Probing the Novelties of Alkalilimnicola ehrlichii strain MLHE-1 with Genomic and Proteomic Approaches

Christine Richey

Follow this and additional works at: https://dsc.duq.edu/etd

Recommended Citation

This Immediate Access is brought to you for free and open access by Duquesne Scholarship Collection. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of Duquesne Scholarship Collection.
PROBING THE NOVELTIES OF

_ALKALILIMNICOLA EHRLICHII_ STRAIN MLHE-1\textsuperscript{T}

WITH GENOMIC AND PROTEOMIC APPROACHES

A Thesis

Submitted to the Bayer School

of Natural and Environmental Sciences

Duquesne University

In partial fulfillment of the requirements for

the degree of Master of Science

By

Christine Richey

July 2008
PROBING THE NOVELTIES OF

ALKALILIMNICOLA EHRLICHII STRAIN MLHE-1T

WITH GENOMIC AND PROTEOMIC APPROACHES

By

Christine Richey

Approved July 15, 2008

John Stolz
Professor of Biology
(Thesis Advisor)

Partha Basu
Professor of Biochemistry
(Committee Member)

Mary Alleman
Associate Professor of Biology
(Committee Member)

Michael Jensen-Seaman
Assistant Professor of Biology
(Committee Member)

David Seybert
Dean, Bayer School
Professor of Chemistry and
Biochemistry

Philip Auron
Chair, Biology
Professor of Biology
ABSTRACT

PROBING THE NOVELTIES OF

ALKALILIMNICOLA EHRLICHII STRAIN MLHE-1\textsuperscript{T}

WITH GENOMIC AND PROTEOMIC APPROACHES

By

Christine Richey

July 2008

Thesis Supervised by Dr. John Stolz

Alkalilimnicola ehrlichii strain MLHE -1\textsuperscript{T} is a gram negative haloalkaliphilic

gammaproteobacterium isolated from Mono Lake, CA. It has the unique ability to grow

both aerobically as a heterotroph (with oxygen and acetate) and anaerobically as a

chemolithoautotroph by coupling arsenite oxidation to nitrate reduction. Its genome

(3.276 MB, 67.53% mole GC) has been sequenced to closure and annotated (2883

putative ORFs). The pathways involved in nitrogen (e.g., denitrification, assimilation),
sulfur, and carbon metabolism (glycolysis, gluconeogenesis), carbon fixation,
tricarboxylic acid cycle, and motility (flagella, pili) are described herein. Though no

homolog for arsenite oxidase was found, two putative arsenate reductases (arr) were

identified along with arsenic resistance genes (ars). One reductase homolog (Mlg0215-
Mlg0218) was expressed only under anaerobic conditions and was shown to exhibit both
arsenate reductase and arsenite oxidase activity. Proteomics, blue native gels, and enzyme activity assays suggest that it functions primarily as the arsenite oxidase.
ACKNOWLEDGMENT

I would like to thank Dr. John Stolz for his guidance, insight, and patience during my time at Duquesne. I would also like to thank Dr. Partha Basu for his advice and critiques. I would also like to thank my other committee members, Dr. Michael Jensen-Seaman and Dr. Mary Alleman, for their roles in my degree completion.

I am most grateful to Dr. Peter Chovanec for all of his hard work and instruction for the proteomic portion of my project. I do not think my project would have been nearly as successful without his presence. Much gratitude is given to Shelley Hoeft of the US Geological Survey for providing cultures and cell pellets. I wish to also extend my thanks to Bryan Crable and Courtney Sparacino-Watkins for helping me develop and implement protocols. The other past and present members of the Stolz and Basu labs have also lended their support and advice throughout the past two years and for that I am appreciative.

I would also like to take the opportunity to thank my fiance, Michael James, for his unwaivering love and support. Another special thank you is extended to the rest of my friends and family for listening to my science babble and having no idea what I was talking about but who supported me anyway.
## TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
</tr>
<tr>
<td>Acknowledgment</td>
</tr>
<tr>
<td>List of Tables</td>
</tr>
<tr>
<td>List of Figures</td>
</tr>
</tbody>
</table>

### 1 Introduction and Specific Aims

1.1 Arsenic | 1 |
1.2 Arsenic in the Environment | 2 |
1.3 Arsenic Transformation by Microbes | 5 |
1.4 Arsenite Oxidation | 8 |
1.5 Mono Lake, California | 10 |
1.6 *Alkalilimnicola ehrlichii* strain MLHE -1<sup>T</sup> | 12 |
1.7 Specific Aims and Rationale | 16 |

### 2 Genome of *Alkalilimnicola ehrlichii* strain MLHE -1<sup>T</sup>

2.1 Background | 18 |
2.2 Materials and Methods | 20 |
2.3 Results | 21 |
   2.3.1 General Characteristics | 21 |
   2.3.2 Relationship to Closest Relatives | 22 |
   2.3.3 Specific Functions | 26 |
      2.3.3.1 Arsenic Transformation | 26 |
      2.3.3.2 Nitrate Metabolism | 30 |
2.3.3.3 Sulfur Metabolism .................................................. 31
2.3.3.4 Carbon Metabolism ................................................. 33
2.3.3.5 Storage ................................................................. 41
2.3.3.6 Motility and Secretion ............................................. 41

2.4 Discussion ........................................................................ 44

3 Proteome of *Alkalilimnicola ehrlichii* strain MLHE -1T

3.1 Materials and Methods ...................................................... 46
   3.1.1 Media Preparation ....................................................... 46
   3.1.2 Cell Harvesting and Fractionation .................................. 47
   3.1.3 Protein Concentration Determination ............................. 48
   3.1.4 TCA Precipitation ....................................................... 48
   3.1.5 SDS-PAGE Gels ........................................................ 49
   3.1.6 2-D SDS-PAGE Gels ................................................ 50
   3.1.7 DIGE ..................................................................... 52
   3.1.8 Blue Native Gels ........................................................ 53
   3.1.9 Digestion .................................................................... 56
   3.1.10 Mass Spectrometry .................................................... 57
   3.1.11 Activity Assays- Vials ................................................. 58
   3.1.12 Activity Assays- Zymograms ...................................... 60

3.2 Results .............................................................................. 62
   3.2.1 2-D SDS-PAGE Gels ................................................ 62
   3.2.2 DIGE ........................................................................ 62
   3.2.3 Blue Native Gels ........................................................ 66
3.2.4 Activity Assays- Vials.................................................................72
3.2.5 Activity Assays- Zymograms.....................................................72
3.3 Discussion......................................................................................76
References..........................................................................................79
Appendix I..........................................................................................84
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>COG Color Guide to Chromosome</td>
<td>24</td>
</tr>
<tr>
<td>2.2</td>
<td>COG Content Comparison</td>
<td>25</td>
</tr>
<tr>
<td>3.1</td>
<td>Protein Identification from 2-D Gels</td>
<td>64</td>
</tr>
<tr>
<td>3.2</td>
<td>Proteins Identified from Cytosolic DIGE</td>
<td>68</td>
</tr>
<tr>
<td>3.3</td>
<td>Proteins Identified from CHAPS-soluble DIGE</td>
<td>70</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

1.1 Map of Arsenic Levels in the United States.................................................................4
1.2 Biochemical Cycling of Arsenic by Microbes.................................................................6
1.3 Phylogenetic Tree of Arsenic Transformers.................................................................9
1.4 Aerial View of Mono Lake, CA....................................................................................11
1.5 Alternate View of Mono Lake, CA................................................................................11
1.6 Arsenic Levels and Cycling in Mono Lake, CA............................................................13
1.7 Micrograph Showing Flagellum....................................................................................13
1.8 Micrographs Showing Internal Storage.........................................................................15
1.9 Phylogenetic Tree of Arsenite Oxidizers......................................................................15
2.1 Chromosome of MLHE-1.............................................................................................23
2.2 Venn Diagram of Relatedness.......................................................................................27
2.3 Arsenic-Related Gene Clusters.....................................................................................29
2.4 Phylogenetic Tree of DMSO Catalytic Subunits..........................................................29
2.5 Nitrogen Metabolism....................................................................................................32
2.6 Gene Clusters for Nitrogen Metabolism.....................................................................32
2.7 Sulfur Metabolism........................................................................................................34
2.8 Sulfur Gene Clusters......................................................................................................34
2.9 Glycolysis and Gluconeogenesis..................................................................................37
2.10 Carbon Fixation...........................................................................................................38
2.11 Methane Metabolism..................................................................................................39
2.12 Gene Clusters for Carbon Metabolism......................................................................39
2.13 TCA Cycle..............................................................................................................40
2.14 Polyhydroxybutyrate Operon..................................................................................40
2.15 Gene Clusters for Flagella......................................................................................43
2.16 Gene Clusters for Pili.............................................................................................43
3.1 Activity Assay Reactions..........................................................................................59
3.2 2-D Gel Comparison................................................................................................63
3.3 DIGE of Cytosolic Proteins.....................................................................................67
3.4 DIGE of CHAPS-Soluble Proteins..........................................................................69
3.5 Blue Native Gel of Aerobic and Anaerobic Membrane Fractions.........................71
3.6 Activity Assays........................................................................................................73
3.7 Zymograms...............................................................................................................74
3.8 Second Dimension Blue Native Gel........................................................................75
Chapter 1

Introduction and Specific Aims

1.1 Arsenic

Arsenic is the thirty-third element on the periodic table and has an atomic weight of 74.9. It is the twentieth most prevalent element in the Earth's crust (0.0002%). It can occur in several oxidation states: As (III), As (V), As (0), and As (-III). In nature, arsenic primarily cycles between arsenite, As (III), and arsenate, As (V). Arsine, As(-III), can be found in trace amounts in gases in very anoxic environments, while the elemental form is very rare. Arsenate is relatively immobile and is common in aqueous aerobic environments, while arsenite is prevalent in anoxic environments and is much more mobile. Organic arsenic tends to be less toxic than inorganic compounds, and since inorganic arsenite is considered 100 times more toxic than arsenate, its mobility poses a huge environmental problem.

Arsenic is typically categorized as a metalloid and has toxicity characteristics similar to heavy metals. Arsenic species can affect DNA repair, DNA methylation, increase radical formation, and activate the protooncogene c-myc in humans. Arsenite alone has the ability to inhibit over 200 enzymes by binding to sulfhydryl groups. Arsenite can react with thiol groups and inhibits pyruvate dehydrogenase by binding to the sulfhydryl groups of dihydrolipoamide. This causes a decrease in the conversion of
pyruvate to acetyl coenzyme A (CoA), which decreases the activity of the citric acid cycle and therefore reduces ATP production. Arsenite also inhibits cellular glucose uptake, gluconeogenesis, fatty acid oxidation, and further production of acetyl CoA. Arsenate resembles inorganic phosphate and substitutes for phosphate in glycolytic and cellular respiration pathways. Since the high-energy phosphate bonds cannot be made, oxidative phosphorylation cannot occur and ADP-arsenate is manufactured instead of ATP\(^5\). Though arsenite is considered more toxic than arsenate, the human body reduces arsenate to arsenite before undergoing detoxification through methylation\(^6\).

1.2 Arsenic in the Environment

Arsenic can accumulate in the environment because of both natural (weathering, volcanic activity, erosion, etc.) and anthropogenic activities\(^2\). Arsenic is usually found in mineral deposits and has a high affinity for pyrite and hydrous iron oxides. It can readily be solubilized into the groundwater depending on solution composition, temperature, and pH. The two biggest triggers are a pH above 8.5 and the onset of reductive iron dissolution. High concentrations of phosphate, bicarbonate, silicate, and organic matter in the groundwater also promote arsenic solubilization. Erosion of black shale, thermal springs, volcanoes, and other geothermal events often lead to the mobilization of the element\(^6\). Anthropogenic sources include mineral extraction, waste processing, pesticides, and poultry and swine feed additives like Roxarsone.

Arsenic is usually found at very low levels in the environment. The ground typically contains 0.2 to 40µg arsenic per gram soil. The atmosphere in urban areas generally contain around 0.02µg per cubic meter air\(^7\). The real threat of arsenic exposure is from ingestion. Wells in Taiwan have been found to contain 170-800µg/L\(^7\). 45% of
wells in 560 villages in West Bengal, India have arsenic concentrations higher than 50µg/L\textsuperscript{8}. The highest concentration found in the region was 3.7mg/L, well over the World Health Organizations standard of 10µg/L. Concentrations greater than 5.0mg/L have been recorded in Thailand. It is estimated that in Bangladesh and India alone 36 million people are being exposed to elevated arsenic levels in their drinking water\textsuperscript{6}.

Arsenic poses a problem in the United States as well. It is estimated that nearly 350,000 people in the country consume water containing over 50µg/L daily. More than 2.5 million consume water with over 25µg/L\textsuperscript{7}. Of 30,000 analyses done, 10% exceeded the WHO standard of 10µg/L\textsuperscript{9}. Most of the contamination is located in the midwest and southwest United States (Figure 1.1). These concentrations are due to natural environmental processes such as erosion, geothermal activity, and evaporation. The activity along the Eastern seaboard is in part due to natural mineral deposits, but manufacturing, pesticides, and lumber treatment should also be noted as sources.

Arsenic poses health concerns for several reasons. It is well known that arsenic can cause skin cancer, but it can also cause liver, lung, kidney, and bladder cancer. Noncarcinogenic effects include such illnesses as diabetes, peripheral neuropathy, and cardiovascular diseases\textsuperscript{4}. Smith \textit{et al.} predict that the arsenic in U.S. water supplies poses as much of a cancer risk as tobacco smoke and radon\textsuperscript{7}. Drinking one liter of water with 50 micrograms of arsenic, the current EPA standard, per day could result in the cancer death of 13 per 1000 people. Mandal \textit{et al.} reports nearly a million people drinking arsenic-contaminated water in West Bengal, and over 200,000 have been diagnosed with arsenic-related diseases so far\textsuperscript{8}.
Figure 1.1  United States map showing arsenic levels in groundwater\textsuperscript{10}. The highest concentrations of arsenic are found in the western portion of the nation and in the northern central states. High concentrations are also found in New England.
1.3 Arsenic Transformation by Microbes

Even though arsenic is toxic to most organisms, there are microbes that use it for growth and metabolism. Bacteria are able to overcome the usual toxic effects by readily exporting the compound from the cell, by preventing its initial uptake, or by chemically altering it. Both arsenite and arsenate can be methylated in the cell to lower their toxicity\textsuperscript{11}.

The first arsenite oxidizing bacteria and arsenate reducing bacteria were isolated from cattle-dipping tanks in 1918\textsuperscript{12}. Since this time, certain bacteria from the Alphaproteobacteria, Betaproteobacteria, and Gammaproteobacteria have been shown to have arsenic transforming abilities. Arsenic speciation, mobility, and toxicity can be greatly influenced by microbial activity through arsenate reduction, arsenite oxidation, methylation, and demethylation. Geochemical evidence suggests that biological oxidation of arsenite coupled with nitrate reduction occurs in arsenic-contaminated freshwater lakes and in the anoxic, subsurface aquifers of Bangladesh\textsuperscript{13,14}. Figure 1.2 depicts a proposed biochemical cycle for arsenite and arsenate as mediated by microbes\textsuperscript{15}.

Arsenite reduction by microorganisms has been studied in depth. \textit{Bacillus arsenicoselenatis} was isolated from Mono Lake, CA in 1998. It grows using dissimilatory arsenate reduction coupled to the oxidation of acetate or lactate in the presence of carbon dioxide\textsuperscript{16}. \textit{Bacillus macyae} sp. nov. is another arsenate-respiring bacterium. It was isolated in 2004 from an Australian gold mine and is a strict anaerobe\textsuperscript{17}. At this time a total of 19 arsenate-respiring prokaryotes have been isolated with various physiologies and phylogenetics. The most well known and studied are
Figure 1.2 Proposed biochemical cycle of arsenic oxidation states as mediated by microbes.

MLHE-1 transfers electrons to nitrate to convert arsenite, As (III), to arsenate, As (IV).

Modified from Oremland et al. (2007)\textsuperscript{15}. 
Chrysiogenes arsenatis, Bacillus selenitireducens strain MLS10, and Shewanella sp. strain ANA-3.

Most organisms that are able to oxidize arsenite to arsenate are heterotrophic and require organic matter for growth\textsuperscript{18}. The oxidation is considered a means of detoxification rather than a pathway for growth and requires oxygen to be present. This transformation of arsenic species can be measured using ion and high-performance liquid chromatography\textsuperscript{15}. *Alcaligenes faecalis* is the most well characterized heterotrophic arsenite oxidizer. *Thermus aquaticus* and *Thermus thermophilus* are two organisms that are able to oxidize arsenite to arsenate when an additional energy source is present\textsuperscript{19}. *Hydrogenophaga* species strain NT-14 was found to also oxidize arsenite when additional energy sources were present\textsuperscript{20}.

Arsenite oxidation under anaerobic conditions is a much more recent discovery. The first chemolithoautotrophic arsenite oxidizer characterized was *Pseudomonas arsenitoxidans*\textsuperscript{21}. Other organisms growing chemoautotrophically were then identified, such as NT-26, an Alphaproteobacteria in the *Rhizobium* clade that was isolated from a gold mine\textsuperscript{18}. When anoxic bottom water of Mono Lake, California was amended with nitrate and nitrite, a novel arsenite oxidizer was isolated\textsuperscript{15}. The bacterium was named *Alkalilimnicola ehrlichii* strain MLHE-1. Nitrate-linked oxidation has been further identified in freshwater lakes, subsurface aquifers, and soil sediments\textsuperscript{13,14,22}. Figure 1.3 shows the relationship between some arsenic-transforming bacteria.

The arsenic genome project was undertaken to better understand the cycling and mobility of arsenic. Four organisms were proposed as candidates for sequencing. The three arsenate respiring bacteria included *Bacillus selenitireducens* strain MLS-10,
*Alkaliphilus oremlandii* OhILAs, and Mono Lake strain MLMS-1. The fourth organism was the arsenite-oxidizing bacterium, *Alkalilimnicola ehrlichii* strain MLHE-1T. The genome sequence of MLHE-1 was completed in September 2006 by the DOE Joint Genome Institute.

1.4 Arsenite Oxidation

The first arsenate oxidase was characterized from *Alcaligenes faecalis*11. It was found to be located on the outer surface of the inner membrane of the cell in the periplasmic space and shuttles electrons to cytochrome c. The enzyme contains one molybdenum, several iron molecules, and an inorganic sulfide. Biophysical studies of this enzyme by Ellis *et al.* showed that it is composed of a large subunit of 825 residues and a small subunit of at least 134 residues23. The large subunit, the catalytic subunit, contains a molybdenum binding site and a [3Fe-4S] cluster, whereas the small subunit harbors a Rieske type [2Fe-2S] site. The large subunit is structurally similar to members of the dimethyl sulfoxide (DMSO) reductase family. It is most closely related to dissimilatory nitrate reductase (NAP) and formate dehydrogenase (FDH)23.

Though two different designations have been proposed for the arsenite oxidase of heterotrophs (Aox) and autotrophs (Aro), their structures appear very similar (but only one has been structurally characterized). For example, the oxidase of the chemolithoautotroph NT-26 is similar in size to that of *A. faecalis*. It too is a molybdoenzyme with a molybdopterin cofactor and two iron-sulfur clusters18. Rhine *et al.* showed that six novel bacterial strains had 72-74% identical to NT-26's arsenite oxidase24. They also found that the autotrophic *Aro* genes form a distinct clade from the heterotrophic *Aox* genes, except for the heterotroph *A. tumefaciens*, which groups with
Figure 1.3 Phylogenetic diversity based on 16S rRNA gene sequences of known arsenic transforming species from both the Archaea and Bacteria domains. Dissimilatory arsenate-respiring prokaryotes (DARPs) are represented by blue circles, chemoautotrophic arsenite oxidizers (CAOs) are indicated by red squares, and heterotrophic arsenite oxidizers (HAOs) are signified by gold triangles. Arsenate-resistant microbes (ARMs) are not shown. Figure used by permission from Oremland and Stolz (2005).
the autotrophs. This supports the belief that the two different oxidases evolved separately from some common ancestor. Figure 1.3 shows a phylogenetic tree of known arsenic transforming microorganisms.

Regulation of arsenite oxidation is complex. Kashyap et al. studied the regulatory controls in Agrobacterium tumefaciens and discovered arsenite-sensing, two-component signal transduction, and quorum sensing at work.\textsuperscript{26} aoxA was regulated by aoxR and aoxB and chlE amplified together. aoxBA mRNA was detected using RT-PCR only in late log phase and stationary phase cells in the presence of arsenite. Kashyap also determined that aoxS-aoxR-aoxA-aoxB is polycistronic and other genes downstream may also be involved with regulation\textsuperscript{26}.

1.5 Mono Lake, California

Soda lakes represent the most natural alkaline environments on Earth but have not yet been extensively characterized in most cases. Mono Lake is a hypersaline, closed basin soda lake located in California that has recently begun to be studied (Figure 1.4 and 1.5). The Owens River and its tributaries fed the lake until they were diverted in 1941 to increase water service to Los Angeles. The evaporation rate soon exceeded the rate of inflow, causing the water level to drop dramatically. In 1994 a bill was passed by the California State Legislature to protect Mono Lake and its tributaries in hopes of restoring the former ecosystem. This reclamation has shown to be difficult.

Because of the calcite rock formations and alkaline sand at the bottom of the lake, in conjunction with low water levels, the pH of Mono Lake is typically 9.8. The salt concentration averages 81 grams/liter, three times saltier than the ocean.\textsuperscript{27} It also has a high concentration (200µM) of dissolved inorganic arsenic due to hydrothermal sources
Figure 1.4 An aerial view of Mono Lake, CA showing its decrease in size\textsuperscript{28}.

Figure 1.5 A closer view of Mono Lake showing evidence of drastic evaporative processes\textsuperscript{29}.
feeding the lake and the high rate of evaporation for the region. There is very little organic arsenic in the water, however. The water at the bottom of the lake has become anoxic and contains high levels of ammonia, methane, and sulfide.

The chemical characteristics of the lake have been controlled by natural events. The lake has expanded and contracted due to climate change throughout the centuries. Currently the lake is five times smaller than it was during the Pleistocene and contains eighteen times less water. Because of the high evaporation rate, elements like sodium, chlorine, sulfur, boron, and arsenic are highly concentrated. The amount of carbonates and phosphate also increased dramatically, while inorganic nitrogen dropped sharply. Because of the high concentrations of these compounds, the amount of arsenite increased in the anoxic bottom waters. Figure 1.6 depicts the arsenic levels and cycling of the element in Mono Lake. Because of the unusual chemical makeup of Mono Lake, it is a prime target for identifying new and unique organisms.

1.6 *Alkaliminnicola ehrlichii* strain MLHE-1

*A. ehrlichii* strain MLHE-1 is a haloalkaliphile that was isolated from Mono Lake, California. MLHE-1 is a gram-negative, motile rod that is approximately 1.5-2.5µm in length and has a diameter of 0.3-0.5µm. The cell has one polar flagellum and several pili as can be seen in Figure 1.7. The bacterium can grow over a salinity range of 15-190g/L, with an optimum of 30g/L. The optimum pH was shown to be 9.3, though the organism tolerated up to 10.4. The optimal temperature for MLHE-1 growth is 30°C.

MLHE-1 can grow using a multitude of electron donors and acceptors when presented with the right combination. It has the unique ability to grow
Figure 1.6  Arsenic cycle as proposed for Mono Lake, CA. MLHE-1 would be considered a CAO (Chemolithotrophic Arsenite Oxidizer) in this model. Figure reproduced with permission of Oremland and Stolz (2003)\textsuperscript{30}.

![Arsenic cycle diagram](image)

Figure 1.7  An electron micrograph of MLHE-1 showing the single polar flagellum. Figure reproduced by permission of Hoeft \textit{et al.} (2007)\textsuperscript{31}.

![Electron micrograph](image)
chemoautotrophically by coupling arsenite oxidation with nitrate reduction\textsuperscript{31}. It can also use hydrogen gas or sulfide as the electron donor when nitrate is used as the electron acceptor\textsuperscript{15}. Thiosulfate can also be used with nitrate when grown autotrophically. Under anaerobic conditions, \textit{A. ehrlichii} obtains its carbon from bicarbonate compounds. The organism uses the Calvin-Benson-Bassham pentose phosphate pathway for this metabolic activity. The bacteria can also grow on organic carbon compounds, such as acetate, but only when oxygen is present in the environment. TEM images of thin sections of cells grown on acetate reveal the presence of polyhydroxybutyrate (PHB) storage granules in the cytoplasm, while cells grown on arsenite lack the PHB granules and instead possess much smaller intracellular inclusions (Figure 1.8).

The Proteobacteria phylum contains the largest number of bacteria and is metabolically diverse. They are all gram-negative and represent some of the most significant species of medical, industrial, and agricultural concern. There are five clusters within the phylum, each designated by a Greek letter: alpha, beta, gamma, delta, or epsilon\textsuperscript{32}. This classification is based on the organism's 16S rRNA sequence. \textit{Alkalilimnicola ehrlichii} falls into the Gammaproteobacteria and forms a distinct clade within the Ectothiorhodospiraceae with other \textit{Alkalilimnicola} species\textsuperscript{33}. Genetic analysis of \textit{A. ehrlichii} revealed that it has a GC content of 67.5\%, which is similar to that of \textit{A. halodurans} and \textit{A. mobile}. DNA-DNA relatedness values between MLHE-1 and these other species were below 50\%, which supports this strain as a new novel species\textsuperscript{31}. Figure 1.9 shows the phylogenetic relationship of MLHE-1 to other arsenite-oxidizing bacteria.
Figure 1.8  Electron micrographs showing MLHE-1 grown heterotrophically on acetate (Right) and autotrophically on arsenite (Left). “P” labels one polyhydroxybutyrate globule, while arrows point to the intracellular inclusions. Pictures reproduced with permission from Hoeft et al. (2007)\textsuperscript{31}.

Figure 1.9  Phylogenetic tree (maximum parsimony) showing the relatedness of MLHE-1 to other \textit{Alkalilimnicola} species, Ectothiorhodospiraceae, and arsenite oxidizers. Figure modified from Hoeft et al. (2007)\textsuperscript{31}.
1.7 Specific Aims and Rationale

**Specific Aim 1:** Complete the annotation and characterization of the *Alkalilimnicola ehrlichii* strain MLHE-1\(^T\) genome.

Rationale: *A. ehrlichii* demonstrates different phenotypes when grown anaerobically versus aerobically. These differences are not limited to the presence or absence of the arsenite oxidase. The nitrate reductases (Nar), carbon monoxide dehydrogenases (Cox), arsenic resistance proteins (Ars), respiratory arsenate reductases (Arr), and RuBisCo were compared between the two cultures. Whole genome comparisons with *Nitrococcus mobilis* and *Halorhodospira halophila* SL1 species were done to see if they shared similar properties, since they are also members of the Ectothiorhodospireae family.

**Specific Aim 2:** To determine the mechanism of arsenite oxidation by *Alkalilimnicola ehrlichii* strain MLHE-1\(^T\).

Hypothesis A: The arsenate reductase (Arr) homolog functions as the arsenite oxidase.

Hypothesis B: Arsenite oxidation in *A. ehrlichii* is catalyzed by a novel enzyme.

Rationale: In order to identify the proteins involved in arsenite oxidation, the proteome of MLHE-1 was examined. The protein profiles of cells that are grown heterotrophically on acetate under aerobic conditions were compared to those grown chemoautotrophically under anaerobic conditions. The anaerobic cultures contained arsenite as the electron donor and nitrate as the acceptor. Two-dimensional gels were run of different cell fractions to look for proteins only expressed under anaerobic conditions. Activity assays were done using methyl viologen and DCIP to locate any proteins showing oxidase abilities. Once the proteins were located, they were identified and characterized using
mass spectrometry. The genes responsible were predicted to be novel sequences or homologs of a respiratory arsenate reductase.
Chapter 2

Genome of *Alkalilimnicola ehrlichii* strain MLHE-1\(^T\)

Specific Aim: Complete the annotation and characterization of the *Alkalilimnicola ehrlichii* strain MLHE-1\(^T\) genome.

2.1 Background

Cells of *Alkalilimnicola ehrlichii* strain MLHE-1T (ATCC BAA-1101) were provided by R.S. Oremland (US Geological Survey, Menlo Park CA). The DNA was extracted by Rishu Bansal (Duquesne University) using the phenol method followed by treatment with RNAse\(^{34}\). The genome was sequenced at the Joint Genome Institute (JGI) using a combination of 4 kb, 8 kb and 40 kb (fosmid) DNA libraries. Draft assemblies were based on 42019 total reads providing 10x coverage of the genome. Sequence assembly and quality assessment was done using the Phred/Phrap/Consed software package (www.phrap.com). Gaps between contigs were closed by editing in Consed, custom primer walk, or PCR amplification (Roche Applied Science, Indianapolis, IN). Open reading frames were assigned consecutive locus tags in the form of “Mlg_#####” based on their position on the chromosome.

Two gene modeling programs, Critica (v1.05) and Glimmer (v3.x), were used by JGI for annotation, and a BLASTP search of the translations vs. Genbank's non-redundant database (NR) was conducted by JGI. Gene product assignments were made based on PRIAM, TIGRFam, Pfam, Intepro profiles, pairwise BLAST vs. Swiss-
Prot/TrEMBL, KEGG, and COG groups. A more detailed description of the specific protocols, programs, and parameters used in the sequencing, closure, and annotation can be found at the JGI web site (http://www.jgi.doe.gov). The complete sequence of the *A. ehrlichii* genome was submitted to GenBank and assigned the accession number NC_008340.

Initial annotation is automated and has limitations. Pathways can be incomplete, subunits of multimeric proteins may be missing, and misannotation can occur. For example, molybdenum-containing enzymes are often annotated as “putative anaerobic dehydrogenases” or “formate dehydrogenases”. The automated annotation may provide some information, but any gene of interest should be further examined to make sure it has been identified correctly.

There are seven main gene groups that need to be more closely inspected. First, the genes involved in arsenic respiration (*aox, arr*) and resistance (*ars*) need to be located, since the organism is capable of arsenite oxidation under anaerobic conditions. Since arsenite oxidation is known to be coupled to nitrate reduction, the nitrogen metabolism genes should also be examined. These included genes for denitrification (*nar, nir, nor, nos*) and assimilation (*nas, nir*). Genes for sulfur metabolism need to also be checked, since MLHE-1 is closely related to the sulfide oxidizer *Halorhodospira halophila*.

MLHE-1 can act as a heterotroph under aerobic conditions so genes involved in carbon metabolism need to be examined. Because carbon fixation can occur by different pathways, the annotation should be scanned for carbon fixation genes (RuBisCo, reverse TCA, acetyl-coA pathways). Genes for carbon monoxide dehydrogenase (*cox*) need to
also be examined. The genes involved with polyhydroxybutyrate storage (pha) should be
investigated, since the PHB granules form inside the cells grown on acetate. Motility
genes for pili and flagella should also be located and studied. Some pili have been shown
in other organisms to bind metal and strip electrons from the substance. Also, the
structures for motility might be associated with chemotaxis.

2.2 Materials and Methods

Basic information about the genome of MLHE-1 was compiled and compared
with the information on JGI for two other sequenced Ectothiorhodospira, Halorhodospira
halophila and Nitrococcus mobilis. A comparison of the MLHE-1 genome with the
genomes of these two organisms was also done based on their automated annotations. A
Venn diagram was produced based on the number of genes found to be held in common.

Manual work was done to verify and clarify the annotation of the Alkaliminicola
ehrlchii genome. BLAST analysis using the NCBI website was used to locate certain
specific genes of interest. The amino acid sequence of both subunits of the arsenite
oxidase (aox) from Alcaligenes faecalis was used to locate putative aox homologs in the
MLHE-1 genome. The A. faecalis amino acid sequence was used with a BLASTP in the
non-redundant protein sequence database (nr). The homologs for arsenite reductase were
found in a similar manner. Other genes of interest were located by searching the
automated annotated genome. When a gene of interest was found, the genes located up
and downstream were blasted using BLASTP and the nr database.

These approaches are independent of the annotation as it actually BLASTs against the
whole genome instead of searching the assignments already given.
KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway maps were found on the JGI website for *A. ehrlichii*. The KEGG maps represent molecular interactions in the cell and use a color scheme where green signifies PRIAM predictions above the cutoff value. Yellow signifies PRIAM predictions below the cutoff value, while gray denotes an Enzyme Commission (EC) number with no proteins found in the ExPASy ENZYME database. White boxes represent no information and were assumed to represent proteins absent in the cell. These pathways were examined to see what, if any, enzymes were missing.

Gene clusters pertaining to the genes of interest (Figures 2.3, 2.6, 2.8, 2.12, 2.14, 2.15, 2.16) were constructed based on the gene ortholog neighborhoods found on the IMG website (http://img.jgi.doe.gov/) and the BLASTP results. The genes were labeled with the appropriate locus tag and their gene identity. The colors of the genes signify their function based on the automated annotation, not the manual assignment.

Further evaluation of MLHE-1’s two arsenate reductases was necessary. Their amino acid sequences were aligned with those of the catalytic subunit of other molybdoenzymes of the DMSO family using the program ClustalX. A neighbor-joining phylogenetic tree was produced using PAUP (version 4.0B10). Appendix I shows the alignment upon which the tree is based.

2.3 Results

2.3.1 General Characteristics

The genome of *Alkalilimnicola ehrlichii* strain MLHE-1 has been sequenced to closure by the Joint Genome Institute (JGI)\(^{36}\). The circular chromosome is 3.276MB in size and has a G+C content of 67.53% (Figure 2.1, Table 2.1). It has an unusually high
percentage (69%) of stop codon UGA usage. There are 2940 total genes with 2883 encoding proteins. There are also 57 RNA encoding genes with the majority of them being for tRNA (48). One 5S rRNA, two identical 16S rRNAs, and three 23S rRNAs are also present in the genome. The chromosome also contains 18 pseudogenes. In total, there are 2245 genes (76.36%) with a predicted function and 638 without.

Clusters of orthologous groups (COGs) are used to classify the proteins encoded in a genome. The genome of MLHE-1 contains 162 COGS for translation, 121 for transcription, and 135 for replication, recombination, and repair. It has 199 proteins involved in signal transduction and 92 for cell motility. 100 proteins encoded in the genome conduct intracellular trafficking and secretion. Over 4.2% of the genome is involved in posttranslational modification, protein turnover, and chaperoning. 216 COGs are responsible for energy production and conversion with 95 being involved with the transportation and metabolism of carbohydrates. Approximately 5% of the genes are for inorganic ion transport and metabolism and roughly 2% encodes proteins for the biosynthesis, transport, and catabolism of secondary metabolites. Table 2.2 shows all of the COGs present in the MLHE-1 genome. It also shows how the clusters of orthologous groups compare to other organisms of the Gammaproteobacteria clade and the kingdom of Bacteria.

2.3.2 Relationship to Closest Relatives

*Alkalilimnicola ehrlichii* strain MLHE-1 is a very unique organism as it shares little in common with its closest sequenced relatives. *Nitrococcus mobilis* Nb-231 and *Halorhodospira halophila* SL-1 are both Gammaproteobacteria in the family Ectothiorhodospiraceae. *H. halophila* is a purple phototroph that has the ability to
Figure 2.1 A map of the chromosome of MLHE-1. The different colors of the circle represent the various COGs possessed by the organism. The color guide can be seen in Table 2.1. The inner black circle shows G+C content around the chromosome. Peaks in this circle might signify instances of horizontal gene transfer. The inner magenta/gold circle shows the G+C skew present. Figure taken from http://genome.ornl.gov/microbial/mlg_mlhe1.
Table 2.1 COG color guide to the chromosome map and gene clusters\textsuperscript{35}.

<table>
<thead>
<tr>
<th>COG Code</th>
<th>COG Function Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>[A]</td>
<td>RNA processing and modification</td>
</tr>
<tr>
<td>[B]</td>
<td>Chromatin structure and dynamics</td>
</tr>
<tr>
<td>[C]</td>
<td>Energy production and conversion</td>
</tr>
<tr>
<td>[D]</td>
<td>Cell cycle control, cell division, chromosome partitioning</td>
</tr>
<tr>
<td>[E]</td>
<td>Amino acid transport and metabolism</td>
</tr>
<tr>
<td>[F]</td>
<td>Nucleotide transport and metabolism</td>
</tr>
<tr>
<td>[G]</td>
<td>Carbohydrate transport and metabolism</td>
</tr>
<tr>
<td>[H]</td>
<td>Coenzyme transport and metabolism</td>
</tr>
<tr>
<td>[I]</td>
<td>Lipid transport and metabolism</td>
</tr>
<tr>
<td>[J]</td>
<td>Translation, ribosomal structure and biogenesis</td>
</tr>
<tr>
<td>[K]</td>
<td>Transcription</td>
</tr>
<tr>
<td>[L]</td>
<td>Replication, recombination and repair</td>
</tr>
<tr>
<td>[M]</td>
<td>Cell wall/membrane/envelope biogenesis</td>
</tr>
<tr>
<td>[N]</td>
<td>Cell motility</td>
</tr>
<tr>
<td>[O]</td>
<td>Posttranslational modification, protein turnover, chaperones</td>
</tr>
<tr>
<td>[P]</td>
<td>Inorganic ion transport and metabolism</td>
</tr>
<tr>
<td>[Q]</td>
<td>Secondary metabolites biosynthesis, transport and catabolism</td>
</tr>
<tr>
<td>[R]</td>
<td>General function prediction only</td>
</tr>
<tr>
<td>[S]</td>
<td>Function unknown</td>
</tr>
<tr>
<td>[T]</td>
<td>Signal transduction mechanisms</td>
</tr>
<tr>
<td>[U]</td>
<td>Intracellular trafficking, secretion, and vesicular transport</td>
</tr>
<tr>
<td>[V]</td>
<td>Defense mechanisms</td>
</tr>
<tr>
<td>[W]</td>
<td>Extracellular structures</td>
</tr>
<tr>
<td>[Y]</td>
<td>Nuclear structure</td>
</tr>
<tr>
<td>[Z]</td>
<td>Cytoskeleton</td>
</tr>
<tr>
<td>[NA]</td>
<td>Not assigned</td>
</tr>
</tbody>
</table>
Table 2.2  COGs contained within the genome of *A. ehrlichii* strain MLHE-1. The last two columns compare the amount of COGs in MLHE-1 with other clades. Modified from the NCBI website\(^{38}\).

<table>
<thead>
<tr>
<th>Description</th>
<th>Number of COGs</th>
<th>% in Genome</th>
<th>% in Gammaproteobacteria</th>
<th>% in Bacteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Translation</td>
<td>162</td>
<td>4.9708</td>
<td>4.2982</td>
<td>4.3421</td>
</tr>
<tr>
<td>RNA Processing and Modification</td>
<td>1</td>
<td>0.0307</td>
<td>0.0311</td>
<td>0.0144</td>
</tr>
<tr>
<td>Transcription</td>
<td>121</td>
<td>3.7128</td>
<td>5.9460</td>
<td>5.9565</td>
</tr>
<tr>
<td>Replication, Recombination, and Repair</td>
<td>135</td>
<td>4.1424</td>
<td>5.6096</td>
<td>4.9420</td>
</tr>
<tr>
<td>Chromatic Structure and Dynamics</td>
<td>3</td>
<td>0.0921</td>
<td>0.0162</td>
<td>0.0279</td>
</tr>
<tr>
<td>Cell Cycle Control, Mitosis, and Meiosis</td>
<td>39</td>
<td>1.1967</td>
<td>0.7877</td>
<td>0.7632</td>
</tr>
<tr>
<td>Nuclear Structure</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Defense Mechanisms</td>
<td>45</td>
<td>1.3808</td>
<td>1.1040</td>
<td>1.2533</td>
</tr>
<tr>
<td>Signal Transduction Mechanisms</td>
<td>199</td>
<td>6.1062</td>
<td>4.3089</td>
<td>3.9486</td>
</tr>
<tr>
<td>Cell Wall/Membrane Biogenesis</td>
<td>182</td>
<td>5.5845</td>
<td>4.6689</td>
<td>4.3540</td>
</tr>
<tr>
<td>Cell Motility</td>
<td>92</td>
<td>2.8230</td>
<td>2.1616</td>
<td>1.1429</td>
</tr>
<tr>
<td>Cytoskeleton</td>
<td>2</td>
<td>0.0614</td>
<td>0.0107</td>
<td>0.0120</td>
</tr>
<tr>
<td>Extracellular Structures</td>
<td>0</td>
<td>0</td>
<td>0.0241</td>
<td>0.0145</td>
</tr>
<tr>
<td>Intracellular Trafficking and Secretion</td>
<td>100</td>
<td>3.0684</td>
<td>2.4173</td>
<td>1.7598</td>
</tr>
<tr>
<td>Posttranslational Modification, Protein Turnover, Chaperones</td>
<td>139</td>
<td>4.2651</td>
<td>3.3036</td>
<td>2.9523</td>
</tr>
<tr>
<td>Energy Production and Conversion</td>
<td>216</td>
<td>6.6278</td>
<td>4.9092</td>
<td>4.7983</td>
</tr>
<tr>
<td>Carbohydrate Transport and Metabolism</td>
<td>95</td>
<td>2.9150</td>
<td>5.1274</td>
<td>4.8607</td>
</tr>
<tr>
<td>Amino Acid Transport and Metabolism</td>
<td>217</td>
<td>6.6585</td>
<td>7.5364</td>
<td>7.2550</td>
</tr>
<tr>
<td>Nucleotide Transport and Metabolism</td>
<td>56</td>
<td>1.7183</td>
<td>1.7399</td>
<td>1.7834</td>
</tr>
<tr>
<td>Coenzyme Transport and Metabolism</td>
<td>138</td>
<td>4.2344</td>
<td>3.1682</td>
<td>2.9938</td>
</tr>
<tr>
<td>Lipid Transport and Metabolism</td>
<td>87</td>
<td>2.6695</td>
<td>2.5132</td>
<td>2.9319</td>
</tr>
<tr>
<td>Inorganic Ion transport and Metabolism</td>
<td>164</td>
<td>5.0322</td>
<td>4.9139</td>
<td>4.5646</td>
</tr>
<tr>
<td>Secondary Metabolites Biosynthesis, Transport, and Catabolism</td>
<td>69</td>
<td>2.1172</td>
<td>1.9177</td>
<td>2.3300</td>
</tr>
<tr>
<td>General Function Prediction Only</td>
<td>349</td>
<td>10.7088</td>
<td>9.9953</td>
<td>10.4301</td>
</tr>
<tr>
<td>Function Unknown</td>
<td>216</td>
<td>6.6278</td>
<td>6.5640</td>
<td>6.1355</td>
</tr>
<tr>
<td>Not in COGs</td>
<td>432</td>
<td>13.2556</td>
<td>16.9267</td>
<td>20.1622</td>
</tr>
</tbody>
</table>
oxidize sulfide to sulfur, which is deposited outside the cell and transformed into sulfate. It is extremely halophilic and produces such compounds as glycine betaine, ectoine, and trehalose to balance the osmotic pressure\textsuperscript{39}. \textit{Nitrooccus mobilis} is a nitrite-oxidizing obligate chemoautotroph found only in marine environments\textsuperscript{40}. MLHE-1 can grow as a heterotroph or as a chemolithoautotroph by oxidizing arsenite. The other two organisms not been shown to possess this ability. \textit{H. halophila} does possess a homolog for arsenate reductase, but \textit{N. mobilis} does not.

\textit{A. ehrlichii} possesses 579 proteins that are unique to itself when compared to \textit{N. mobilis} and \textit{H. halophila}. It shares 310 solely with \textit{N. mobilis} and 346 solely with \textit{H. halophila}. All three organisms share 1634 proteins in common. The Venn diagram in Figure 2.2 shows this relationship. The GC content of MLHE-1 is more similar to that of \textit{H. halophila} (67.98\%) than to \textit{N. mobilis} (59.97\%). \textit{N. mobilis} only has one 16S rRNA, while MLHE-1 and \textit{H. halophila} each have two. The only organism of the three that has pseudogenes is MLHE-1. Both MLHE-1 and \textit{H. halophila} have a high usage of the stop codon UGA (69\% and 58.5\% respectively). This can be compared to \textit{Alkaliphilus oremlandii} OhILAs (8\%), \textit{Shewanella} sp. ANA-3 (15\%), and \textit{Geobacter metallireducens} (45\%). Data for \textit{Nitrooccus mobilis} was not available for comparison.

2.3.3 Specific Functions

2.3.3.1 Arsenic Transformation

The arsenite oxidizing ability of MLHE-1 is unique to it as the other known \textit{Alkalilimnicola} species seem to lack it. However, no arsenite oxidase genes (\textit{aoxA, aoxB}) were found in the sequenced genome when a BLASTP was done of the nr database for the organism using the homolog from \textit{A. faecalis}. Instead, a BLAST search revealed two
Figure 2.2 A Venn diagram showing the relationship of MLHE-1 to its two closest sequenced relatives, *Nitrocoecus mobilis* and *Halorhodospira halophila*. MLHE-1 has 1634 proteins in common with both organisms, 310 in common only with Nb-231, and 346 only in common with SL-1. It has 579 proteins that neither of the organisms possesses.
homologs for dissimilatory arsenate reductase (arrDABC and arrCABD), though the bacteria was previously shown to not grow on arsenate under tested conditions\textsuperscript{31}. The two operons for arsenate reductase are located at Mlg_0214 to Mlg_0217 and from Mlg_2425 to Mlg_2428, respectively, but are on opposite strands. MLHE-1 also possesses a homolog for arsenic resistance (arsBADCR), which has an operon from Mlg_2709 to Mlg_2713\textsuperscript{41}. There is an additional arsC at Mlg_0826 and several other arsR genes scattered throughout the genome (Mlg_0623, Mlg_0756, Mlg_1306, and Mlg_1776). Figure 2.3 shows the different gene clusters. Figure 2.4 shows a phylogenetic tree of the catalytic subunits of the DMSO reductase family of molybdoenzymes. The arr homologs of MLHE-1 cluster together with \textit{H. halophilus}, but are quite different than those in other organisms.

Dissimilatory arsenate reductases are expressed only under anaerobic conditions and are repressed when oxygen and nitrate are present\textsuperscript{42}. These reductases are periplasmic or membrane-bound heterodimers consisting of a large molybdopterin subunit with an iron-sulfur center and a small subunit that possesses four iron-sulfur centers. This smaller subunit is not homologous with the Rieske-type subunit of an arsenite oxidase\textsuperscript{43}. The larger subunit, ArrA, is catalytic and binds arsenate to convert it to arsenite. ArrB serves as a conduit for electrons stemming from c-type cytochromes in the respiratory chain. ArrC is the membrane-anchoring subunit, and ArraD is a chaperone-like regulator. MLHE-1 possesses two operons each with these four subunits of the dissimilatory arsenate reductase, but they are positioned on opposite strands and do not have the same gene order. The arrA Mlg_216 is slightly larger than Mlg_2426. Genes for arsenic resistance (arsBADCR) were also identified in the genome\textsuperscript{31}. The ars
Figure 2.3 The gene clusters for the two arsenate reductases and the arsenic resistance protein.

The reductases are found on separate strands at opposite ends of the genome. The genes for the subunits are of different sizes and are found in a different order between the two homologs.

Figure 2.4 A phylogenetic tree (neighbor joining) of representative enzymes (catalytic subunit) in the DMSO reductase family of molybdoenzymes based on amino acid sequence. Both homologs of arrA from A. ehrlichii, as well as homologs from H. halophila and M. magnetotacticum, form a sub-branch with the respiratory arsenate reductase (arrA) and not arsenite oxidase (aoxB). BisC is a biotin sulfoxide reductase, DorA is a DMSO reductase, FdhG is a formate dehydrogenase, NapA is a periplasmic nitrate reductase, NarG is a respiratory nitrate reductase, NasA is an assimilatory nitrate reductase, PsrA is a polysulfide reductase, SerA is a respiratory selenate reductase, TorA is a trimethylamine oxide reductase, and TtrA is a tetrathionate reductase.
system is the most widespread arsenic resistance mechanism that has been found in bacteria, yeast and some lower eukaryotes. The *ars* operon possesses three essential genes (*arsC*, *arsR* and *arsB*) but may have other genes present, like *arsA*, *arsD*, and *arsH* depending on the species. In the case of *A. ehrlichii*, there are two additions, *arsA* and *arsD*. *ArsC* is the protein responsible for reducing arsenate to arsenite by accepting electrons from glutathione. The arsenite is then removed from the cell by *ArsB*, an arsenite specific transmembrane protein. *ArsR* and *ArsD* both serve as regulatory proteins, but *ArsR* is nearly 10 times more sensitive to arsenic concentrations and binds the operator site with a much greater affinity. *ArsA* is an ATPase that can hydrolyze ATP as arsenite is expelled, and *ArsH* (absent in MLHE-1) has an unknown function.

It should also be noted that MLHE-1 has two Rieske-type genes, but neither are involved in arsenite oxidation based on JGI BLASTP results. Mlg_1558 codes for a Rieske-1 iron sulfur protein for a carbon dioxide dehydrogenase. When BLAST is used, the arsenite oxidase small subunit is produced as one possibility but the scores are very low (identity=32%, gaps=20%). Another Rieske 2F-2S subunit is found at Mlg_2454 and is associated with an amidase, and BLAST does not produce any arsenite oxidase (*aoxB*) in its list of results.

2.3.3.2 Nitrate metabolism

MLHE-1 possesses a unique set of genes for nitrogen metabolism. The genome has operons for respiratory nitrate reductase (*narKLXK₂GHJI*), assimilatory nitrate reductase (*nas*), nitric oxide reductase (*norDQBC*), and nitrous oxide reductase (*nosRZDFYL*). Nitric oxide reductases transform nitric oxide generated from the nitrite reductase into nitrous oxide. This nitrous oxide is then turned into dinitrogen by the
nitrous oxide reductase. However, since MLHE-1 lacks a nitrite reductase (*nirK* or *nirS*), the end product is nitrite, not nitrous oxide. Attempts at growth on nitric oxide and nitrous oxide were unsuccessful, so the reductases for these compounds are thought to be nonfunctional. The nitrogen pathways are pictured in Figure 2.5.

The operon for respiratory nitrate reductase is located at position Mlg_1000 to Mlg_1007. The protein encoded for by these genes is responsible for the conversion of nitrate to nitrite. The nitric oxide reductase genes have an operon at Mlg_2116 to Mlg_2120, but the first gene, *norE*, is on the reverse strand. The nitrous oxide reductase can be found at Mlg_1073 to Mlg_1078. There is an additional *nosL* at Mlg_1081.

The assimilatory nitrate reductase reduces nitrate to ammonia, foregoing the nitrite intermediate. The assimilatory nitrate reductase large subunit is positioned at Mlg_1700 and is followed by the small subunit and the alpha subunit apoprotein gene. Mlg_1703 is a nitrate transporter and is followed by the gene for a nitrate ABC inner membrane subunit. Nitrate transport ATP-binding subunits C and D are coded for by Mlg_1705 with an assimilatory nitrite reductase directly after. Mlg_1707 is uroporphyrinogen-III synthase. Mlg_1708 is on the opposite strand and codes for a probable nitrate regulatory protein. The gene clusters can be seen in Figure 2.6.

2.3.3.3 Sulfur metabolism

MLHE-1 has many genes for sulfur metabolism. The pathways for sulfur metabolism can be seen in Figure 2.7. Two small operons (Mlg_0623-Mlg_0625, Mlg_0633-Mlg_0635) are found in close proximity to each other. The first contain the genes for a rhodanese-related sulfurtransferase (*pspE*), cysteine desulferase (*nifS*), and the intracellular sulfur oxidation protein DsrE. Another operon for an intracellular sulfur
Figure 2.5 Nitrogen metabolism. Green boxes signify PRIAM predictions above the cutoff value. Yellow signifies PRIAM predictions below the cutoff value, while gray denotes an Enzyme Commission (EC) number with no proteins found in the ExPASy ENZYME database. White boxes represent no information. The red box marks the missing respiratory nitrite reductase (NirK/NirS) necessary for the conversion of nitrite to nitric oxide.

Figure 2.6 The gene clusters involved in nitrogen metabolism.
oxidation protein is found spanning from Mlg_1653 to Mlg_1664 (dsrABEE'HCMK LJOP). dsrN is found at Mlg_1666, and a sulfate transporter (sulP) is at Mlg_1668. Another operon of sulfur genes immediately follows from Mlg_1672 to Mlg_1682. The first gene of this group is sirA, a redox protein regulator for disulfide bond formation. This gene is followed by cysD and fcsD, a gene for sulfide dehydrogenase. The next three genes are for DMSO reductase. napC and cysG are next on the strand and are followed by another pspE, soxY, and soxZ. Another dsrC is found at Mlg_099741.

A fifth operon (Mlg_1260-Mlg_1269) contains a conglomeration of sulfur-related proteins. Mlg_1260 to Mlg_1264 are cysNDGHI. cysN and cysD are the alpha and beta subunits of a sulfate adenylyltransferase. cysG is a siroheme synthase for the reduction of sulfite to sulfide, and cysH is a phosphoadenylylsulfate reductase. The operon continues on the opposite strand. Mlg_1264 (cysI) is a nitrite/sulfite reductase beta subunit and is on the opposite strand. Mlg_1266 a cytochrome c followed by a sulfide dehydrogenase. Mlg_1269 is a putative sulfur oxidation protein (soxZ). An additional cysJ is found at Mlg_189241.

The genome lacks genes necessary to convert sulfite into hydrogen sulfide. The genome also lacks those required to transform sulfide into sulfur as MLHE-1’s relative, H. halophila SL-1 does as part of its metabolism.

2.3.3.4 Carbon Metabolism and Fixation

MLHE-1 has several means of carbon metabolism, though it has very few carbon compounds on which it grows successfully. It has major components of glycolysis, the reductive carboxylate cycle, methane metabolism, and the citric acid (TCA) cycle. These can be seen in Figures 2.9, 2.10, 2.11, and 2.13, respectively.
Figure 2.7  Sulfur metabolism. Green boxes depict PRIAM predictions above the cutoff value. Yellow signifies PRIAM predictions below the cutoff value, while gray denotes an Enzyme Commission (EC) number with no proteins found in the ExPASy ENZYME database. White boxes represent no information. The red box highlights the lack of predicted ability to transform sulfite into hydrogen sulfide.

Figure 2.8  The operons involved in sulfur metabolism.
Even though MLHE-1 cannot metabolize glucose it has many of the genes needed for the process. There is no gene for hexokinase based on the KEGG pathway, but a glucokinase is present at Mlg_0008. Glucose-6-phosphate isomerase is found at Mlg_0290. The gene for 6-phosphofructokinase is absent, which may be the reason glucose cannot be metabolized. The rest of the genes appear intact. There are two genes for fructose bisphosphate aldolases (Mlg 2089, Mlg_2844), one gene for triosephosphate isomerase (Mlg_1972), and three genes for glyceraldehyde-3-phosphate dehydrogenases(Mlg_1514, Mlg_2260, and Mlg_2840). Mlg_2842 is phosphoglycerate kinase, Mlg_1323 is phosphoglycerate mutase, and Mlg_1839 is enolase. Pyruvate kinase is located at Mlg_2843, while two pyruvate dehydrogenase subunits, E1 and E2, are found at Mlg_0269 and Mlg_22595, respectively. A phosphomannomutase is at Mlg_2845.

Genes that are involved with RuBisCo form an operon with those of glycolysis. This RuBisCo group of genes can be found from Mlg_2833 to Mlg_2839 and includes phosphoribulokinase (prkB), ribulose-phosphate 3-epimerase (rpe), RuBisCo transcriptional regulator (ccbR), RuBisCo large subunit (cbbL), RuBisCo small subunit (cbbS), cbbQ, and cbbO. Another RuBisCo-like gene, rbcL, is found at Mlg_1168. Transketolase is located at Mlg_2841.

Genes for methane metabolism are also found within the genome. An ethanol dehydrogenase at Mlg_2729 and a catalase/peroxidase at Mlg_1522 convert methanol to formaldehyde, but no gene to convert methane to methanol is present according to the KEGG pathway. The ethanol dehydrogenase may also be able to convert formaldehyde into formate. A serine hydroxymethyltransferase is present at Mlg_0373 to take 5,10-
methylenetetrahydrofolate to serine. This compound can also be reduced by methylenetetrahydrofolate reductase to 5-methyltetrahydrofolate. A tungsten-containing formate dehydrogenase encoded for by Mlg_2513 may convert formate into carbon dioxide. The genes for carbon monoxide dehydrogenase form an operon from Mlg_1562 to Mlg_1567. These genes include coxM, coxS, coxL, coxD, coxE, and coxF. An additional coxF is found at Mlg_1569 with a mobA preceeding it. A nifU nitrogen-fixing domain finishes the cluster at position Mlg_1570. All of the gene clusters can be seen in Figure 2.12.

The genes for proteins involved in the citric acid (TCA) cycle (Figure 2.13) are scattered throughout the genome. Citrate synthase is located at Mlg_2752 while two genes for aconitase are found at Mlg_1453 and Mlg_1454 but on opposite strands from each other. Three genes for isocitrate dehydrogenase are present. The first is located at Mlg_0262 and is NADP dependent. The genes for orotate phosphoribosyltransferase and phosphoenolpyruvate carboxylase immediately precede it. The second dehydrogenase is located at Mlg_0664 and is NAD-dependent. The third is NADP-dependent and located at Mlg_1121. Mlg_2608 and Mlg_2609 are the subunits for 2-oxoglutarate dehydrogenase. Mlg_2547 is the succinyl-CoA synthetase and the next gene downstream is the ligase. Mlg_1332 through Mlg_1335 are the genes for succinate dehydrogenase (sdhCDAB). Genes for fumarase can be found at Mlg_1359 and Mlg_2412. Two malate dehydrogenases are also present in the genome at Mlg_0246 and Mlg_0487 and are isolated from any operon. 
Figure 2.9 Glycolysis and gluconeogenesis pathways in MLHE-1<sup>35</sup>. The red box surrounds missing elements. Note that no enzymes are present to bring glucose into the cell and that conversion of fructose-6-phosphate to fructose-1,6-bisphosphate also is impaired.
Figure 2.10 Carbon fixation pathways in MLHE-1. Though the pathway is generic and was meant for phototrophs, most of it still pertains to MLHE-1. Missing enzymes of importance are denoted by red boxes. Note the lack of an enzyme to convert sedoheptulose-1,7-bisphosphate to sedoheptulose-7-phosphate. Enzymes responsible for converting 1,3-bisphosphate glycerate and xylose-5-phosphate into glyceraldehydes-3-phosphate are also absent.
Figure 2.11 Methane metabolic pathway. The red box denotes holes in the pathway. No enzymes have been identified that are capable of converting methane to methanol. The rest of the pathway is mostly intact.

Figure 2.12 Gene clusters involved in carbon metabolism. The genes for RubisCo are located just upstream of those involved in glycolysis. The cox genes could be involved in methane metabolism.
Figure 2.13  The TCA cycle in MLHE-1 appears intact.

Figure 2.14  The small operon of polyhydroxybutyrate genes.
2.3.3.5 Storage
MLHE-1 uses polyhydroxybutyrate globules as storage mechanisms when it is
grown heterotrophically on acetate. These genes are mostly scattered throughout the
genome. Mlg_0443, Mlg_0748, and Mlg_2239 code for a polyhydroxybutyrate syntase
(PhaC), and Mlg_2380 encodes a polyhydroxybutyrate polymerase domain protein. A
small operon spans from Mlg_2485 to Mlg_2487. The first gene, phaC, is on the forward
strand, while phaB, which encodes a NADPH-linked acetyl-CoA reductase, and phaR are
on the reverse. phaR is the repressor gene to shut down the biosynthesis of
polyhydroxybutyrate. No phaA genes for ketothiolase were identified in the genome
from the automated annotation but the organism is still capable of the synthesis and
storage of the molecule.

2.3.3.6 Motility and Secretion
MLHE-1 has one polar flagellum and multiple pili. The genes responsible for
their production are scattered throughout the genome. There are 3 operons associated
with the flagellum, and 15 genes form the first operon from Mlg_0698 to Mlg_0714, with
the exception of Mlg_0699, which is the methyltransferase fkbM. The gene order is fliC,
fkbM (on the reverse strand), flgL, flaG, flaG', fliD, fliS, fliT, fleQ, fleT, fliE, fliF, fliG,
fliH, fliI, fliJ, and fliK. Another operon is located from Mlg_0888 to Mlg_0903. The gene
order is flgN, flgM, flgA, cheW, cheR, flgB, flgC, flgD, flgE, flgF, flgG, flgH, flgI, flgJ,
flgK, and flgL. The first three genes are on the opposite strand from the rest of the
operon. A third operon (Mlg_0974-0985) is composed of fliL, fliM, fliN, fliO, fliP, fliQ,
fliR, flhB, flhA, flhF, flhG, and fliA. Just downstream of this operon (Mlg_0986 to
Mlg_0992), are a cluster of genes involved with chemotaxis, with the exception of cpaE.
They are \textit{cheY}, \textit{cheZ}, \textit{cheA}, \textit{cheB}, \textit{cpaE}, \textit{cheW}, and \textit{cheW'^1}. All of the flagellar genes are pictured in Figure 2.15.

The pili genes are also dispersed throughout the genome and most are contained in operons. The first spans from Mlg\_0252 to Mlg\_0256. It is composed of \textit{pilA}, \textit{fimT}, \textit{pilV}, \textit{pilW}, and \textit{pilX} in that order. The latter three are pseudopilin genes and may be involved in secretion. An additional \textit{pilA} is found just before the operon at Mlg\_0249. The second major operon exists at Mlg\_0355 to Mlg\_0360 and contains both pilin and chemotaxis genes. The gene order is \textit{pilG}, \textit{pilH}, \textit{pilI}, \textit{pilJ}, \textit{cheA}, and a chemosensory pili system gene, \textit{chpC}. \textit{pilG}, \textit{pilH}, \textit{pilI}, \textit{pilJ} are involved in twitching motility. The third operon contains six genes involved in pseudopilin. It begins at Mlg\_0855 and ends with Mlg\_0861. The components are \textit{pilE} (on reverse strand), \textit{fimT}, \textit{fimT'}, \textit{pilV}, \textit{pilW}, \textit{pilX}, and \textit{pilX'}. The fourth operon consists of only three genes, \textit{pilB}, \textit{pilC}, and \textit{pilD} at Mlg\_2079 to Mlg\_2081. These genes are involved in Type II secretion processes. This cluster is preceded by the flagellar motor proteins \textit{motA} at Mlg\_2075, \textit{motB} at Mlg\_2076, and \textit{cheY} at Mlg\_2077. The fifth operon is the largest and spans from Mlg\_2151 to 2165. It contains genes for pilus assembly and secretion. The genes include, in order, \textit{citB}, \textit{citA}, \textit{cheY}, \textit{tadG}, two hypothetical protein genes, \textit{tadD}, \textit{tadB}, \textit{tadB'}, \textit{cpaF}, \textit{cpaE}, \textit{cpaC}, \textit{cpaB}, a Flp/Fap pilin component gene, and \textit{cpaA}. The sixth and final operon for pili and fimbrial assembly spans from Mlg\_2755 to Mlg\_2760 and includes \textit{pilM}, \textit{pilN}, \textit{pilO}, \textit{pilP}, and \textit{pilQ}, and an adhesin gene. There is an additional \textit{pilT} and \textit{pilU} at Mlg\_0342 and Mlg\_0343, respectively\textsuperscript{41}. The gene clusters are shown in Figure 2.16.
Figure 2.15 Gene assembles used for flagella. Note the chemotaxis che genes within the clusters.

Figure 2.16 Gene clusters involved with pili. Genes for chemotaxis (che) are again present within these clusters.
2.4 Discussion

*Alkalilimnicola ehrlichii* has a very unique genome. It uses a normally rare stop codon, UGA, nearly 69% of the time. It possesses two identical 16S rRNAs. It has very little in common with *H. halophila*, and even less with *N. mobilis* though both are Gammaproteobacteria in the clade Ectothiorhodospiraceae. It has 1980 genes in common with *H. halophila* and 1944 genes with *N. mobilis*. It has 579 genes that cannot be found in the other two genomes. MLHE-1 also has 18 pseudogenes that neither of the other organisms possesses.

The organism has a variety of pathways for carbon, nitrogen, and sulfur metabolism, but many are lacking key enzymes. MLHE-1 is unable to metabolize nitrite to nitric acid because it lacks a nitrate reductase. MLHE-1 cannot metabolize sulphur like *H. halophila* because of two missing genes. It is unable to utilize glucose in glycolysis, because it lacks the gene for 6-fructophosphokinase. Fructose 6-phosphate cannot be converted to fructose 1,6-bisphosphate, and so the cycle is disrupted. The citric acid cycle, however, seems to have a full set of genes needed for operation. Calvin cycle proteins, like RuBisCo, and methane metabolism proteins, like carbon monoxide dehydrogenase, are also present within this bacteria’s genome.

The motility genes in the *A. ehrlichii* genome seem normal. There is no evidence that the organism uses its pili to bind to metal to strip electrons. There is evidence that the organism might be chemotaxic toward arsenic since several genes for chemotaxis were found in and around operons for both flagella and pili. Future experiments in the laboratory could test this hypothesis.

The most interesting thing about the genome of this arsenite oxidizer is its lack of
an arsenite oxidase. No homolog for either of the subunits is present. Instead, there are
two gene clusters for arsenate reductase. Each has a different gene order, different size
subunits, and they are located on opposite strands. There is also an operon for arsenic
resistance on the reverse strand.
Chapter 3

Proteome of *Alkalilimnicola ehrlichii* strain MLHE-1\(^T\)

Specific Aim: To determine the mechanism of arsenite oxidation by *Alkalilimnicola ehrlichii* strain MLHE -1\(^T\).

3.1 Materials and Methods

3.1.1 Media Preparation

Media for *Alkalilimnicola ehrlichii* strain MLHE-1\(^T\) was prepared by combining 10.6 grams of sodium carbonate (Na\(_2\)CO\(_3\)), 4.2 grams of sodium bicarbonate (NaHCO\(_3\)), 0.1 grams of ammonium sulfate ((NH\(_4\))\(_2\)SO\(_4\)), 0.025 grams of magnesium sulfate (MgSO\(_4\)·7H\(_2\)O), 0.15 grams of monobasic potassium phosphate (KH\(_2\)PO\(_4\)), 0.08 grams of dibasic potassium phosphate (K\(_2\)HPO\(_4\)), and 90 grams of sodium chloride (NaCl) in one liter of nanopure water. 5 mL of SL-10 trace element mix was also added. The mix contains 10 mL of hydrochloric acid (7.7M), 1.5 grams of ferric chloride (FeCl\(_3\)-6H\(_2\)O), 0.07 grams of zinc chloride (ZnCl\(_2\)), 0.1 gram of manganese chloride (MnCl\(_2\)-4H\(_2\)O), 0.006 grams of boric acid (H\(_3\)BO\(_3\)), 0.19 grams of cobalt chloride (CoCl\(_2\)-6H\(_2\)O), 0.002 grams of cupric chloride (CuCl\(_2\)-2H\(_2\)O), 0.024 grams of nickelous chloride (NiCl\(_2\)-6H\(_2\)O), and 0.036 grams of sodium molybdate dihydrate (Na\(_2\)MoO\(_4\)-2H\(_2\)O). The media was adjusted to pH 9.8 and dispensed for autoclaving into culture bottles or tubes. Anaerobic media was degassed for five minutes with an 80:20 ratio of nitrogen to carbon dioxide,
while aerobic media was sealed without degassing. All media was then autoclaved. After cooling, sodium acetate was added to the aerobic media using sterile filtration to bring the concentration to 5mM. Sodium arsenate and sodium nitrate was filtered into the anaerobic media once the stock solutions were degassed to bring the final concentration of each to 5mM.

3.1.2 Cell Harvesting and Fractionation

Cells were grown in 22mL culture tubes and 150mL septum-sealed bottles and transferred to new media weekly. In order to accumulate enough material, aerobic cultures were also grown in one liter bottles. When sufficient growth was apparent, the material was centrifuged at 7,000RPM (5204xg) for 10 minutes to collect the cell pellet. The pellet was then homogenized in Buffer A, which consisted of 10 mM Tris and 1 mM EDTA disodium salt. The suspension was then French pressed three times to ensure that the majority of the cells had lysed. 1 mL of protease inhibitor cocktail (Sigma, St. Louis, MO) and 2 mL of DNAse inhibitor was added to every 20 mL of cell suspension. The suspension was then centrifuged at 7,000RPM (5204xg) for 30 minutes at 4ºC. The pellet contained unlysed cells, while the supernatant held the cell lysate. The supernatant was then placed in ultracentrifuge tubes for ultracentrifugation. The samples were spun at 45,000RPM (272403xg) for 1.5 hour at 4ºC. The cytosolic and periplasmic fraction was contained in the supernatant, while membrane and particulates were restricted to the pellet. A CHAPS-soluble membrane fraction was derived from the pellet by placing the sample in 2% CHAPS overnight.
3.1.3 Protein Concentration Determination

Bradford assays were performed to determine the protein concentration in each cell fraction. A stock solution of bovine serum albumin (BSA) was by dissolving 1mg in 1 mL water. 0.5µL, 1µL, 2µL, 5µL, and 10µL of this stock solution were added to 1mL of Bio-Rad Quick Start Protein Assay Dye Reagent. 1µL of each cell fraction was also suspended in 1mL of the dye reagent. The samples were incubated at room temperature for 5 minutes. The absorbance was read on the spectrophotometer at 595nm. A graph depicting the relationship between BSA concentration and absorbance was plotted, and a best-fit line was added. The equation of the line was used with the absorbance of the unknowns in order to calculate the protein concentration of each sample.

3.1.4 TCA Precipitation

TCA precipitation was used to concentrate samples. 700µg of protein sample was placed in a 1.5 microcentrifuge tube. 1mL of nanopure water was added along with 150µL of 100% trichloroacetic acid. The samples were placed on ice and refrigerated for at least 10 minutes. The samples were then spun at 12,000RPM (15294xg) for five minutes. The supernatant was discarded and the process was repeated once more. Upon discard of the supernatant, 1mL of chilled (-20°C) 100% acetone was added to the pellet. The pellet was resuspended and then centrifuged again at 12,000RPM (15294xg) for five minutes. The supernatant was discarded and any remaining acetone was allowed to evaporate from the sample. If the sample was to be used for a SDS-PAGE gel, 50µL of 50mM Tris-HCl buffer (pH 8)and 50µL of 2x sample buffer were added. The sample buffer consisted of 1.5mL of 20% SDS, 1mL of 100% glycerol, 1.25mL of nanopure water, and 1.25 mL of 0.5M Tris-HCl buffer (pH 6.8). If the samples were intended for
use for 2-D gels, the pellets were resuspended in 100µL rehydration buffer. This buffer contained 10.5g urea, 3.8g thiourea, 1.0g CHAPS, and 0.07g dithiothreitol in 25mL of nanopure water.

3.1.5 SDS-PAGE Gels

12% Tris-HCL Bio-Rad Readyprep gels (Hercules, CA) were used to run SDS-PAGE. The plastic strip was removed from the bottom of the gel, and the gel was placed inside the Bio-Rad gel apparatus. The comb was removed, and running buffer was added. The buffer consists of 22.5 grams glycine, 4.5 grams Tris, and 1.0 gram SDS in 1 L distilled water. 30µg of protein was placed in a microcentrifuge tube, and 5µL of Laemmli buffer was added. The entire contents were then loaded into each well on the gel. 5µL of Bio-Rad Precision Plus Protein Standards were loaded into the first lane for comparison upon completion. The gel was run at 75 volts until the dye front reached the bottom of the gel. The gel was then placed into Coomassie stain and left overnight. The Coomassie stain was made by combining 0.12g Coomassie Brilliant Blue R-250 with 125mL methanol, 50mL glacial acetic acid, and 325mL distilled water. The Coomassie was poured off after sufficient staining, and the gel was destained with a 25% methanol, 8% glacial acetic acid solution for several hours. If less protein was used, silver staining was necessary because of its higher sensitivity. In order to silver stain, the gels were placed in a fixing solution of 45% methanol and 10% glacial acetic acid for twenty minutes. They were then quickly rinsed twice with deionized water and then washed with water for an additional twenty minutes. The water was then replaced with 0.02% sodium thiosulfate for two minutes. Two water washes for one minute each then followed. A solution of 0.1% silver nitrate was then poured over the gel, and it was
stored at 4°C for 45 minutes. The gel was again rinsed with water for one minute before the developer was added. The developer consisted of 12.5g of sodium carbonate and 150µL of formalin in 500mL of deionized water. The developer was left on the gel until bands began to resolve. The gel was then washed extensively with water to cease the reaction. The gels were photographed and stored in water in airtight plastic Ziploc bags.

3.1.6 2-D SDS-PAGE Gels

2-D SDS-PAGE gels were used to separate and visualize the proteins present in the cytosolic cell fractions. The first dimension is done on Bio-Rad broad range (pH 3-10) 11cm IPG ReadyStrips (Hercules, CA), which were stored at -80°C. First, paper wicks were placed over the electrodes in the channel on the IEF focusing tray. 10µL of nanopure water was pipetted onto each wick. 100µg of protein was added to 200µL of rehydration buffer in the center of the channel. The buffer was made 25mL at a time and frozen. It contained 10.5g urea, 3.8g thiourea, 1.0g CHAPS, and 0.07g dithiothreitol and brought to 25mL with nanopure water. 2µL of carrier ampholytes (Bio-Rad Bio-Lyte 3-10) was added to the protein and buffer solution immediately before placing the strip into the channel. The plastic sheet was carefully pulled away from the IPG strip using forceps, and it was placed in the proper orientation into the tray on top of the protein sample. Forceps were used to press the strip down gently to remove any air bubbles. The strip was then overlaid with 2mL of mineral oil to prevent evaporation during the active rehydration. The focusing tray was then placed in the PROTEAN IEF cell. Active rehydration was achieved by running the rehydration program at 50V for 12-16 hours at 20°C. The cell was programmed to start electrophoresis as soon as rehydration was complete. The program started at 0V and ended at 8,000V over 35,000V-hours. Rapid
voltage ramping was used, and the temperature was maintained at 20°C. When electrophoresis was completed, the strips were immediately removed from the apparatus. The mineral oil was drained, and the strips were placed gel side up in a clean rehydration tray. They were again overlaid with mineral oil and the top of the tray was secured by wrapping the container in plastic wrap. Strips were stored in this manner at -80°C for up to two weeks before running the second dimension.

IPG strips were removed from the freezer and allowed to thaw for 10 minutes on the benchtop. Equilibration buffers were prepared during this time in the amount of 50mL. Both equilibration buffers contained 18.02g urea, 1.0g SDS, 2.27g Tris, 10mL glycerol, and nanopure water to bring the volume to 50mL. 0.1g dithiothreitol was added to equilibration buffer I, while 0.25g iodoacetamide was added to buffer II. The pH of both was brought to 8.3. Once the IPG strips were thawed, excess mineral oil was blotted off without touching the gel side. They were then placed in a new rehydration tray with 4mL of equilibration buffer I. The tray was placed on an orbital shaker for 10 minutes, and then the first buffer was carefully poured off. 4mL of equilibration buffer II was then added, and the tray was placed on the shaker for an additional 10 minutes. The second buffer was decanted, and the strips were dipped into running buffer for 5 seconds. Excess liquid was blotted off with a Kimwipe before applying the IPG strips to the second dimension gel.

A 12% acrylamide gel with a 4% stacker was used for the second dimension. The 12% solution was made by combining 7.5mL of 1.5M Tris-HCl buffer with 1% SDS and pH 8.8, 9mL of 40% Bis-Acrylamide, 13.5mL of nanopure water, and 150µL of 20% ammonium persulfate. The solution was degassed for 5 minutes, and 15µL of TEMED
was added. The gel was then poured leaving several centimeters for the stacking gel. The stacking gel was made by combining 1.5mL 40% Bis-Acrylamide, 3.75mL of .5M Tris-HCl buffer with 1% SDS and pH 6.8, 9.5mL of nanopure water, and 150µL of 20% ammonium persulfate. The solution was degassed for 5 minutes, and 15µL of TEMED was added before adding the solution to the top of the gel leaving about one centimeter at the top. The gel was allowed to set before the IPG strip was attached. The strip was positioned on the gel with the plastic against the back of the glass. A plastic spacer was placed at the positive side of the strip to make a well for protein standards. Overlay agarose was melted and pipetted over the IPG strip. Forceps were used to push down the strip to remove any bubbles underneath it. Once the overlay agar set, the spacer was removed and the gels were attached to the gel apparatus. It was then placed in the tank and running buffer was added. Standards were loaded and the gel was run at 100V until the dye front reached the bottom of the gel. The gels were then Coomassie and/or silver stained as described previously.

Additional gels were run by Dr. Peter Chovanec (Duquesne University) on the GE Amersham system in a similar manner as described above.  

3.1.7 DIGE

Two-dimensinal fluorescence difference gel electrophoresis (DIGE) was done based on the protocol set forth by Tonge et al., Yan et al., and GE Healthcare. Three different dyes, Cy2, Cy3, and Cy5 were used for this portion of the experiment. The CyDyes were reconstituted with high quality anhydrous dimethylformamide (DMF) to make a stock solution of 1.0mM. 2µL of the stock CyDye was then diluted with 3µL of DMF to bring the amount of CyDye to 4pmol per 1µL solution.
The cytosolic cell fractions from both aerobically and anaerobically grown cultures of *A. erlichii* were labeled. 50µg of aerobic protein was placed in a microcentrifuge tube with 1µL of Cy5. The same amount of anaerobic protein was placed in another tube with 1µL of Cy3 dye. 25µg of the anaerobic fraction was combined with 25µg of the aerobic fraction, and 1µL C2 was added to the contents of this container. The samples were then briefly vortexed and centrifuged before placing them on ice in the dark. After 30 minutes, 1µL of 1mM lysine was pipetted into each tube to stop the reaction. The tubes were kept in the dark on ice for an additional 10 minutes before an equal amount of 2x sample buffer was added. The three tubes were stored in the dark on ice for another 10 minutes before their contents were combined. The labeled sample was then used to load an IPG strip, which was then processed as described previously to run the first and second dimensions of a gel. Once the second dimension was complete, the gel was viewed using the Typhoon 8600 imager. The Cy3-labeled (anaerobic) proteins fluoresced green at 590nm, while the Cy5-labeled (aerobic) spots fluoresced red at 680nm. The internal standard, Cy2, was not able to be visualized using this particular imager.

3.1.8 Blue Native Gels

Blue native gels were used to isolate membrane protein complexes using a modified protocol from Stenberg et al.⁴⁹ A 10% acrylamide gel was used with a 4% stacker. The 10% gel was made by combining 3.34mL of 30% acrylamide, 50µL of 20% ammonium persulfate, 6.67mL of gel buffer, and 10µL of TEMED. The gel buffer was made by dissolving 3.28g of 6-aminohexanoic acid and 25g Bis-Tris in 100mL of nanopure water. The pH was adjusted to 7.0. The gel solution was poured into a Bio-
Rad mini-gel casting apparatus until 1.5cm remained at the top. The 4% gel was made by adding 0.67mL of 30% acrylamide, 25µL of 20% ammonium persulfate, and 5µL of TEMED to 4.33mL of gel buffer. This solution was carefully pipetted onto the top of the 10% gel to avoid mixing. A 1mm comb was inserted into the gel, and it was left to solidify. The anode buffer was a 50mM (10.46g/L) Bis-Tris solution with a pH of 7.0, while the cathode buffer was a 50mM tricine, 15mM Bis-Tris solution (pH 7.0). The cathode buffer also contained 0.05% (w/v) Coomassie G250 and 0.03% (w/v) TritonX or DDM depending on the samples being run.

50mL of ACA750 sample buffer was made by combining 4.919g of 6-aminohexanoic acid, 0.523g Bis-Tris, and 0.0093g of EDTA disodium salt in nanopure water, and the pH was adjusted to 7.0. The buffer was aliquotted into 10mL tubes. 0.1mL TritonX was added to one tube, while a 1%DDM buffer was also created. 0.25g of Coomassie G250 was also added to each 10mL tube. Approximately 100µg of protein sample was suspended in 20µL of either the 1% TritonX or DDM buffer. The microcentrifuge tubes were held at 4°C for at least twenty minutes before centrifuging at 11,000RPM (12883xg) and 4°C for 30 minutes. The supernatant was collected and 20µL was loaded into the appropriate well on the gel. A high molecular weight standard was also loaded (GE Healthcare HMW calibration kit for native electrophoresis), after it was reconstituted with 100µL of ACA750 buffer. The gel was run at 125V under refrigeration until the dye front reached the bottom of the gel. Coomassie was used overnight to further stain the gels, and the destaining was done as for any SDS-PAGE gel. Any bands of interest were extracted and sent to the University of Pittsburgh Proteomics Facility for LC/MS/MS mass spectrometry.
The second dimension of the gel began with excising bands of interest from the gel. This protocol was done because cutting the entire lane and placing it at the top of a second gel proved too difficult and inefficient. Bands were excised from the blue native gel using a sterile razor blade or pipette tip and placed into a small weighing dish. To extract the proteins from the gel, the bands were cut into tiny pieces and then placed in a 300K Nanosep concentrator column (Pall, Ann Arbor, MI). 50µL of equilibration buffer (2% SDS, 250mM Tris-HCl, pH 6.8) was added to the column before centrifuging at 12,000 RPM (15294xg) for 20 minutes. An additional 20µL of buffer was then added to the top of the column and centrifuged again for an additional 5 minutes. The solution in the bottom of the tube was then transferred to a second Nanosep column with a 10K filter. The column was centrifuged at 12,000 RPM (15294xg) for 10 minutes. The proteins remaining in the filter were collected by adding 20µL of Laemmli buffer and vortexing. The liquid was extracted and loaded into wells on an SDS-PAGE gel.

The second dimension gels were typical SDS-PAGE gels. A 10% gel was made by combining 2.5mL of 1.5M Tris-HCl buffer (pH 8.8, 1% SDS), 3.3mL of 30% acrylamide, 4.2mL of nanopure water, 25µL of 20% ammonium persulfate, and 5µL of TEMED. The 4% stacker contained 0.508 mL of .5M Tris-HCl buffer (pH 6.8, 1% SDS), 0.267mL of 30% acrylamide, 1.225mL of nanopure water, 5µL of 20% ammonium persulfate, and 1µL of TEMED. Both solutions were degassed for 5 minutes before pouring the gel. A 10-well comb was placed at the top of the gel, and it was allowed to solidify. Once the gel had set, it was placed into the gel apparatus and covered with the normal running buffer. The comb was removed, and the samples were loaded. On some occasions BioRad precast gels were used in place of freshly poured ones. 5µL of Bio-
Rad Precision Plus Protein Standards was also loaded onto the gel. The gel was run at 125 volts until the dye front reached the bottom of the gel. The gels were Coomassie or silver stained as described previously.

3.1.9 Digestion

Protein spots of interest were excised from the gel using a pipette tip or a clean, sterile razor blade. The pieces were placed into individual centrifuge tubes and washed with 100µL of water for 5 minutes. The samples were then spun down and the liquid removed. Digestion proceeded according to the protocol set forth by Shevchenko et al.\textsuperscript{50}.

The gel pieces were rehydrated in 100µL of 100mM ammonium bicarbonate (NH\textsubscript{4}HCO\textsubscript{3}), and after 10 minutes, 100µL of acetonitrile was added to each tube. The tubes were placed on a shaker for 15 minutes before they were again spun down and the liquid removed. An additional 150µL of acetonitrile was then added to shrink the gel pieces. After 10 minutes, all liquid was again removed using a pipette. The samples were dried using a vacuum centrifuge. 100µL of digestion buffer containing 50mM NH\textsubscript{4}HCO\textsubscript{3}, 5mM CaCl\textsubscript{2}, and 12.5ng/µL of trypsin was added into each tube. The samples were kept on ice for 30 minutes with more buffer being added halfway through if necessary. Any unabsorbed buffer was removed after this time, and 20µL of the same buffer without trypsin was added to keep the pieces wet during enzymatic cleavage. The samples were left on a 37°C heatblock overnight.

Peptides were extracted on the second day. 10µL of nanopure water was added to the digest, which was then centrifuged briefly and incubated for 15 minutes at 37°C on a shaking platform. The gel pieces were spun down yet again, the water removed, and 100µL of acetonitrile added. The tubes were incubated again for 15 minutes at 37°C on a
shaking platform. After centrifuging, the supernatant was removed, and 40µL of 5% formic acid was added. Then the tubes were vortexed and incubated for 15 minutes at 37ºC. The samples were spun down again, and an equal amount of acetonitrile was added to the contents. An additional 15 minute incubation at 37ºC was done. The gel pieces were spun down for the final time and the supernatant containing the extracted proteins were collected.

3.1.10 Mass Spectrometry

Alpha-cyano-4-hydroxycinnamic acid was recrystallized and used to prepare the matrix solution by dissolution at a concentration of 10 mg/mL in 70% acetonitrile and 0.1% TFA. 0.5µL of peptide solution was deposited onto the MALDI target, rapidly followed by 0.5µL of this matrix solution.

All mass spectrometry was done at the University of Pittsburgh Genomics and Proteomics Core laboratory. MALDI-TOF mass spectrometry was conducted on samples isolated in SDS-PAGE gels and the second dimension of blue native gels by Dr. Peter Chovanec (Duquesne University). The MALDI-TOF M/S analysis was performed on a Voyager-DE STR MALDI-TOF mass spectrometer from Applied Biosystems in the reflector mode. External mass calibration was achieved utilizing calibration mixture of the Sequazyme peptide mass standard kit also from Applied Biosystems. The proteins were identified by searching in the NCBInr database using the program Mascot (www.matrixscience.com). Protein identities were assigned if proteins score was significant within a maximum error spread of 1 Da and the estimated molecular weight and pI of the candidate protein were in agreement with the assignment.
3.1.11 Activity Assay - Vials

Activity assays were done under anaerobic conditions in order to see if arsenite oxidase and arsenate reductase activity could be seen. The chemical reactions for these equations can be seen in Figure 3.1. 10mM Tris buffer (pH 9.8) was made and filtered. The reductase assay required the creation of a 900mM sodium arsenate dibasic heptahydrate stock in the 10mM Tris buffer. A 300µM methyl viologen (ε=13.1 mM⁻¹cm⁻¹) stock was also made. 50µL of the arsenate stock was added to 50mL of the methyl viologen solution. The bottle was degassed for 5 minutes with nitrogen. A 0.575M sodium hydrosulfite (dithionite) stock was made using degassed Tris buffer. Degassing continued as the dithionite was added and was maintained for an additional 5 minutes. All of the bottles were capped using rubber stoppers and aluminum seals. The headspace of each was degassed for 5 minutes, and 50µL of the dithionite stock was transferred into the arsenate/methyl viologen solution, which caused the color to turn to a dark blue. 100µL of the cell lysate was placed into smaller vials and degassed for 10 minutes. Using a degassed syringe, 2.8mL of the reaction mixture was injected into the vials containing the protein samples. The vials were allowed to sit on the benchtop for up to one hour to see if a color change indicating reductase activity took place. When the color disappeared, additional dithionite was added to recycle the reaction.

The oxidase activity was done somewhat differently. A 60µM 2,6-dichlorophenolindophenol (DCIP) (ε= 23mM⁻¹cm⁻¹) and 20mM sodium arsenite solution was made using the 10mM Tris buffer. 200µM of methyl viologen could also be substituted for the DCIP. The solution was degassed, capped, and sealed before degassing the headspace with nitrogen. 2.8mL of this solution was injected into
Figure 3.1 Schematics of arsenite oxidase and arsenate reductase activity assays. A. Arsenite oxidation in the presence of DCIP. B. Arsenite oxidation in the presence of methyl viologen. C. Arsenite reduction using methyl viologen as the electron acceptor.
anaerobic vials containing 100µL of the cell lysate, which had been degassed as previously described. The vials were allowed to sit for up to one hour to observe any color change indicating oxidase activity.

The same oxidase activity assay using DCIP was also carried out in an anaerobic cuvette in a spectrophotometer to examine the rate of enzymatic activity. The wavelength was set at 600nm and a blank containing the arsenite/DCIP solution was used. Approximately 100µL of degassed cell lysate was injected into the anaerobic cuvette using a degassed syringe at time zero. Readings were taken for 45 minutes before the data was plotted on a graph of absorbance versus time using Excel. The data was further analyzed using Prism software.

3.1.12 Activity Assays- Zymograms

In gel assays were done on SDS-PAGE gels and modified blue native gels. The blue natives were run as previously described, with the exception of omitting Coomassie G250 from both the ACA750 sample buffer and cathode buffer, as it was found to inhibit activity. When electrophoresis was complete, the gel was placed in a plastic dish and placed in the anaerobic dry box. Degassed solutions of 10mM oxidized methyl viologen or 5mM DCIP was poured onto the gel and left to sit for 30 minutes. The solution was then replaced with a 20mM arsenite solution and allowed to incubate. Arsenite activity was determined by the presence of a blue band on the methyl viologen gel or a clear band on the DCIP gel. Since this assay is very oxygen sensitive, the gel was placed into a securely closed Ziploc bag before removing it from the dry box. The gel was quickly taken to a gel scanner to be photographed while still in the bag. Attempts to document
the gel outside of the sealed bag were unsuccessful. The activity completely vanishes after approximately one minute of air exposure.

The same gel was then used for the reductase activity. The gel was placed in a degassed 10mM methyl viologen solution that had been reduced completely with hydrogen sulfite (1.5g in 100mL). Nitrogen was bubbled into the dish during the 5 minute incubation. The solution was poured off and replaced with a 20mM arsenite solution. Nitrogen was again allowed to keep bubbling into the container housing the gel. Reductase activity resulted in clear bands being present on the dark blue gel. The gel was scanned before the methyl viologen reoxidized and the entire gel became clear. The gel was then Coomassie stained overnight before taking the final picture of the gel.

Activity assays were also performed on the second dimension of blue native gels. Once the first dimension completed electrophoresis, the gel underwent the arsenate reductase assay as described above. The highlighted complexes were quickly cut from the gel and processed as described above. Once the second electrophoresis was complete, the gel underwent the oxidase activity assay and a second arsenate reductase assay before Coomassie staining.

Only the SDS-PAGE gels and blue native second dimension gels were used for protein identification. The photographs from the gels were compared to isolate the band responsible for activity. The bands were extracted using a sterile razor blade or pipette tip and trypsin digested as described previously. MALDI-TOF mass spectrometry was then done on the sample by Dr. Peter Chovanec (Duquesne University) at the University of Pittsburgh Proteomics Facility.
3.2 Results

3.2.1 2-D Gels

Two-dimensional gels of aerobic and anaerobic cytosolic proteins from MLHE-1 showed many differences (Figure 3.2). Most of the protein identifications (Table 3.1) were made from the aerobic gel by Dr. Peter Chovanec (Duquesne University). They included citric acid cycle proteins such as acetyl-CoA synthetase, malate dehydrogenase, isocitrate dehydrogenase and glutamate dehydrogenase. Glycolytic proteins were also found, such as fructose-1,6-bisphosphate aldolase. Phosphoribulokinases from the Calvin cycle were also discovered. Superoxide dismutase was also identified in the aerobic gel. Hypothetical proteins Mlg_0233, Mlg_1954, and Mlg_2396 were also located in the gel, so they can now be described as proteins of unknown function. Only three identifications were made from the anaerobic gel using MALDI-TOF mass spectrometry. Surprisingly, nitrous oxide reductase was identified. Phosphoribulokinase and a gram-negative type porin were also identified.

3.2.2 DIGE

2-D Fluorescence Difference Gel Electrophoresis (DIGE) was also used to compare proteins. This eliminated any gel-to-gel variation so a better comparison could be made on sight. Gels of cytosolic proteins (Figure 3.3) revealed the presence of nitrous oxide reductase, RuBisCo, fructose bisphosphate aldolase, and triosephosphate isomerase exclusively in the anaerobic sample. The aerobic cytosolic fraction exclusively contained acetyl-CoA synthetase, citrate synthase, 2-oxoglutarate dehydrogenase, as well as two nitrate transporters. Chaperonin GroEL, hypothetical protein Mlg_1954, and superoxide dismutase were found in both fractions. Also in both were malate dehydrogenase, acetyl-
Figure 3.2  Two-dimensional gel comparison.  Gel A (Left) shows the spots excised and identified from the aerobic cytosolic cell fraction.  Gel B (Right) shows those from the equivalent anaerobic fraction.  The actual identities of the protein spots can be found in Table 3.1.
Table 3.1 Proteins Identified from 2-D Gels.

<table>
<thead>
<tr>
<th>Spot No.</th>
<th>Accession No.</th>
<th>Protein Name</th>
<th>MV</th>
<th>pI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>gi</td>
<td>114321906</td>
<td>Type IV pilus secretin PilQ</td>
<td>76260</td>
</tr>
<tr>
<td>2</td>
<td>gi</td>
<td>114321006</td>
<td>Surface antigen (D15)</td>
<td>89790</td>
</tr>
<tr>
<td>3</td>
<td>gi</td>
<td>114319616</td>
<td>Translation elongation factor 2</td>
<td>77123</td>
</tr>
<tr>
<td>4</td>
<td>gi</td>
<td>114320678</td>
<td>Catalase/peroxidase HPI</td>
<td>80941</td>
</tr>
<tr>
<td>5</td>
<td>gi</td>
<td>114319179</td>
<td>L-glutamine synthetase</td>
<td>52400</td>
</tr>
<tr>
<td>6</td>
<td>gi</td>
<td>114319185</td>
<td>Chaperonin GroEL</td>
<td>58349</td>
</tr>
<tr>
<td>7</td>
<td>gi</td>
<td>114319834</td>
<td>Acetyl-CoA synthetase</td>
<td>72011</td>
</tr>
<tr>
<td>8</td>
<td>gi</td>
<td>114320799</td>
<td>Malate synthase G</td>
<td>80850</td>
</tr>
<tr>
<td>9</td>
<td>gi</td>
<td>114320279</td>
<td>NADP-dependent isocitrate dehydrogenase</td>
<td>83553</td>
</tr>
<tr>
<td>10</td>
<td>gi</td>
<td>114320442</td>
<td>Formate dehydrogenase, alpha subunit</td>
<td>107710</td>
</tr>
<tr>
<td>11</td>
<td>gi</td>
<td>114321272</td>
<td>Glutamate dehydrogenase</td>
<td>50032</td>
</tr>
<tr>
<td>12</td>
<td>gi</td>
<td>114321754</td>
<td>Dihydrolipamide dehydrogenase</td>
<td>50972</td>
</tr>
<tr>
<td>13</td>
<td>gi</td>
<td>114319849</td>
<td>Beta-ketothiolase</td>
<td>41014</td>
</tr>
<tr>
<td>14</td>
<td>gi</td>
<td>114321899</td>
<td>Citrate synthase</td>
<td>48234</td>
</tr>
<tr>
<td>15</td>
<td>gi</td>
<td>114320856</td>
<td>Nitrate transporter</td>
<td>52393</td>
</tr>
<tr>
<td>16</td>
<td>gi</td>
<td>114319648</td>
<td>Malate dehydrogenase</td>
<td>35364</td>
</tr>
<tr>
<td>17</td>
<td>gi</td>
<td>114319754</td>
<td>Gram-negative type porin</td>
<td>39336</td>
</tr>
<tr>
<td>18</td>
<td>gi</td>
<td>114320262</td>
<td>Superoxide dismutase</td>
<td>22739</td>
</tr>
<tr>
<td>19</td>
<td>gi</td>
<td>114319755</td>
<td>Gram-negative type porin</td>
<td>37500</td>
</tr>
<tr>
<td>20</td>
<td>gi</td>
<td>114321093</td>
<td>Polynucleotide phosphorylase/polyadenylase</td>
<td>78356</td>
</tr>
<tr>
<td>21</td>
<td>gi</td>
<td>114319942</td>
<td>Ton-B dependent receptor</td>
<td>85552</td>
</tr>
<tr>
<td>22</td>
<td>gi</td>
<td>114321621</td>
<td>Phosphoribosylformylglycin-amidine synthase</td>
<td>140375</td>
</tr>
<tr>
<td>23</td>
<td>gi</td>
<td>114321829</td>
<td>Isocitrate lyase</td>
<td>59539</td>
</tr>
<tr>
<td>24</td>
<td>gi</td>
<td>114319834</td>
<td>Acetyl-CoA synthetase</td>
<td>72011</td>
</tr>
<tr>
<td>25</td>
<td>gi</td>
<td>114321543</td>
<td>Hypothetical protein Mlg_2396</td>
<td>65631</td>
</tr>
<tr>
<td>26</td>
<td>gi</td>
<td>114319350</td>
<td>Branched-chain amino acid ABC transporter</td>
<td>49022</td>
</tr>
<tr>
<td>27</td>
<td>gi</td>
<td>114321103</td>
<td>Hypothetical protein Mlg_1954</td>
<td>44327</td>
</tr>
<tr>
<td>28</td>
<td>gi</td>
<td>114319605</td>
<td>Elongation factor Tu</td>
<td>43461</td>
</tr>
<tr>
<td>29</td>
<td>gi</td>
<td>114319705</td>
<td>Ketol-acid reductoisomerase</td>
<td>36371</td>
</tr>
<tr>
<td>30</td>
<td>gi</td>
<td>114321736</td>
<td>Putative nitrate transporter</td>
<td>53258</td>
</tr>
<tr>
<td>31</td>
<td>gi</td>
<td>114320423</td>
<td>Sulfide dehydrogenase</td>
<td>47815</td>
</tr>
<tr>
<td>32</td>
<td>gi</td>
<td>114321681</td>
<td>NADPH-glutathione reductase</td>
<td>48616</td>
</tr>
<tr>
<td>Spot No.</td>
<td>Accession No.</td>
<td>Protein Name</td>
<td>MV</td>
<td>pI</td>
</tr>
<tr>
<td>---------</td>
<td>------------------</td>
<td>--------------------------------------------------------------------------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>33</td>
<td>gi</td>
<td>114321681</td>
<td>NADPH-glutathione reductase</td>
<td>48616</td>
</tr>
<tr>
<td>34</td>
<td>gi</td>
<td>114321983</td>
<td>Ribulose bisphosphate carboxylase</td>
<td>52870</td>
</tr>
<tr>
<td>35</td>
<td>gi</td>
<td>114321751</td>
<td>2-methylcitrate synthase</td>
<td>41772</td>
</tr>
<tr>
<td>36</td>
<td>gi</td>
<td>114321991</td>
<td>Fructose-1,6-bisphosphate aldolase</td>
<td>38809</td>
</tr>
<tr>
<td>37</td>
<td>gi</td>
<td>114320199</td>
<td>TonB-dependent copper receptor</td>
<td>73613</td>
</tr>
<tr>
<td>38</td>
<td>gi</td>
<td>114321980</td>
<td>Phosphoribokinase</td>
<td>33722</td>
</tr>
<tr>
<td>39</td>
<td>gi</td>
<td>114319437</td>
<td>Phospho-2-dehydro-3-heoxyheptonate aldolase</td>
<td>39332</td>
</tr>
<tr>
<td>40</td>
<td>gi</td>
<td>114319410</td>
<td>Thioredoxin reductase</td>
<td>34375</td>
</tr>
<tr>
<td>41</td>
<td>gi</td>
<td>114321426</td>
<td>NADH-dependent enoyl-[acyl-carrier-protein] reductase</td>
<td>28027</td>
</tr>
<tr>
<td>42</td>
<td>gi</td>
<td>114321018</td>
<td>2,3,4,5-tetrahydropyridine-2,6-dicarboxylate N-succinyltransferase</td>
<td>29183</td>
</tr>
<tr>
<td>43</td>
<td>gi</td>
<td>114321011</td>
<td>Ribosome recycling factor</td>
<td>21055</td>
</tr>
<tr>
<td>44</td>
<td>gi</td>
<td>114321360</td>
<td>Ubiquinol-cytochrome c reductase, iron-sulfur subunit</td>
<td>22340</td>
</tr>
<tr>
<td>45</td>
<td>gi</td>
<td>114320479</td>
<td>Phosphoglycerate mutase</td>
<td>26866</td>
</tr>
<tr>
<td>46</td>
<td>gi</td>
<td>114322013</td>
<td>UspA domain protein</td>
<td>30384</td>
</tr>
<tr>
<td>47</td>
<td>gi</td>
<td>114319895</td>
<td>Response regulator receiver protein</td>
<td>31141</td>
</tr>
<tr>
<td>48</td>
<td>gi</td>
<td>114320404</td>
<td>Nucleoside diphosphate kinase</td>
<td>15614</td>
</tr>
<tr>
<td>49</td>
<td>gi</td>
<td>114319395</td>
<td>Hypothetical protein Mlg_0233</td>
<td>15123</td>
</tr>
<tr>
<td>50</td>
<td>gi</td>
<td>114319193</td>
<td>YfaZ family protein</td>
<td>20405</td>
</tr>
<tr>
<td>51</td>
<td>gi</td>
<td>114321755</td>
<td>2-oxoglutarate dehydrogenase, E2 component</td>
<td>45511</td>
</tr>
<tr>
<td>52</td>
<td>gi</td>
<td>114321238</td>
<td>Fructose-bisphosphate aldolase</td>
<td>38290</td>
</tr>
<tr>
<td>53</td>
<td>gi</td>
<td>114320412</td>
<td>Adenylate kinase</td>
<td>23896</td>
</tr>
<tr>
<td>54</td>
<td>gi</td>
<td>114319605</td>
<td>Elongation factor Tu</td>
<td>43461</td>
</tr>
<tr>
<td>55</td>
<td>gi</td>
<td>114319773</td>
<td>Phosphoribosylamine-glycine ligase</td>
<td>46136</td>
</tr>
<tr>
<td>56</td>
<td>gi</td>
<td>114320233</td>
<td>Nitrous-oxide reductase</td>
<td>72356</td>
</tr>
<tr>
<td>57</td>
<td>gi</td>
<td>114319755</td>
<td>Gram-negative type porin</td>
<td>37500</td>
</tr>
<tr>
<td>58</td>
<td>gi</td>
<td>114321980</td>
<td>Phosphoribokinase</td>
<td>33572</td>
</tr>
</tbody>
</table>
CoA acetyltransferase, adenylate kinase, glutamate dehydrogenase and amino transferase proteins. A complete list of identified proteins can be found in Table 3.2.

DIGE gels were also completed with proteins that were CHAPS-soluble (Figure 3.4). These gels contained similar proteins to those identified in the cytosolic fractions. Nitrous oxide reductase and RuBisCo were identified in the anaerobic fraction. Also of interest was NarB and ArrA, neither of which were identified in the cytosolic gels. Some of the aerobic proteins identified were isocitrate lyase, acetyl-CoA synthetase, and citrate synthase. Both fractions contained the hypothetical proteins Mlg_1954 and Mlg_2396. A complete list of identified proteins from these gels can be found in Table 3.3.

3.2.3 Blue Native Gels

Blue native gels were run for both aerobic and anaerobic MLHE-1 samples. When Triton-X was used as the detergent, a prominent band of approximately 67kDa was observed in the anaerobic lanes (Figure 3.5). This band was less readily seen in gels that used DDM as the detergent. The band was excised, trypsin digested, and LC/MS/MS data was collected at the University of Pittsburgh's proteomic facility. Five possible protein identities were produced which were ArrA (Mlg_0216), ArrB (Mlg_0215), CheW (Mlg_0991), protein of unknown function, and acetyl-CoA synthetase. However, based on the molecular weight of the band, it was concluded that only ArrA and ArrB, the top two hits, were present. These two proteins are the primary components of the first dissimilatory arsenate reductase found in the genome. No reductase encoded by the second operon was identified on the blue native gels.
Figure 3.3 Cytosolic DIGE 2-D gel. Cy3-labeled anaerobic proteins appear green in color, while Cy5-labeled aerobic proteins fluoresce red in color. Proteins present in both fractions in relatively equal amounts appear yellow. The Cy2 internal standard is not visible. The Cy3 dye was viewed at 590nm, while the Cy5 was visible at 690nm. Numbers represent spots that were identified by mass spectrometry and their identities can be found in Table 3.2.
Table 3.2 Proteins Identified from Cytosolic DIGE 2-D Gels.

<table>
<thead>
<tr>
<th>Spot No.</th>
<th>Accession No.</th>
<th>Protein Name</th>
<th>MV</th>
<th>pl</th>
<th>Predominant Fraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>gi</td>
<td>114319754</td>
<td>Porin, Gram-negative type</td>
<td>39336</td>
<td>4.6</td>
</tr>
<tr>
<td>2</td>
<td>gi</td>
<td>114320233</td>
<td>Nitrous-oxide reductase</td>
<td>71881</td>
<td>5.4</td>
</tr>
<tr>
<td>3</td>
<td>gi</td>
<td>114319834</td>
<td>Acetyl-coenzyme A synthetase</td>
<td>72011</td>
<td>5.2</td>
</tr>
<tr>
<td>4</td>
<td>gi</td>
<td>114321143</td>
<td>Hydroxyacylglutathione hydrolase</td>
<td>27717</td>
<td>5.6</td>
</tr>
<tr>
<td>5</td>
<td>gi</td>
<td>114320199</td>
<td>TonB-dependent copper receptor</td>
<td>73613</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>gi</td>
<td>114319185</td>
<td>Chaperonin GroEL</td>
<td>58439</td>
<td>4.8</td>
</tr>
<tr>
<td>7</td>
<td>gi</td>
<td>114320799</td>
<td>Malate synthase G</td>
<td>80850</td>
<td>5.2</td>
</tr>
<tr>
<td>8</td>
<td>gi</td>
<td>114321899</td>
<td>Citrate synthase</td>
<td>48234</td>
<td>5.3</td>
</tr>
<tr>
<td>9</td>
<td>gi</td>
<td>114320856</td>
<td>Nitrate transporter</td>
<td>52393</td>
<td>6.2</td>
</tr>
<tr>
<td>10</td>
<td>gi</td>
<td>114319849</td>
<td>Acetyl-CoA acetyltransferase</td>
<td>41014</td>
<td>5.9</td>
</tr>
<tr>
<td>11</td>
<td>gi</td>
<td>114319648</td>
<td>Malate dehydrogenase</td>
<td>35364</td>
<td>5</td>
</tr>
<tr>
<td>12</td>
<td>gi</td>
<td>114321736</td>
<td>Putative nitrate transporter component</td>
<td>53258</td>
<td>5.4</td>
</tr>
<tr>
<td>13</td>
<td>gi</td>
<td>114321980</td>
<td>Phosphoribulokinase</td>
<td>33722</td>
<td>5.7</td>
</tr>
<tr>
<td>14</td>
<td>gi</td>
<td>114321991</td>
<td>Fructose-bisphosphate aldolase</td>
<td>38809</td>
<td>5.3</td>
</tr>
<tr>
<td>15</td>
<td>gi</td>
<td>114321103</td>
<td>Hypothetical protein Mlg_1954</td>
<td>44327</td>
<td>4.6</td>
</tr>
<tr>
<td>16</td>
<td>gi</td>
<td>114319350</td>
<td>Branched-chain amino acid ABC transporter</td>
<td>49022</td>
<td>4.3</td>
</tr>
<tr>
<td>17</td>
<td>gi</td>
<td>114319605</td>
<td>Translation elongation factor Tu</td>
<td>43461</td>
<td>4.9</td>
</tr>
<tr>
<td>18</td>
<td>gi</td>
<td>114321121</td>
<td>Triosephosphate isomerase</td>
<td>26488</td>
<td>4.8</td>
</tr>
<tr>
<td>19</td>
<td>gi</td>
<td>114320412</td>
<td>Adenylate kinase</td>
<td>23896</td>
<td>5</td>
</tr>
<tr>
<td>20</td>
<td>gi</td>
<td>114320262</td>
<td>Superoxide dismutase</td>
<td>22739</td>
<td>5.3</td>
</tr>
<tr>
<td>21</td>
<td>gi</td>
<td>114320404</td>
<td>Nucleoside diphosphatase kinase</td>
<td>15614</td>
<td>5.4</td>
</tr>
<tr>
<td>22</td>
<td>gi</td>
<td>114321755</td>
<td>2-oxoglutarate dehydrogenase E2 component</td>
<td>45539</td>
<td>5.2</td>
</tr>
<tr>
<td>23</td>
<td>gi</td>
<td>114321018</td>
<td>2,3,4,5-tetrahydropyridine-2,6-dicarboxylate N-succinyltransferase</td>
<td>29183</td>
<td>5.4</td>
</tr>
<tr>
<td>24</td>
<td>gi</td>
<td>114321272</td>
<td>Glutamate dehydrogenase</td>
<td>50032</td>
<td>6</td>
</tr>
<tr>
<td>25</td>
<td>gi</td>
<td>114320423</td>
<td>Aminotransferase</td>
<td>46184</td>
<td>5.5</td>
</tr>
<tr>
<td>26</td>
<td>gi</td>
<td>114321823</td>
<td>Cold-shock DNA-binding protein family</td>
<td>7038</td>
<td>6</td>
</tr>
<tr>
<td>27</td>
<td>gi</td>
<td>114322013</td>
<td>UspA domain protein</td>
<td>30384</td>
<td>5.9</td>
</tr>
</tbody>
</table>
Figure 3.4 CHAPS-soluble DIGE 2-D gel. Cy3-labeled CHAPS-soluble anaerobic proteins appear green in color, while Cy5-labeled aerobic proteins are red in color. Proteins present in both fractions in relatively equal amounts appear yellow. The Cy2 internal standard is not visible. The Cy3 dye was viewed at 590nm, while the Cy5 was visible at 690nm. Numbers represent spots that were identified by mass spectrometry and their identities can be found in Table 3.3.
Table 3.3 Proteins Identified from CHAPS-Soluble DIGE 2-D Gels.

<table>
<thead>
<tr>
<th>Spot No.</th>
<th>Accession No.</th>
<th>Protein Name</th>
<th>MV</th>
<th>pl</th>
<th>Predominant Fraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>gi</td>
<td>114319754</td>
<td>Porin, Gram-negative type</td>
<td>39336</td>
<td>4.6</td>
</tr>
<tr>
<td>2</td>
<td>gi</td>
<td>114319754</td>
<td>Porin, Gram-negative type</td>
<td>39336</td>
<td>4.6</td>
</tr>
<tr>
<td>3</td>
<td>gi</td>
<td>114320937</td>
<td>Addiction module toxin, Txe/YoeB family</td>
<td>10159</td>
<td>9.7</td>
</tr>
<tr>
<td>4</td>
<td>gi</td>
<td>114321103</td>
<td>Hypothetical protein Mlg_1954</td>
<td>44327</td>
<td>4.6</td>
</tr>
<tr>
<td>5</td>
<td>gi</td>
<td>114321543</td>
<td>Hypothetical protein Mlg_2396</td>
<td>65631</td>
<td>4.3</td>
</tr>
<tr>
<td>6</td>
<td>gi</td>
<td>114319859</td>
<td>Flagellin domain protein</td>
<td>51444</td>
<td>4</td>
</tr>
<tr>
<td>7</td>
<td>gi</td>
<td>114319185</td>
<td>Chaperonin GroEL</td>
<td>58439</td>
<td>4.8</td>
</tr>
<tr>
<td>8</td>
<td>gi</td>
<td>114319179</td>
<td>L-glutamine synthetase</td>
<td>52400</td>
<td>4.8</td>
</tr>
<tr>
<td>9</td>
<td>gi</td>
<td>114319942</td>
<td>TonB-dependent receptor</td>
<td>85552</td>
<td>5</td>
</tr>
<tr>
<td>10</td>
<td>gi</td>
<td>114320678</td>
<td>Catalase/peroxidase HPI</td>
<td>80941</td>
<td>4.8</td>
</tr>
<tr>
<td>11</td>
<td>gi</td>
<td>114321093</td>
<td>Polyrribonucleotide nucleotidyltransferase</td>
<td>78356</td>
<td>5</td>
</tr>
<tr>
<td>12</td>
<td>gi</td>
<td>114320039</td>
<td>TonB-dependent receptor</td>
<td>86122</td>
<td>5.3</td>
</tr>
<tr>
<td>13</td>
<td>gi</td>
<td>114320279</td>
<td>Isocitrate dehydrogenase, NADP-dependent</td>
<td>83553</td>
<td>5.5</td>
</tr>
<tr>
<td>14</td>
<td>gi</td>
<td>114320233</td>
<td>Nitrous-oxide reductase</td>
<td>71881</td>
<td>5.4</td>
</tr>
<tr>
<td>15</td>
<td>gi</td>
<td>114319378</td>
<td>Arsenic reductase, alpha subunit</td>
<td>94067</td>
<td>6.2</td>
</tr>
<tr>
<td>16</td>
<td>gi</td>
<td>114321756</td>
<td>2-oxoglutarate dehydrogenase E1 component</td>
<td>1E+05</td>
<td>5.9</td>
</tr>
<tr>
<td>17</td>
<td>gi</td>
<td>114320442</td>
<td>Formate dehydrogenase alpha subunit</td>
<td>1E+05</td>
<td>6</td>
</tr>
<tr>
<td>18</td>
<td>gi</td>
<td>114320162</td>
<td>Respiratory nitrate reductase beta subunit</td>
<td>60575</td>
<td>5.6</td>
</tr>
<tr>
<td>19</td>
<td>gi</td>
<td>11431829</td>
<td>Isocitrate lyase</td>
<td>59539</td>
<td>5.1</td>
</tr>
<tr>
<td>20</td>
<td>gi</td>
<td>114319834</td>
<td>Acetyl-coenzyme A synthetase</td>
<td>72011</td>
<td>5.2</td>
</tr>
<tr>
<td>21</td>
<td>gi</td>
<td>114321899</td>
<td>Citrate synthase</td>
<td>48234</td>
<td>5.3</td>
</tr>
<tr>
<td>22</td>
<td>gi</td>
<td>114320199</td>
<td>TonB-dependent copper receptor</td>
<td>73613</td>
<td>5</td>
</tr>
<tr>
<td>23</td>
<td>gi</td>
<td>114319648</td>
<td>Malate dehydrogenase (NAD)</td>
<td>35364</td>
<td>5</td>
</tr>
<tr>
<td>24</td>
<td>gi</td>
<td>114319705</td>
<td>Ketol-acid reductoisomerase</td>
<td>36371</td>
<td>5.1</td>
</tr>
<tr>
<td>25</td>
<td>gi</td>
<td>114321983</td>
<td>Ribulose 1,5-bisphosphate carboxylase large subunit</td>
<td>53237</td>
<td>6</td>
</tr>
<tr>
<td>26</td>
<td>gi</td>
<td>114321754</td>
<td>Dihydrolipoamide dehydrogenase</td>
<td>50972</td>
<td>6</td>
</tr>
<tr>
<td>27</td>
<td>gi</td>
<td>114321272</td>
<td>Glutamate dehydrogenase (NADP)</td>
<td>50032</td>
<td>6</td>
</tr>
</tbody>
</table>
Table 3.3 (Continued)

<table>
<thead>
<tr>
<th>Spot No.</th>
<th>Accession No.</th>
<th>Protein Name</th>
<th>MV</th>
<th>pI</th>
<th>Predominant Fraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>28</td>
<td>gi</td>
<td>114319849</td>
<td>Acetyl-CoA acetyltransferases</td>
<td>41014</td>
<td>5.9</td>
</tr>
<tr>
<td>29</td>
<td>gi</td>
<td>114321980</td>
<td>Phosphoribulokinase</td>
<td>33722</td>
<td>5.7</td>
</tr>
<tr>
<td>30</td>
<td>gi</td>
<td>114321755</td>
<td>2-oxoglutarate dehydrogenase E2 component</td>
<td>45539</td>
<td>5.2</td>
</tr>
</tbody>
</table>

Figure 3.5 A blue native gel of membrane proteins. Lane A (Left) contains aerobic proteins, while Lane B (Right) contains anaerobic ones. There is dominant band (highlighted by the red box) at approximately 67kDa that only appears in the anaerobic fraction.
3.2.4 Activity Assays

The activity assays done in vials were positive for arsenite oxidase activity when strict anaerobic conditions were met. When methyl viologen was used as the electron acceptor, vials containing the reaction mixture plus anaerobic cell lysate slowly changed from colorless to blue. Vials that did not contain any cell lysate had no change in appearance (Figure 3.6A). When DCIP was used as the electron acceptor, the vials containing cell lysate turned clear, while the control remained dark blue (Figure 3.6B). Figure 3.6C shows the arsenate reducase activity of the sample. The time period for the color change to occur was dependent on how much cell lysate was added to the sample. Once the solutions were exposed to oxygen, the transformation ceased and the original color was regained within a few minutes.

3.2.5 Zymograms

Blue native gels without Coomassie and native SDS-PAGE gels were used for in-gel assays to show both arsenite oxidation and arsenate reduction. The band seen previously in blue native gels of anaerobic membrane fractions showed arsenate reductase activity as it should since two of the components of the complex are known to be ArrA and ArrB. Surprisingly, this same band at 67 kDa showed oxidase activity under strict anaerobic conditions as well (Figures 3.7A and 3.7B).

Oxidase activity was also seen on the second dimension of the blue native (Figure 3.7C) and in quasi-native SDS-PAGE gels (Figure 3.7D). Again the same single band also showed arsenate reduction abilities. When the second dimension of the blue native gel was Coomassie stained, an additional two bands were seen (Figure 3.8). All three bands were excised, trypsin digested, and MALDI-TOF M/S data was collected. The top
Figure 3.6 Arsenite oxidase activity assays. The three vials at the left (A) show the oxidation of arsenite coupled to the reduction of methyl viologen. The solution changes from clear to blue as arsenite oxidation occurs. The first vial is the control, while the remaining ones contain the reaction mixture plus cell lysate from an anaerobic culture. The two vials in the middle (B) show the oxidation of arsenite coupled to the reduction of DCIP. The solution changes from blue to clear as the reaction proceeds. The two vials on the right (C) show the reductase assay coupled to the oxidation of methyl viologen.
Figure 3.7 Zymograms. The first lane in each panel shows the arsenite oxidase activity. The second lane shows the arsenate reductase activity, and the third lane shows the Coomassie staining of the bands. Panel A shows a blue native gel that used DCIP as the electron acceptor for the oxidase reaction. Panel B is also a blue native, but the oxidation was coupled to the reduction of methyl viologen. Panel C is the second dimension of a blue native using methyl viologen as the electron acceptor for arsenite oxidation. Panel D shows an SDS-PAGE gel that also used methyl viologen for the oxidation reaction.
Figure 3.8 Second dimension of a blue native gel. The first two lanes show a band exhibiting arsenite oxidase activity. The two lanes to the right show the Coomassie stained gel and the presence of two other bands. The top band was identified as ArrA and the bottom band as ArrB. The active band in the middle was found to be heterodimer of these two subunits. The dimer should have been higher than ArrA, but incomplete denaturation caused it to run differently and fall in between the subunits.
band was identified with a score of 102 as ArrA and the bottom band was found to be ArrB with a score of 82. Protein scores greater than 75 are significant (p<0.05), so these identifications are credible. The middle band, the one which showed activity at approximately 50kDa, had two hits. It was identified as ArrA with a score of 83, but also as ArrB with a score of 44.

3.3 Discussion

*Alkalilimnicola ehrlichii* strain MLHE-1$^T$ is a very unique organism. Up to this point, all sequenced arsenite oxidizers have possessed genes for an arsenite oxidase. In MLHE-1 however, these genes are absent. Instead it appears that the organism uses one of its two respiratory arsenate reductases to transform arsenite. This is supported by the zymograms showing arsenite oxidation and MALDI-TOF mass spectrometry. The oxidase has been identified as a heterodimer of ArrA (Mlg_0216) and ArrB (Mlg_0215). The subunits individually show no activity and thus only function in combination. As the bacteria cannot grow on arsenate, the enzyme is primarily used for the conversion of arsenite to arsenate. The other arsenate reductase seems to be nonfunctional as no other bands of activity were visible during the reductase or oxidase assays. Since the assays were run on multiple gel types and multiple protein fractions, one can assume that no other proteins are potential candidates for the arsenite oxidase.

Attempts at the assay outside the anaerobic chamber produced negative results each time. The zymograms done inside the glove box produced positive results consistently, but quickly changed back to their original color once removed from the chamber. The oxidase assay must be conducted under strict anaerobic conditions in order to determine whether or not the enzyme in question is functional. The reductase assay
did not require anaerobic conditions. These observations support that MLHE-1 only oxidizes arsenite under the strictest anaerobic conditions, which is part of its native habitat in the anoxic bottom waters of Mono Lake.

Another mystery is the appearance of nitrous oxide reductase in the two-dimensional gels. Since the organism lacks nirK and nirS, it was assumed that the end product of nitrate metabolism was nitrite. This would render the nitric oxide reductases and nitrous oxide reductases useless since no nitric oxide would be produced to further transform. The nitrate reductase may be further reducing its product so that the pathway may continue. The organism cannot grow on nitric oxide or nitrous oxide, and so nitrate is the essential starting material when the organism is grown anaerobically.

Future studies should be done to identify more anaerobic proteins from two-dimensional gels in order to complete the proteomic profiling for this organism. In addition, the arsenite reductases from other organisms should be assayed to see if they also possess the reversibility of the MLHE-1 homolog. Since other organisms have operons of different gene orders and different components, it will be interesting to see if a certain combination is needed to generate the oxidase capabilities. The other arsenate reductase that MLHE-1 possesses should be examined more to see if it indeed is nonfunctional or why it lacks the oxidoreductase abilities of the larger operon. Further work still needs to be done on the reversible arsenate reductase. More activity assays will be done in order to determine the kinetics of the arsenite oxidation process. During the activity assay, samples will be taken at different time intervals. The samples will then be processed using the molybdenum blue method, a colormetric determination of arsenic$^{51}$. This procedure will allow the amount of arsenate being produced to be
measured over time. Since the oxidase assay is so oxygen sensitive, however, great care must be taken to maintain an anaerobic environment during the process.
REFERENCES


36. http://img.jgi.doe.gov/cgi-bin/pub/main.cgi?section=TaxonDetail&page=taxon Detail&taxon_oid=637000005#information


APPENDIX I

ClustalX sequence alignment of arrA homologs