Yes
NADH oxidase	No
Peroxiredoxin	No
Predicted NADH:ubiquinone oxidoreductase (RnfC subunit-containing)	No
Prx-like thiol:disulfide oxidoreductase	No
Putative Sec-containing anaerobic dehydrogenase	No
Sarcosine reductase A	No
Sarcosine reductase B	No
Selenophosphate synthase (SelD)	Yes
SelW-like selenoprotein	Yes
Split soret cytochrome c precursor 1	No
Split soret cytochrome c precursor 2	No
Thiol:disulfide interchange protein	No
Thioredoxin	No

Formate dehydrogenase N α subunit.

One of the formate dehydrogenase α subunits in MLS10 is a formate dehydrogenase N α subunit, which catalyzes the oxidation of formate to CO2 during the anaerobic respiration of nitrate in *E. coli* (Ruiz-Herrera and DeMoss 1969). MLS10grows readily using nitrate as a substrate for anaerobic respiration (Switzer Blum et al. 1998). As in MLS10, the formate dehydrogenase N α subunit is a selenoprotein in *E. coli* (Berg et al. 1991). It remains unknown whether formate dehydrogenase N is expressed when anaerobic bacteria are grown on electron acceptors other than nitrate.

NAD+ dependent formate dehydrogenase.

The other formate dehydrogenase α subunit is an NAD+ dependent formate dehydrogenase α subunit that catalyzes the reduction of CO2 to formate in the Wood-Ljungdahl pathway of acetogenic bacteria (Andreesen and Ljungdahl 1974; Yamamoto et al. 1983). The genome of MLS10 contains the genes necessary to undergo acetogenesis via the Wood-Ljungdahl pathway (Lucas et al. 2013), yet it has not yet been demonstrated that MLS10 is physiologically capable of using the Wood-Ljungdahl pathway for energy generation. Many non-acetogenic bacteria exploit the Wood-Ljungdahl pathway for both autotrophic carbon assimilation (Oelgeschläger and Rother 2008; Stock and Rother 2009) and as an electron sink (White et al. 2012), suggesting that the ability of MLS10 to exploit the Wood-Ljungdahl pathway cannot automatically be invoked as proof of its ability to generate energy via acetogenesis.

D-proline reductase.

The other selenoprotein in the genome of MLS10 that could potentially be involved in energy generation is the D-proline reductase. The D-proline reductase is a selenoprotein involved in purine fermentation that catalyzes the reduction of D-proline to 5-aminovalerate (Kabisch et al. 1999). Purine fermentation is a selenium-dependent fermentative pathway, which requires the D-proline selenoprotein and a xanthine dehydrogenase that incorporates an inorganic selenium co-factor (Self and Stadtman 2000). It has not yet been demonstrated that MLS10 is capable of generating energy via purine fermentation, however, the genome of MLS10 shows that MLS10possesses both a xanthine dehydrogenase and a D-proline reductase.

Methionine sulfoxide reductase A.

Methionine sulfoxide reductases (Msrs) are antioxidant defense proteins that reduce the methionine sulfoxide residues that form on proteins when methionine residues are exposed to reactive oxygen species, and there are two Msrs that are specific to each of the two isoforms of methionine, methionine-S (MsrA) and methionine-R (MsrB) (Brot et al. 1981; Kryukov et al. 2002). Msrs are found in organisms from all three domains of life, and is thought to be one of earliest antioxidant defense proteins to have evolved in organisms (Delaye et al. 2007). The genome encodes both Msrs in MLS10 (Lucas et al. 2013), however only the MsrA is a selenoprotein.

Selenophosphate synthase (SelD), HesB-like and SelW-like selenoproteins.

The selenophosphate synthase, as noted above, is integral to the formation of selenocysteine, and, additionally, is required for the formation of the tRNA nucleoside

selenouridine (Veres and Stadtman 1994). The function of selenouridine *in vivo* remains unknown, but Kramer and Ames (Kramer and Ames 1988) have hypothesized that the selenouridine nucleoside serves to suppress nonsense mutations in lysine, glutamate and glutamine codons. The HesB-like and SelW-like selenoproteins are named after the HesB and SelW proteins, which facilitate the formation of Fe-S clusters in proteins and serve as an antioxidant protein in eukaryotes respectively (Kryukov and Gladyshev 2004). The physiological function of these two selenoproteins remains unknown.

Phylogenies of MLS10 Selenoprotein Genes

Bayesian, maximum likelihood, maximum parsimony, and neighbor-joining phylogenies of the seven selenoprotein genes were constructed using the protein sequences (except for the selenophosphate synthase phylogenies, which were constructed with the nucleotide sequence data). I chose to construct a phylogeny of *SelD* gene sequence because the Blastx returns yielded only Bacilli homologs with one Clostridia homolog. The nucleotide sequence data would thus provide information on third base pair synonymous mutations that would be necessary for phylogenetic inference programs to construct the most plausible evolutionary scenarios, given the narrow phylogenetic range of the Blastx returns.

To construct the phylogenies, I chose a subset of 12-15 taxa toincorporate into each selenoprotein phylogeny I constructed from the Blastp returns that were returned from the MLS10 selenoprotein query I inputted into the Blastp search. I chose taxa that would most closely reflect the proportion of different phyla and subphyla homologs that were included in the Blastp returns for each MLS10 selenoprotein. After a phylogeny

was constructed, I explored many possibilities for the taxon or taxa that would be the most logical root. In all cases, the two most important criteria I considered when rooting the phylogeny was how well a particular root would be most consistent with our understanding of the evolutionary relationships among prokaryotes (Fig 1) and would minimize instances of horizontal gene transfer and Sec loss events. These criteria would ensure that I chose the most parsimonious phylogeny possible.

The Bayesian, maximum likelihood, and neighbor joining phylogenies yielded very similar topologies. I chose to focus on the Bayesian phylogenies, because in rare instances the other two phylogenetic methods would yield topologies that would require either a horizontal gene transfer event between distantly related taxa that inhabit markedly different habitats, or would require multiple instances of multiple Sec gain and loss events in the phylogenies, evolutionary scenarios which are unlikely. The maximum parsimony method yielded topologies that were largely consistent with the other methods, but more frequently required invoking unlikely horizontal gene transfer and Sec loss events. Therefore, the phylogenies generated for each selenoprotein gene using the the other three methods can be found in Appendix B.

Given the prevalence of horizontal gene transfer among prokaryotic taxa, I established criteria for determining when it would be appropriate to invoke horizontal gene transfer to explain the evolutionary relationships of selenoprotein genes in my bacterial taxa. I used the phylogenetic work of Wu et al. (2009) to determine the evolutionary relationships between prokarytotic taxa based on 31 housekeeping genes, as

a framework for what I would expect to see if horizontal gene transfer has not occurred in the evolution of these selenoprotein genes (Fig 1).

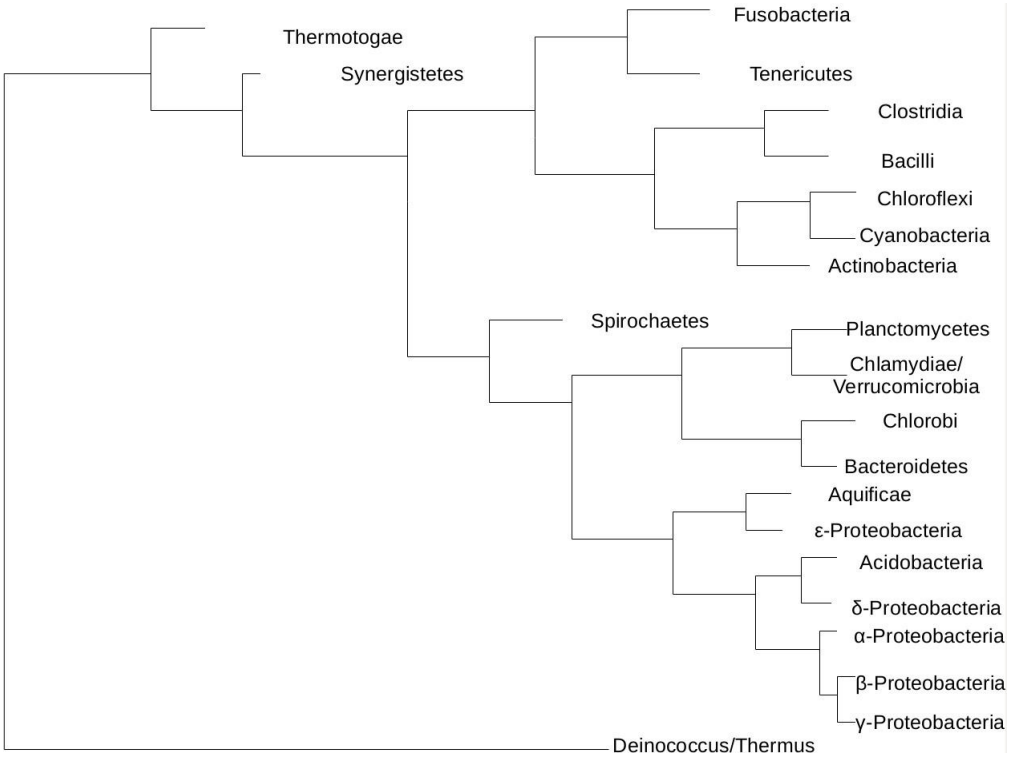


Figure 1. Phylogeny Modified from Wu et al. (2009) Showing the Relationships Between Bacterial Phyla Based on a Maximum Likelihood Phylogeny of 31 Housekeeping Genes. Note: Branch lengths do not correspond to evolutionary distances between taxa. The Firmicutes phylum is further split into the Clostridia and Bacilli classes.

Distantly related taxa that inhabit similar environmental niches would be likely candidates for a potential horizontal gene transfer event. Additionally, information concerning the environmental niches inhabited by Sec-utilizing taxa can help determine what specific environmental conditions may select for Sec utilization. I thus determined the environmental niche of each taxa using information available from IMG (Markowitz et al. 2006) (Tables 2-8). For oxygen requirement, I designated a particular taxon as a

facultative anaerobe if the genome of the taxon contained genes for both aerobic and anaerobic respiration. The designation “facultative anaerobe” thus does not imply to what extent a particular bacteria physiologically exploits aerobic and anaerobic respiration.

Table 2. Environmental Niches Inhabited by Bacteria that Possess Homologs of the D-proline Reductase.

Species	Phylum	Sec synthesizer?	Environment	Oxygen Req	pH	Halophile?	Temperature
Alkaliphilus oremlandii	Firmicutes	Yes	Anoxic river sediment	Anaerobe	Mild alkaliphile	No	Mesophile
Bacillus chagannorensis	Firmicutes	Yes	Soda lake	Facultative anaerobe	Alkaliphile	Yes	Mesophile
Bacillus selenitireducens	Firmicutes	Yes	Soda lake	Facultative anaerobe	Alkaliphile	Yes	Mesophile
Bacillus sp. 1NLA3E	Firmicutes	Yes	Nitrate and uranium contaminated soil	Facultative anaerobe	Unknown	No	Mesophile
Chloroflexus sp. Y-400-fl	Chloroflexi	Yes	Hot spring	Anaerobe	Alkaliphile	No	Mesophile
Clostridium botulinum str. F Langeland	Firmicutes	Yes	Intestinal pathogen	Anaerobe	Mild acidophile/neutrophile	No	Mesophile
Clostridium sticklandii	Firmicutes	Yes	Anoxic river sediment	Anaerobe	Neutrophile	No	Mesophile
Geobacter sp. M18	Proteobacteria	Yes	Sediments	Facultative anaerobe	Neutrophile	No	Mesophile
Geobacter uraniireducens	Proteobacteria	Yes	Uranium contaminated sediments	Facultative anaerobe	Neutrophile	No	Mesophile
Ilumatobacter coccineum	Actinobacteria	Yes	Sea sand	Facultative anaerobe	Neutrophile	Yes	Mesophile
Ktedonobacter racemifer	Chloroflexi	Yes	Sediments	Facultative anaerobe	Acidophile	No	Mesophile
Propionibacterium acidipropionici	Actinobacteria	No	Dairy products	Anaerobe	Unknown	No	Mesophile
Psychrilyobacter atlanticus	Fusobacteria	Yes	Anoxic marine sediments	Anaerobe	Neutrophile	No	Psychorphile
Selenomonas bovis	Firmicutes	No	Mammalian digestive system	Anaerobe	Acidophile	No	Mesophile
Thermosediminibacter oceani	Firmicutes	Yes	Anoxic deep sea sediments	Anaerobe	Neutrophile/mild alkaliphile	No	Thermophile

Table 3. Environmental Niches Inhabited by Bacteria that Possess Homologs of the Formate Dehydrogenase N α Subunit.

Species	Phylum	Sec synthesizer?	Environment	Oxygen Req	pH	Halophile?	Temperature
Bacillus selenitireducens	Firmicutes	Yes	Soda lake	Facultative anaerobe	Akaliphile	Yes	Mesophile
Bacillus sp. 1NLA3E	Firmicutes	Yes	Nitrate and uranium contaminated soil	Facultative anaerobe	Unknown	No	Mesophile
Bacillus vireti	Firmicutes	Yes	Soil	Facultative anaerobe	Neutrophile	No	Mesophile
Carboxydothermus hydrogenoformans	Firmicutes	Yes	Thermal springs	Anaerobe	Unknown	No	Hyperthermophile
Desulfatibacillum alkenivorans	Proteobacteria	Yes	Anoxic freshwater sediment	Anaerobe	Neutrophile	No	Mesophile
Desulfotomaculum carboxydivorans	Firmicutes	Yes	Sludge from reactor	Anaerobe	Neutrophile	No	Mesophile
Desulfovibrio africanus str. Walvis Bay	Proteobacteria	Yes	Anoxic marine sediment	Anaerobe	Neutrophile	No	Mesophile
Escherichia coli str K-12 substr W3110	Proteobacteria	Yes	Mammalian digestive system	Facultative anaerobe	Neutrophile	No	Mesophile
Geobacter bemidjiensis	Proteobacteria	Yes	Freshwater sediment	Facultative anaerobe	Neutrophile	No	Mesophile
Geobacter metallireducens	Proteobacteria	Yes	Freshwater sediment	Facultative anaerobe	Neutrophile	No	Mesophile
Hydrogenobacter thermophilus	Aquificae	Yes	Thermal spring sediments	Facultative anaerobe	Neutrophile	No	Thermophile
Syntrophothermus lipocalidus	Firmicutes	Yes	Sludge from reactor	Anaerobe	Neutrophile	No	Thermophile
Thermodesulfobacterium geofontis	Thermodesulfobacteria	Yes	Thermal springs	Anaerobe	Neutrophile	No	Hyperthermophile
Thermus scotoductus str SA-01	Deinococcus/Thermus	No	Groundwater from mine	Anaerobe	Neutrophile	No	Thermophile

Table 4. Environmental Niches Inhabited by Bacteria and Archaea that Possess Homologs of the NAD+ Dependent Formate Dehydrogenase α Subunit.

Species	Phylum	Sec synthesizer?	Environment	Oxygen Req	pH	Halophile?	Temperature
Acetohalobium arabaticum	Firmicutes	Yes	Saline lakes	Anaerobe	Neutrophile	Yes	Mesophile
Aciduliprofundum boonei	Euryarchaeota	No	Hydrothermal vents	Facultative anaerobe	Acidophile	No	Thermophile
Bacillus selenitireducens	Firmicutes	Yes	Soda lake	Facultative anaerobe	Alkaliphile	Yes	Mesophile
Bacillus sp. 1NLA3E	Firmicutes	Yes	Nitrate and uranium contaminated soil	Facultative anaerobe	Unknown	No	Mesophile
Bacillus vireti	Firmicutes	Yes	Soil	Facultative anaerobe	Neutrophile	No	Mesophile
Caldisericum exile	Caldiserica	No	Thermal springs	Anaerobe	Neutrophile/Mild acidophile	No	Thermophile
Carboxydothermus hydrogenoformans	Firmicutes	Yes	Thermal springs	Anaerobe	Unknown	No	Hyperthermophile
Desulfomonile tiedjei	Proteobacteria	Yes	Sludge	Facultative anaerobe	Neutrophile	No	Mesophile
Desulfovibrio salexigens	Proteobacteria	Yes	Sediments	Facultative anaerobe	Neutrophile	No	Mesophile
Desulfitobacterium hafniense str TCE-1	Firmicutes	Yes	Sludge	Facultative anaerobe	Neutrophile	No	Mesophile
Methanocaldococcus fervens	Euryarchaeota	Yes	Hydrothermal vents	Anaerobe	Neutrophile	No	Thermophile
Methanosphaerula palustris	Euryarchaeota	No	Anoxic sediments	Anaerobe	Mild acidophile	No	Mesophile
Thermosediminibacter oceani	Firmicutes	Yes	Anoxic deep sea sediments	Anaerobe	Neutrophile/Mild alkaliphile	No	Thermophile
Thermoanaerobacter kivui	Firmicutes	Yes	Anoxic freshwater sediments	Anaerobe	Neutrophile/Mild acidophile	No	Thermophile

Table 5. Environmental Niches Inhabited by Bacteria and Archaea that Possess Homologs of the HesB-like Protein.

Species	Phylum	Sec synthesizer?	Environment	Oxygen Req	pH	Halophile?	Temperature
Bacillus massiliosenegalensis	Firmicutes	Yes	Human digestive tract	Facultative anaerobe	Neutrophile	No	Mesophile
Bacillus selenitireducens	Firmicutes	Yes	Soda lake	Facultative anaerobe	Alkaliphile	Yes	Mesophile
Bacillus sp. ZYK	Firmicutes	Yes	Sediments	Facultative anaerobe	Neutrophile	No	Mesophile
Bacillus sp. 1NLA3E	Firmicutes	Yes	Nitrate and uranium contaminated soil	Facultative anaerobe	Unknown	No	Mesophile
Calditerrivibrio nitroreducens	Deferribacteres	Yes	Thermal springs	Anaerobe	Neutrophile	No	Thermophile
Desulfonatronovibrio hydrogenovorans	Proteobacteria	Yes	Soda lake	Anaerobe	Alkaliphile	Yes	Mesophile
Desulfosporosinus meridiei	Firmicutes	Yes	Groundwater (oil contaminated)	Anaerobe	Neutrophile	No	Mesophile
Desulfovibrio desulfuricans	Proteobacteria	Yes	Tar sands	Facultative anaerobe	Neutrophile	No	Mesophile
Geobacter lovleyi	Proteobacteria	Yes	Freshwater sediments	Facultative anaerobe	Neutrophile	No	Mesophile
Geobacter sulfurreducens	Proteobacteria	Yes	Terrestrial sediments (oil contaminated)	Facultative anaerobe	Unknown	No	Mesophile
Halobacillus halophilus	Firmicutes	No	Salt marsh soils	Anaerobe	Neutrophile	Yes	Mesophile
Methanocella arvoryzae	Euryarchaeota	Partial	Anoxic terrestrial sediments	Anaerobe	Neutrophile	No	Mesophile
Methanosarcina acetivorans	Euryarchaeota	No	Deep sea marine sediments	Facultative anaerobe	Neutrophile	No	Mesophile
Peptoclostridium difficile	Firmicutes	Yes	Human digestive tract	Anaerobe	Neutrophile	No	Mesophile
Planococcus antarcticus	Firmicutes	Yes	Cyanobacterial communities in freshwater lakes	Facultative anaerobe	Neutrophile	No	Psychrophile

Table 6. Environmental Niches Inhabited by Bacteria, Archaea, and Eukaryotes that Possess Homologs of the Methionine-S Sulfoxide Reductase.

Species	Phylum	Sec synthesizer?	Environment	Oxygen Req	pH	Halophile?	Temperature
Alkaliphilus oremlandii	Firmicutes	Yes	Anoxic freshwater sediment	Anaerobe	Neutrophile	No	Mesophile
Aminobacterium colombiense	Synergistetes	Yes	Industrial wastewater	Anaerobe	Neutrophile	No	Mesophile
Bacillus cellulosilyticus	Firmicutes	Yes	Terrestrial sediments	Facultative anaerobe	Alkaliphile	No	Mesophile
Bacillus selenitireducens	Firmicutes	Yes	Soda lake sediments	Facultative anaerobe	Alkaliphile	Yes	Mesophile
Bacillus sp. 72	Firmicutes	No	Unknown	Facultative anaerobe	Unknown	Unknown	Unknown
Carnobacterium jeotgali	Firmicutes	No	Human foods	Facultative anaerobe	Mild alkaliphile	No	Mesophile
Caenorhabditis briggsae	Nematoda	Yes	Terrestrial sediments	Aerobe	Neutrophile	No	Mesophile
Desulfomonile tiedjei	Proteobacteria	Yes	Sludge	Facultative anaerobe	Neutrophile	No	Mesophile
Drosophila willistoni	Arthropoda	No	Rainforests	Aerobe	Neutrophile	No	Mesophile
Haloarcula marismortui	Euryarchaeota	No	Saline lakes	Facultative anaerobe	Neutrophile	Yes	Mesophile
Haloterrigena turkmenica	Euryarchaeota	No	Saline terrestrial sediments	Facultative anaerobe	Neutrophile	Yes	Mesophile
Mesotoga prima	Thermotogae	No	Marine sediments	Anaerobe	Neutrophile	No	Mesophile
Natrialba chahannaoensis	Euryarchaeota	Yes	Soda lake sediments	Facultative anaerobe	Mild alkaliphile	Yes	Mesophile
Planococcus antarcticus	Firmicutes	Yes	Cyanobacterial communities in freshwater lakes	Facultative anaerobe	Neutrophile	No	Psychrophile
Salinicoccus luteus	Firmicutes	No	Saline terrestrial sediments	Facultative anaerobe	Mild alkaliphile	Yes	Mesophile

Table 7. Environmental Niches Inhabited by Bacteria that Possess Homologs of the Selenophosphate Synthase.

Species	Phylum	Sec synthesizer?	Environment	Oxygen Req	pH	Halophile?	Temperature
Bacillus azotoformans	Firmicutes	Yes	Terrestrial sediments	Facultative anaerobe	Neutrophile	No	Mesophile
Bacillus cellulosilyticus	Firmicutes	Yes	Terrestrial sediments	Facultative anaerobe	Alkaliphile	No	Mesophile
Bacillus nealsonii	Firmicutes	No	Mammalian digestive tract	Anaerobe	Neutrophile	No	Mesophile
Bacillus selenitireducens	Firmicutes	Yes	Soda lake sediments	Facultative anaerobe	Alkaliphile	Yes	Mesophile
Bhargavaea cecembensis	Firmicutes	No	Marine sediments	Facultative anaerobe	Neutrophile	No	Mesophile
Carnobacterium alterfunditum	Firmicutes	No	Anoxic freshwater sediments	Facultative anaerobe	Neutrophile	No	Psychorphile
Desulfosporosinus meridiei	Firmicutes	Yes	Groundwater (oil contaminated)	Anaerobe	Neutrophile	No	Mesophile
Ornithinibacillus scapharcae	Firmicutes	Yes	Associated with mullusks	Facultative anaerobe	Mild alkaliphile	No	Mesophile
Paenibacillus dendritiformis	Firmicutes	No	Terrestrial sediments	Facultative anaerobe	Mild acidophile	No	Mesophile
Salsuginibacillus kocurii	Firmicutes	Yes	Soda lake sediments	Facultative anaerobe	Mild alkaliphile	Yes	Mesophile
Thermicanus aegyptius	Firmicutes	Yes	Terrestrial sediments	Facultative anaerobe	Neutrophile	No	Thermophile
Tuberibacillus calidus	Firmicutes	Yes	Compost piles	Facultative anaerobe	Neutrophile	No	Thermophile

Table 8. Environmental Niches Inhabited by Bacteria that Possess Homologs of the SelW-like Protein.

Species	Phylum	Sec synthesizer?	Environment	Oxygen Req	pH	Halophile?	Temperature
Acetobacter tropicalis	Proteobacteria	No	Plant symbiote	Facultative anaerobe	Neutrophile	No	Mesophile
Adhaeribacter aquaticus	Bacteroidetes	No	Freshwater biofilms	Facultative anaerobe	Neutrophile	No	Mesophile
Agrobacterium tumefaciens	Proteobacteria	Partial	Plant pathogen	Facultative anaerobe	Neutrophile	No	Mesophile
Alkaliphilus metalliredigens	Firmicutes	Yes	Industrial wastewater	Anaerobe	Alkaliphile	Yes	Mesophile
Bacillus cellulosilyticus	Firmicutes	Yes	Terrestrial sediments	Facultative anaerobe	Alkaliphile	No	Mesophile
Bacillus selenitireducens	Firmicutes	Yes	Soda lake sediments	Facultative anaerobe	Alkaliphile	Yes	Mesophile
Desulfotomaculum ruminis	Firmicutes	Yes	Mammalian digestive tract	Anaerobe	Neutrophile	No	Mesophile
Haliangium ochraceum	Proteobacteria	Yes	Marine sediments	Facultative anaerobe	Neutrophile	Yes	Mesophile
Methylobacillus glycogenes	Proteobacteria	No	Plant symbiote	Facultative anaerobe	Neutrophile	No	Mesophile
Methylotenera versatilis	Proteobacteria	No	Freshwater sediments	Facultative anaerobe	Mild acidophile	No	Psychrotolerant
Psychromonas arctica	Proteobacteria	No	Marine	Facultative anaerobe	Neutrophile	No	Psychrotolerant
Streptomyces acidiscabies	Actinobacteria	No	Plant pathogen	Facultative anaerobe	Neutrophile	No	Unknown
Thermaerobacter subterraneus	Firmicutes	Yes	Thermal springs	Aerobic	Mild alkaliphile	No	Thermophile
Thioalkalivibrio nitratireducens	Proteobacteria	Yes	Soda lakes	Facultative anaerobe	Alkaliphile	Yes	Mesophile

Overall, the phylogeny of the MLS10 D-proline reductase selenoprotein homologs (Fig 2) does not mirror the phylogenetic relationships between taxa that I would expect (Fig 1) given vertical descent, as none of the bacterial phyla form the expected monophyletic clades, with the exception of the δ-Proteobacteria. For the Actinobacteria, Fusobacteria, and Chloroflexi representatives included in this phylogeny, this may be a result of the small number of representatives from these taxa included in the phylogeny. The most striking feature of the D-proline reductase phylogeny is the large phylogenetic distance between the Bacilli and Clostridia classes in the Firmicutes phylum, with the Bacilli being part of a different clade. The close relationship between the δ-Proteobacteria D-proline reductase homologs and the *Chloroflexus* sp. Y-400-fl D-proline homolog in the outer branches of the phylogeny suggest that the δ-Proteobacteria D-proline reductase homologs were inherited via horizontal gene transfer from the Gram-positive bacteria given the significant evolutionary divergence between the δ-Proteobacteria and the other phyla included in the phylogeny.

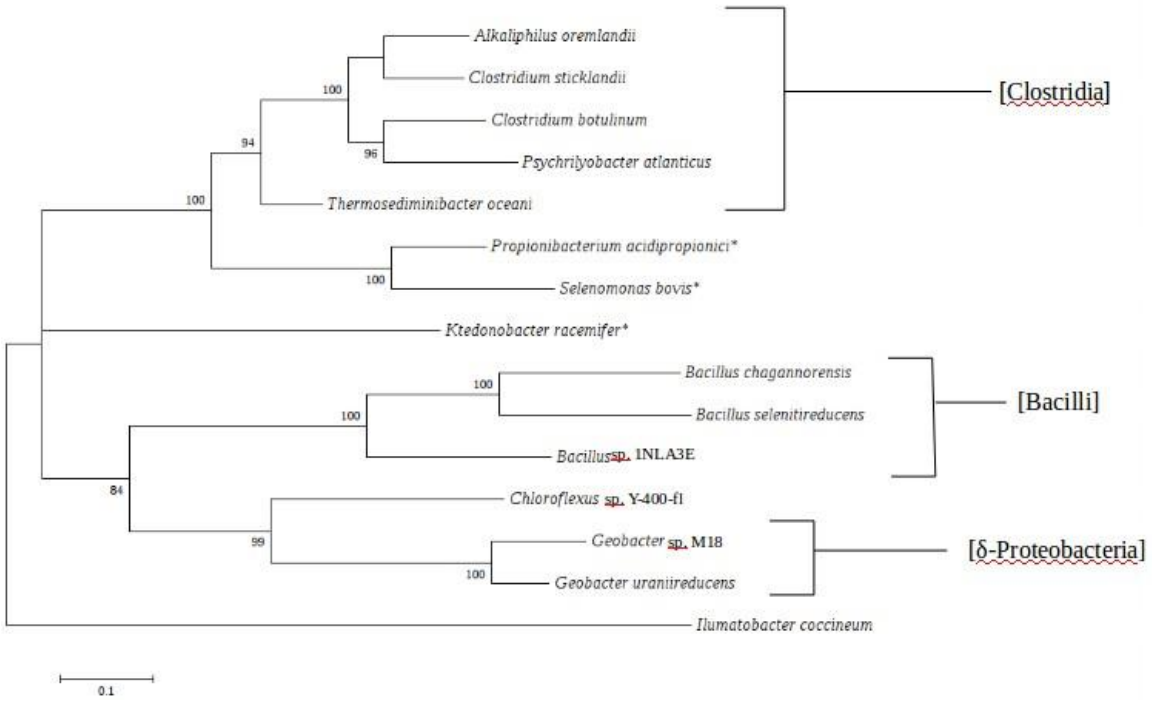


Figure 2. Bayesian Phylogeny of the MLS10 D-proline Reductase Homologs. Branch lengths correspond to evolutionary distance. An * denotes that a homologs contains a Cys residue, instead of a Sec residue.

Given that only three taxa lack a Sec residue, the incorporation of a Sec residue appears to be the ancestral state of the D-proline reductase gene. Most taxa inhabit river or marine environments or sediments (Table 2). Of the three taxa that live in association with animals (*Clostridium botulinum*, *Propionibacterium acidipropionici*, and *Selenomonas bovis*), two taxa (*Propionibacterium acidipropionici*, and *Selenomonas bovis*) lack the ability to synthesize and incorporate Sec (Table 2). Purine fermentation, including the selenium-dependent purine fermentation pathway, has thus far only been documented in the Clostridia (Kabisch et al. 1999; Self and Stadtman 2000). This phylogeny (Fig 2) suggests either that purine fermentation is much more widespread than is recognized in the current literature, or that the D-proline reductase has a different

function in these taxa and that purine fermentation is a derived trait in the Clostridia. The deep branch between the Clostridial and Bacilli/δ-Proteobacterial clades raises the latter possibility.

The formate dehydrogenase N α subunit homologs phylogeny (Fig 3), in contrast to the D-proline reductase phylogeny, follows more closely the expected relationships between bacterial taxa I would expect given vertical descent (Fig 1). The Bacilli, for example, form a monophyletic clade with their sister class in the Firmcutes, the Clostridia. As with the D-proline reductase phylogeny, this phylogeny supports a scenario of horizontal gene transfer between the δ-Proteobacteria and the Gram-positive bacteria, and this horizontal gene transfer appears to have occurred multiple times between the Clostridia and the δ-Proteobacteria. This is supported by the habitats occupied by both taxa, because both taxa inhabit anaerobic habitats in marine anoxic saltwater, freshwaters, and sediments. This phylogeny suggests that, though the use of Cys was ancestral in the bacteria, the use of a Sec residue appeared to have evolved early in the evolution of the bacteria, before the Gram-positivebacteria diverged from the Gram-negative bacteria, and has been conserved in bacterial taxa.

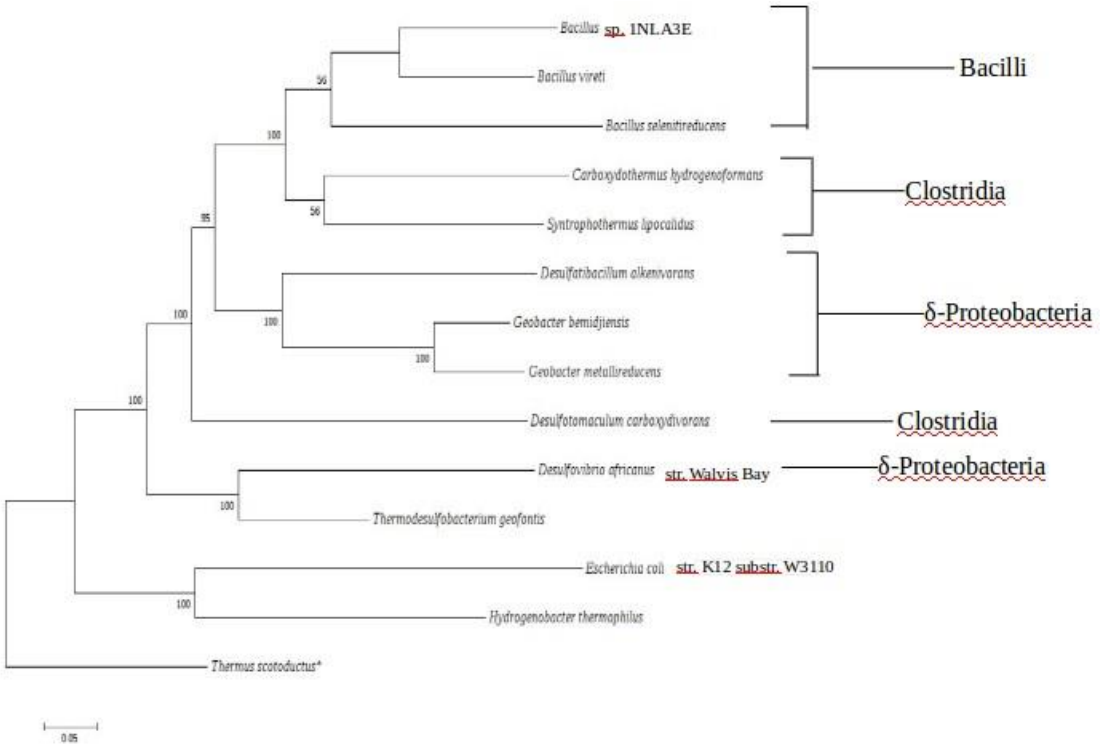


Figure 3. Bayesian Phylogeny of the MLS10 Formate Dehydrogenase N α Subunit Homologs. Branch lengths correspond to evolutionary distances.An * denotes that a homologs contains a Cys residue, instead of a Sec residue.

The phylogeny of the NAD+ dependent formate dehydrogenase α subunit (Fig 4) is similar to the formate dehydrogenase N α subunit. The Bacilli and the Clostridia classes are closely related to one another, as would be expected given that the bacili and the Clostridia are both members of the Firmicutes phylum. In comparison with the results I would expect given vertical descent, the NAD+ dependent formate dehydrogenase α subunit phylogeny shows evidence of horizontal gene transfer between the δ-Proteobacteria and the Clostridia. Additionally there is evidence for a horizontal gene transfer event between the methanogenic Archaea and the Caldiserica phylum, as the Caldiserica phylum is phylogenetically closely related to the Proteobacteria phlum

(Mori et al. 2009), making its divergence from the bacteria relatively recent. This inference is further supported by the shared habitat of *Caldisericum exile* and the methanogenic Archaea- marine hydrothermal vents (Table 4).

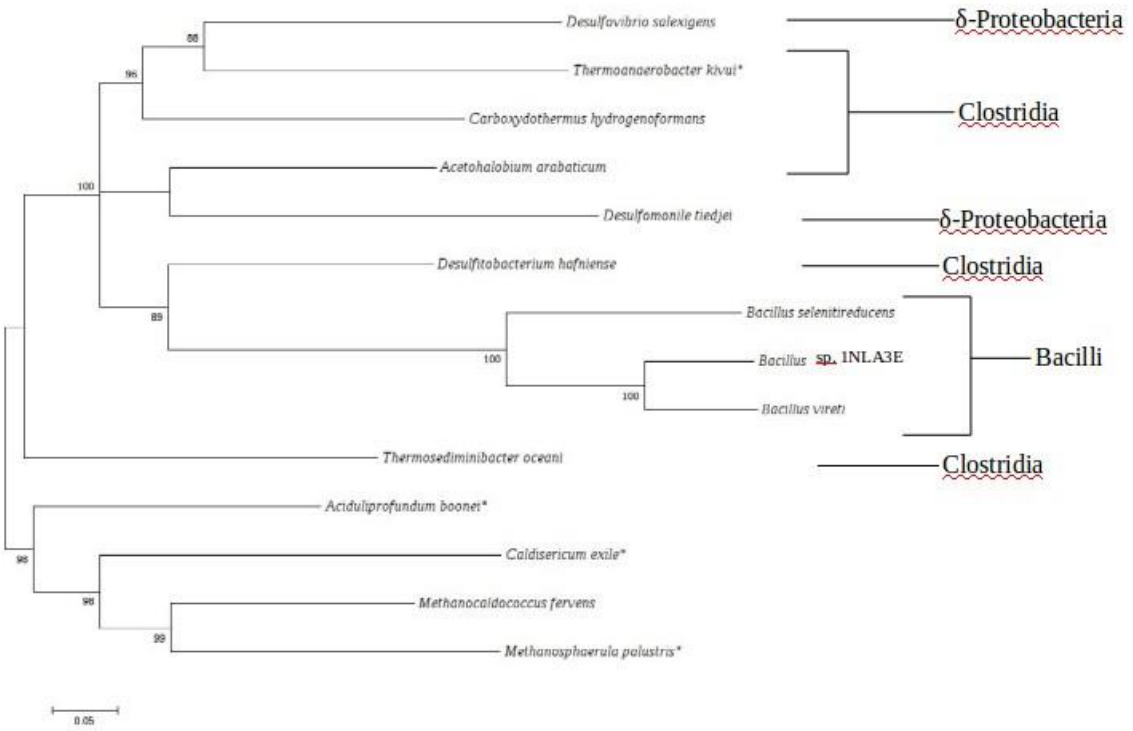


Figure 4. Bayesian Phylogeny of the MLS10 NAD+ Dependent Formate Dehydrogenase α Subunit Homologs. Branch lengths correspond to evolutionary distances. An * denotes that a homologs contains a Cys residue, instead of a Sec residue.

One taxon, *Methanocaldococcus fervens*, incorporates Sec, suggesting that the NAD+ dependent formate dehydrogenase α subunit selenoprotein is involved in both methanogenesis and acetogenesis. Nonetheless, two of the three archaeans use Cys, rather than Sec, making it difficult to infer whether Sec use is an ancestral or derived trait in the Archaea, though this phylogeny strongly suggests that Sec use in this selenoprotein is the ancestral trait in the bacteria. The ability to synthesize and incorporate Sec seems

strongly associated with the incorporation of Sec in this α subunit- three of the four taxa that use a Cys residue cannot synthesize Sec.

As discussed above, the extent to which the Bacilli and the δ-Proteobacteria generate energy via acetogenesis, as with purine fermentation, remains unknown, as it is possible that these taxa could exploit the Wood-Ljungdahl pathway for other purposes. In contrast to the D-proline reductase phylogeny, however, the phylogeny of this formate dehydrogenase α subunit does not suggest the possibility of differences between the physiological function of this α subunit (for example, autotrophic carbon assimilation, rather than acetogenesis) in the Bacilli and the δ-Proteobacteria taxa compared to the Clostridia.

Interpretation of the phylogeny of the HesB-like selenoprotein (Fig 5) is complicated by the presence of multiple polytomies and the unknown physiological function of the HesB-lke selenoprotein. The most striking feature of the phylogeny is that the Bacilli homologs of the HesB-like selenoprotein are much more closely related to δ-Proteobacteria homologs than Clostridia homologs, in contrast to what I would expect to observe given vertical descent (Fig 1). The larger number of δ-Proteobacteria taxa than Clostridia, suggests that a lateral gene transfer event from the δ-Proteobacteria to the Firmicutes may be a more likely evolutionary scenario than vice versa. As with the phylogeny of the NAD+ dependent formate dehydrogenase α subunit, this phylogeny contains representatives from the methanogenic Archaea. Cys use appears to be the ancestral condition for the Archaea, while Sec use appears to be the ancestral condition for the bacteria given that most bacteria that possess the Sec trait are more basal taxa.

The Bacilli clade is marked by extensive Sec loss, with no common environmental characteristics differentiating Bacilli taxa that utilize Sec and Bacilli taxa that utilize Cys.

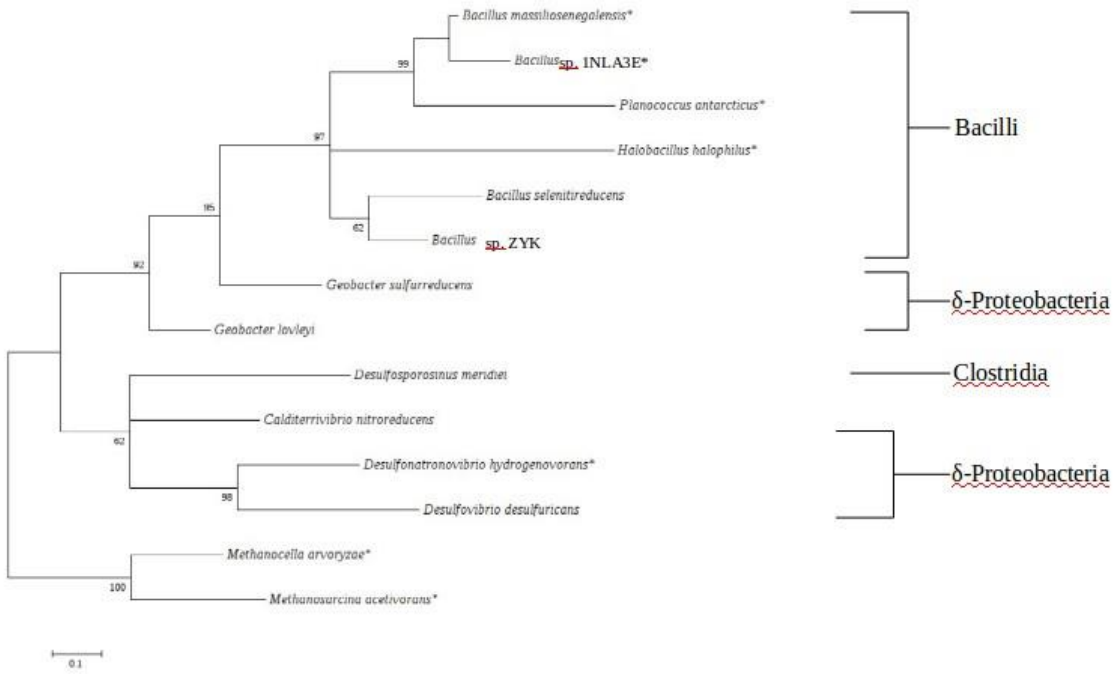


Figure 5. Bayesian Phylogeny of the MLS10 HesB-like Protein Homologs. Branch lengths correspond to evolutionary distances.An * denotes that a homologs contains a Cys residue, instead of a Sec residue.

As with the D-proline reductase phylogeny, the phylogeny of the MsrA selenoprotein (Fig 6) does not correspond to the evolutionary relationships between taxa expected for vertical descent (Fig 1). This is surprising given the wide phylogenetic distribution and deep antiquity of MsrA homologs (Delaye et al. 2007). Not only are Bacilli MsrA homologs distantly related to the Clostridial homolog, but the Bacilli MsrA constitutes a deeply branched clade in the MsrA phylogeny, a clade that is characterized by extensive loss of the Sec residue in MsrA homologs. This is unusual given that the Bacilli diverged from the Firmicutes phylum comparatively recently in the evolution of bacteria(Fig 1), particularly when compared to the evolutionary time represented in this phylogeny, which stretches back in time to before the divergence of the three domains of life from the last universal common ancestor.

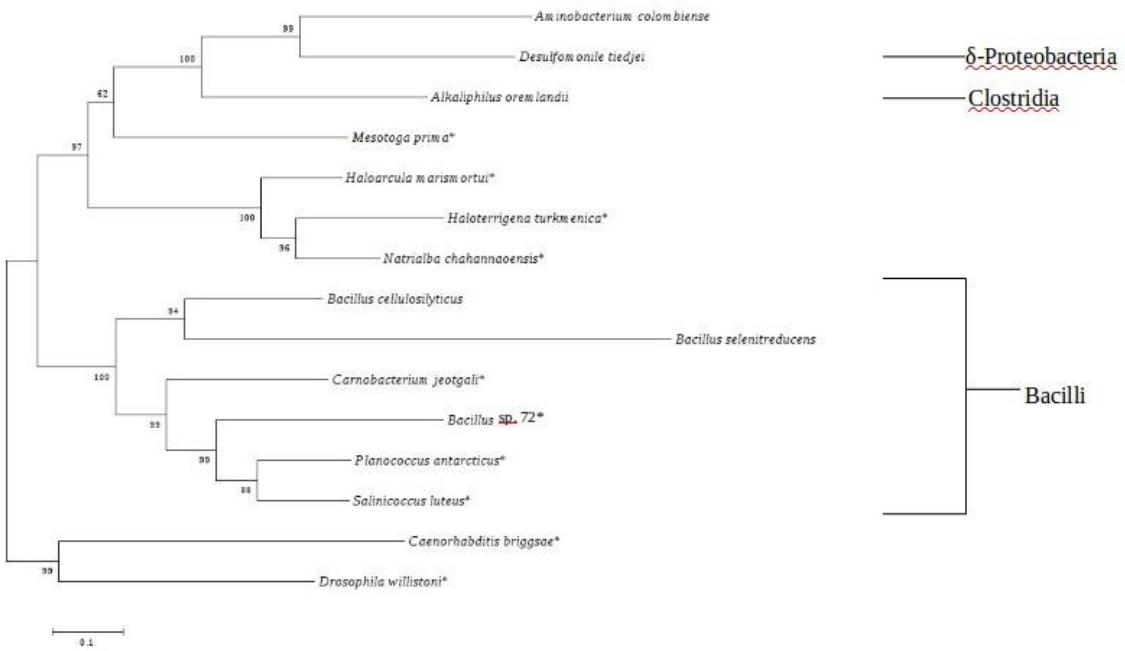


Figure 6. Bayesian Phylogeny of the MLS10 Methionine-S Sulfoxide Reductase Homologs. Branch lengths correspond to evolutionary distances.An * denotes that a homologs contains a Cys residue, instead of a Sec residue.

One feature of the phylogeny is that the eukaryotes and the archaeans could not both be rooted as an outgroup with respect to the included bacterial taxa requiring that either the eukaryotes or the archaeans be placed within the bacteria. Given that the halophilic archaeans inhabit anaerobic environments similar to the bacteria included in the phylogeny (Table 6) suggests that a lateral gene transfer event between the bacteria and the halophilic archaeans is a far more likely scenario than between the bacteria and the eukaryotes. Additionally, this phylogeny, unlike the NAD+ depednent formate dehydrogenase α subunit and the HesB-like selenoprotein phylogenies, includes representatives from the halophilic Archaea, rather than the methanogenic Archaea. Finally, neither the eukaryotes nor the archaeans incorporate the Sec residue, suggesting

that Cys use may be the ancestral trait for the MsrA protein, and the use of Sec is a derived trait within the bacteria. The ability to incorporate a Sec residue into the MsrA protein is associated with the ability to synthesize and incorporate Sec residues. Seven out of ten taxa that incorporate a Cys residue in MsrA homologs lack the ability to synthesize Sec.

The phylogeny of the MLS10 *SelD* selenoprotein genes (Fig 7) only includes representatives from the Firmicutes phylum, and is the selenoprotein phylogeny that appears to be the most like a phylogeny of a Bacilli housekeeping gene. As expected, the root of the phylogeny is a member of the Clostridia class, with all of the representatives from the Bacilli forming a monophyletic clade. There are two deeply branched clades in the Bacilli. One clade contains only taxa that incorporates a Sec residue in the SelD protein, and the other clade contains only taxa that incorporates a Cys residue in the SelD protein. Given that the only taxon from the Clostridia incorporates a Sec residue, it appears that using a Sec residue is ancestral with respect to the Bacilli. Incorporating a Sec residue into the SelD selenoprotein is associated with the ability to incorporate Sec. Four of the five taxa that incorporate a Cys residue in the SelD selenoprotein cannot synthesize and incorporate Sec.

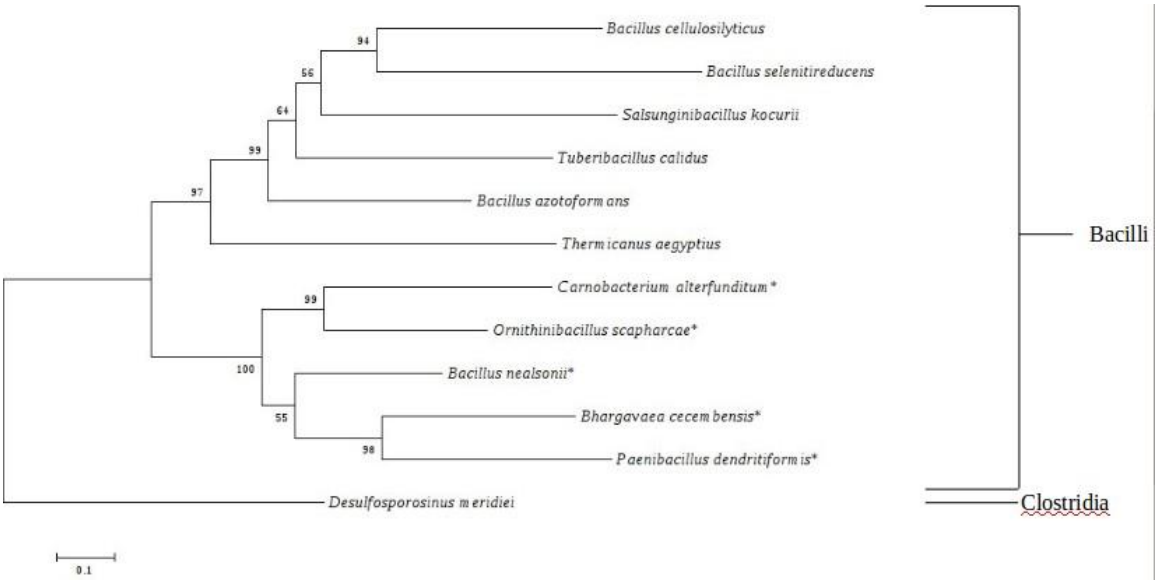


Figure 7**.** Bayesian Phylogeny of the MLS10 Selenophosphate Synthase Gene Homologs. Branch lengths correspond to evolutionary distances.An * denotes that a homologs contains a Cys residue, instead of a Sec residue.

The SelW-like selenoprotein phylogeny (Fig 8) suggests that SelW-like selenoprotein homologs constitute a deeply branched clade in the evolution of the SelW-like protein, with the use of Sec in the SelW-like selenoprotein being the ancestral trait. The evolutionary relationships between SelW-like homologs match the expected relationships between bacterial taxa I would expect given vertical descent (Fig 1). The Bacilli and the Clostridia form a monophyletic clade, as do most of the Proteobacteria. There are two exceptions that suggest horizontal gene transfer events. First, the δ-Proteobacteria SelW-like homolog is more closely related to the Firmicutes phylum than to the Proteobacteria phylum. Second, the SelW-like protein homolog from the Actinobacteria is more closely related to the Proteobacteria rather than the other Gram-positive bacteria. The presence of a Sec residue in the SelW-like protein is associated

with the ability to synthesize Sec. Seven out of eight taxa that incorporate a Cys residue cannot synthesize Sec.

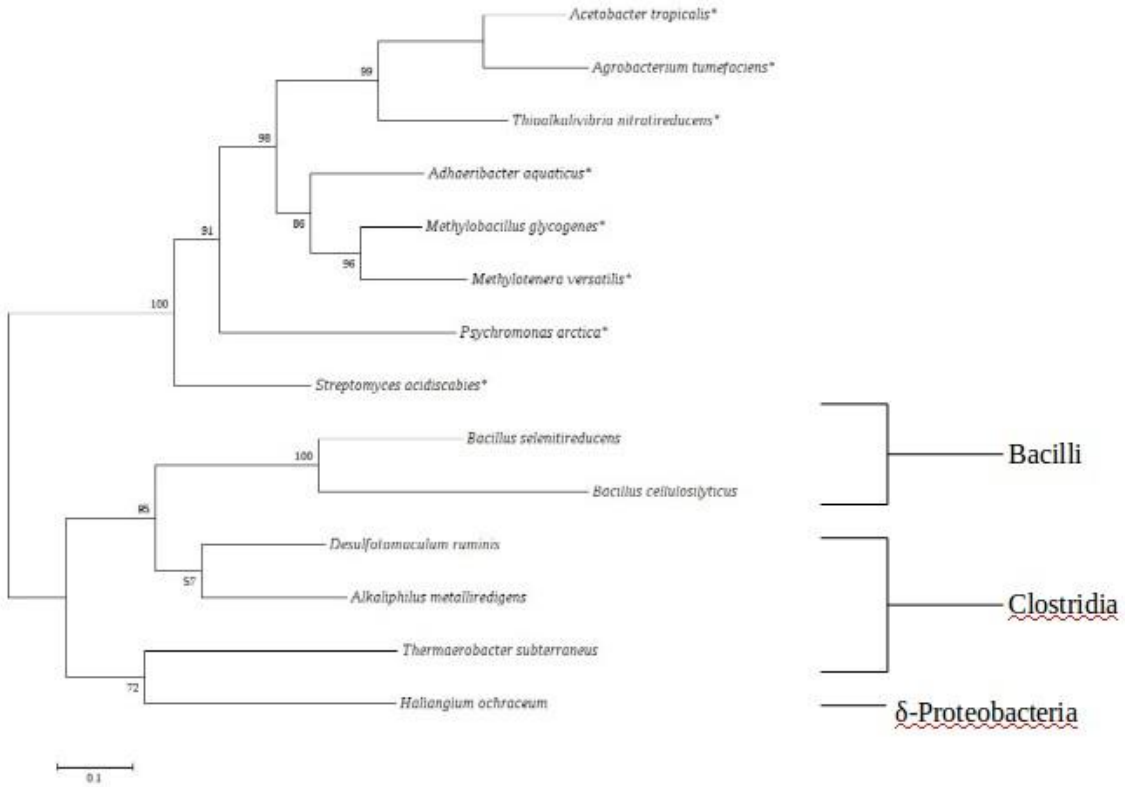


Figure 8. A Bayesian Phylogeny of the MLS10 SelW-like Protein Homologs. Branch lengths correspond to evolutionary distances.An * denotes that a homologs contains a Cys residue, instead of a Sec residue.

SDS-PAGE Results of MLS10 Soluble and Insoluble Protein Fractions

Figure 9a shows the protein profile of the soluble and insoluble fractions of MLS10 cells grown on arsenate, fumarate, nitrate, and selenite from a 7% acrylamide SDS-PAGE gel. The arsenate, fumarate, and nitrate insoluble fractions demonstrate banding patterns that are consistent with previous research on these respiratory pathways. The arsenate insoluble fractions show expression of a protein that is estimated at 108.37 kDa and is not observed in the other three fractions. This band's molecular weight is

consistent with the molecular weight of the respiratory arsenate reductase α subunit of MLS10 reported by Afkar et al. (2003), suggesting that the most likely identity of this band is the arsenate reductase. A band of protein expression was detected in the insoluble fumarate fraction that is estimated at 83.75 kDa, which is consistent with previous reports of the molecular weight of the fumarate reductase of *Wolinella succinogenes* (Lancaster et al. 1999). A dark band of expression was detected in the nitrate insoluble fractions that is estimated at approximately 121.32 kDa, which is consistent with previous research on the molecular weight of the NarG respiratory nitrate reductase α subunit(Blümle and Zumft 1991), suggesting that the darker band of the nitrate-grown insoluble fractions could be attributed to the nitrate reductase α subunit.

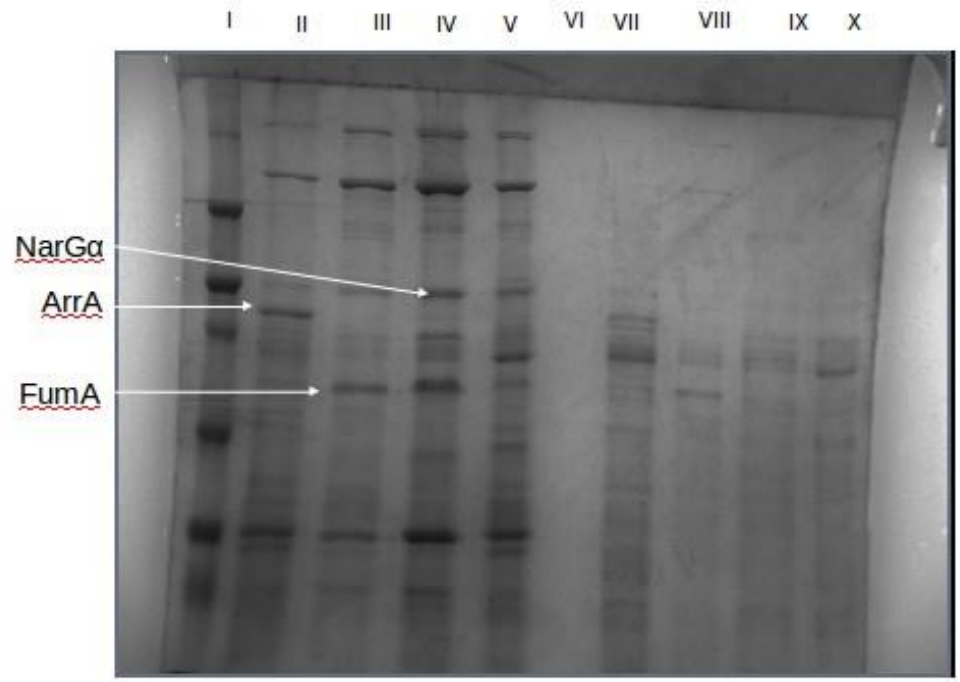


Figure 9a. 7% Acrylamide SDS-PAGE Gel of Arsenate-grown, Fumarate-grown, Nitrate-grown, and Selenite-grown MLS10 Insoluble and Soluble Proteins. Lane I: ladder. Lane II: arsenate-grown MLS10 insoluble proteins. Lane III: fumarate-grown MLS10 insoluble proteins. Lane IV: nitrate-grown MLS10 insoluble proteins. Lane V: selenite-grown MLS10 insoluble proteins. Lane VI: blank. Lane VII:arsenate-grown soluble proteins. Lane VIII: fumarate-grown soluble proteins. Lane IX: nitrate-grown soluble proteins. Lane X: selenite-grown soluble proteins. The molecular weights of the protein ladder bands are (in descending order) 260 kDa, 160 kDa, 110 kDa, 80 kDa, 60 kDa, 50 kDa, and 40 kDa. The likely candidates for the α subunits of the arsenate reductase (ArrA), fumarate reductase (FumA), and the respiratory nitrate reductase (NarGα) are highlighted.

The MLS10 selenite-grown insoluble and soluble fractions show banding patterns (Fig 10) that are distinct from the fractions obtained from MLS10 cells grown under other electron acceptors. Bands representing proteins estimated at 149.58 kDa, 127.32 kDa, 101.61 kDa, 98.39 kDa, 73.62 kDa, 69.03 kDa, and 64.72 kDa were found that were unique to the selenite insoluble protein fractions. Additionally, several bands were shared with some fractions. A particularly prominent protein band was detected in the gel with an estimated weight of 87.9 kDa that, while shared with the nitrate-grown insoluble

fraction, was significantly darker, suggesting that either the selenite-grown insoluble fractions were expressing a protein or proteins that were unique to the selenite-grown MLS10 cells, or that proteins shared with the nitrate-grown MLS10 cells were expressed at significantly higher levels. An additional band was detected in the selenite-grown fraction with an estimated molecular weight of 48.43 kDa that was shared with the arsenate insoluble fractions. Two prominent bands were detected in the selenite-grown soluble fractions that were also unique to selenite-grown MLS10 cells. These bands had estimated molecular weights of 111.93 kDa and 95.27 kDa, respectively

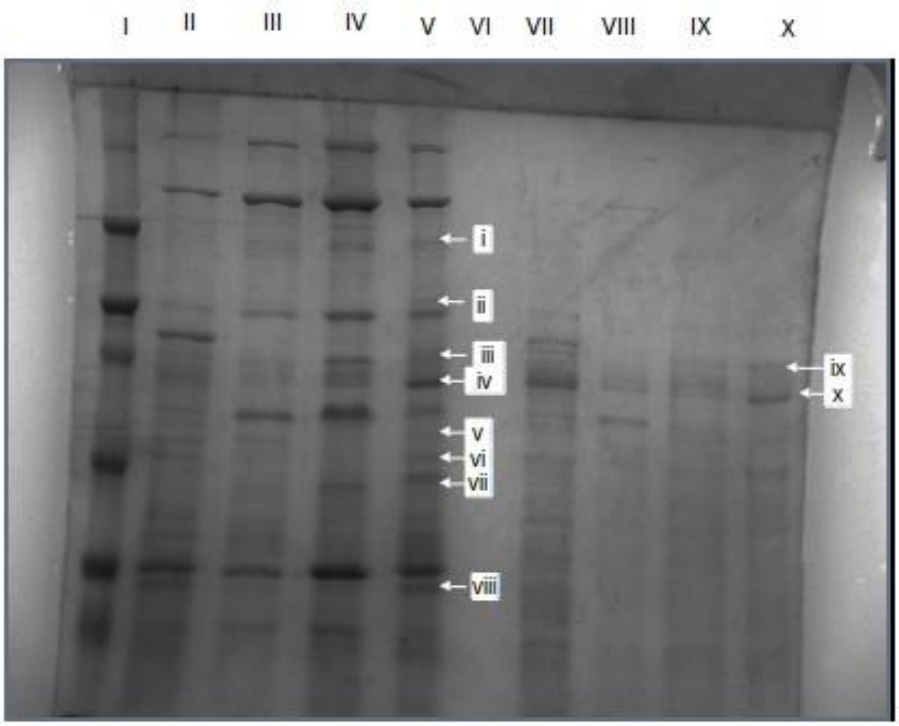


Figure 9b. 7% Acrylamide SDS-PAGE Gel of Arsenate-grown, Fumarate-grown, Nitrate-grown, and Selenite-grown MLS10 Insoluble and Soluble Proteins. Lane I: ladder. Lane II: arsenate-grown MLS10 insoluble proteins. Lane III: fumarate-grown MLS10 insoluble proteins. Lane IV: nitrate-grown MLS10 insoluble proteins. Lane V: selenite-grown MLS10 insoluble proteins. Lane VI: blank. Lane VII:arsenate-grown soluble proteins. Lane VIII: fumarate-grown soluble proteins. Lane IX: nitrate-grown soluble proteins. Lane X:selenite-grown soluble proteins. Lower case roman numerals highlight protein expression patterns that differentiate selenite-grown MLS10 cells from cells grown with other electron acceptors. The estimated molecular weights of the bands are i: 149.58 kDa, ii: 127.32 kDa, iii: 101.61 kDa, iv: 87.9 kDa, v: 73.62 kDa, vi: 69.03 kDa, vii: 64.72 kDa, viii: 48.43 kDa, ix: 111.93 kDa, and x: 95.27 kDa.

Cytochrome Content of MLS10 Cells Grown on Arsenate, Fumarate, Nitrate, and Selenite

No cytochromes were detected in any of the fractions of the arsenate-grown, fumarate-grown, nitrate-grown, or selenite-grown MLS10 cells in the pyridine hemochrome assay. I attempted the pyridine hemochrome assay with amounts of protein up to 400 μg (data not shown).

Native In-gel Assays of MLS10 Cells Grown on Arsenate, Fumarate, Nitrate, and Selenite

I ran a native gel under oxic conditions using CHAPS detergent. Only the insoluble arsenate-grown protein fractions demonstrated enzymatic activity (Figs. 10 and 11). The arsenate-grown insoluble fraction demonstrated reductase activity with both arsenate and selenite, suggesting that the arsenate reductase of MLS10 was mediating the reduction of selenite. This is consistent with Afkar et al. (2003), who noted that the arsenate reductase of MLS10 possesses both arsenate and selenite reductase activity.

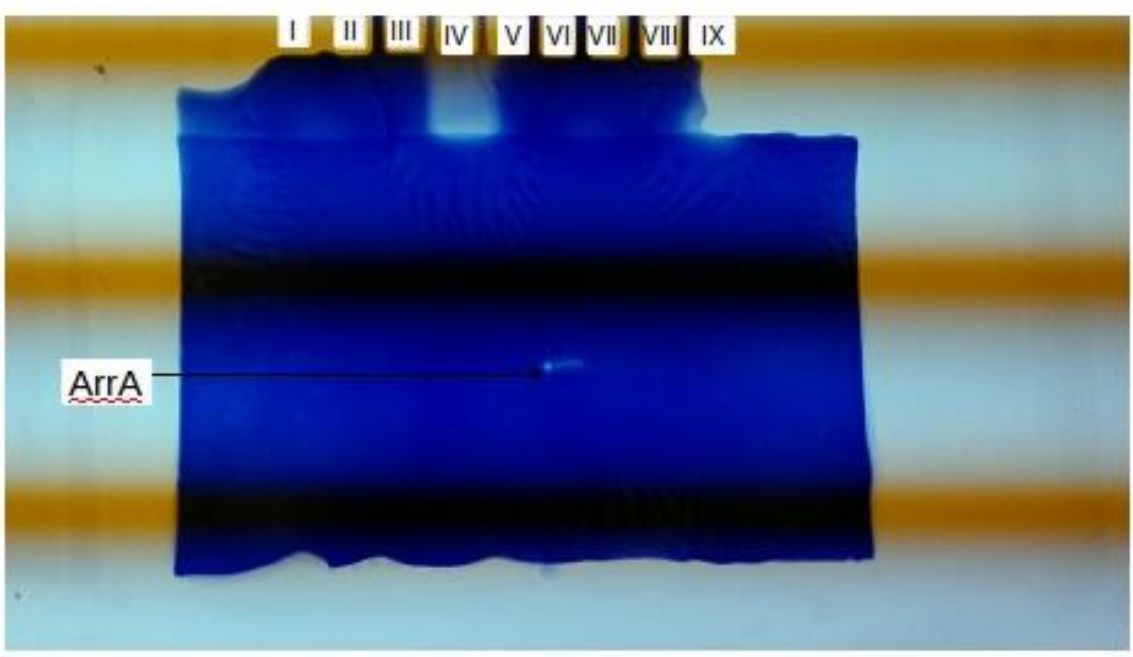


Figure 10. 7% Acrylamide Native Gel Containing 10% CHAPS Detergent Soaked in Reduced Methyl Viologen. The blue background represents reduced methyl viologen. The gel was soaked in a 5mM selenite solution. The white band representing the MLS10 ArrA homolog highlighted. Lane I: arsenate-grown MLS10 insoluble proteins. Lane II: fumarate-grown MLS10 insoluble proteins. Lane III: nitrate-grown MLS10 insoluble proteins. Lane IV: selenite-grown MLS10 insoluble proteins. Lane V: blank. Lane VI:arsenate-grown soluble proteins. Lane VII: fumarate-grown soluble proteins. Lane VIII: nitrate-grown soluble proteins. Lane IX:selenite-grown soluble proteins.

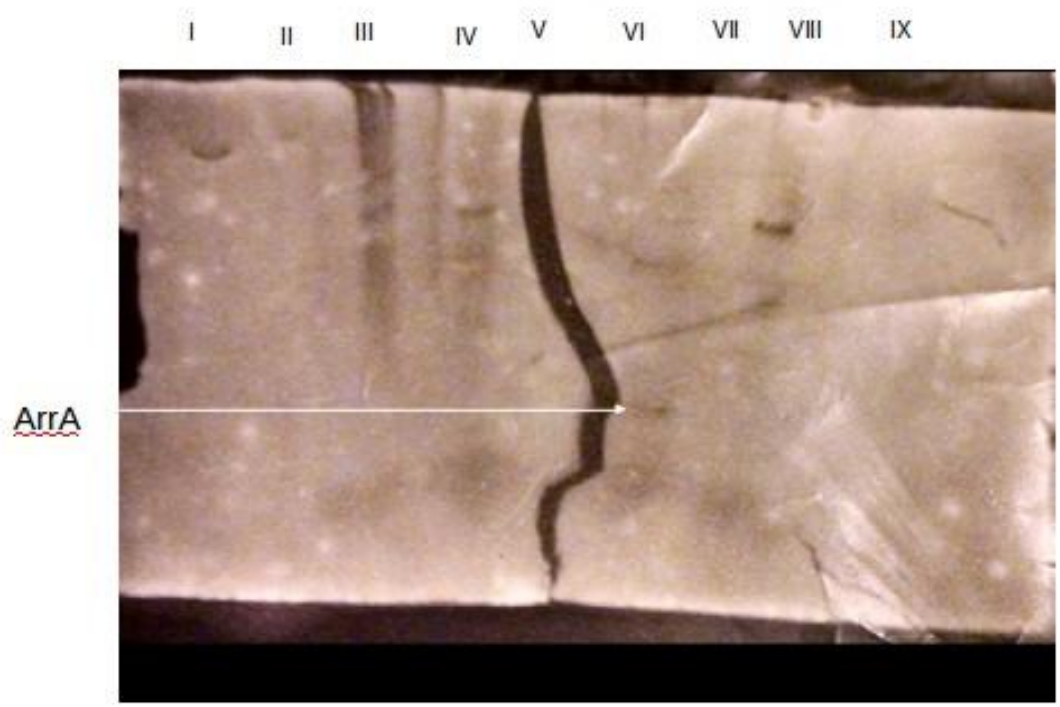


Figure 11. Image of Native 10% CHAPS Gel from Gel Used in the Native In-gel Assay. The dark band that represents the arsenate reductase of MLS10 is highlighted, confirming that selenite and arsenate reduction is being mediated by an enzyme. Lane I: arsenate-grown MLS10 insoluble proteins. Lane II: fumarate-grown MLS10 insoluble proteins. Lane III: nitrate-grown MLS10 insoluble proteins. Lane IV: selenite-grown MLS10 insoluble proteins. Lane V: blank. Lane VI:arsenate-grown soluble proteins. Lane VII: fumarate-grown soluble proteins. Lane VIII: nitrate-grown soluble proteins. Lane IX:selenite-grown soluble proteins.

CHAPTER IV

DISCUSSION

Selenium is a Strong Selective Force in the Evolution of the Bacilli that has Previously Been Unrecognized

My research has demonstrated that selenium metabolism traits (either via incorporation into selenoproteins or into selenouridine) are widespread in the Bacilli and that selenium has been a strong selective force in the evolution of the Bacilli. The Bacilli utilize Sec to a much greater extent than has previously been realized. I found that the genome of MLS10 contains seven selenoproteins, thus expanding the phylogenetic distribution of selenoprotein-rich organisms into the Bacilli. Additionally, my search for homologs of MLS10 selenoproteins yielded many Bacilli homologs, with thirteenBacilli that are capable of incorporating and synthesizing Sec included in the seven selenoprotein phylogenies I constructed. More broadly, the *SelD* gene phylogeny I constructed has demonstrated that Bacilli from many genera metabolize selenium, as the *SelD* gene is likewise required for the synthesis of the tRNA nucleoside selenouridine. These results are a start contrast to findings of Zhang et al. (2006), when only one Bacilli genome out of nineteen available genomes contained selenium metabolism traits.

The Content of the Selenoproteome of MLS10 is Similar to the Content of the Selenoproteomes of the Clostridia

My annotation of the MLS10 selenoproteome revealed that the complement of selenoproteins utilized by MLS10 is very similar to the complement of selenoproteins

exploited by the Clostridiaand the δ-proteobacteria, with three selenoproteins involved in energy generation (the D-proline reductase, the formate dehydrogenase N α subunit, and the NAD+ dependent formate dehydrogenase α subunit), one selenoprotein involved in antioxidant defense (methionine sulfoxide redutase A), one selenoprotein involved in selenocysteine and selenouridine synthesis (selenophosphate synthase), and two selenoproteins whose physiological function remains unknown (the HesB-like and SelW-like proteins). Contrary to my prediction, no novel selenoprotein candidates were found in the genes encoding molybopterin oxidoreductases or cytochromes, despite the many terminal electron acceptors MLS10 can utilize as substrates for anaerobic respiration (Blum et al. 1998).

The formate dehydrogenase N α subunit is the only selenoprotein in the MLS10 genome that is known to be involved in anaerobic respiration. The other two selenoproteins, the NAD+ dependent formate dehydrogenase α subunit and the D-proline reductase, are involved in fermentative pathways (acetogenesis and purine fermentation, respectively) a process of energy generation that has been associated with and extensively studied in the Clostridia. While the ability of MLS10to generate energy via anaerobic respiration has been elucidated, nothing is known about the extent to which MLS10is able to generate energy via fermentative pathways such as acetogenesis or purine fermentation. The genome of MLS10 possesses at least some genes for purine fermentation (a D-proline reductase gene and a xanthine dehydrogenase gene) and the full complement of genes needed for the Wood-Ljundahl pathway of acetogenesis. However, physiological confirmation of the ability of MLS10 to generate energy via

these processes is needed. Purine fermentation depends upon the xanthine dehydrogenase incorporating an inorganic selenium co-factor, and many microorganisms use the Wood-Ljungdahl pathway to autotrophically incorporate carbon, rather than to generate energy.

The lack of selenoproteins involved in anaerobic respiration in MLS10, and the fact that the composition of the selenoproteome of MLS10 is similar to the composition of the selenoprotome of many Clostridia, suggests that my prediction that Sec use varies in bacterial taxa depending on the environment a particular bacteria inhabits needs to be amended. These observations suggest that the advantage of incorporating Sec residues depends significantly on the oxidoreductase and the physiological process catalyzed. Thus, this suggests that patterns of Sec use do not vary significantly among bacterial taxa because Sec residues are only advantageous in a few specific physiological processes.

Selenoproteins in MLS10 and Other Bacilli are Ancestral and Closely Related to Clostridia Homologs

The selenoprotein phylogenies I constructed are the first phylogenies that have been constructed for these selenoproteins, with the exception of the formate dehydrogenase N α subunit phylogeny constructed by Zhang et al. (2006). These are the first selenoprotein phylogenies to include representatives from the Bacilli. The selenoprotein phylogenies I constructed demonstrate that all of the selenoproteins in MLS10 are ancestral to the Bacilli, and Bacilli selenoproteins are closely related to homologs in the Clostridia and other Gram-positive bacteria. My prediction that novel selenoprotein candidates might be found in genes involved in anaerobic respiration in MLS10 was incorrect, and thus no selenoproteins were found that were either derived in

the Bacilli, or shared only with more distantly related taxa. This suggests that the evolutionary history of selenoproteins in the Bacilli is very similar to the evolutionary history of selenoproteins in other Gram-positive bacteria.

My prediction that selenoproteins in MLS10 are ancestral was supported in each phylogeny, though horizontal gene transfer and Sec loss events complicate the interpretation of my phylogenies in several instances. The formate dehydrogenase α subunit, the *SelD*, and the SelW-like selenoprotein phylogenies all clearly support a scenario where the use of Sec is ancestral, and that the Bacilli homologs cluster closely with the Clostridia homologs, as would be expected given vertical descent. The D-proline reductase, the HesB-like, and the MsrA selenoprotein phylogenies, while supporting a scenario where Sec use is ancestral to the Bacilli, suggest that horizontal gene transfer and Sec loss events also shape the evolution of selenoproteins in the Bacilli. The D-proline reductase and HesB-like phylogenies raise the possibility that selenoprotein homologs in the Bacilli were inherited via horizontal gene transfer. For the D-proline reductase phylogeny, the Bacilli homologs could have been inherited from other Gram-positive bacteria, given the divergence between Bacilli and Clostridiahomologs. A more in-depth phylogeny of the D-proline reductase selenoprotein in the Gram-positive bacteria would be required to resolve this. For the HesB-like selenoprotein phylogeny, the Bacilli homologs seem to have been inherited from the δ-proteobacteria, but the presence of a Clostridia homolog suggests that this horizontal gene transfer event predated the divergence between the Bacilli and the Clostridia. The deeply branched Bacilli clade in the MsrA phylogeny is difficult to

interpret. A horizontal gene transfer event is difficult to invoke, because there are no other taxa included in the clade that would suggest a potential horizontal gene transfer event. It is possible that the deep evolutionary divergence between the Bacilli and the other bacteria represent a gene duplication event, with the Bacilli homologs representing a paralogue of the MsrA gene and the other bacteria and halophilic Archaea representing another. A more extensive phylogeny of the MsrA protein in prokaryotes would be required to determine if this explanation is plausible. In the HesB-like and MsrA phylogenies, the Bacilli clade appears to be undergoing, or to have undergone, a significant Sec loss event.

Horizontal Gene Transfer has Significantly Influenced the Evolution of Selenoproteins in the Gram-positive Bacteria and the δ-proteobacteria

My findings are significant for our understanding of the evolution of selenoproteins in the bacteria because it suggests a compelling explanation for why the Gram-positive Firmicutes and the δ-proteobacteria are both selenoprotein rich taxa. Both taxa utilize a similar array of selenoproteins because extensive horizontal gene transfer indelibly links the evolution of selenoproteins in these taxa. My findings therefore suggest new directions for the study of selenoproteins in the bacteria. The selenoproteomes of the Gram-positive bacteria and the δ-proteobacteria need to be systematically characterized to understand which selenoproteins are shared between the two taxa and which selenoproteins, if any, are exclusive to the Gram-positive bacteria and to the δ-proteobacteria. This information, coupled with knowledge of the physiological function of these selenoproteins, will provide insight into the shared selective pressures that made the active evolution of an unusually large number of selenoproteins

advantageous in these taxa. This will ultimately lead to more refined hypotheses concerning the specific advantage Sec incorporation offers to biological organisms in all three domains of life.

My results show compelling evidence that extensive horizontal gene transfer characterizes the evolution of selenoproteins in the Gram-positive bacteria and the δ-proteobacteria. Selenoprotein phylogenies feature topologies in which the δ-proteobacteria consistently clustered with the Clostridia or other Gram-positive bacteria in every phylogeny that included representatives from the δ-proteobacteria. This feature was consistent even when representatives from other subphyla in the Proteobactera were included. This finding is surprising because the phylogeny of bacteria based on housekeeping genes (Fig 1) demonstrates that these taxa are distantly related. Moreover, the fact that this topology was consistent in all the phylogenetic methods (Bayesian, maximum likelihood, maximum parsimony, and neighbor-joining) suggests that this clustering reflects a genuine evolutionary insight, rather than a shortcoming of a particular phylogenetic method.

It is difficult to construct a plausible explanation of these findings that does not invoke horizontal gene transfer. For example, it is possible that these phylogenies reflect an evolutionary scenario where the δ-proteobacteria selenoprotein homologs cluster with the Gram-positive bacteria because these homologs were subsequently lost in the other bacterial taxa. For rare selenoproteins, this would indeed be a scenario that would yield topologies similar to the topologies in my results. However, the crucial difference is that given the deep evolutionary divergence between these two taxa (Fig 1), I would expect

that the δ-proteobacteria homologs would cluster together in a deeply branched clade from the Gram-positive bacteria. Instead, my results show topologies with theδ-proteobacteria homologs interspersed with Gram-positive bacteria homologs. The most parsimonious interpretation of such topologies is that the phylogeny reflects an evolutionary scenario where δ-proteobacteria and Gram-positive selenoprotein homologs have undergone multiple horizontal gene transfer events in the evolution of selenoproteins.

My research shows additional evidence that the inheritance of many selenoproteins was specifically inherited in the δ-proteobacteria from the Gram-positive bacteria, but that active horizontal gene transfer events from the δ-proteobacteria to the Gram-positive bacteria has also occurred in the evolution of rare selenoproteins. The most compelling example is the phylogeny of the formate dehydrogenase Nα subunit (Fig 3). Zhang et al. (2006) found that this formate dehydrogenase α subunit was the most ubiquitous selenoprotein in bacterial taxa, and was suggested to be a key driver in the conservation of Sec utilizing traits. The formate dehydrogenase Nα subunit phylogeny offers compelling evidence that the α subunit was inherited in the δ-proteobacteria from the Firmicutes phylum, as the δ-proteobacteria homologs clustered with the Gram-positive clade, rather than with the Clostridia homologs clustering with the Proteobacteria clade. The phylogenies of selenoprotein synthesis proteins (Sel A and Sel B) constructed by Zhang et al. (2006) offer support for this interpretation, as a similar topology was found in Sec synthesis proteins, suggesting that the ability to exploit Sec residues and the most conserved selenoproteins in the δ-proteobacteria was inherited from the Gram-

positive bacteria, particularly the Firmicutes phylum. Many of the more rare selenoprotein phylogenies in my results also suggested that these selenoproteins were inherited in the δ-proteobacteria from the Gram-positive bacteria via horizontal gene transfer, as the D-proline reductase (Fig 2), NAD+ dependent formate dehydrogenase α subunit (Fig 4) phylogenies show the Gram-positive taxa occupied basal positions in both phylogenies, suggestingan origin in the Gram-positive bacteria. Nonetheless, inheritance in the Gram-positive bacterica, particularly the Firmicutes phylum, via horizontal gene transfer in the δ-proteobacteria for other rare selenoproteins also seems to have occurred, as the HesB-like protein phylogeny (Fig 5) places the δ-proteobacterial homologs as the most basal bacterialtaxa in the phylogeny.

SDS-PAGE Gels Suggest that Selenite Respiration is an Inducible Pathway in MLS10, but Future Work is Needed to Test this Hypothesis

My work on selenite respiration in MLS10 has offered some support for my hypothesis that selenite respiration is an ancient respiratory pathway. The results from the SDS-PAGE gel have revealed that both the insoluble and soluble fractions of MLS10 cells grown on selenite as an electron acceptor differ from MLS10 cells grown on arsenate, fumarate, and nitrate as electron acceptors, which has demonstrated that MLS10 expresses proteins that are induced only when grown in the presence of selenite. The in-gel enzyme assays further corroborated these results, by offering compelling evidence that selenite respiration is not mediated by the arsenate reductase, a possibility speculated upon by Afkar et al. (2003). My in-gel enzyme assays have demonstrated that the arsenate reductase maintains functionality even under oxic conditions. The fact that only the arsenate grown MLS10 cells demonstrated arsenate and selenite reductase activity

suggests that the enzyme that mediates selenite respiration in MLS10 differs from the arsenate reductase that mediates arsenate respiration. This is further corroborated by my finding that while MLS10 is capable of growing on arsenate without the addition of cysteine-HCl as a reducing agent, growth on selenite is only achieved with the addition of cysteine-HCl, thus strongly suggesting that the enzymes that mediate selenite respiration are very sensitive to oxygen.

Future work is needed to test this hypothesis. No cytochrome data was obtained from MLS10 cells grown on any electron acceptor using the pyridine hemochrome assay, which has been successfully applied to bacteria that produce a sufficient quantity of cytochromes for a spectrophotometer to detect the cytochrome absorbance spectra in the midst of cellular debris (Liu and Peck 1981; Stolz et al. 1997). Ammonium sulfate precipitation, coupled with ion and gel filtration chromatography has been demonstrated to be an effective technique for identifying the cytochromes involved in respiratory pathways in bacteria that do not produce a sufficient quantity of cytochromes to be detected with a spectrophotometer, as it isolates the cytochromes from the bacterial proteins that prevent the spectrophotometer from obtaining absorbance spectra (Tsapin et al. 2001; Bell et al. 2009). This allows for a spectrophotometer to detect the absorbance spectra of the cytochromes needed to identify the cytochrome type (if cytochromes are exploited by MLS10 to respire selenite) involved in selenite respiration. The detection of distinct α, β, and soret peaks in a purified fraction of MLS10 proteins will be evidence that MLS10 exploits cytochromes in the electron transport chain during selenite

respiration. If no peaks are detected in any purified fraction, this will suggest that, as with arsenate respiration, MLS10 does not exploit cytochromes to respire selenite.

Ultimately, supporting my hypothesis that selenite respiration is an ancient metabolic pathway will consist of the identification of a selenite reductase that is induced only when MLS10 cells are grown on selenite, and the identification of genes involved in selenite respiration in other bacterial taxa. Several methods are available to purify the selenite reductase of MLS10. In-gel enzyme assays under anoxic conditions in an anaerobic glove box preserves the enzymatic activity of oxygen sensitive enzymes, and the bands of reductase activity can be excised from the gel to identify candidates for the selenite reductase in MLS10. This technique was successfully applied by Richey et al. (2009) to identify the arsenite oxidase of *Alkalilimnicola ehrlichii*. An additional approach to identifying the selenite reductase involves using mass spectrometry to identify the proteins MLS10 cells expresses when grown on arsenate, fumarate, nitrate, and selenite as terminal electron acceptors. This will allow for the identification of a subset of proteins that MLS10 only expresses when grown on selenite. This approach was successfully applied by Thomas et al. (Thomas et al. 2014) to characterize the proteins involved in the ability of *Alkaliphilus oremlandii* to generate energy via the respiration of the organoarsenical roxarsone. Because the selenite reductase pathway has not been elucidated yet, it must be demonstrated that a candidate for the selenite reductase actually reduces selenite *in vivo.* Both methods would therefore require that candidates for the selenite reductase gene be transformed into *E. coli*, and then batches of *E. coli* would be grown in media containing selenite as a terminal electron acceptor, as

was done by Saltikov and Newman (Saltikov and Newman 2003) to identify the respiratory arsenate reductase that mediates arsenate respiration in *Shewanella oneidensis*. The ability of *E. coli* to grow on selenite would positively identify an enzyme as the selenite reductase. An additional approach would involve the purification of the selenite reductase from MLS10 cells using ion exchange and high performance liquid chromatography to obtain pure quantities of the selenite reductase (Afkar et al. 2003). This method would not require transformation of the selenite reductase into *E. coli* because the selenite reductase would be purified from fractions of MLS10 proteins that demonstrate selenite reductase activity, thus offering confirmation that the purified enzyme actually mediates the reduction of selenite *in vivo*.

BIBLIOGRAPHY

Afkar, E., J. Lisak, C. Saltikov, P. Basu, R. S. Oremland, and J. F. Stolz. 2003. The respiratory arsenate reductase from *Bacillus selenitireducens* strain MLS10. FEMS Microbiol. Lett. 226:107–112.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. J. Mol. Biol. 215:403–410.

Altschul, S. F., T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25:3389–3402.

Andreesen, J. R., and L. G. Ljungdahl. 1974. Nicotinamide adenine dinucleotide phosphate-dependent formate dehydrogenase from *Clostridium thermoaceticum*: purification and properties. J. Bacteriol. 120:6–14.

Bell, P. D., Y. Xin, and R. E. Blankenship. 2009. Purification and characterization of cytochrome c6 from *Acaryochloris marina*. Photosynth. Res. 102:43–51.

Berg, B. L., J. Li, J. Heider, and V. Stewart. 1991. Nitrate-inducible formate dehydrogenase in *Escherichia coli* K-12. I. Nucleotide sequence of the fdnGHI operon and evidence that opal (UGA) encodes selenocysteine. J. Biol. Chem. 266:22380–22385.

Blum, J. S., J. F. Stolz, A. Oren, and R. S. Oremland. 2001. *Selenihalanaerobacter shriftii* gen. nov., sp. nov., a halophilic anaerobe from Dead Sea sediments that respires selenate. Arch. Microbiol. 175:208–219.

Blümle, S., and W. G. Zumft. 1991. Respiratory nitrate reductase from denitrifying *Pseudomonas stutzeri*, purification, properties and target of proteolysis. Biochim. Biophys. Acta BBA - Bioenerg. 1057:102–108.

Bradford, M. M. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. 72:248–254.

Brot, N., L. Weissbach, J. Werth, and H. Weissbach. 1981. Enzymatic reduction of protein-bound methionine sulfoxide. Proc. Natl. Acad. Sci. U. S. A. 78:2155–215

Chambers, I., J. Frampton, P. Goldfarb, N. Affara, W. McBain, and P. R. Harrison. 1986The structure of the mouse glutathione peroxidase gene: the selenocysteine in the active site is encoded by the "termination" codon, TGA. EMBO J. 5:1221–1227

Delaye, L., A. Becerra, L. Orgel, and A. Lazcano. 2007. Molecular evolution of peptide methionine sulfoxide reductases (MsrA and MsrB): on the early development of a mechanism that protects against oxidative damage. J. Mol. Evol. 64:15–32.

Douady, C. J., F. Delsuc, Y. Boucher, W. F. Doolittle, and E. J. P. Douzery. 2003. Comparison of Bayesian and maximum likelihood bootstrap measures of phylogenetic reliability. Mol. Biol. Evol. 20:248–254.

Edgar, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32:1792–1797.

Forchhammer, K., and A. Böck. 1991. Selenocysteine synthase from *Escherichia coli*. Analysis of the reaction sequence. J. Biol. Chem. 266:6324–6328.

Fuchs, G. 2011. Alternative Pathways of Carbon Dioxide Fixation: Insights into the Early Evolution of Life? Annu. Rev. Microbiol. 65:631–658.

Gursinsky, T., J. Jäger, J. R. Andreesen, and B. Söhling. 2000. A selDABC cluster for selenocysteine incorporation in *Eubacterium acidaminophilum*. Arch. Microbiol. 174:200–212.

Heider, J., C. Baron, and A. Böck. 1992. Coding from a distance: dissection of the mRNA determinants required for the incorporation of selenocysteine into protein. EMBO J. 11:3759–3766.

Huber, R., M. Sacher, A. Vollmann, H. Huber, and D. Rose. 2000. Respiration of arsenate and selenate by hyperthermophilic archaea. Syst. Appl. Microbiol. 23:305–314.

Kabisch, U. C., A. Gräntzdörffer, A. Schierhorn, K. P. Rücknagel, J. R. Andreesen, and A. Pich. 1999. Identification of d-Proline Reductase from *Clostridium sticklandii* as a Selenoenzyme and Indications for a Catalytically Active Pyruvoyl Group Derived from a Cysteine Residue by Cleavage of a Proprotein. J. Biol. Chem. 274:8445–8454.

Kim, H.-Y., Y. Zhang, B. C. Lee, J.-R. Kim, and V. N. Gladyshev. 2009. The selenoproteome of Clostridium sp. OhILAs: characterization of anaerobic bacterial selenoprotein methionine sulfoxide reductase A. Proteins 74:1008–1017.

Krafft, T., A. Bowen, F. Theis, and J. M. Macy. 2000. Cloning and sequencing of the genes encoding the periplasmic-cytochrome B-containing selenate reductase of *Thauera selenatis*. DNA Seq. J. DNA Seq. Mapp. 10:365–377.

Kramer, G. F., and B. N. Ames. 1988. Isolation and characterization of a selenium metabolism mutant of *Salmonella typhimurium*. J. Bacteriol. 170:736–743.

Kryukov, G. V., and V. N. Gladyshev. 2004. The prokaryotic selenoproteome. EMBO Rep. 5:538–543.

Lancaster, C. R. D., A. Kröger, M. Auer, and H. Michel. 1999. Structure of fumarate reductase from *Wolinella succinogenes* at 2.2 |Å resolution. Nature 402:377–385.

Leinfelder, W., K. Forchhammer, B. Veprek, E. Zehelein, and A. Böck. 1990. In vitro synthesis of selenocysteinyl-tRNA(UCA) from seryl-tRNA(UCA): involvement and characterization of the selD gene product. Proc. Natl. Acad. Sci. U. S. A. 87:543–547.

Leinfelder, W., K. Forchhammer, F. Zinoni, G. Sawers, M. A. Mandrand-Berthelot, and A. Böck. 1988.*Escherichia coli* genes whose products are involved in selenium metabolism. J. Bacteriol. 170:540–546.

Liu, M. C., and H. D. Peck. 1981. The isolation of a hexaheme cytochrome from *Desulfovibrio desulfuricans* and its identification as a new type of nitrite reductase. J. Biol. Chem. 256:13159–13164.

Lucas, S., A. Copeland, A. Lapidus, T. Glavina del Rio, E. Dalin, H. Tice, D. Bruce, L. Goodwin, S. Pitluck, D. Sims, T. Brettin, J. C. Detter, C. Han, F. Larimer, M. Land, L. Hauser, N. Kyrpides, G. Ovchinnikova, and J. Stolz. 2013. *Bacillus selenitireducens* MLS10 chromosome, complete genome.

Markowitz, V. M., F. Korzeniewski, K. Palaniappan, E. Szeto, G. Werner, A. Padki, X. Zhao, I. Dubchak, P. Hugenholtz, I. Anderson, A. Lykidis, K. Mavromatis, N. Ivanova, and N. C. Kyrpides. 2006. The integrated microbial genomes (IMG) system. Nucleic Acids Res. 34:D344–D348.

Martin W, Baross J, Kelley D, and Russell MJ. 2008. Hydrothermal vents and the origin of life. Nat. Rev. Microbiol. 6:805–14.

McEwan, A. G., J. P. Ridge, C. A. McDevitt, and P. Hugenholtz. 2002. The DMSO Reductase Family of Microbial Molybdenum Enzymes; Molecular Properties and Role in the Dissimilatory Reduction of Toxic Elements. Geomicrobiol. J. 19:3–21.

Mori, K., K. Yamaguchi, Y. Sakiyama, T. Urabe, and K. Suzuki. 2009. *Caldisericum exile*gen. nov., sp. nov., an anaerobic, thermophilic, filamentous bacterium of a novel bacterial phylum, Caldiserica phyl. nov., originally called the candidate phylum OP5, and description of Caldisericaceae fam. nov., Caldisericales ord. nov. and Caldisericia classis nov. Int. J. Syst. Evol. Microbiol. 59:2894–2898.

Oelgeschläger, E., and M. Rother. 2008. Carbon monoxide-dependent energy metabolism in anaerobic bacteria and archaea. Arch. Microbiol. 190:257–269.

Pearce, C. I., R. A. D. Pattrick, N. Law, J. M. Charnock, V. S. Coker, J. W. Fellowes, R. S. Oremland, and J. R. Lloyd. 2009. Investigating different mechanisms for biogenic selenite transformations: *Geobacter sulfurreducens*, *Shewanella oneidensis* and *Veillonella atypica*. Environ. Technol. 30:1313–1326.

Richey, C., P. Chovanec, S. E. Hoeft, R. S. Oremland, P. Basu, and J. F. Stolz. 2009. Respiratory arsenate reductase as a bidirectional enzyme. Biochem. Biophys. Res. Commun. 382:298–302.

Ronquist, F., M. Teslenko, P. van der Mark, D. L. Ayres, A. Darling, S. Höhna, B. Larget, L. Liu, M. A. Suchard, and J. P. Huelsenbeck. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61:539–542.

Rother, M., A. Resch, R. Wilting, and A. Böck. 2001. Selenoprotein synthesis in archaea. BioFactors Oxf. Engl. 14:75–83.

Ruiz-Herrera, J., and J. A. DeMoss. 1969. Nitrate reductase complex of *Escherichia coli* K-12: participation of specific formate dehydrogenase and cytochrome b1 components in nitrate reduction. J. Bacteriol. 99:720–729.

Saltikov, C. W., and D. K. Newman. 2003. Genetic identification of a respiratory arsenate reductase. Proc. Natl. Acad. Sci. 100:10983–10988.

Schröder, I., S. Rech, T. Krafft, and J. M. Macy. 1997. Purification and characterization of the selenate reductase from *Thauera selenatis*. J. Biol. Chem. 272:23765–23768.

Self, W. T., and T. C. Stadtman. 2000. Selenium-dependent metabolism of purines: A selenium-dependent purine hydroxylase and xanthine dehydrogenase were purified from *Clostridium purinolyticum* and characterized. Proc. Natl. Acad. Sci. U. S. A. 97:7208–7213.

Stock, T., and M. Rother. 2009. Selenoproteins in Archaea and Gram-positive bacteria. Biochim. Biophys. Acta 1790:1520–1532.

Stolz, J. F., P. Basu, J. M. Santini, and R. S. Oremland. 2006. Arsenic and selenium in microbial metabolism. Annu. Rev. Microbiol. 60:107–130.

Stolz, J. F., D. J. Ellis, J. S. Blum, D. Ahmann, D. R. Lovley, and R. S. Oremland. 1999. *Sulfurospirillum barnesii* sp. nov. and*Sulfurospirillum arsenophilum* sp. nov., new members of the Sulfurospirillum clade of the epsilon Proteobacteria. Int. J. Syst. Bacteriol. 49 Pt 3:1177–1180.

Stolz, J. F., T. Gugliuzza, J. Switzer Blum, R. Oremland, and F. M. Murillo. 1997. Differential cytochrome content and reductase activity in *Geospirillum barnesii* strain SeS3. Arch. Microbiol. 167:1–5.

Stolz, J. F., and R. S. Oremland. 1999. Bacterial respiration of arsenic and selenium. FEMS Microbiol. Rev. 23:615–627.

Switzer Blum, J., A. Burns Bindi, J. Buzzelli, J. F. Stolz, and R. S. Oremland. 1998. *Bacillus arsenicoselenatis*, sp. nov., and *Bacillus selenitireducens*, sp. nov.: two haloalkaliphiles from Mono Lake, California that respire oxyanions of selenium and arsenic. Arch. Microbiol. 171:19–30.

Takai, K., H. Hirayama, Y. Sakihama, F. Inagaki, Y. Yamato, and K. Horikoshi. 2002. Isolation and metabolic characteristics of previously uncultured members of the order aquificales in a subsurface gold mine. Appl. Environ. Microbiol. 68:3046–3054.

Tamura, K., D. Peterson, N. Peterson, G. Stecher, M. Nei, and S. Kumar. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28:2731–2739.

Thomas, J. A., P. Chovanec, J. F. Stolz, and P. Basu. 2014. Mapping the protein profile involved in the biotransformation of organoarsenicals using an arsenic metabolizing bacterium. Met. Integr. Biometal Sci. 6:1958–1969.

Tormay, P., R. Wilting, J. Heider, and A. Böck. 1994. Genes coding for the selenocysteine-inserting tRNA species from *Desulfomicrobium baculatum* and *Clostridium thermoaceticum*: structural and evolutionary implications. J. Bacteriol. 176:1268–1274.

Tsapin, A. I., I. Vandenberghe, K. H. Nealson, J. H. Scott, T. E. Meyer, M. A. Cusanovich, E. Harada, T. Kaizu, H. Akutsu, D. Leys, and J. J. Van Beeumen. 2001. Identification of a Small Tetraheme Cytochrome c and a Flavocytochrome c as Two of the Principal Soluble Cytochromes c in Shewanella oneidensis Strain MR1. Appl. Environ. Microbiol. 67:3236–3244.

Veres, Z., and T. C. Stadtman. 1994. A purified selenophosphate-dependent enzyme from *Salmonella typhimurium* catalyzes the replacement of sulfur in 2-thiouridine residues in tRNAs with selenium. Proc. Natl. Acad. Sci. U. S. A. 91:8092–8096.

White, D., J. Drummond, and C. Fuqua. 2012. C1 Metabolism. Pp. 358–382 *in* The Physiology and Biochemistry of Prokaryotes.Oxford University Press.

Widdel, F., G.-W. Kohring, and F. Mayer. 1983. Studies on dissimilatory sulfate-reducing bacteria that decompose fatty acids. Arch. Microbiol. 134:286–294.

Wu, D., P. Hugenholtz, K. Mavromatis, R. Pukall, E. Dalin, N. N. Ivanova, V. Kunin, L. Goodwin, M. Wu, B. J. Tindall, S. D. Hooper, A. Pati, A. Lykidis, S. Spring, I. J. Anderson, P. D'haeseleer, A. Zemla, M. Singer, A. Lapidus, M. Nolan, A. Copeland, C. Han, F. Chen, J.-F. Cheng, S. Lucas, C. Kerfeld, E. Lang, S. Gronow, P. Chain, D. Bruce, E. M. Rubin, N. C. Kyrpides, H.-P. Klenk, and J. A. Eisen. 2009. A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea. Nature 462:1056–1060.

Yamamoto, I., T. Saiki, S. M. Liu, and L. G. Ljungdahl. 1983. Purification and properties of NADP-dependent formate dehydrogenase from *Clostridium thermoaceticum*, a tungsten-selenium-iron protein. J. Biol. Chem. 258:1826–1832.

Zhang, Y., and V. N. Gladyshev. 2005. An algorithm for identification of bacterial selenocysteine insertion sequence elements and selenoprotein genes. Bioinforma. Oxf. Engl. 21:2580–2589.

Zhang, Y., H. Romero, G. Salinas, and V. N. Gladyshev. 2006. Dynamic evolution of selenocysteine utilization in bacteria: a balance between selenoprotein loss and evolution of selenocysteine from redox active cysteine residues. Genome Biol. 7:R94.

Zinoni, F., A. Birkmann, W. Leinfelder, and A. Böck. 1987. Cotranslational insertion of selenocysteine into formate dehydrogenase from *Escherichia coli* directed by a UGA codon. Proc. Natl. Acad. Sci. U. S. A. 84:3156–3160.

Zinoni, F., J. Heider, and A. Böck. 1990. Features of the formate dehydrogenase mRNA necessary for decoding of the UGA codon as selenocysteine. Proc. Natl. Acad. Sci. U. S. A. 87:4660–4664.

APPENDIX A

MLS10 SELENOPROTEIN SEQUENCES

Note: The selenocysteine residue (U) in each protein is highlighted in yellow.

>D-proline reductase Bsel_0470 and Bsel_0469
MLKQIKHWLFKLSANRTINNRDRTNAFTTPVKPMTNWNVAFLTTAGVHLKSQEG
FDVDAGDPSVRLIPSDTDPDQLMITHTHYDTEEADKDTGAVFPLEALKKLAEEGR
IGSVAKTHYGMMGYIPETDRLDQESIPVILKQLKKEHVDVLLLSPGUYICHQSVG
LIQQAAEQAGIATASVTHLPDLTEKVSVPRALHIKFPLGRTFGQAGRSDLQEKITV
DLLEAVQNRTEDDEKIQKLPYRWRRD

>Formate dehydrogenase N α subunit Bsel_1218
MLEVSRRQFLKLSGATAATLAVVELGFDPNKAQAESRTLKTESSVITPTICPYCSV
GCGILVHVKDEDVVYTEGDPDHPINRGSLCSKGTSIRQLYTSDRRVQKPMYRAPG
SDQWEERDWDWTLDRIAEKIKQTRDESFEVTADGMPVYRTEAIASLGGAALENE
ECYMIQKFMRGMGATFIEHQARIUHSSTVAGLAPSFGRGAMTNHWNDIQHADVI
FVIGGNPAENHPISMKYVQKAKDKGAKLIVVDPRFTRTAQLSDVYAPLRSGTDIP
VMGGLMNYALQNGLYHEEYVRHYTNATFLVHDDFDFDDGLFTGYDEDSRSYDK
ATWTFQRDEDGEILTDETMQDPRCVFQLLKKHYERYDAETVSAMAGMTVDDFN
RVAETFCSTGATDKTGTIMYAMGTTQHTVGSQNVRSYAMLQLLLGNVGRPGGG
VNAMRGECNVQGSTDFALLFHLMSGYIGAPTQSANHASLAAYNENETPASGFWS
NKPKFLASLLKAYYGENATPENDFLYDYFPKGQKNYSHISLFESMHNEEIKGLIT
WGQNPMVGGPNANFEREAMTKLDWFVSMDLWETETAAFWKDNAGSNPADIDT
EVFMLPACGPYEKEGSVSNSGRWMQYRWKALEPKHDSKSDAWLVNNLAKRLK
ALYEGEQSEAAKPIQALDWNFGDGDYPDVDLVCREINGYDLKTGKTITNFTHLK
DDGTTSSGNWIYSGFYPDAGPGEDKNLAKRRDDEDTGMENYLNWSFAWPVNRR
NLYNRAGADPQGNPWSSNKETIWWDGEQWTGHDVPDFGANNDPAGPGGKNPFI
MIPHGKGGLFTDGTADGPFPEHYEPYESPIPNAFSSQELNPAVHIWGGDHNKRGK
FADFPIVATTYRLTEHWQSGSMTRQLEWPSELMPHMFVEISQELADEKGIQEKDK
VMVSTARGEIEALAMITKRFKPYTIRGEKIHHLGMPWHYGYEGIATGSIANHLTS
HIGDANTMIPEYKAFLCDVRRVEA

>NAD+ dependent formate dehydrogenase α subunit Bsel_2064
MTETKTIATTCAYCGTGCGLLVDVEDNRIVKVKGNRNAAVNEGQTCIKGAFGYH
YIHSNDRLTAPLIRKEGVLTKVSWDEAISYVADKLTQIKERFGPESFSMFACERATN
ETNYITQKFTRAVMGTNNIDGCNRTUHAPSVAGLATVFGNGAPTSSIMDVDHSDV
LLLIGSNTTDAHPIIANRMKKAAKKGLKIIVVDPRKIAMTKSADQHLQIKVGSDIA
LMNGMMRVMIKEGLYNESFVSKNAVNFEALKDQVESYTLEKTEEITGVPKADIE
SAARTYAEADRSMIAYTLGITEHHCGVNNVFDIANMALLTGHIGREGTGIMPLRG
QNNVQGAGDMGCLPNMLPGATPVSDDSFRSKLEKEWGVSLNPHVGQTQTGML
ERMETGEMKSLFVIGENPIVADVHRNHTTKLFQNLDFLVVQDLFLTETAELADVV
LPAKGWAEVEGTYTNTDRRVQKVNKAVSAPGEALDDWDVLSRLATEMGYPMH
YEHAEQIWNELRDAVPHLFGGMDYSRLTEGQSLQYPCPDVNHPGTSLLHTEFHE
SENRSAPFTPVSYTEPVEMPDAEFPFTLTTGRRYEPYNTNTQTRYYPDTLKRKQTE
ETVDMHPSDAKRLNIDDGEMVTVSSRRGTVQVKARVTDEVQEELVFMSFHWKE
TPTNVLTINEFDPISGTAEYKACAVKIEKQS

>Hes-B like protein Bsel_2426
MNITDKAKDFIQNILDENNASNIKVYVAGMGUGGPQLGLALDEPSQTDIVEEINGI
KVAFEQHVHGQTSNMALDYQETPQGSGLVMTGNESDCC

>Methionine sulfoxide reductase A Bsel_1564
MNHLQNVIFGSGUFWGPDGRFGQLEGVTATEVGYAGGDMPEPTYRQMRDHTEV
VRVTFDSDRISFERLLDEFWQHHSGKQHGYGGTQYQSLLITETTEQLQMAKRMI
QRYRETENREIETVLTMNKPFTSAEMYHQKYMLRNRSRSWQELLDQFDSEEACI
RSTFTARLNALACGELTKQELRSMLNYSVDFSGEREIFTSFLGRMKW

>Selenophosphate synthase Bsel_0346
MLKNILSGGNKPLTQTTKKAGUGCKIGPADLTQVLRHLPEGTKNENLLVGLDTSD
DGGVFKLTDDLAIVQSIDYFTPICDDPYMFGQIAAANALSDIYAMGGKPVTALNIV
GYPIKKMPPETLAEILRGGADKIQESGAVLAGGHSIDDQEPKYGLSVTGTVHPDAI
FKNVGAKTGDKLVLTKPLGAGIITTAIKFGKASEQEQKDVMTAMATLNKTAAETL
ADFHPHAVTDVTGFGLTGHGFEMASGSNVTLHISYKDVPVINGTLSHARNKVIPG
GGRENRDYLLEHVEHAPHIELADQLILSDSITSGGLLVSLPADEADAYVEAYNRA
QDTFKAAVIGHVTDFEGHAIKIR

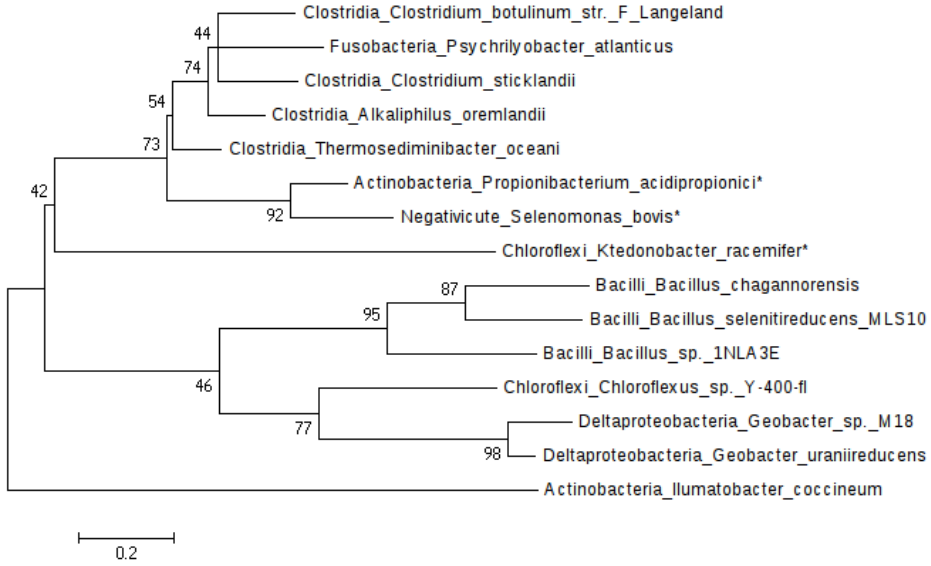
>SelW-like protein Bsel_1870
MSLHVSIEFCMQUNYAPKAASLAEDIFEDMRHDVSKLDLIPSSGGVFEVKVNNQ
LIFSKFETDQFPDHMEIINTLQSRKQYG

APPENDIX B

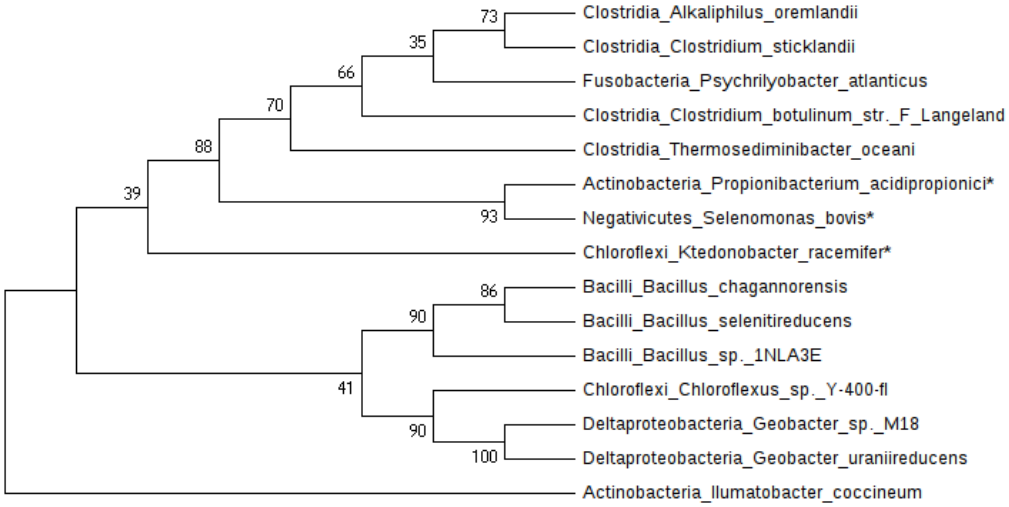
SELENOPROTEIN MAXIMUM LIKELIHOOD, MAXIMUM PARSIMONY, AND NEIGHBOR-JOINING PHYLOGENIES

D-proline reductase phylogenies

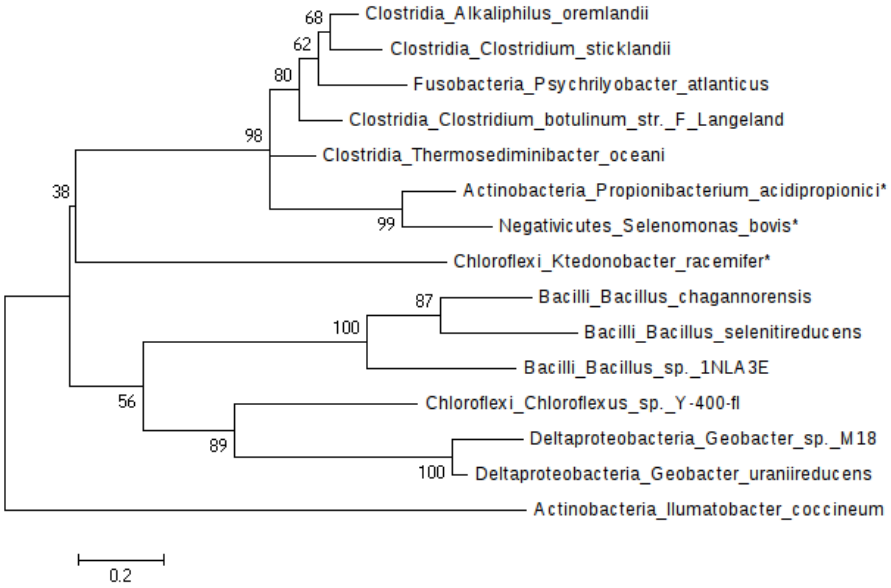
A. Maximum likelihood phylogeny



B. Maximum parsimony phylogeny

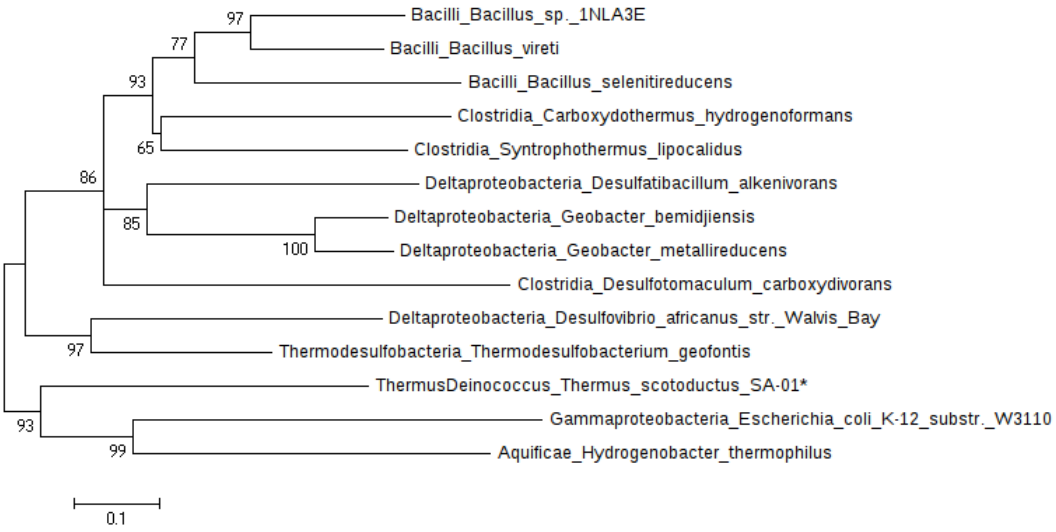


C. Neighbor-joining phylogeny

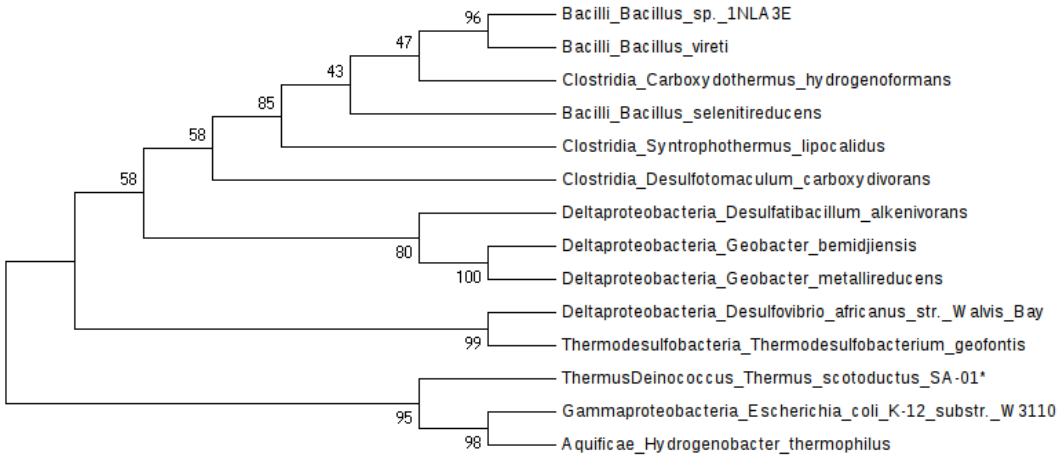


Formate Dehyrogenase N α subunit phylogenies

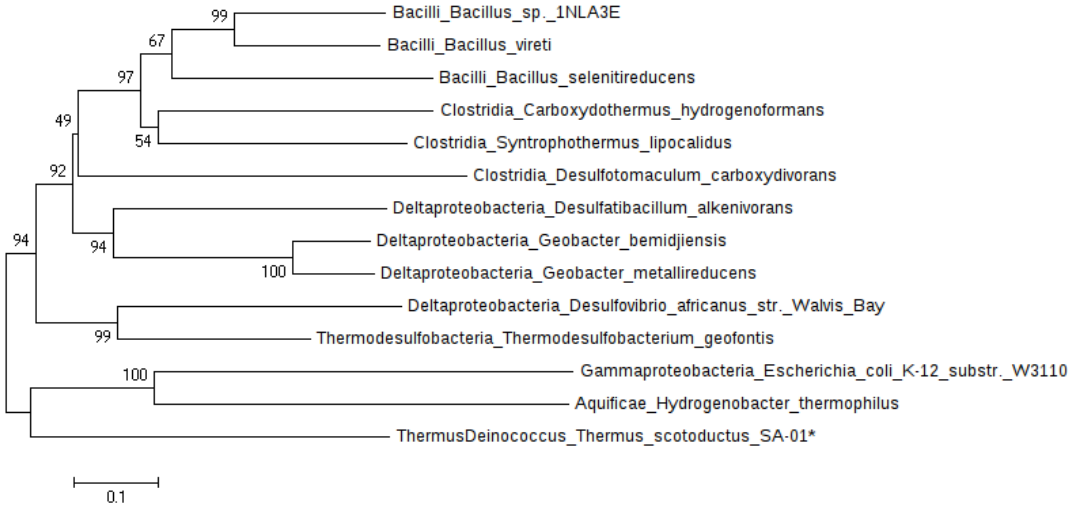
A. Maximum likelihood phylogeny



B.Maximum parsimony phylogeny

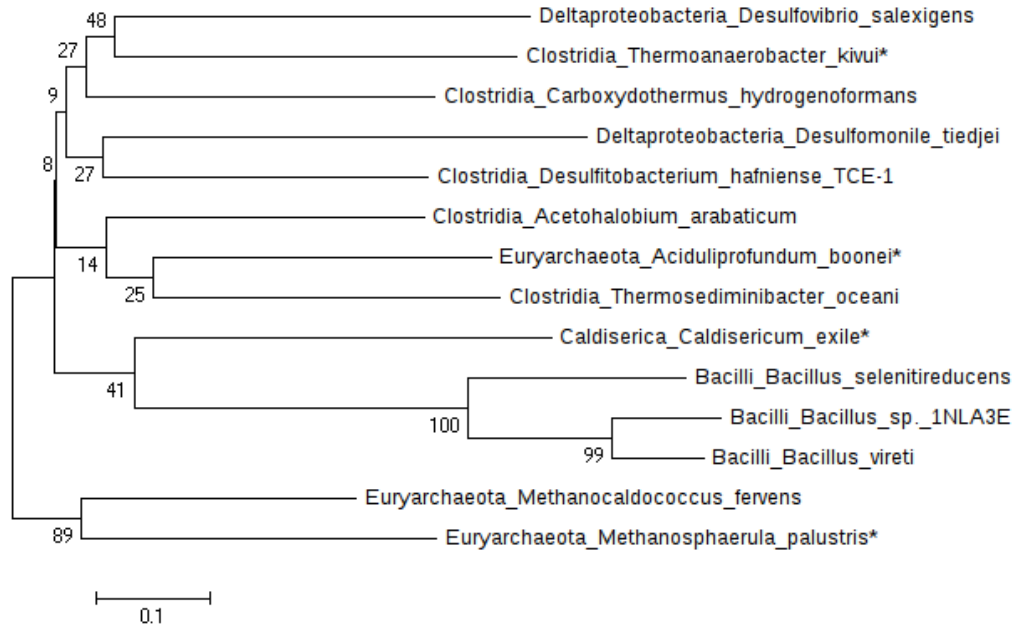


C. Neighbor-joining phylogeny

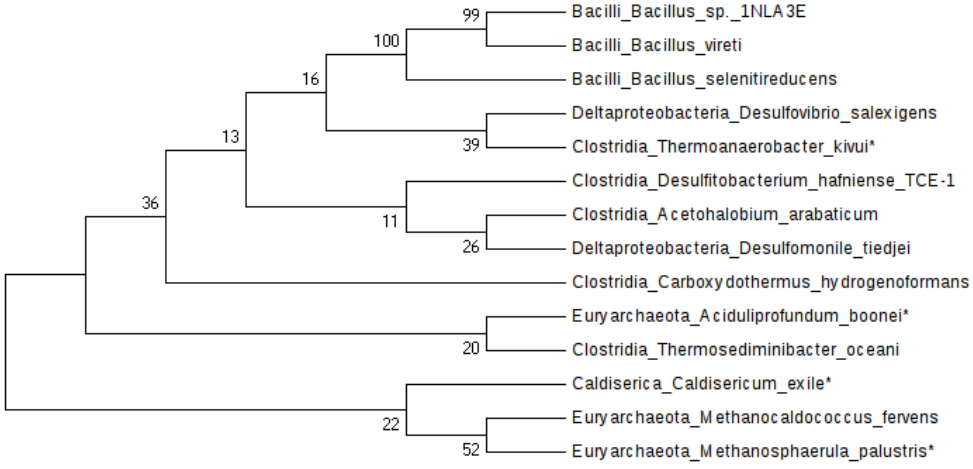


NAD+ dependent Formate Dehydrogenase α subunit phylogenies

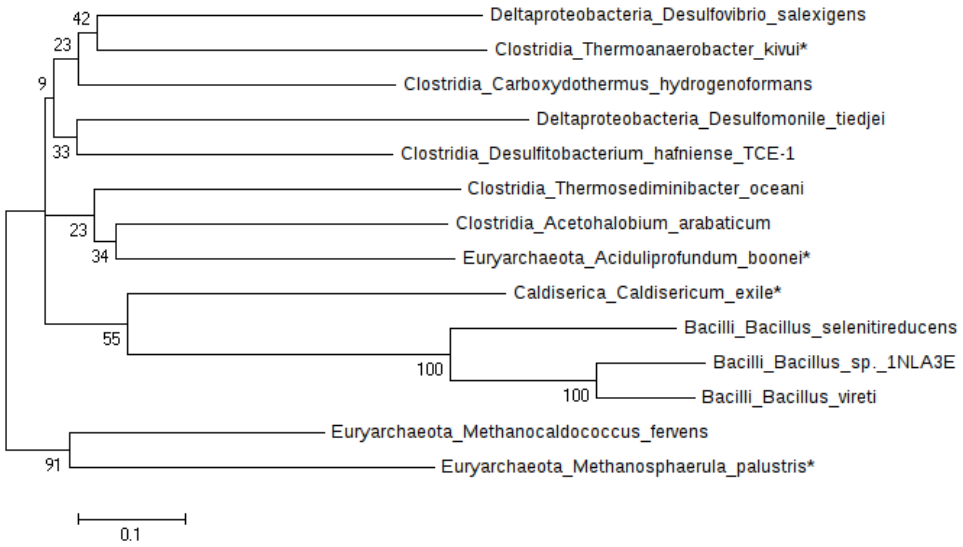
A.Maximum likelihood phylogeny



B. Maximum parsimony phylogeny

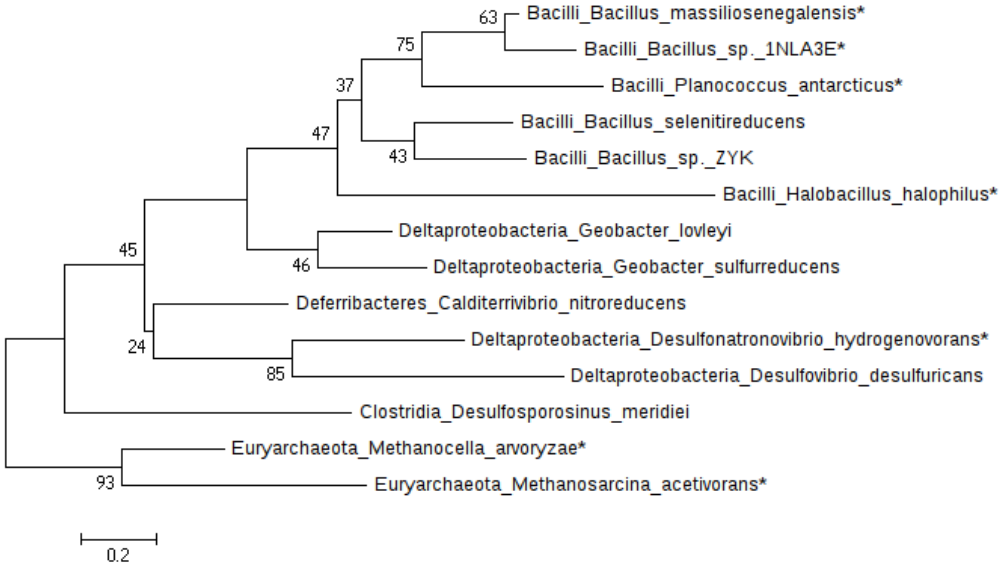


C. Neighbor-joining phylogeny

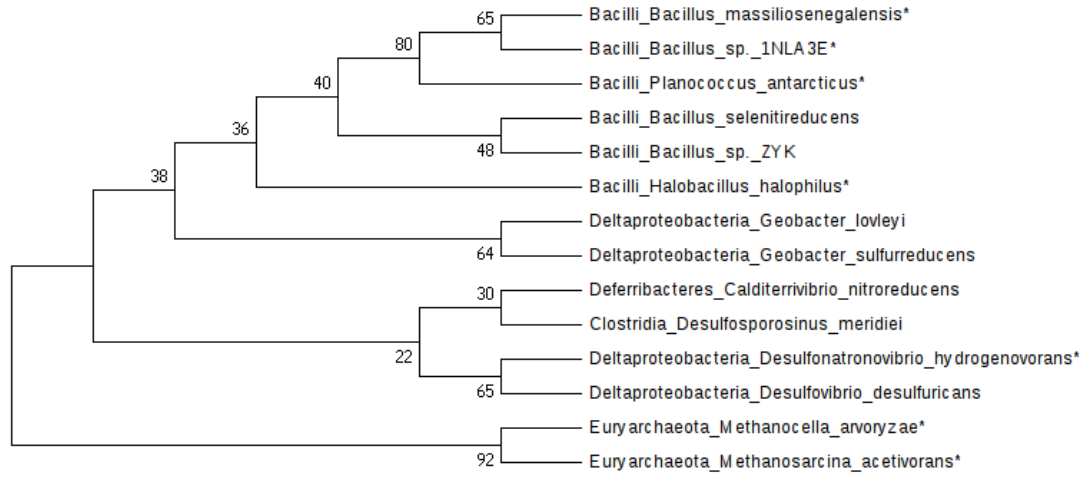


HesB-like protein phylogenies

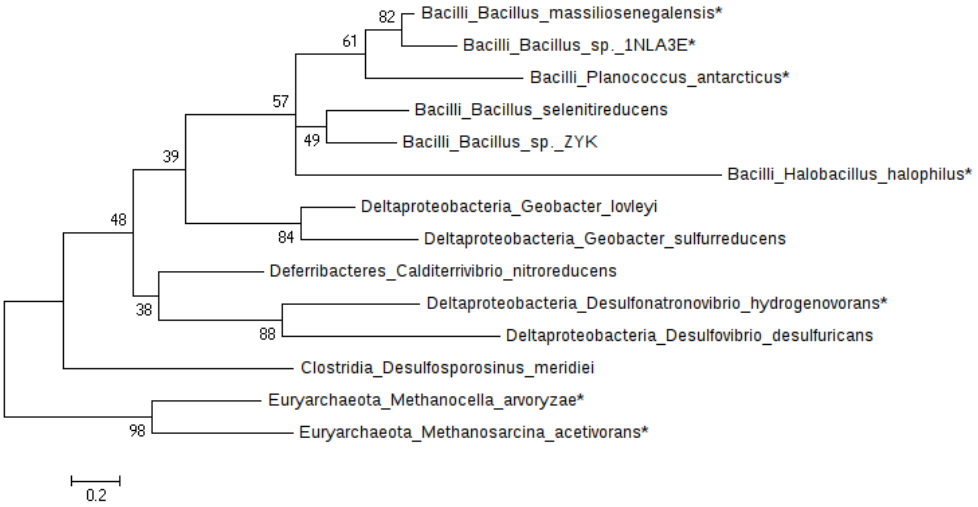
A. Maximum likelihood phylogeny



B. Maximum parsimony phylogeny

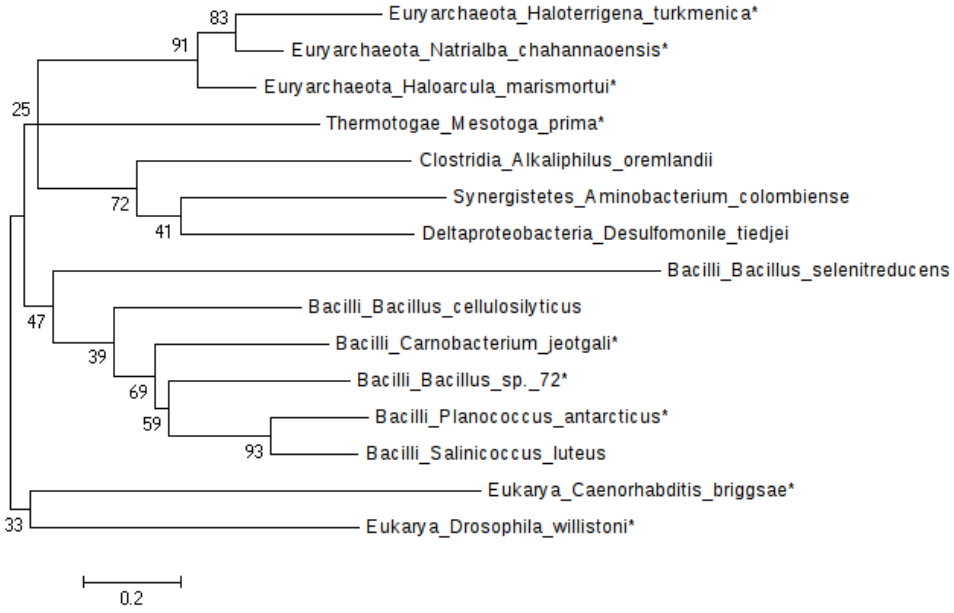


C. Neighbor-joining phylogeny

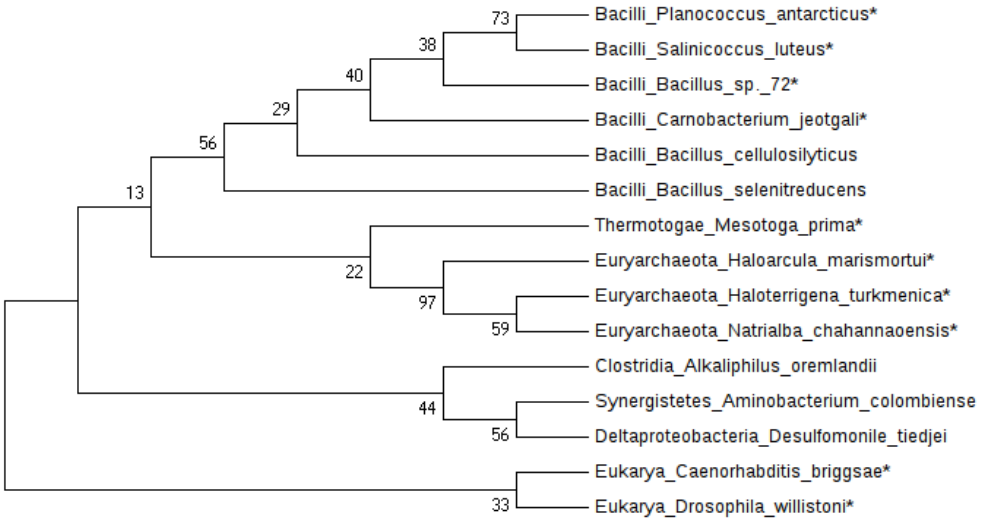


Methonine sulfoxide reductase A phylogenies

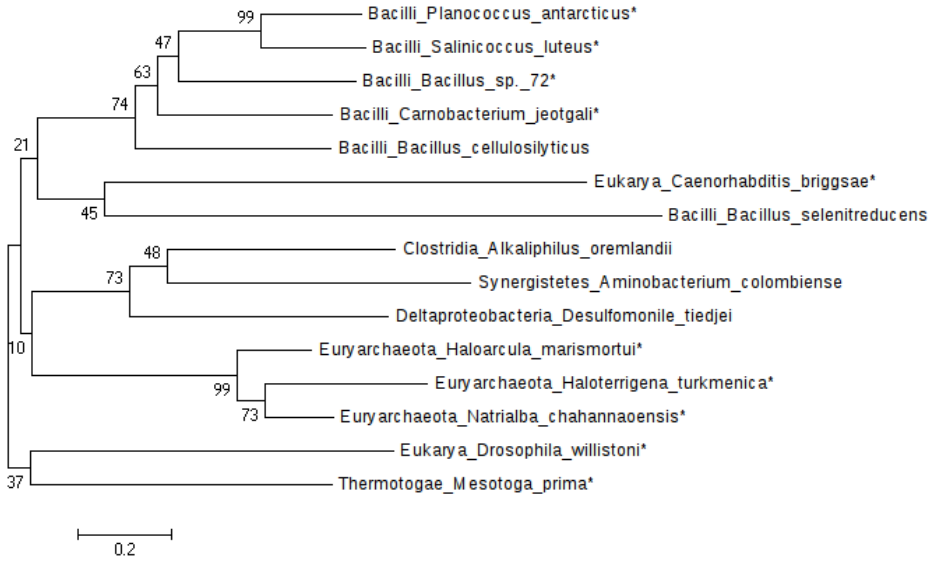
A. Maximum likelihood phylogeny



B. Maximum parsimony phylogeny

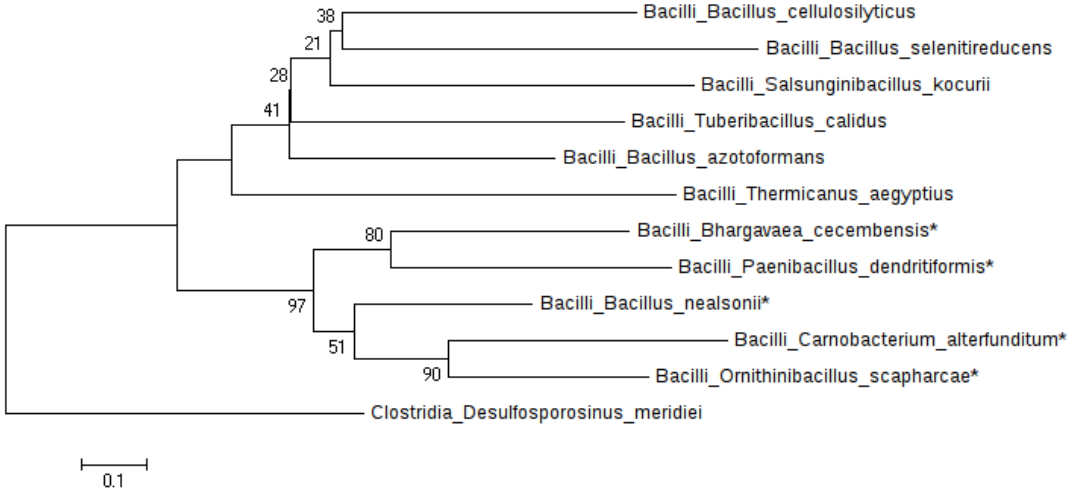


C. Neighbor-joining phylogeny

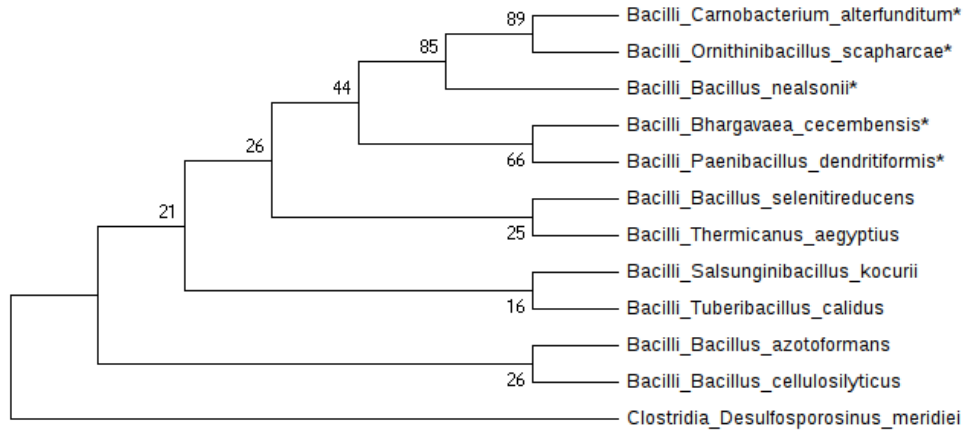


Selenophosphate synthase phylogenies

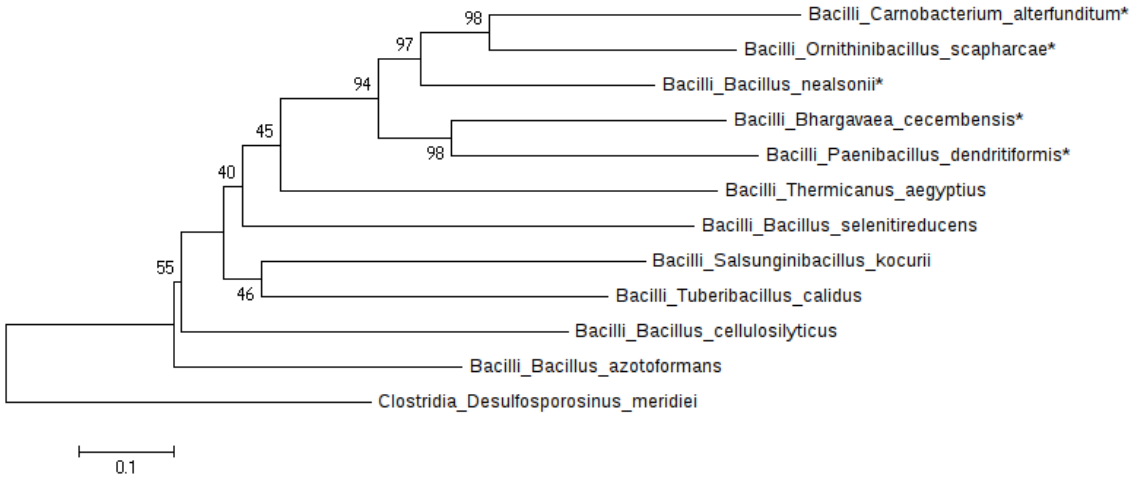
A. Maximum likelihood phylogeny



B. Maximum parsimony phylogeny

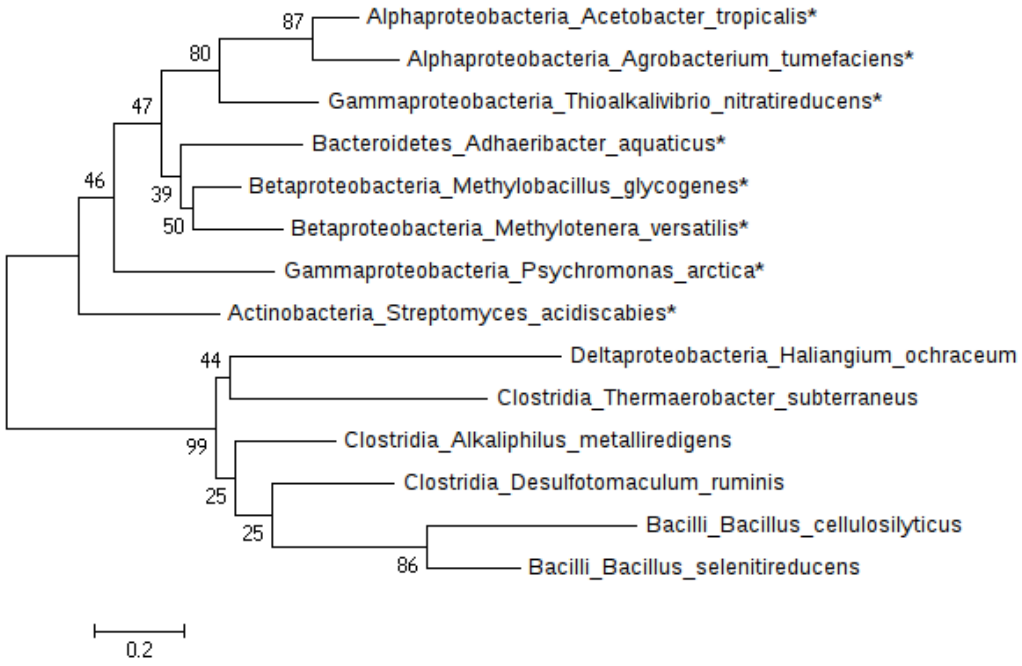


C. Neighbor-joining phylogeny

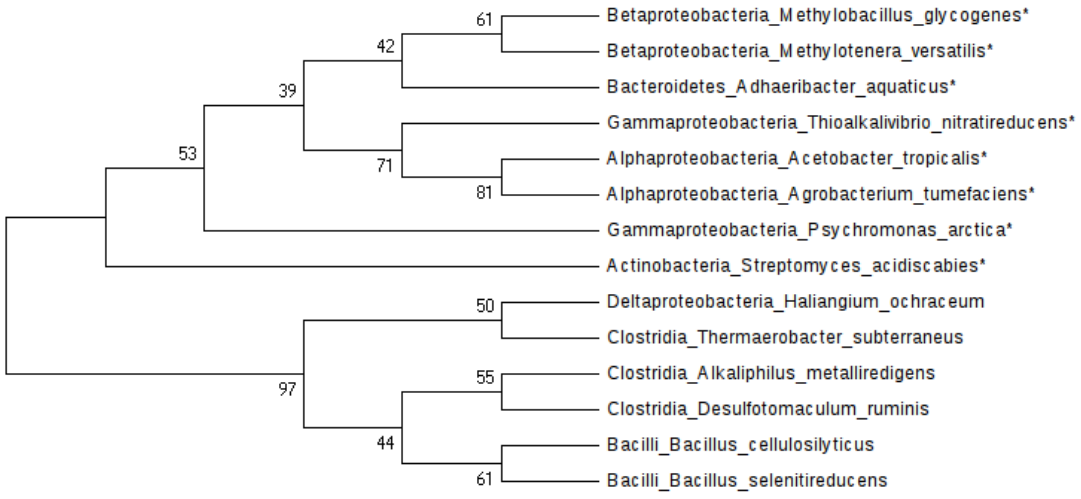


SelW-like protein phylogenies

A. Maximum likelihood phylogeny



B. Maximum parsimony phylogeny



C. Neighbor-joining phylogeny