2007

Characterization of arsenic resistant bacteria and a novel gene cluster in 'Bacillus' sp. CDB3

N. Somanath Bhat
University of Wollongong

Characterization of Arsenic Resistant Bacteria and A Novel Gene Cluster in *Bacillus* sp. CDB3

A thesis submitted in (partial) fulfillment of the requirements for the award of the degree of

DOCTOR OF PHILOSOPHY (PhD)

From

THE UNIVERSITY OF WOLLONGONG

by

N. SOMANATH BHAT

Master of Science (*Microbiology*), University of Bangalore, India
Master of Science (*Biotechnology*), University of Wollongong, Australia.

SCHOOL OF BIOLOGICAL SCIENCES

-2007-

CERTIFICATION

I, N. Somanath Bhat, declare that this thesis, submitted in fulfillment of the requirements for the award of Doctor of Philosophy, in the School of Biological Sciences, University of Wollongong, is wholly my own work unless otherwise referenced or acknowledged. The document has not been submitted for qualifications at any other academic institution.

N. Somanath Bhat

July 2007

TABLE OF CONTENTS

Chapter 2: Characterization and Identification of Arsenic Resistant Bacterial Strains Isolated from Cattle Dip Sites

Chapter 3: Genetic Characterization and Sequence Analysis of CDB3 *ars* Gene Cluster 1

Chapter 6: Functionality of the CDB3 *arsD* Gene

Chapter 7: Transcriptional Regulation of CDB3 *ars* Gene Cluster 1

LIST OF PUBLICATIONS

- **Baldeo. K. Chopra*, Somanath Bhat*, Irina P. Mikheenko, Zhiqiang Xu, Ying Yang, Xi Luo, Hancai Chen, Lukas van Zwieten, Ross McC. Lilley and Ren Zhang (2007)**. "The Characteristics of Rhizosphere Microbes and their Interaction with Plants in Arsenic contaminated Soils from Cattle dip-sites" *Science of the Total Environment*. 378:331-342.

- **Somanath Bhat et. al.,** "A novel *ars* operon of *Bacillus* sp. (CDB3) isolated from cattle dip-sites confers increased resistance to arsenate and arsenite in *Escherichia coli*" **(in preparation)**

- **Somanath Bhat et. al.,** "Metalloregulation of repressor protein ArsD of *Bacillus* sp. CDB3 isolated from cattle dip sites" **(in preparation)**

- **Somanath Bhat et. al.,** "Phylogenetic study of arsenite efflux pumps among prokaryotes and eukaryotes" **(in preparation)**

* Equal first authors

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

aa	amino acid
ABC	ATP binding cassette
Amp	ampicillin
ANGIS	Australian National Genomic Information Service
ars	arsenic resistance
As	arsenic
AsIII	trivalent arsenic (arsenite)
AsV	pentavalent arsenic (arsenate)
ATP	adenosine triphosphate
bp	base pair
CDB	cattle dip bacteria
cDNA	complementary DNA
Cm	chloramphenicol
CSPD	(disodium 3-(4-methoxyspiro-[1,2-dioxetane-3-2′(5′chloro)-tricyclo[3.3.1.1$^{3.7}$]decane]-4-yl)phenyl phosphate)
C-terminal	carboxy-terminal
Cys	cysteine
d. H_2O	distilled water
DIG	digoxigenin
DMA	dimethylarsenate
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
dNTP	deoxynucleotide triphosphate
EDTA	ethylenediamine-tetraacetic acid
g	gravity, 9.8 ms^{-2}
GSH	glutathione
Grx	glutaredoxin
h	hour
His	histidine
Km	kanamycin
kb	kilobase pairs

kDa	kilo daltons
kV	kilo volts
LB	Luria-Bertani
M	molar
MBD	metalloid-binding domain
mg	milligram
min	minute
mL	millilitre
mM	millimolar
MMA	monomethylarsenate
MSD	macromolecular structure database
mRNA	messenger ribonucleic acid
NaAc	sodium acetate
NBD	nucleotide binding domain
NBS	nucleotide binding site
ng	nanogram
nm	nanometre
N-terminal	amino-terminal
OD	optical density
ONPG	*o*-nitrophenyl-β-D-galactopyranoside
ORF	open reading frame
ori	origin of replication
O/P	operator/ promoter
P_{arsR}	promoter in front of the *arsR* gene
P_{arsD}	promoter in front of the *arsD* gene
P_{trxB}	promoter in front of the *trxB* gene
PBS	phosphate buffer saline
PCR	polymerase chain reaction
PEG	polyethylene glycol
p-loop	phosphate-binding loop
pmol	picomole
pro	promoter
RNA	ribonucleic acid
RNase	ribonuclease

rpm	revolutions per minute
SAP	shrimp alkaline phosphatase
Sb	antimony
SbIII	trivalent antimony (antimonite)
SD	standard deviation
SDS	sodium dodecyl sulphate
SD sequence	Shine-Dalgarno sequence
sec	seconds
sp.	species
spp	species
Strep	streptomycin
TAE	Tris/acetate buffer
TBE	Tris/borate buffer
Taq	*Thermus aquaticus* DNA polymerase
TE	Tris/EDTA buffer
Tet	tetracycline
TMD	Trans Membrane Domain
Trx	thioredoxin
UV	ultraviolet
V	volts
Val	valine
X-gal	5-bromo, 4-chloro, 3-indolyl-β-D-galactopyranoside
<	less than
μ	micro
Ω	ohms

ABSTRACT

Arsenic is a toxic metalloid of the nitrogen family which is found in both natural environments and sites contaminated in a number of ways. Arsenate acts as a phosphate analog which interferes with phosphate uptake and utilization and arsenite disrupts enzymatic function. Micro organisms have evolved a variety of mechanisms in coping with toxicity of arsenic and the best known detoxification pathway involves the arsenic resistance (*ars*) cluster. To date, a number of *ars* clusters have been characterized at the molecular level.

Contamination of cattle dip-site soils with arsenic along with other toxicants such as DDT represents a major pollution problem in agricultural sites and has been shown to have distinct effects on the soil microbial populations. Little is known of the micro organisms at these sites. This thesis investigated the identity of some of the rhizosphere bacteria isolated from arsenic contaminated dip-site soils of north-eastern New South Wales and functionally characterized some of the *ars* genes cloned from one strain.

Five bacterial strains isolated from cattle dip-sites (referred to as CDB) were identified based on morphological, biochemical and 16S rDNA sequence characters. These bacterial strains belonged to four different genera, CDB1- *Arthrobacter* sp., CDB2- *Ochrobactrum* sp., CDB3 & CDB4- *Bacillus* spp and CDB5- *Serratia* sp. The arsenic resistance profiles of these bacteria were quite different. The highest resistance to arsenite was by CDB5 followed by CDB4, CDB3, CDB2 and CDB1. CDB2 was exceptionally tolerant to arsenate, exhibiting normal growth on agar

containing 200 mM arsenate, an almost saturated concentration, while all the other strains showed a minimum inhibitory concentration to arsenate of 75 mM.

From *Bacillus* sp. CDB3, a novel *ars* gene cluster was cloned and sequenced (Luo, 2006), which revealed eight intact open reading frames organised in a unique gene order, *arsRYCDATIP* when compared to other *ars* gene clusters. This gene cluster was found to exist on the chromosome, unlike the *arsRDABC* clusters which exist on plasmids of Gram negative bacteria.

Bioinformatic analysis was undertaken to examine any novel sequence characters. Alignment of the CDB3 ArsD with known ArsDs showed that the CDB3 ArsD protein lacks the four C-terminal cysteine residues (Cys112-Cys113 and Cys119-Cys120), which have been demonstrated to be required for induction in *E. coli* pR773 ArsD, suggesting the regulatory mechanism by CDB3 ArsD may be different from that of *E. coli* pR773. This was the case, when the ArsD of CDB3 was mutated and compared to the intact ArsD as presented in this thesis, the *E. coli* cells bearing mutant ArsD showed sensitivity to arsenic. The membrane bound AsIII pump encoded by the second ORF is a YqcL type protein but may still associate with an ATPase encoded by the fifth ORF. Phylogenetic analysis of arsenite efflux pumps also revealed the existence of a sub-group of the ArsB group proteins, adding to the diversity of the arsenite pump protein family. Further investigation was done to determine whether the CDB3 ArsA, (which is known to couple to ArsB) can couple to YqcL and vice versa. Results indicated that CDB3 ArsA can couple to both YqcL and *E. coli* ArsB resulting in elevated resistance to the host, confirming the functionality of ArsA. No data on ArsA coupling with YqcL is available and this thesis was the

first instance to show the interaction of YqcL with ArsA in extruding arsenite out of the cell.

Transcriptional regulation of CDB3 *ars* gene cluster was performed by reporter gene and northern analysis. Northern analysis showed that the mRNA transcription could read through the whole cluster, indicating that the CDB3 *ars* gene cluster can be transcribed under the control of a single promoter. However, some short transcripts were also detected. These results showed that the *ars* gene cluster can be transcribed as whole cluster, but under certain conditions it may be transcribed as two or more sub-operons. This was further studied by the reporter gene analysis which indicated active promoters in front of *arsR*, *arsD* and *arsT*. Gel mobility shift assays also displayed the existence of a promoter in front of *arsD*, where the purified ArsD protein was shown to bind to its own promoter (between *arsC* and *arsD*) in addition to the *arsR* promoter. It appeared that the regulation of expression of the CDB3 *ars* gene cluster 1 is a very interesting but complicated mechanism.

ACKNOWLEDGEMENTS

I would like to extend my sincere thanks to my supervisor Dr. Ren Zhang, for his able guidance, patience, motivation and support at all times.

I would like to thank, Luo Xi, Zhiqiang Xu, for providing the cloned *ars* gene cluster from CDB3, Chenwei Su for construction of reporter gene constructs, and their previous work in this area of research.

I would like to thank my co-workers for their assistance in some experiments and thesis proof reading. My sincere thanks to Jake Nicholas Matic, Sherin Alex and Stephanie Ruiz for their help in proof reading the entire thesis, Ying Yang, for helping me in 16S rDNA identification, Mrs. Laura Mothersdill, Mr. Alan Adolfson, Ms. Linda Dietch and Mrs. Joy Williams for sequencing and staff and students in the School of Biological Sciences for their assistance and guidance.

I would like to thank Dr. Andrew Aqulina for collaborating in the study of ArsD function and Mass spec identification of the purified protein, Dr. Mark Dowton for collaborating and providing able guidance in to the study of arsenite efflux pumps and proof reading the manuscript.

I thank Prof. Ross Lilley for advice and proof reading the manuscripts for publication.

This thesis is dedicated to my family; my mother Dr. (Mrs) H. Shanta, my father Mr. T. S. Nagaraj and my brother N. Manjunath Bhat who have been an inspiration and have provided constant support and guidance through out. I thank them to have believed in my ability and supported me in my study overseas.

I would like to thank all my relatives and friends. Last but not the least, I would like to thank my wife Mrs. Pragati Verma for her motivation and support when much needed.

Chapter 1

Introduction (Literature Review)

1.1 History and Chemical Forms of Arsenic

Arsenic is a chemical element in the periodic table denoted by the symbol **As** (Figure 1.1). It is a poisonous metalloid with atomic number 33 that has many allotropic forms; yellow, black and grey are a few that are quite oftenly seen. Arsenic and its compounds are used as insecticides, pesticides, herbicides, medicines and in various alloys.

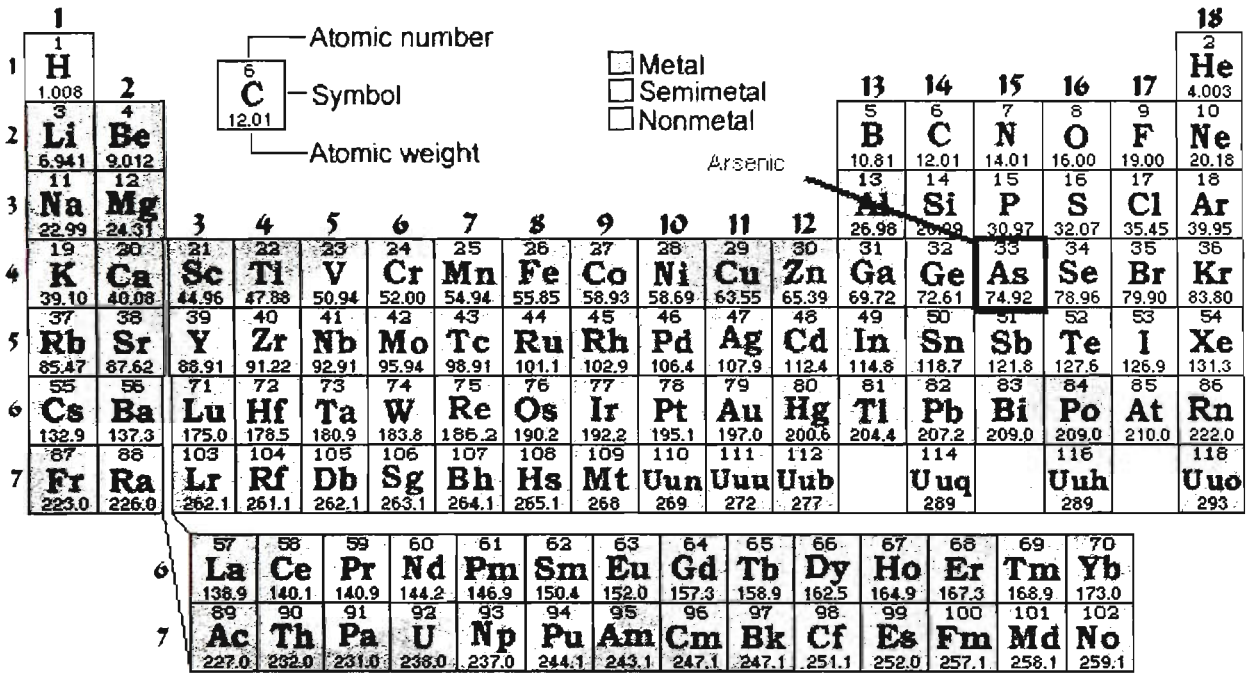
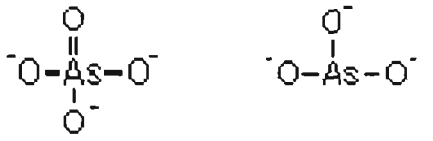


Figure 1.1: Periodic table showing the Arsenic metalloid as As

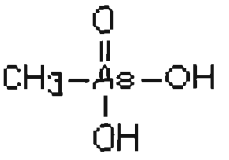
The word *arsenic* has been borrowed from the Persian word *Zarnikh* which means "yellow orpiment". Elemental arsenic is found in two solid forms: yellow and gray/metallic, with specific gravities of 1.97 and 5.73, respectively. In its inorganic form, arsenic primarily exists in two redox states: the reduced form, arsenite (AsIII), and the oxidized form, arsenate (AsV). Both states are toxic to most organisms. Arsenite (specifically the arsenite ion, AsO_3^{3-}) interferes with sulfhydryl groups in amino acids and can disrupt protein structure. Arsenate (specifically the arsenate ion, AsO_4^{3-}) is a

phosphate (PO_4^{3-}) analog and can interfere with phosphate uptake and a variety of cellular processes that involve phosphate (Tamaki and Frankenberger, 1992). In most environments, arsenite is generally thought to be more soluble and mobile form, which increase its potential toxicity. However, arsenate is the thermodynamically favorable form in most aerobic systems (Cullen and Reimer, 1989). Because of the severe toxicity of both forms of inorganic arsenic, issues of arsenic mobility and toxicity are important on both regional and global scales.

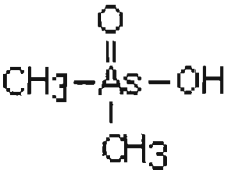
The speciation of arsenicals is of importance because of the differing levels of toxicity exhibited by various species (Kumaresan and Riyazuddin, 2001). The major arsenic species found in clinical and environmental samples are arsenite, arsenate, arsenous acids (H_2AsO^{3-}), arsenic acids (H_2AsO^{4-}), monomethylarsonate (MMA), dimethylarsinate (DMA), arsenobetaine (AB) and arsenocholine (AC) (Figure 1.2). These forms exhibit various oxidation states (-III, 0, III, V) and chemical complexity. Among the arsenic compounds in the environment, arsenite is 10 times more toxic than arsenate and 70 times more toxic than the methylated species such as DMA and MMA (Kumaresan and Riyazuddin, 2001).



Arsenate **Arsenite**



Monomethylarsonic acid (MMA)



Dimethylarsenic acid (DMA)

Figure 1.2: The chemical structure of some common arsenic compounds arsenate, arsenite, MMA and DMA

1.2 Ecology of Arsenic

Arsenic has been used since ancient times and nevertheless certain prokaryotes use arsenic oxyanions for energy generation, either by respiring AsV or by oxidizing AsIII. Arsenic cycling contributes to the oxidation of organic matter thus, enabling microbial reactions which results in mobilizing arsenic from solid to the aqueous phase, ensuing in contaminated drinking water (Oremland and Stolz, 2003).

Arsenic which is widely distributed in spite of its low abundance (0.0001 ppm is normally associated with the metals such as, Cu, Pb and Ag (Nriagu, 2002; Oremland and Stolz, 2003) and may exist in four oxidation states; As (-III), As (O), As (III) and As (V). Arsenate is the pre-dominant form of inorganic arsenic in aqueous aerobic environments (as $H_2AsO_4^{-1}$ and $HAsO_4^{-2}$) and is adsorbed strongly to the surface of several common

minerals, like ferrihydrite, alumina etc., while, arsenite is adsorbed to a lesser extent, which results in further mobility (Smedley and Kinniburgh, 2002). In natural waters, a number of methylated organo arsenicals (e.g.: methylarsonic, methylarsonus and dimethyl arsenic acids) are found as a result of urinary excretions of animals including humans or as breakdown or excretory products from aquatic biota. A recent review illustrates more information in regards to the detection of various organo arsenicals found in nature (Francesconi and Kuehnelt, 2002; Oremland and Stolz, 2003).

In contrast to restricted sources of anthropogenic arsenic pollution, naturally occurring arsenic is very largely distributed in many subsurface drinking water aquifers all around the world. It is these 'natural' sources that are of the most concern to human health (Welch et. al., 2000; Nordstrom, 2002; Oremland and Stolz, 2003).

Elevated levels of arsenic mobility in natural environments is a major concern in areas that are rich in arsenic: for example the arsenic exposure in West Bengal, India (Das et. al., 1996) and Bangladesh (Nickson et. al., 1998), Yellowstone National Park, USA (Stauffer and Thompson, 1984), have resulted in millions of people exposed to potentially high toxic levels of arsenic (Jackson et. al., 2003) which make them unsuitable for human consumption or recreation (Nimick et. al., 1998). Several other geothermal areas in the USA (Wilkie and Hering, 1998), Japan (Tanaka, 1990), Canada (Koch et. al., 1999), and New Zealand (Aggett and O'Brien, 1985) have also been shown to naturally contain high levels of arsenic compounds (Jackson et. al., 2003). Arsenic can be a waste product as a result of a number of activities including mining and ore refining (Harrington et. al., 1998; Smith et. al., 1998), fossil fuel combustion (Smith et. al., 1998), and wood treatment (Bull, 2000; Bull and Harland, 2001). High toxic levels of arsenic

can be detrimental to both human and ecological communities neighboring on such industries (Jackson et. al, 2003).

1.3 Microbial Transformations of Arsenic

A number of micro organisms have been identified to-date which demonstrates arsenic redox activity transforming arsenic between the two states as oxidized arsenate or reduced arsenite. In aerobic respiration, dissimilatory arsenate reducers use arsenate as the terminal electron acceptor, while arsenite is used as the electron donor for autotrophic growth by chemoautotrophic arsenite oxidizers. However, both these processes appear to be limited to a small group of bacteria (Jackson et. al., 2003).

A number of bacteria are capable of reducing arsenate as a form of detoxification. Recently organisms capable of using arsenate as the terminal electron acceptor in anaerobic respiration were discovered (Jackson et. al, 2003). Even though thermodynamic considerations suggest that dissimilatory reduction of arsenate could provide enough energy for microbial growth (Laverman et. al., 1995), it is likely that the overall toxicity of arsenic has restricted the distribution of this process amongst bacteria (Jackson et. al., 2003).

In comparison to the dissimilatory arsenate reducers, which use arsenate as the terminal electron acceptor in anaerobic respiration, few bacteria use arsenite as the electron donor for chemoautotrophic growth (Jackson et. al., 2003). Enough energy for growth could be produced as a result of the oxidation reaction (Anderson et. al., 1992). One of the bacteria isolated from mine waters, draining a gold-arsenic deposit in the USSR described as *Pseudomonas arsenitoxidans* could grow aerobically through the

chemoautotrophic oxidation of AsIII (Ilyaletdinov and Abdrashitova, 1981; Jackson et. al, 2003), although very few organisms appear to grow in this way.

Many more micro organisms are capable of reducing arsenate to arsenite via the function of arsenate reductase, a component of an arsenic detoxification system fairly widespread in bacteria, archaea, and eukarya. The various microbial arsenic redox processes can significantly affect the distribution and mobility of arsenic within natural environments, and might influence the toxicity of arsenic to other organisms.

1.4 Uses of Arsenic

The use of arsenic has been in combination with other materials in pigments, poison gases and insecticides such as, Paris green, calcium arsenate and lead arsenate and is also well known for its former use as a rat poison.

Arsenic as chromated copper arsenate (CCA) is utilized as a wood preservative. Lead arsenate and arsenic trioxide have been used for controlling pests and termites. Sodium arsenite solutions have been utilized for dipping cattle and sheep to control ticks and lice and also as a weed control in Australia, to eradicate ticks since early 1900s until 1955 (For tickicide use see 1.9 for details).

Arsenic compounds, in particular Orpiment and Realgar, were used as coloring agents and alloys in Greco-Roman times in ornamental painting and as cosmetics. In the late 19^{th} century, arsenic compounds were widely used as coloring agents, for example, in wallpapers and toys (Azcue and Nriagu, 1994). In the 20^{th} century, and indeed today, arsenic has been and still is widely used in agriculture, as pesticides, wood preservatives, and tickicides for cattle and sheep, in medicine for humans and animals, to treat sleeping

sickness and syphilis and in electronics and the metallurgy industry including manufacture of glassware, catalysts and anti-fouling paints.

The agricultural industry utilizes arsenic in the forms of monosodium methyl arsenate (MSMA), disodium methyl arsenate (DSMA), arsenic acid and dimethyl arsenic acid (cacodylic acid). MSMA and DSMA and dimethyl arsenic acid are used for the control of weeds in cotton fields as an herbicide application. In addition calcium arsenate has been used in the past for cotton leaf worms. In fact, arsenic acid is used as a desiccant for the defoliation of cotton ball prior to harvesting (Azcue and Nriagu, 1994).

Arsenic has a long history of medical applications; before penicillin was developed an arsenic compound was used to treat syphilis. Other main uses of arsenic in antiquity were pharmaceutical and medicinal, for example, as a remedy for asthma and coughs, and for a wide variety of health problems in Indian medicine (Azcue and Nriagu, 1994). Arsenic compounds were used to treat malaria, tuberculosis, diabetes and skin diseases. Also it is stated that until the mid 1980's arsenic compounds were used to treat sleeping sickness (Reynolds, 1999). Arsenic trioxide (As_2O_3) was also used in the treatment of acute promyleocytic leukemia (APL) (Shen et. al., 1997).

Though arsenic is a highly poisonous substance it has numerous applications and has been used widely for many years.

1.5 Toxicity of Arsenic

Exposure to the metalloid arsenic is a daily occurrence because of its environmental pervasiveness. Arsenic which is found in several different chemical forms and oxidation states causes acute and chronic adverse health effects, including cancer. The metabolism of arsenic has an important role in its toxicity. It includes reduction to a trivalent state and oxidative methylation to a pentavalent state. The trivalent arsenicals, including those methylated, have more potent toxic properties than the pentavalent arsenicals (Hughes, 2002). The toxicity of the inorganic arsenicals decreases as the oxidation state increases: arsonite (oxidation state +1) is more toxic than arsenite (+3). In turn, arsenite is much more toxic, soluble and mobile than arsenate (+5) (Nriagu, 1984). Since arsenate is structurally and electrically similar to phosphate, it is capable of entering metabolic pathways in the place of inorganic phosphate molecules (Booth and Guidotti, 1997). The mode of toxicity of arsenate is to partially block protein synthesis and interfere with protein phosphorylation when phosphate concentration is low (Tamaki and Frankenberger, 1992). Arsenite, on the other hand, is toxic primarily because AsIII acts like a metal, forming strong metal-thiol bonds with vicinal (adjoining) cysteines of proteins, inhibiting enzyme activities (Rosen, 1999). For instance, arsenite is known to inhibit enzymatic incision during the nucleotide repair pathway by interfering with signal transduction (Rosen, 1999).

Inorganic arsenic is considered to be a potential human carcinogen. Humans are exposed to it from soil, water, air and food. In the process of arsenic metabolism, inorganic arsenic is methylated to monomethylarsonic acid then, to dimethyl-arsenic acid followed by excretion through urine. Thus, arsenic exposure may cause DNA hypo-

methylation due to continuous methyl depletion, facilitating aberrant gene expression that result in carcinogenesis (Roy and Saha, 2002). Though arsenic is non-mutagenic it interacts synergistically with geno-toxic agents in the production of mutations and also induces chromosome abnormalities and cell proliferation. Epidemiological investigations in arsenic endemic regions of West Bengal (India) (Roy and Saha, 2002), indicated that inorganic arsenicals have the potential to cause skin and lung cancers in humans.

Due to chronic adverse effects after a few years of continued low level of arsenic exposure, many skin ailments appear, e.g. hypo pigmentation (white spots), hyper pigmentation (dark spots), collectively called melanosis by some physicians, and keratosis (break up of the skin on hands and feet). After a latency of about 10 years, skin cancers appear. After a latency of 20 - 30 years, internal cancers - particularly bladder and lung appear. These have all been seen in Taiwan and in Chile. By July 2000, 1700 cases of melanosis had already been identified in Bangladesh based on a survey of only 5% of the villages.

Inorganic arsenic exposure is associated with a variety of human tumors, particularly those of the skin. Arsenic disrupts cellular levels of p53 and mdm2 (murine double minute oncogene) which is thought to be a potential mechanism of carcinogenesis. The antitumor protein p53 plays a critical role in DNA repair. Hamadeh et. al., (1999) have proposed the disruption of the p53-mdm2 loop regulating the cell cycle arrest as a model for arsenic related skin carcinogenesis and tumors with elevated mdm2 levels.

Exposures to arsenic large enough to cause acute toxic effects would be easily recognized and the source of exposure would be found and eliminated, but in fact low

doses of arsenic over a long period of time results in the development of cancer. This demands a serious effort to trace all possible sources towards exposure to this so called poison.

1.6 Arsenic Resistance in Micro organisms

1.6.1 Mechanism of Arsenic resistance

A relatively large number of micro organisms are capable of resisting the toxic effects of arsenic, using methods such as arsenite oxidation (to produce the less toxic arsenate), extrusion of arsenic from the cell, and minimizing the uptake of arsenic from the environment.

1.6.1.1 Oxidation

Arsenite oxidization has been observed in strains identified as *Bacillus, Achromobacter* and *Pseudmonas* isolated from cattle dips. The oxidation however, resulted in no release of energy for growth, suggesting that these were not chemoautotrophs (Jackson et. al., 2003). However, in *Alcaligenes faecalis* isolated from soil, pre-exposure to AsIII resulted in arsenite oxidation, confirming to the existence of an inducible enzyme mechanism. Respiratory inhibitors can affect oxidation, which adds to the confirmation that cytochrome system may be required for arsenite oxidation, along with the action of the enzyme arsenite oxidase (Jackson et. al., 2003).

1.6.1.2 Uptake Minimization

In certain cases arsenic toxicity is avoided by minimizing the uptake of arsenic into the cell. For instance, *Escherichia coli* possess two different phosphate transport systems: the *Pit* and the *Pst* systems. The *Pit* system has low phosphate specificity and therefore will also transport arsenate, while the *Pst* system, shows high phosphate

specificity (Rosenberg et. al., 1977). As a result, it is assumed that organisms which lack the *Pit* system have increased resistance to arsenate (Cervantas et. al., 1994; Jackson et. al., 2003) than those organisms utilizing the more specific *Pst* system. However, in several plant species, arsenate uptake and toxicity can be substituted by increasing the amount of phosphate in the environment (Abedin et. al., 2002; Jackson et. al., 2003), a similar effect can be noted when bacterial cultures are incubated at different concentrations of arsenate and phosphate (Jackson et. al., 2003).

1.6.1.3 Extrusion

The best characterized, and probably the most widespread, arsenic resistance mechanism in micro organisms is the extrusion from the cells by the function of *ars* (arsenic resistance) gene system. Previous studies have shown that, arsenite and arsenate resistance in certain bacteria is mediated by plasmid borne *ars* genes, for example in *E. coli*, *Staphylococcus aureus*, and *Pseudomonas aeruginosa*. Basically, the *ars* gene clusters consists of at least three genes; *arsR* - a regulatory gene, *arsB* - a gene encoding for an dedicated arsenite-specific transmembrane pump, and *arsC* - a gene encoding for an arsenate reductase (Jackson and Dugas, 2003). Toxic effect of arsenic is reduced by the action of the ArsB which is an membrane protein that pumps arsenite out of the cell; however, arsenate must first be reduced to arsenite by the soluble arsenate reductase ArsC. *arsR* codes for a repressor protein that regulates *ars* operon expression (Figure 1.3). In some bacteria the operon contains other genes: *arsA* produces an oxyanion-stimulated ATPase which couples ATP hydrolysis thereby extruding arsenicals and antimonials by the action of ArsB protein; Another regulatory protein *arsD* is identified in fewer arsenic resistant species which encodes for a regulatory protein that controls the

upper level of *ars* expression by means of controlling the over expression of ArsB which itself is toxic to cells; *arsT* is a member of the flavoprotein pyridine nucleotide-disulfide oxidoreductase family which catalyzes the NADPH-dependent reduction of the active site disulfide in thioredoxin and some other genes including *arsH* and *arsO* with uncertain functions but essential for arsenic resistance in some bacteria (Neyt et. al., 1997). Thus, resistance to both arsenite and arsenate can be provided by the series of genes, *arsRBC*, with increased efficiency at extruding arsenite if *arsA* is also expressed (Figure 1.3). Apart from plasmids of *S. aureus* and *S. xylosus*, the *ars* operon and its role in arsenic resistance has been identified in several other bacteria, either on the plasmid or chromosome. Within the division Proteobacteria, *ars*-mediated resistance has been observed in several bacterial species such as; *Escherichia coli*, *Pseudomonas aeruginosa*, a number of *Yersinia* species, *Acidiphilum multivorum*, *Thiobacillus ferrooxidans*, *Acidithiobacillus caldus* and *Acidothiobacillus ferrooxidans* (Dopson et. al., 2001; Butcher et. al., 2000; Butcher and Rawlings, 2002). Gram-positive bacteria possessing the *ars* operon system has been identified which includes the Staphylococci and *Bacillus subtilis*. In higher organisms such as the yeast *Saccharomyces cerevisiae*, the *ACR* (or *ARR*) cluster is homologous to *ars* and also confers arsenic resistance, suggesting a similar mechanism employed as prokaryotes. A search for *ars* genes in GenBank reveals a greater similarity to putative *ars* genes in bacteria and archaea that have not been characterized yet (see Chapter 3). Thus, the *ars* system seems to be a common mechanism adapted by micro organisms thus evading toxic effects of arsenite and arsenate.

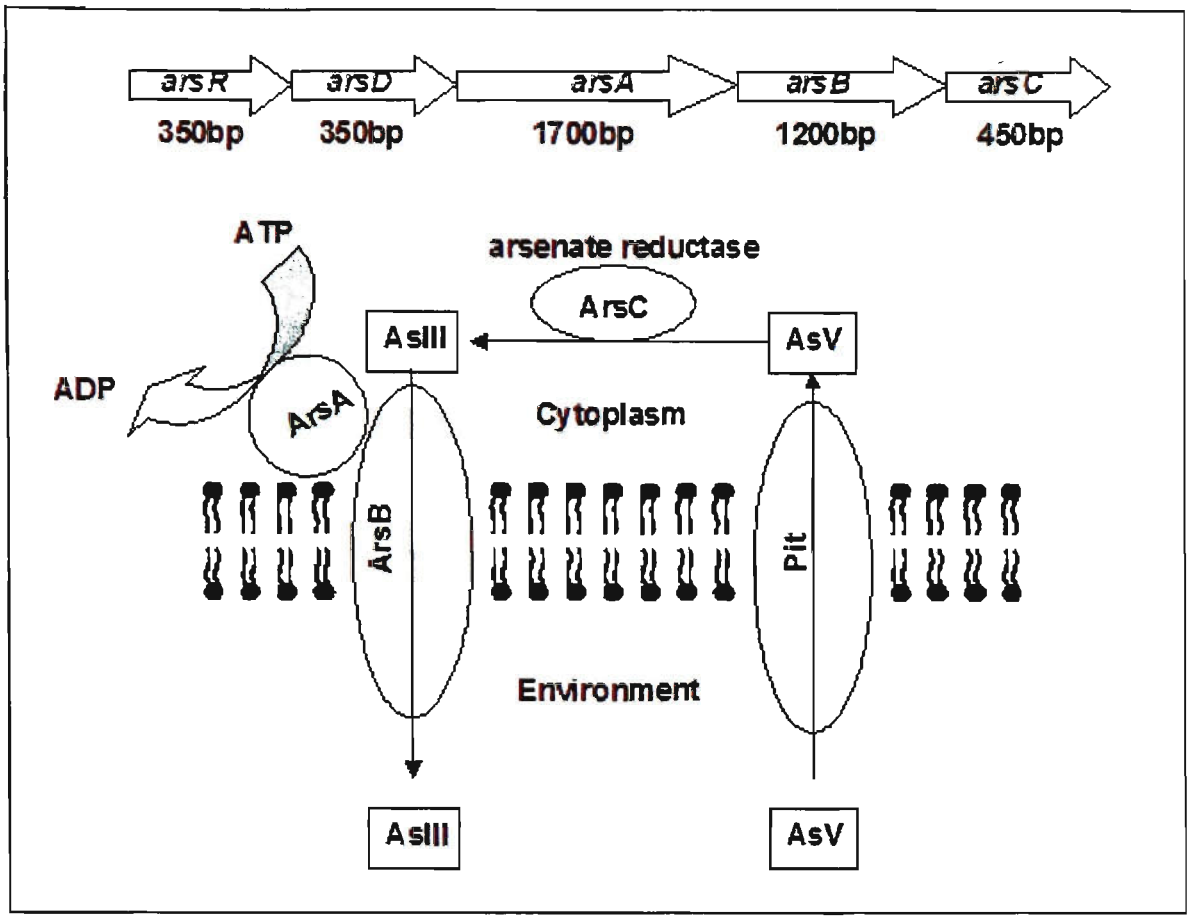


Figure 1.3: Schematic illustration of the genes of the *ars* operon from *E. coli* R773 and the actions of the proteins involved in arsenic resistance. Arsenate can be taken up by bacterial cells via non specific phosphate transport, such as the *Pit* system. Arsenate reductase (ArsC) converts arsenate to arsenite, which is extruded from the cell via the ArsB protein. *arsA* codes for an ATPase which enhances arsenite extrusion in some organisms.

1.6.2 Arsenic Resistant Species Characterized at Molecular Level

To date a number of arsenic resistant species having a genetic determinant encoded by an *ars* cluster have been characterized to the molecular level to help understand the regulatory mechanism of arsenic detoxification. The following list details some of the known bacterial species and yeast possessing an arsenic resistant determinant from the well characterized model species (*E. coli* R773 *arsRDABC*) to that of the divergent and less known ones.

***Escherichia coli*:** High level resistance to arsenic in *E. coli* is conferred by the *arsRDABC* cluster of plasmid R773. Most of the studies related to arsenic resistance and regulation has come from this plasmid. The *arsR* and *arsD* genes encode trans-acting repressors that bind to the *ars* operon and control the expression of the operon. The *arsA* and *arsB* genes encode the ATPase and membrane translocase subunits of an arsenical pump and the *arsC* encodes an arsenate reductase, which is required to catalyze the reduction of arsenate to arsenite prior to extrusion (Chen et. al., 1986).

***Pseudomonas aeruginosa*:** The *P. aeruginosa* chromosomal *ars* cluster is the second functionally characterized *ars* cluster of bacteria identified so far. It has three ORF's namely *arsR*, *arsB* and *arsC*. Among the three *ars* polypeptides, the putative ArsB protein is highly conserved, consistent with what has been observed for *E. coli* R773 ArsB. ArsR and ArsB proteins share more homology with their respective proteins from Gram-negative *ars* operons while the ArsC is more homologous with the polypeptides of Gram-positive *ars* operons (Cai et. al., 1998).

***Staphylococcus*:** The arsenate and arsenite resistance region of *S. xylosus* pSX267 has been sequenced. This has revealed the presence of three consecutive open reading frames,

named *arsR*, *arsB* and *arsC*. These are homologous to the *ars* genes that have been sequenced so far. The arsenic cluster in *S. aureus* pI258 has only the *arsR*, *arsB* and *arsC* genes whose gene products are similar to those of *S. xylosus* pSX267 and *E. coli* pR773 (Ji and Silver, 1992a; Rosenstein et. al., 1992).

***Bacillus subtilis*:** The *B. subtilis* skin element confers resistance to arsenate and arsenite. The *ars* cluster in the skin element contains four genes in the order *arsR*, *Orf2*, *arsB*, and *arsC*. The products of *arsR* and *arsC* are homologous to products of *ars* genes in *S. aureus* and *E. coli*, ArsR and ArsC, respectively. The *B. subtilis arsB* product is also homologous to the *S. aureus ars*B product. However, the ArsB protein exhibits only limited sequence identity (24%) to the *S. aureus* ArsB. In addition, there is no homolog of the *Orf2* product in *S. aureus* (Sato and Kobayashi, 1998).

***Saccharomyces cerevisiae*:** An *ars* cluster in the eukaryotic monocellular organism *S. cerevisiae* has also been cloned and studied. Homology searches revealed a cluster of three new open reading frames named *ACR1*, *ACR2* and *ACR3*. The hypothetical product of the *ACR1* gene is similar to the transcriptional regulatory proteins, encoded by *YAP1*, and *YAP2* genes from *S. cerevisiae*. Disruption of the *ACR1* gene conduces to an arsenite and arsenate hypersensitivity phenotype. The *ACR2* gene is indispensable for arsenate but not for arsenite resistance. The hypothetical product of the *ACR3* gene shows high similarity to the hypothetical membrane protein encoded by *B. subtilis Orf1* of the skin element and weak similarity to the ArsB membrane protein of the *S. aureus* arsenical-resistance cluster. Over expression of the *ACR3* gene confers an arsenite- but not an arsenate-resistance phenotype. The presence of *ACR3* together with *ACR2* on a multi copy plasmid expands the resistance phenotype into arsenate. These findings suggest that

all three novel genes: *ACR1*, *ACR2* and *ACR3* are involved in the arsenical-resistance phenomenon in *S. cerevisiae* (Bobrowicz et. al., 1997).

***Acidiphilium multivorum*:** The arsenic resistance (*ars*) cluster from pKW301 of *A. multivorum* AIU 301 has been cloned and sequenced. This DNA sequence contains five genes in the following order; *arsR, arsD, arsA, arsB, arsC* similar to that found in *E. coli* R773 *ars* operon. The ArsA, ArsB, and ArsC proteins of pKW301 are highly homologous to the corresponding proteins of *E. coli* pR773, suggesting that the functions of the former proteins are identical to those of the latter proteins. On the other hand, ArsR of pKW301 is less similar to the ArsR proteins described previously. In ArsR of pKW301, the three cysteine residues (Cys32, Cys34 and Cys37) known to form a pyramidal cage to bind arsenite are conserved (Shi et. al., 1996), but the amino acid sequence of the helix-turn-helix region including the His and Ser residues are not conserved. Nevertheless, this exceptionally small protein (85 a.a.) can function as a repressor since it has been demonstrated that the expression of the *ars* genes depends on the presence of sodium, and inactivation of ArsR by the frameshift mutation abolishes the dependence of *ars* gene expression on arsenite (Suzuki et. al., 1998).

***Halobacterium* sp.:** The halophilic archea *Halobacterium* pNRC100 has five genes, three of which *arsDAR* are organized divergently from downstream *arsRC* genes. There is no *arsB* on pNRC100, although a putative arsenite carrier, which exhibits no sequence similarity to ArsB is present in the chromosome (Ng et. al., 1998; Li et. al., 2001).

***Thiobacillus ferrooxidans*:** The *ars* cluster of *T. ferrooxidans* has a very unusual arrangement. No *arsD* gene or *arsA*-like gene (ATPase subunit) is found in the immediate vicinity of the arsenic resistance genes, and only *arsC*, *arsB*, and *arsH*-like genes have

been identified based on initial sequence comparisons. More careful analysis resulted in identification of a putative regulator between the *arsB* and *arsC* genes. However, the predicted protein exhibited only relatively weak homology to the ArsR proteins produced by previously described *ars* operons (Butcher et. al., 2000).

***Yersinia enterocolitica*:** All pathogenic *yersiniae* possess a 70 kb virulence plasmid called pYV which encodes secreted anti host proteins called Yops as well as a type III secretion machinery that is required for Yop secretion. Genes encoding Yop synthesis and secretion are tightly clustered in three quadrants of the pYV plasmid. It has been shown that in the low-virulence strains of *Y. enterocolitica*, the fourth quadrant of the plasmid contains a new class II transposon; Tn2502. Tn2502 confers arsenite and arsenate resistance. This resistance involves four genes; three are homologous to the *arsRBC* genes present on the *E. coli* chromosome, but no homolog of the fourth one, instead *arsH*, has been found. ArsH is a 26.4 kDa protein whose role is not exactly known. It is thought that it acts as a regulator similar to ArsD (Neyt et. al., 1997).

***Streptomyces* sp.:** An arsenic resistance gene cluster from the large linear plasmid pHZ227 of *Streptomyces* sp. strain FR-008, was identified recently containing two novel genes, *arsO* (for a putative flavin-binding monooxygenase) and *arsT* (for a putative thioredoxin reductase), which are co-activated and co-transcribed with *arsR1-arsB* and *arsC*, respectively. Deletion of the *ars* gene cluster on pHZ227 in *Streptomyces* sp. strain FR-008 resulted in sensitivity to arsenic. The pHZ227 ArsB protein shows greater homology to the yeast arsenite transporter Acr3p, while the ArsC appears to be a bacterial thioredoxin-dependent ArsC-type arsenate reductase with four conserved cysteine thioredoxin-requiring motifs (Wang et. al., 2006).

Resistance to arsenical compounds in bacteria can be plasmid or chromosomally mediated. Three types of *ars* operons have been well documented three - (*arsRBC*), four- (*arsRBCH*) and five- (*arsRDABC*) gene arsenic resistance determinants (Ryan and Colleran 2002). Table 1.1 lists the currently specifically characterized *ars* operons / clusters all published by March 2007. Figure 1.4 illustrates the various arrangements of these arsenic resistance genes.

Table 1.1: Characterized bacterial *ars* operons

Bacteria	Gram staining	Operon location*	Structure	Reference
Archeal Halobacterium sp. Strain NRC-1		pNRC100	***arsDARC***	(Ng et. al., 1998)
Acidithiobacillus caldus	–	Chromosome	***arsHBRC***	(Dopson et. al., 2001)
Acidithiobacillus caldus	–	Transposon	***arsDADA...B***	(Groot et. al., 2003)
Acidothiobacillus ferrooxidans	–	Chromosome	***arsCRBH***	(Butcher et. al., 2000)
Acidiphilium multivorum	–	pKW301	***arsRDABC***	(Suzuki et. al., 1998)
Bacillus subtilis	+	SKIN element chromosome	***arsR(orf2)BC***	(Sato and Kobayashi 1998)
Escherichia coli	–	pR733	***arsRDABC***	(Chen et. al., 1986)
Escherichia coli	–	pR46	***arsRDABC***	(Bruhn et. al., 1996)
Escherichia coli	–	Chromosome	***arsRBC***	(Carlin et. al., 1995)
Pseudomonas aeruginosa	–	Chromosome	***arsRBC***	(Cai et. al., 1998)
Pseudomonas aeruginosa	–	pUM310	***arsRBC***	(Cervantes and Chavez, 1992)
Pseudomonas fluorescens	–	Chromosome	***arsRBC***	(Prithivirajsingh et. al., 2001)
Serratia marcescens	–	pR478	***arsRBCH***	(Ryan and Colleran 2002)
Staphylococcus aureus	+	pI258	***arsRBC***	(Ji and Silver 1992b)
Staphylococcus xylosus	+	pSX267	***arsRBC***	(Rosenstein et. al., 1992)
Streptomyces sp. strainFR-008		pHZ227	***arsRBOCT***	(Wang et. al., 2006)
Yersinia enterocolitica	–	pYVe227	***ars HRBC***	(Neyt et. al., 1997)

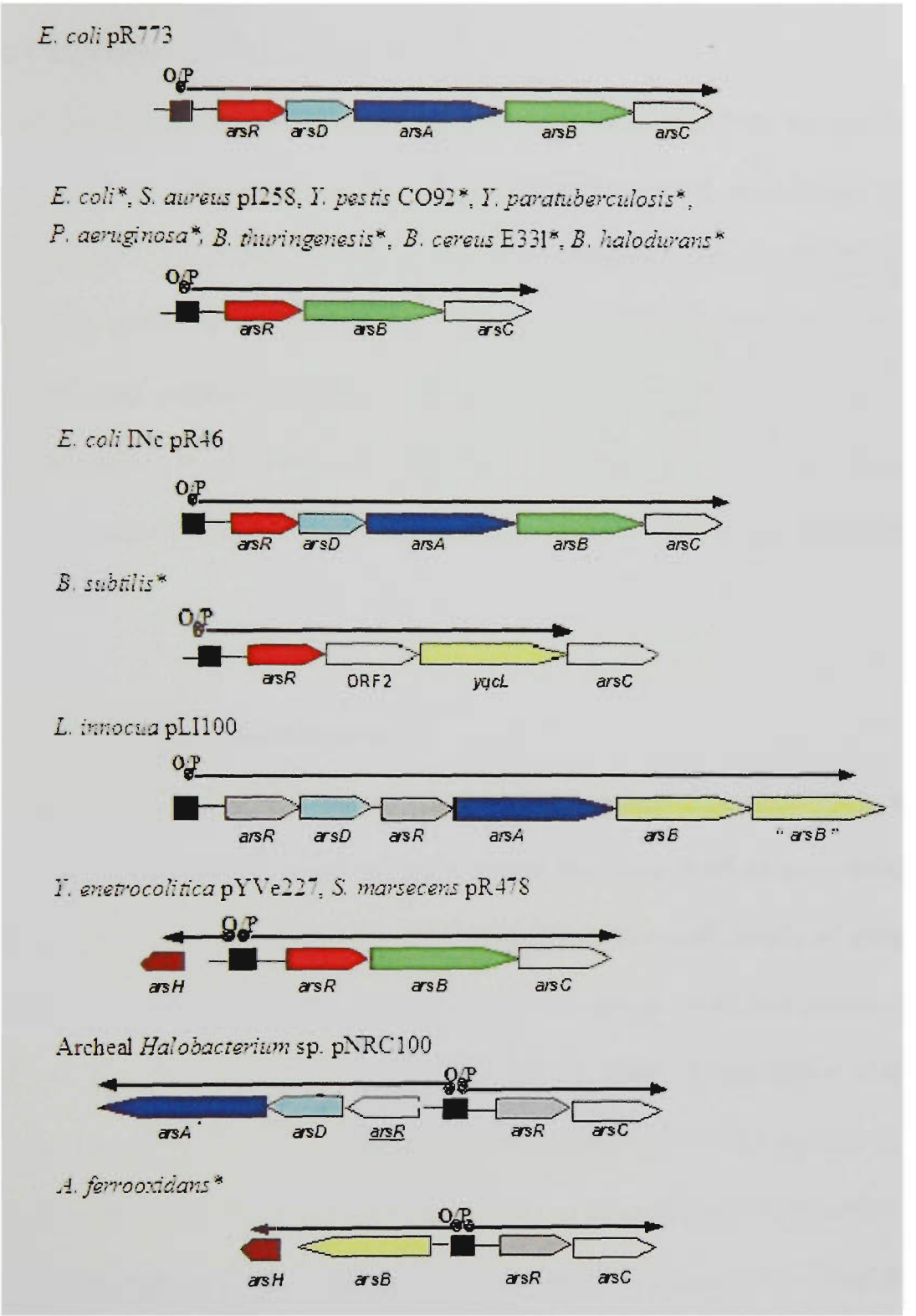


Figure 1.4: Schematic illustration of various *ars* operons/clusters identified in bacteria. Black box: O/P, operator/promoter sites of repressor regulation and initiation of mRNA synthesis. (a) Model species (*E. coli* pR773) *arsRDABC* operon. (b) Three gene *ars* operons and (c) four or five gene *ars* operons identified in different bacteria. Chromosomally located *ars* operons are indicated with * and plasmid borne operons indicated next to the species.

1.7 Proteins Encoded by the *ars* Operons

Arsenic resistance in bacteria encoded by the *ars* operon has been known for many years (Cai et. al., 1998). Much of the study about regulation of *ars* operons has come from *E. coli* pR773 and R46 plasmids in which the five genes are transcribed as a single polycistronic mRNA in the order *arsRDABC* (Owolabi & Rosen, 1990; Bruhn et. al., 1996). Arsenic and antimony have been found to regulate the expression of the *ars* operon on pR773 (Silver et. al., 1981; San Francisco et. al., 1990). The first two genes *arsR* and *arsD* of the R773 operon encodes two inducer dependent trans-acting repressors (Cai et. al., 1998).

1.7.1 ArsR: trans-acting repressor protein

The ArsR, a 117 residue protein of the well characterised *E. coli* R773 is an inducer - dependent trans-acting repressor that controls the basal level of *ars* operon expression (San Francisco et. al., 1990; Wu & Rosen, 1991). The *arsR* family of genes encodes the protein responsible for the regulation of the *ars* operon. ArsR is a protein of approximately 13 kDa and is functional as a dimer (Shi et. al., 1994). Transcription of the *ars* operon comes to a halt when the dimer binds to the operator site of the operon with high affinity. Wu and Rosen (1993) have demonstrated that in the presence of an inducer such as arsenite and antimonite, ArsR binds to the inducer and dissociates from the operator region allowing transcription to proceed.

The members of the ArsR family of repressor proteins have at least 3 domains; 1) a metal binding domain, 2) a DNA-binding domain and 3) a dimerization domain (Xu and Rosen, 1997). The ArsR sequence **$ELC_{32}VC_{34}DL$** has been proposed to form a

portion of the metal binding domain and a putative helix-turn-helix DNA binding motif from residues 38 to 54 in ArsR (Xu and Rosen, 1997). When the ArsR protein is exposed to arsenite for example, it undergoes a conformational change as shown in Figure 1.5.

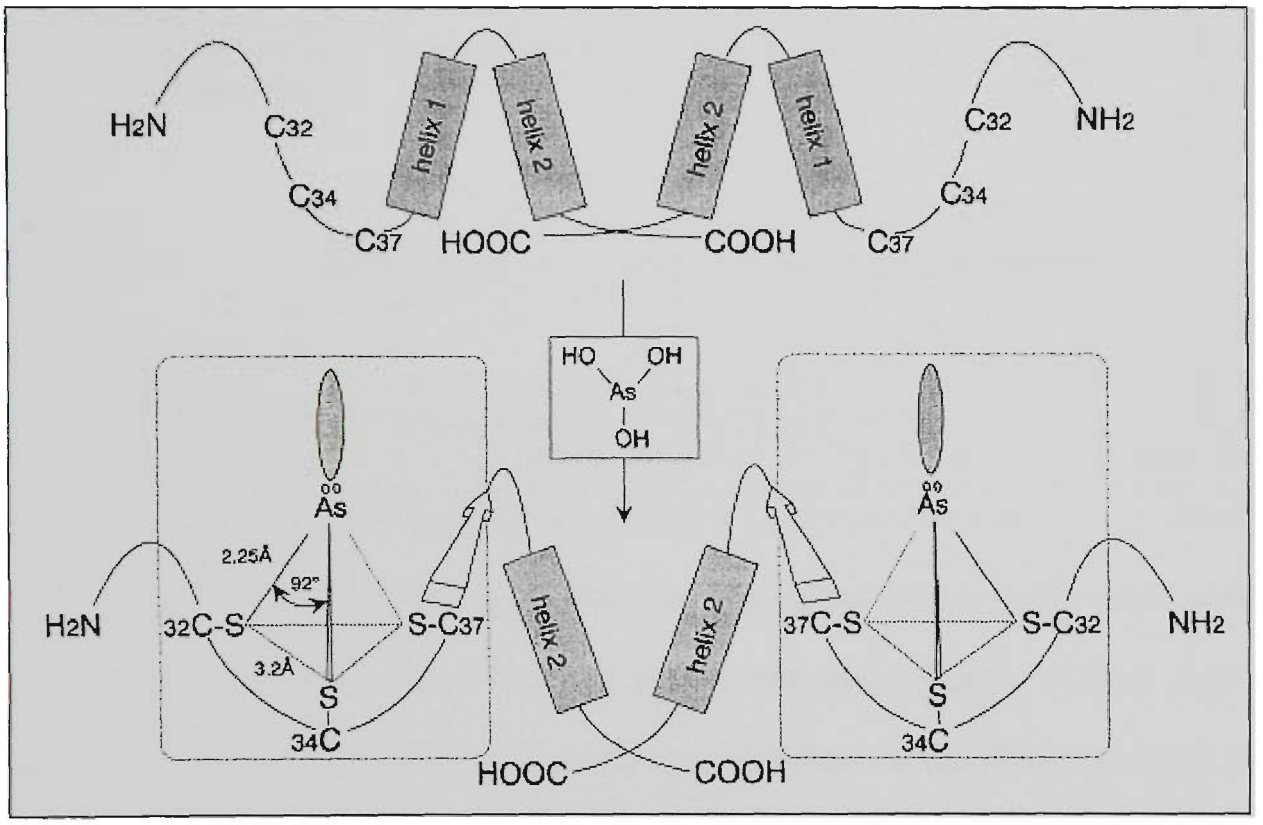


Figure 1.5: Model for arsenic-ArsR complex in *E. coli* R773. ArsR protein in the absence of inducer binds to the operator/promoter region through a helix-turn-helix domain, repressing transcription of the *ars* operon. While in the presence of AsIII, Cys-32 and Cys-34 are bound by arsenite, enabling a change in conformation of DNA-binding domain by reducing the interaction between ArsR and operator. Arsenite is bound in a pyramidal cage formed by the three thiolates of Cys-32, Cys-34 and Cys-37. (Adapted from Shi et. al., 1996)

Through DNase I foot print analysis; Wu and Rosen (1993b) have shown that the binding site of the ArsR repressor is between -64 and -31 nucleotides of the R773 *ars* operon. Xu et. al., (1996), showed that the specific bases to which the protein dimer attaches is **TCAT-N_7-TTTG** (Figure 1.6) and was same in the *E. coli* chromosomal operon. These two ArsR dimers were found to be interchangeable as they could bind to each other's promoters (Xu et. al., 1996). A similar method was used to identify the

homologous *ars* repressor from Gram-positive *S. xylosus* pSX267 (Rosenstein et. al., 1994), but the region **TCAT-N$_7$-TTTG** site was not found in *S. xylosus* suggesting that, it is a Gram-negative consensus sequence.

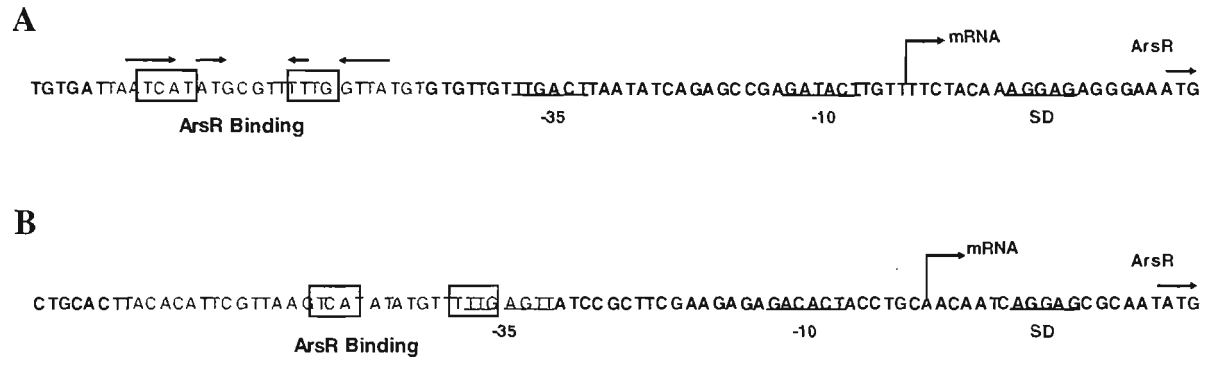


Figure 1.6: Regulatory region of the *E. coli* pR773 and chromosomal *ars* operon. A. Promoter region of the *E. coli* R773 *ars* operon. The contact points between the R773 ArsR repressor and DNA are enclosed in boxes. B. Promoter region of the *E. coli* chromosomal *ars* operon. The shaded sequence indicates the binding site for ArsR as defined by DNase I footprinting. Presumed -10 and -35 promoter elements and Shine-Dalgarno sequence sites are underlined. Adapted from Wu and Rosen (1993a) and Xu et. al., (1996)

However, Sato & Kobayashi, (1998) have shown that the promoter region of *B. subtilis* skin element *ars* operon has an 8 bp inverted repeat (Figure 1.7) that might function as a regulatory site, but the inverted repeat exhibits no significant similarity to those in the *ars* operons of R773 (*E. coli*), pI258 (*S. aureus*) and pSX267 (*S. xylosus*).

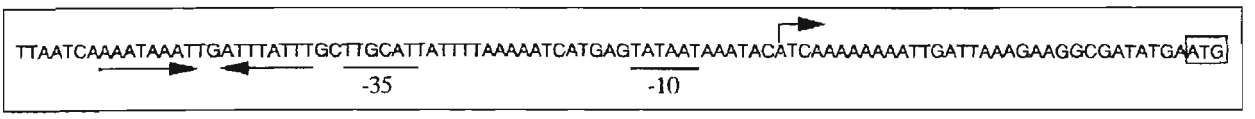


Figure 1.7: Structure of the *ars* operon and nucleotide sequence of the upstream region of the *ars* operon of *B. subtilis*. The site of initiation of transcription is indicated by an arrow. Putative -10 and -35 RNA polymerase-binding regions of the *ars* promoter are underlined, and the putative translation initiation codon of *arsR* is boxed. The converging arrows indicate the region of inverted repeats. Adapted from Sato & Kobayashi, 1998

1.7.2 ArsD: Secondary regulatory protein and chaperone

ArsD protein is a secondary regulator of transcription of the *ars* operon (Sato & Kobayashi, 1998). Like ArsR, ArsD is a small protein with 120 residues but does not exhibit any significant sequence similarity to ArsR (Chen and Rosen, 1997). ArsD is a 13-kDa protein and like ArsR, forms a dimer binding to a very similar site on the *ars* promoter as ArsR, but with two different orders of magnitude which demonstrates ArsR would bind to the *ars* promoter preferentially repressing the *ars* expression (Chen and Rosen, 1997). Shi et. al. (1994), have shown that *in vivo* repression by ArsR can be fully relieved with 10 μM sodium arsenite, while ArsD repression requires approximately 100 μM sodium arsenite, suggesting that ArsR has higher affinity for inducer than ArsD (Chen and Rosen, 1997). Since its affinity for inducer is less than that of ArsR, the relatively low level of inducer present in the cell would not prevent its binding. Likewise, exposure to high levels of environmental metalloid would cause dissociation of ArsD, effecting further expression of the *ars* genes and increased synthesis of the Ars extrusion pump. Synthesis of high levels of the pump is itself toxic (Wu & Rosen, 1993a; Bruhn et. al., 1996), and has led to the hypothesis that one role for ArsD is to prevent over expression of ArsB (Li et. al., 2001), so that there is a balance between detoxification of the metalloid and the expression of the pump genes. Thus, the action of the two repressors ArsR and ArsD forms a homeostatic regulatory circuit that maintains the level of ars expression within a narrow range, with ArsR controlling basal level of expression and ArsD controlling maximal expression in R773 (Chen and Rosen, 1997).

Li et. al., (2001) have shown that, the vicinal cysteine pairs in the protein sequence are responsible for the metalloid sensing by ArsD. It was demonstrated that the

pR773 ArsD bound to the promoter as a homodimer, represses the expression of the *ars* genes. ArsD has three vicinal cysteine pairs; Cys12-Cys13, Cys112-Cys113 and Cys119-Cys120 (Li et. al., 2001). The dimeric repressor contains four independent inducer binding domains, two in each subunit and each monomer has two vicinal cysteine pairs Cys12-Cys13 and Cys112-Cys113 to coordinate with metalloids.

The mutation or deletion of Cys119-Cys120 had no effect on repression or metalloid responsiveness *in vivo* or *in vitro* (Li et. al., 2001). In contrast, mutation of the Cys12-Cys13 and Cys112-Cys113 pairs resulted in loss of ArsD function both *in vivo* and *in vitro* indicating that these cysteine pairs are required for ArsD activity *in vivo*.

SbIII is a better inducer than AsIII for ArsD, the binding of SbIII to the four binding sites occurs sequentially (Figure 1.8). Since ArsD can bind to the metalloids very tightly, it is also presumed that it might also act as a chaperone to deliver the free toxic metalloid within the cytoplasm to the pump (Li et. al., 2002).

ArsD of pR773 was shown to bind to the promoter as a homodimer, repressing the expression of the *ars* genes. When the cysteines (Cys12-Cys13 and Cys112-Cys113) are bound by inducer (AsIII), the confirmation of the repressor changes, thus, decreasing the affinity of ArsD for DNA, hence ArsD dissociates from the *arsRDABC* operon as shown in Figure 1.9.

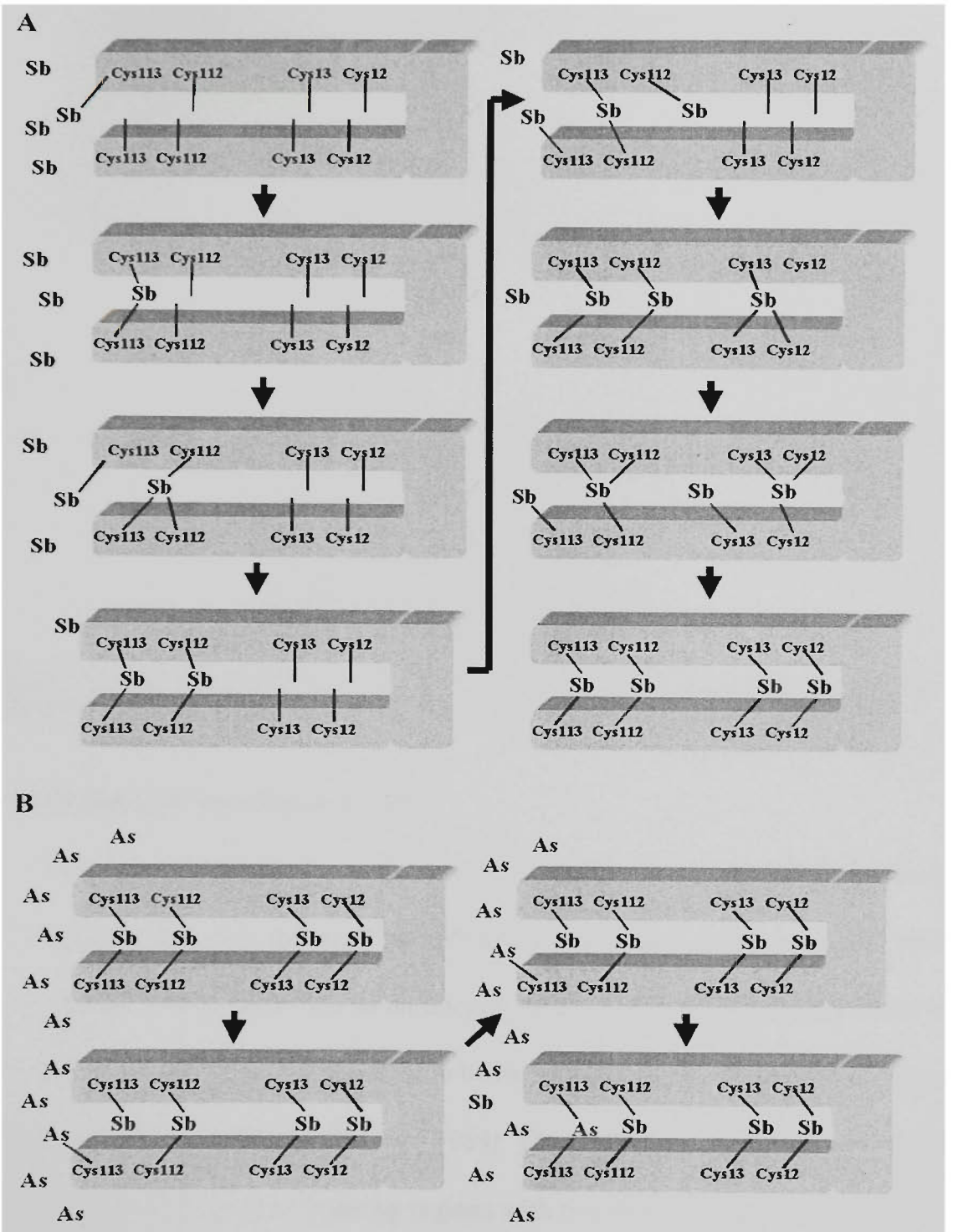


Figure 1.8: A model for metalloid binding to ArsD. A. the metalloid-binding sites are composed of pairs of equivalent cysteine residues donated by each subunit of the ArsD dimer. The binding of metalloids, such as SbIII, to the four binding sites occurs sequentially. The first SbIII is bound by the thiol side chains of the Cys113 residues in each subunit. This SbIII can then be transferred to the next binding site, composed of the Cys112 residues, and will envisage a transient state in which SbIII is bound by both Cys113 and Cys112 residues with either two or three protein ligands. Subsequently, the SbIII ions in the Cys113 and Cys112 binding sites can be transferred to the Cys13 and Cys12 binding sites. B. AsIII can displace SbIII from the fast exchange Cys113 binding site. The bonds between the thiol groups of the cysteine residues and the metalloids are shown as straight lines. SbIII and AsIII ions have been abbreviated to Sb and As, respectively. (Adapted from Li et. al., 2002)

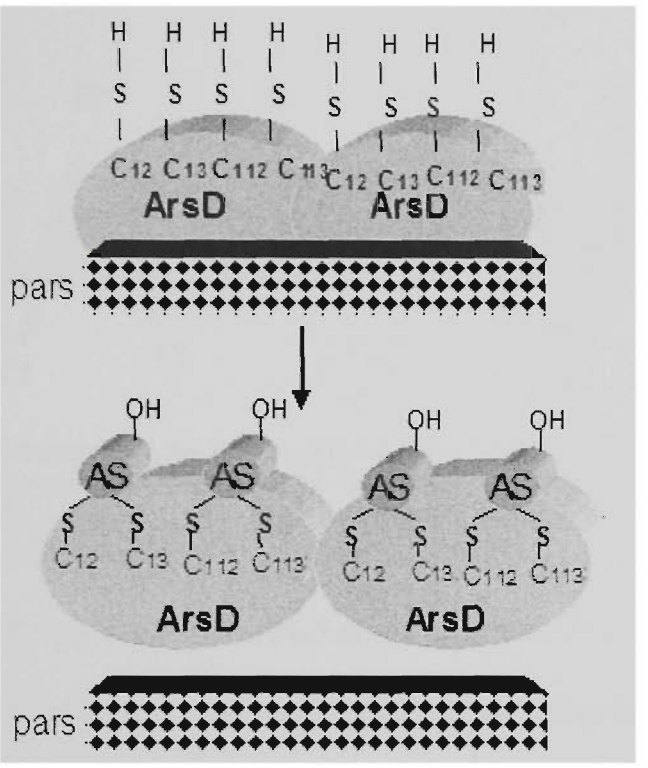


Figure 1.9: Model of inducer-binding domains of the ArsD repressor. ArsD binds to the *arsRDABC* promoter (pars) as a homodimer of two 120-residue monomers (Li et. al., 2001).

1.7.3 ArsB: Integral membrane protein

The arsenical resistance operon has been shown to consist of three structural genes, of which two, *arsA* and *arsB* are required for resistance and transport of arsenite (Tisa and Rosen, 1990). Identified as an integral membrane protein localized in the inner membrane, the ArsB has been postulated to be the anion channel component of the pump (Chen et. al., 1986; San Francisco et. al., 1989). The ArsB from *E. coli* is approximately 45 kDa and has 12 membrane spanning regions with five cytoplasmic and six periplasmic loops. Out of the five cytoplasmic loops C1, C2 and C4 have a net positive charge which may serve as cytoplasmic anchors for membrane spanning regions (Wu et. al., 1992) as shown in Figure 1.10.

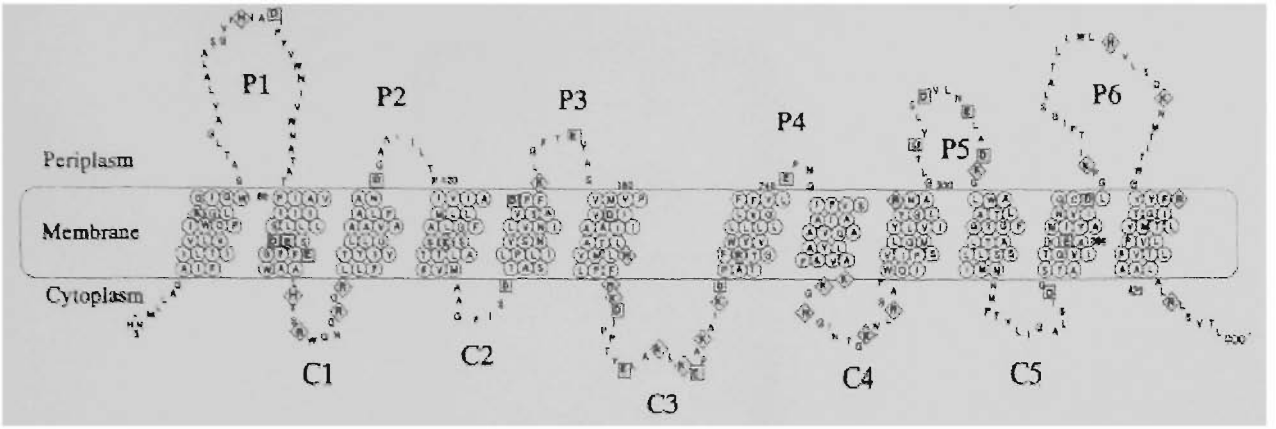


Figure 1.10: Topological structure of the *E. coli* plasmid R773 ArsB efflux pump protein. ArsB protein contains 12 membrane-spanning α-helices joined by six periplasmic loops (P1-P6) and five cytoplasmic loops (C1-C5). The positive and negative charged residues are indicated with diamond- and square-shapes, respectively. (Adapted from Wu et. al., 1992)

Analysis of the residues in the cytoplasmic loop has shown that C3 has relatively high net charge residues. Based on the topological model proposed by Wu et. al., 1992, the cytoplasmic C3 loop has 22 residues, 10 of which are charged, and therefore C3 would be a logical region to interact with the ArsA protein. However, there are other two cytosolic loops, C1 and C4 which contain 4 and 5 charged residues, which could be possible candidates for ArsA interaction (Wu et. al., 1992).

Arsenical and antimonial resistance in bacteria is attributed to the extrusion of oxyanions out of the cells. The ArsB protein forms an oxyanion conductive pathway, whereby, it transports arsenite molecules out of the cell (Kuroda et. al., 1997). The ArsB is shown to function in the presence and absence of the ArsA ATPase. However, the ArsA-B transporter complex is a more efficient system (Dey & Rosen, 1995) Figure 1.11.

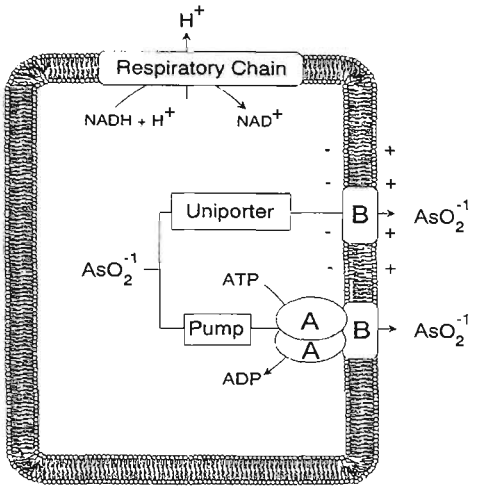


Figure 1.11: Dual energy coupling of the Ars transporter. ArsB functions physiologically in either of two modes: as a potential driven secondary carrier or as a subunit of an obligatory ATP-coupled pump. In cells lacking an ArsA, ArsB translocates arsenical and antimonial oxyanions, with energy derived from the proton-pumping respiratory chain. In the presence of both, ArsA-ArsB complex pumps out arsenic ions utilizing the energy of ATP. (Adapted from Kuroda et. al., 1997)

Most of the arsenite membrane pump proteins identified so far in bacteria, such as those of *S. aureus* pI258 and *S. xylosus* pSX267 are very similar in sequence to the *E. coli* pR773 ArsB, while that of *B. subtilis,* designated as YqcL, shows very limited homology (Sato & Kobayashi, 1998). Hydropathy profiling showed that the *B. subtilis* ArsB protein has 10 membrane spanning regions, suggesting a similar function to that of other ArsB proteins which have 12 membrane spanning regions (Sato & Kobayashi, 1998). Bobrowicz et. al., (1997) revealed that 3 genes *ACR1*, *ACR2* and *ACR3* of *S.cerevisiae* are essential for arsenical resistance, in which the *ACR3* gene has high similarity to *B. subtilis* ArsB (YqcL), suggesting that they possess similar proteins for extrusion of arsenite. Interestingly, both *ACR3* and YqcL were reported to transport arsenite but not antimonite (Bobrowicz et. al., 1997; Sato & Kobayashi, 1998).

1.7.3.1 Efflux mechanisms of arsenic resistance in micro organisms

In bacteria there are two basic mechanisms of arsenic extrusion systems. Firstly, carrier mediated efflux by arsenite carrier protein, and secondly, by arsenite translocating ATPase (Dey and Rosen, 1995) As shown in Figure 1.12, arsenate is transported into bacterial and yeast cells as oxyanions comparable to those of phosphate carried out by phosphate transport membrane systems, for example the Pit and Pst systems of *Escherichia coli* and Pho87p of *S. cerevisiae*; (Mukhopadhyay et. al., 2002). As mentioned earlier, to resist the toxic effects of arsenic to cells, bacteria have developed a fine regulatory mechanism consisting of genes that specifically confer resistance to inorganic arsenic, both AsV and AsIII. Whereas in yeast *S. cerevisiae,* there are two parallel ways of removing toxic arsenite from the cytosol, firstly, arsenate is reduced to arsenite by Acr2p and then pumped out by Acr3p; and secondly, arsenite and antimonite (SbIII) are transferred to $As(GS)_3$ (and presumably $Sb(GS)_3$) and pumped from the cytoplasm into the vacuole by another yeast protein, Ycf1p, which is a close homologue of the human MRP, and has been shown to be a vacuolar glutathione *S*-conjugate with a broad range of substrate specificity (Ghosh et. al., 1999; Mukhopadhyay et. al., 2002).

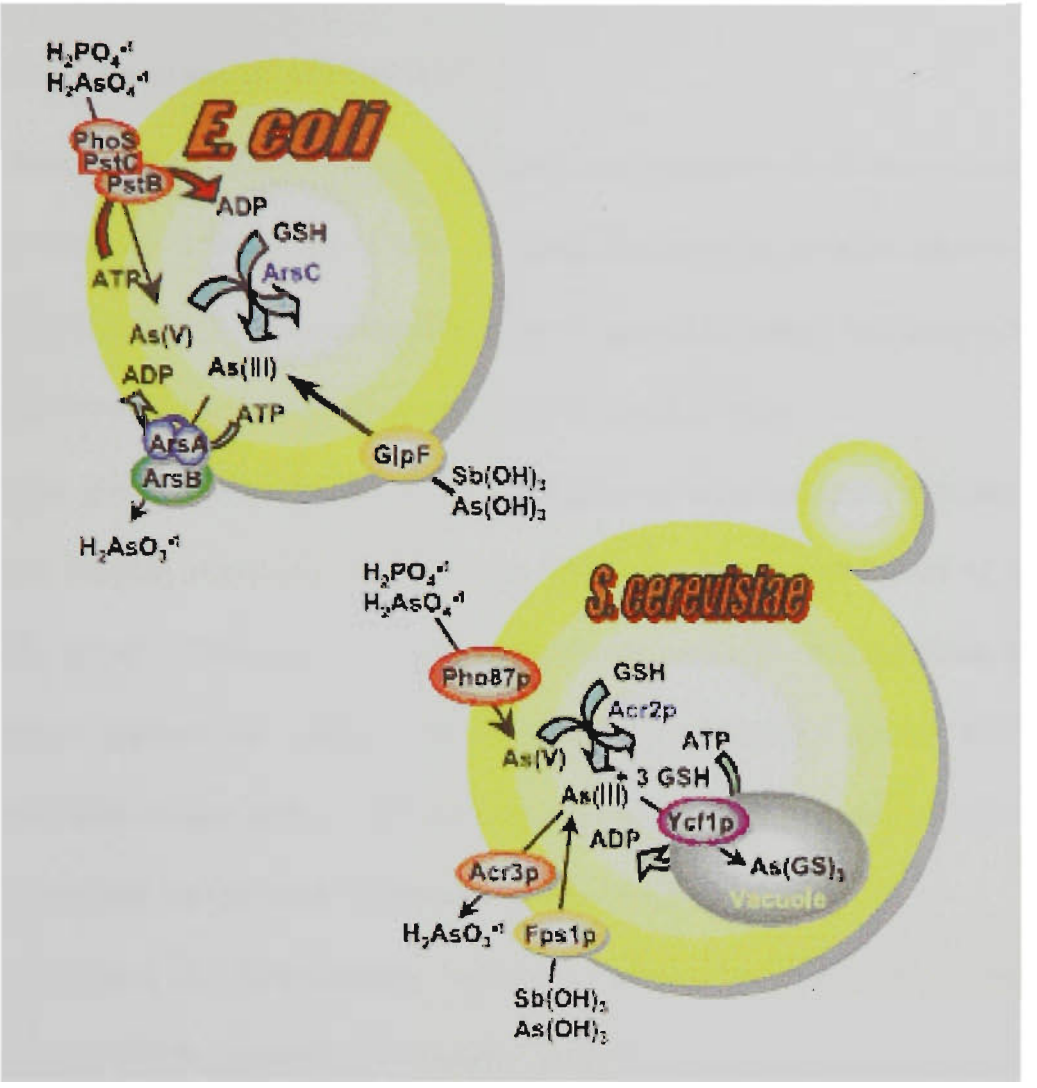


Figure 1.12: Efflux mechanism underlying transport of arsenite and arsenate in *E. coli* and *S. cerevisiae*. ArsC and Acr2p, the cytoplasmic arsenate reductases. ArsB and Acr3p, the potential-driven membrane arsenite efflux proteins (in bacteria coupled with ArsA to form an ATPase). GlpF and Fps1p, the glycerol transport proteins that also transport arsenite. Pit and Pho87p, the potential-coupled phosphate (and arsenate) uptake transporters. PstB, PstC, PhoS, the three-component Pst ATP-coupled phosphate (and arsenate) uptake system. Ycf1p, the As(III)-3 GSH adduct carrier that transports the adduct complex into the yeast cell vacuole compartment, functioning as an ATPase. GSH, glutathione. (Adapted from Rosen, 2002)

1.7.4 ArsA: Oxyanion-stimulated ATPase

Resistance to toxic oxyanions of arsenic and antimony in *Escherichia coli* results from active efflux of these anions out of the cell. Extrusion is an active process mediated by an ATP-dependent pump composed of two types of subunits, the integral membrane ArsB protein and the catalytic ArsA subunit (Dey et. al., 1994).

The ArsA ATPase is the catalytic subunit of a novel arsenite pump with two nucleotide binding consensus sequences in the N and C terminal halves of the protein (Walmsley et. al., 1999). ArsA is structurally similar to NifH, the Fe-protein of bacterial nitrogenases (Zhou et. al., 2000). The *ars* operon of conjugative plasmid R773 encodes an ATP-coupled afflux pump which actively transports arsenicals and antimonials out of the cell (essential for survival) (Walmsley et. al., 1999).

The ArsA 63 kDa catalytic subunit is arranged into two homologous halves (Figure 1.14) the N-terminal A1 domain (residues 1-282) and C-terminal A2 domain (residues 321-583), which are connected by a flexible linker (residues 283-320) (Ramaswamy and Kaur, 1998; Zhou et. al., 2000). Each domain contains the consensus sequence for phosphate binding loop (P-loop) of an ATP binding motif GKGGVGKTS (Walmsley et. al., 1999; Rensing et. al., 1999). Site-directed mutagenesis of these sequences indicates that both nucleotide binding domains (NBDs) are required for enzymatic activity and it is likely that the two sites interact to produce a catalytically competent protein (Karkaria et. al., 1990; Kaur and Rosen, 1992; Li et. al., 1996). Additionally, both A1 and A2 domains have a 12-residue consensus sequence, termed as DTAP, which may act as a signal transduction domain that relays the communication

between the ATP-binding site and the allosteric metalloid-binding site (Zhou and Rosen, 1997).

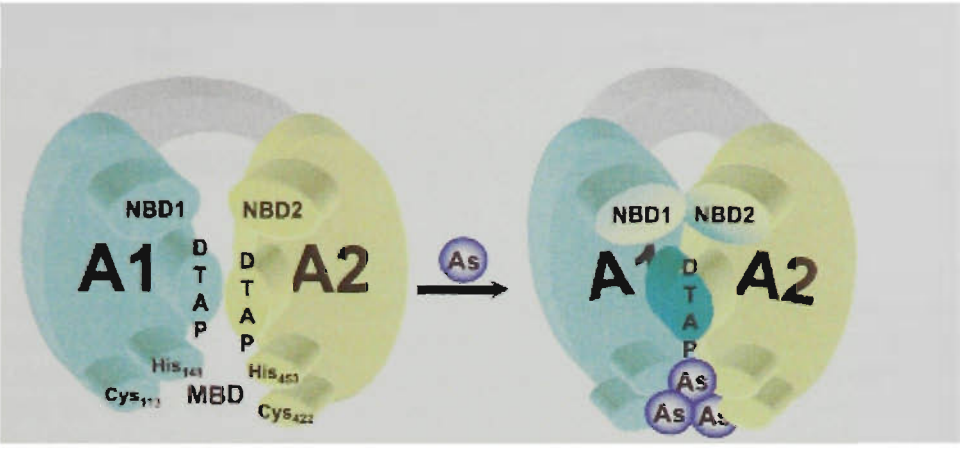


Figure 1.13: Model for allosteric activation of ArsA by AsIII. In the absence of AsIII activator, the A1 and A2 halves of ArsA are loosely interacting, with only the A1 NBD exhibiting a basal rate of ATP hydrolysis (unisite catalysis). While binding of SbIII or AsIII to the MBD, thus brings the two halves of ArsA together, completing the NBDs and accelerating catalysis (multisite catalysis). The NBD, MBD and DTAP domains are indicated. (Rosen, 2002)

The interface between A1 and A2, but at the opposite end of the molecule with respect to the NBDs, is a site in which three AsIII or SbIII ions bind (Zhou et. al., 2000). Three cysteines (Cys113, Cys172 and Cys422), two histidines (His148, His453) and serine (Ser420) ligate the ions. Each As/SbIII is coordinated by one residue from A1 and one residue from A2 thus binding of each of the three metalloids tightens the interaction between A1 and A2 possibly triggering ATP hydrolysis (Zhou et. al., 2000; Bhattacharjee and Rosen, 1996). The catalytic and allosteric sites are located distant from each other in the enzyme, requiring a mechanism to transmit the information of metal occupancy to the ATP hydrolysis site. This function is performed by a 12 residue sequence in A1 and A2 which contains the consensus D142/447TAPTGH148/453, corresponding to the switch II region of other nucleotide-binding proteins (Zhou et. al., 2000) (Figure 1.14).

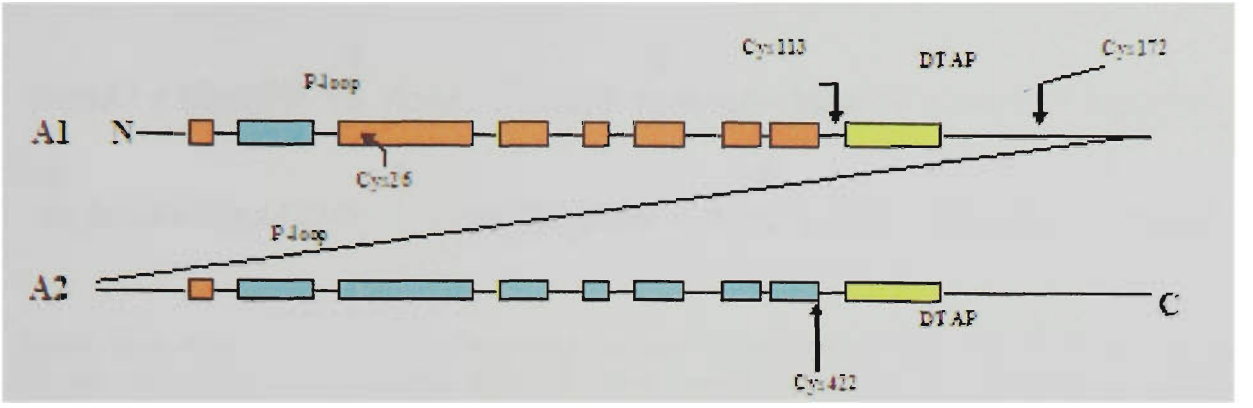


Figure 1.14. Domains in the ArsA ATPase. ATPase is represented linearly, with the homologous A1 and A2 halves aligned and connected by a 25-residue linker sequence. The boxes in the two halves indicate the regions of greatest sequence similarity. The red boxes indicate the location of the P-loops of the NBDs and the blue boxes indicate the DTAP signal transduction domains. The location of the four cysteine residues in ArsA is indicated; Cys113, Cys172 and Cys422 comprise the allosteric metal-binding domain. (Bhattacharjee et. al., 2000)

The reaction mechanism of ArsA ATPase as shown by Rosen et. al. (1999) demonstrated that the interaction between the two NBDs which is linked to the movement of the DTAP domain is facilitated by binding to an allosteric effector. Transient kinetic experiments have shown that the mechanism of the ATPase reaction catalyzed by ArsA is a multistep process (Figure 1.15). In this scheme, ArsA undergoes several conformational changes during binding of substrate (MgATP) and release of products (MgADP and P_i). In the steady state, the $ArsA^7$ conformation of the enzyme builds up, with a slow isomerization back to the $ArsA^1$ form. Since, binding of substrate and release of products are all faster than this isomerization, it becomes the rate-limiting step in the reaction in the absence of activator. Binding of AsIII or SbIII at the allosteric site increases the rate of $ArsA^7$ conversion to $ArsA^1$, accelerating catalysis. The most notable feature of this mechanism is the conformational change to its original state which is also the rate limiting step of the reaction, subsequent to hydrolysis and product release (Step 7).

$$\text{ArsA}^1 + \text{MgATP} \overset{1}{\leftrightarrows} \text{ArsA}^2\text{-MgATP} \overset{2}{\leftrightarrows} \text{ArsA}^3\text{-MgATP} \overset{3}{\leftrightarrows} \text{ArsA}^4\text{-MgATP}$$

$$\overset{4}{\leftrightarrows} \text{ArsA}^5\text{-MgADP}\cdot P_i \overset{5}{\leftrightarrows} \text{ArsA}^6\text{-MgADP} + P_i \overset{6}{\leftrightarrows} \text{ArsA}^7 + \text{MgADP} \overset{7}{\leftrightarrows} \text{ArsA}^1$$

Figure 1.15: Reaction mechanism of the ArsA ATPase. A minimal multistep kinetic scheme for the reaction catalyzed by ArsA has been proposed, where ArsA undergoes several conformational changes during binding of substrate (MgATP) and release of products (MgADP and P_i) (Adapted from Rensing et. al., 1999)

1.7.5 ArsC: Arsenate reductase

Resistance to arsenate is conferred by the *ars* operon. The product of the *arsC* gene, arsenate reductase (ArsC) is required to catalyze the reduction of arsenate to arsenite prior to extrusion (Martin et. al., 2001; Demel et. al., 2004). ArsC is a 141 residue protein (14.8 kDa) with a redox active cysteine residue in the active site (Gladysheva et. al., 1994; Liu and Rosen, 1997).

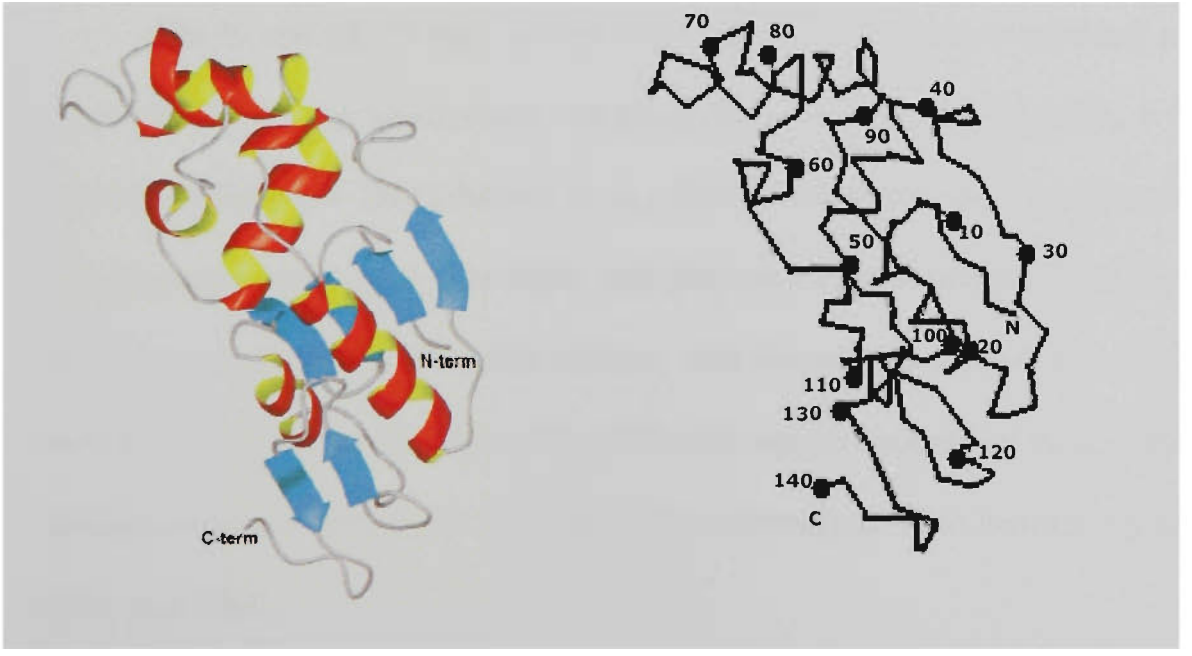


Figure 1.16: Ribbon structure of ArsC. (Adapted from Martin et. al., 2001)

Figure 1.16 shows the relative location of residues 3-140, including the 4 residues that are known to be required for arsenate resistance namely, Cys-12, Arg-60, Arg-94 and Arg107 (Martin et. al., 2001).

Micro organisms use arsenate as a terminal electron acceptor in anaerobic respiration (Saltikov et. al., 2003). Arsenate reductases (ArsC) catalyze the reduction of arsenate to arsenite and participate in the arsenic detoxification (Messens et. al., 2002; Guo et. al., 2005). Arsenate enters the cell via the phosphate transport system (Pst) (Demel et. al., 2004). Three families of arsenate reductases have evolved independently and have been identified (Mukhopadhyay et. al., 2002), the *E. coli* plasmid R773 ArsC arsenic modifying enzyme (Demel et. al., 2004), ArsC from *Staphylococcus aureus* pI258 (Ji and Silver, 1992a) and the Acr2p arsenate reductase from eukaryote *Saccharomyces cerevisiae* represents the third family (Mukhopadhyay et. al., 2000).

The *E. coli* pR773 ArsC is well characterized and obtains its reducing equivalent from glutathione and glutaredoxin (Mukhopadhyay et. al., 2000) $H_3ASO_4 + 2GSH \rightarrow H_3ASO_3 + GS\text{-}SG + 2H_2O$ (Martin et. al., 2001), likewise the yeast enzyme obtains its reducing equivalents from glutathione and glutaredoxin. In *S. aureus* pI258 the oxidized ArsC is regenerated via a coupled reaction with thioredoxin, thioredoxin reductase and NADPH (Messens et. al., 2002). The pI258 of *S. aureus* encodes an ArsC with catalytic residues similar to pR773 but the ArsC utilizes thioredoxin (Trx) instead of glutaredoxin (Ji et. al., 1994).

pI258 arsenate reductase is related to the low molecular weight phosphotyrosine phosphatases (Gladysheva et. al., 1994; Demel et. al., 2004) which includes a p-loop with the characteristic CX_5R sequence motif flanked by a β-strand and an α-helix (Messens et.

al., 2002). Acr2p from *S. cerevisiae* also contains this motif (HCX_5R phosphatase) but is homologous to the human cell cycle control phosphatase Cdc25a (Mukhopadhyay et. al., 2003; Martin et. al., 2001). In contrast, the R773 ArsC does not exhibit phosphatase activity and has known paralogs (Demel et. al., 2004; Messens et. al., 2002). *E. coli* R773 ArsC has a distinct HX_3CX_5R motif and partially resembles crambin and glutaredoxin (Martin et. al., 2001). ArsC of R773 has only one cystiene residue (Cys12) while pI258 utilizes three cystiene residues (Cys10, Cys82 and Cys89) in arsenate reduction (Mukhopadhyay et. al., 2002).

The mechanism of *E. coli* ArsC involves five steps (Figure 1.17); step 1 involves nucleophillic attack by Cys12 on an arsenate non-covalently bound with the release of OH^-. In step 2 this intermediate is glutathionated, with the release of H_2O. Step 3 involves binding of Grx (Glutaredoxin) and arsenate is reduced to arsenite, resulting in dihydroxy monothiol arsenite intermediate and subsequent release of GrsX-SG and GSH. In step 4, monohydroxy intermediate with a positively charge arsenic is formed (Martin et. al., 2001). Finally in step 5, addition of OH^- releases free arsenite. (Demel et. al., 2004).

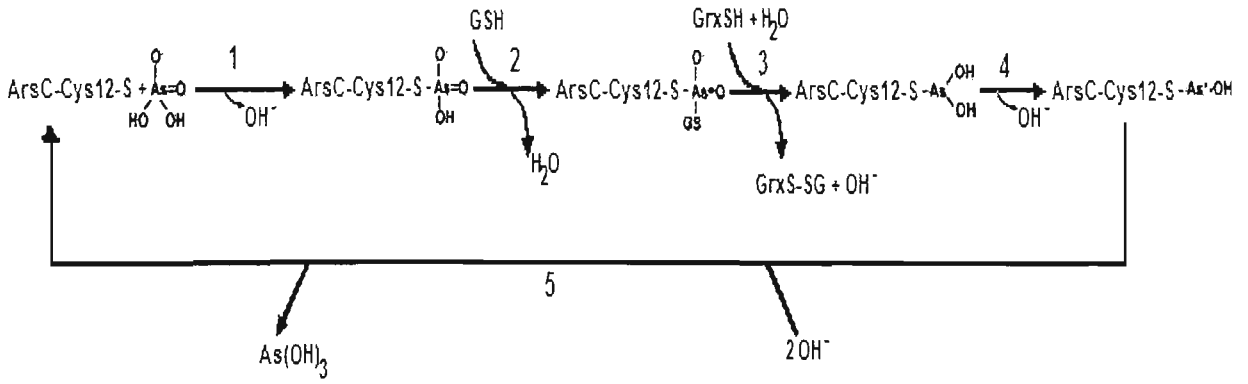


Figure 1.17: Reaction mechanism of the R773 ArsC arsenate reductase. (Demel et. al., 2004)

The *S. aureus* pI258 ArsC has been extensively studied, indicating three redox cysteine residues (Cys10, Cys82 and Cys89) are critical for arsenate reduction (Guo et. al., 2005). A disulfide bridge (Cys82-Cys89) is formed after a single catalytic reaction cycle, converting the enzyme into the inactive form. ArsC is subsequently regenerated by thioredoxin which converts the enzyme into the reduced form (Figure 1.18) (Guo et. al., 2005).

Similarly, the *arsC* gene of *B. subtilis* skin element shares 64% identity with *S. aureus arsC*. The conserved CX_5R anion binding motif containing Cys10 known as p-loop is proposed to be the arsenate binding site. Four Cys residues (Cys10, Cys15, Cys82 and Cys89) are conserved (Guo et. al., 2005).

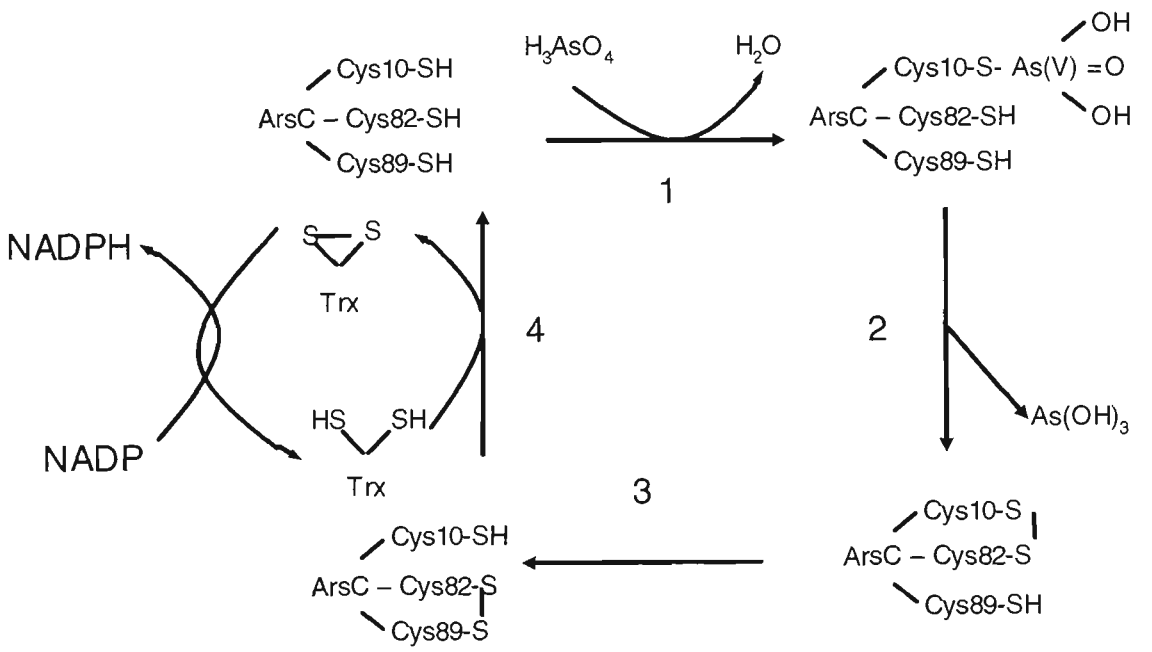


Figure 1.18: Proposed pathway of arsenate reduction by thioredoxin coupled arsenate reductase from *S. aureus* pI258. H_3AsO_4: arsenate; $As(OH)_3$: arsenite; Grx-SH: reduced glutathione; GrxS-SG: S-S bridged oxidized dimer of glutathione; HS-Trx-SH: reduced thioredoxin; Trx-S-S: oxidized thioredoxin. (Adapted from Mukhopadhyay et. al., 2002)

1.8 ArsR-ArsD Metalloregulatory Circuit Model

In *E. coli* pR773, the ArsR and ArsD pair regulates different levels of expression of the *arsRDABC* operon. In the absence of arsenite, a basal level of synthesis produces a small amount of ArsR that binds to the operator of the *ars* operon. The position of the operator occupies part of promoter. Therefore, the ArsR blocks the RNA polymerase binding to the promoter of *ars* operon, suppressing transcription. If a low concentration of arsenite is present, the cysteine thiolates of ArsR bind to arsenite, forming the ArsR-arsenite complex. The combination will change the conformation of ArsR, inducing ArsR to dissociate from the DNA strand and start transcription. Most ArsR and ArsD at this moment will bind to arsenite to form complexes. Thus, insufficient ArsR and ArsD molecules are available to bind to the operator and stop transcription. Due to the higher affinity of arsenite for ArsR than for ArsD, the ArsR protein will bind to arsenite before ArsD (Chen and Rosen, 1997). When the concentration of free ArsD increases sufficiently, the ArsD protein will bind to the operator to stop transcription; otherwise, the high concentration of ArsB will damage the membrane and harm the cell. If a high concentration of arsenite is present, the ArsD-arsenite complex dissociates from the DNA, resulting in the further increase of *ars* transcription (Figure 1.19). In this mechanism, the action of the two repressors forms a homeostatic regulatory circuit that maintains the level of *ars* expression within a narrow range, with ArsR controlling basal level of expression and ArsD controlling maximal expression.

Figure 1.19: Schematic illustration of ArsR and ArsD regulating the expression of *ars* operon. **(a)** In a normal circumstance, ArsR binds to the operator site of the *ars* operon to repress the transcription. At low concentrations of arsenite, ArsR binds to arsenite and changes the conformation. ArsR-arsenite complex dissociates from the DNA. RNA polymerase binds to the promoter, resulting in transcription of the *ars* operon. **(b)** When the concentration of ArsD increases to a higher level, ArsD binds to the operator of the *ars* operon and inhibits the transcription. At high concentrations of arsenite, ArsD binds to arsenite and dissociates from the DNA. RNA polymerase binds to the promoter and transcription starts again.

1.9 Background and Aims of This Thesis

Micro organisms have evolved a number of mechanisms to cope with arsenic toxicity, and some organisms even benefit from the presence of arsenic. In Australia, cattle ticks (*Boophilus microplus*) jeopardize the beef and dairy industries and a cattle-dipping program for eradication and containment using arsenic as the major pesticide commenced in the early 1900s, Fifty years later, arsenic was replaced by DDT [2,2-bis(*p*-chlorophenyl)-1,1,1-trichloroethane] which was used until 1962 (Edvantoro et. al.,

2004). According to NSW department of primary industries (DPI) latest review in 2006, much less persistent tickicides have been used to dip cattle in the active sites that are still required by cattle tick program while, majority of the sites are on the list for decommissioning.

In one region of Australia alone (North Eastern New South Wales and Southern Queensland), at least 1600 cattle dips were constructed (Edvantoro et. al., 2003). Disposal of used pesticides, by burying in the vicinity of the dip bath, have left a legacy of high level, localized contaminated sites. A recent study of soil contaminants in 12 dip sites found that 80% of soil samples had concentrations between 1 and 1000 mg arsenic.kg^{-1} soil while the remaining 20% ranged from 1000 to 4540 mg arsenic.kg^{-1}, all co-contaminated with DDT (Van Zwieten et. al., 2003). These sites present both a challenge and an opportunity for developing new techniques using micro organisms and/or plants for future remediation.

Microbes in the soil are important in providing nutrients to plant roots. Soil bacteria degrade organic compounds and modify the inorganic products. Soil fungi provide a large surface area for scavenging soil-bound nutrients such as inorganic phosphate and, through ectomycorrhizal and endomycorrhizal associations with roots, transport these to the plant. Toxic compounds in soils are often modified by microbes (Van Zwieten et. al., 2003), but many such toxins may hinder growth of soil microbes and impair their ability to promote plant growth. Additionally, soil fungi associated with roots have the potential to either increase or ameliorate the uptake of inorganic contaminants by plants. Consequently, mycorrhizal fungi in polluted soils are crucial in maintaining diverse populations of indigenous vegetation and act as a barrier to the

uptake of toxic heavy metals by plants (Leyval et. al., 1997). Sharples et. al. (2000) presented evidence that the ericoid mycorrhizal fungus *Hymenoscyphus ericae* acts as a filter to maintain low arsenic uptake rates by roots of the plant *Calluna vulgaris* when growing in arsenic-contaminated soil. In a study of evolved arsenate resistance in cultivars of the grass *Holcus lanatus*, Gonzalez-Chavez et. al. (2002) found that colonization by the arbuscular-mycorrhizal fungus *Glomus* suppressed high-affinity arsenate and phosphate transport into the roots.

Several recent publications have addressed the microbial population in cattle dip-site soils and their interaction with the contaminants. The co-contamination of these soils with arsenic and DDT has been shown to have distinct effects on the soil microbial populations. Edvantoro et. al. (2003) showed that in contaminated dip-site soils, fungal counts, microbial biomass, carbon and respiration rates were significantly lower compared to uncontaminated controls. However, the bacterial population was not significantly different. They attributed the effects on the fungal population to arsenic rather than DDT. Van Zwieten et. al. (2003) measured the concentrations of DDT and two products of its natural slow degradation in dip-site soils, for which soil bacteria are responsible. They found that, in soils with high arsenic concentrations, the breakdown of DDT was impaired. In a study of arsenic loss from similar soils, Edvantoro et. al. (2004) made the symmetrical finding that volatilization of arsenic, attributed to microbial activity, was inhibited by the DDT contaminant.

While some recent advances have been made in understanding the effects of arsenic/DDT co-contamination on microbial populations of soils in dip sites, little is known about the identity of these organisms or their interaction with plants growing on

these sites. To study the arsenic resistance mechanism and search for solutions to cleaning up the sites, a plant and microbial survey on two of the sites were carried out by former members of our group in collaboration with the NSW department of agriculture (Morris, 1998; Chopra et. al., 2007) with the following findings.

1. Nearly 50 plant species, mostly weeds, were recorded in the sites. No difference in comparison to adjacent areas as noted

2. Six dominant fungal strains (*Aspergillus, Penicillium, Curvularia, Fusarium, Cephalosporium* and *Trichoderma*), two soil-borne pathogens, *Rhizoctonia* and *Pythium* and two mycorrhizae strains were observed

3. Five soil bacterial strains designated as CDB1, CDB2, CDB3, CDB4 and CDB5 were isolated and tentatively identified.

In addition, two *ars* gene clusters have been cloned from the CDB3 strain (Luo, 2006).

The practical direction of this research is to develop novel method(s) in solving the arsenic contamination problem. While there has been no efficient means to do it, bioremediation has been thought to be a potential application. Heavy metal accumulation by plants is a well-characterized technology and plants which can be supplemented with additional genetic information of bacterial origin are a useful process for bioremediation of contaminated soils (Pieper and Reineke, 2000). Our group (Chopra et. al., 2007) presented a glasshouse trial demonstration using *Agrostis tenuis* (bentgrass) growing in soil with high total, but low soluble arsenic content, that microbial associations result in enhanced arsenic accumulation, without inhibition of plant growth. It was observed that

the fungal populations were more than one fold lower in contaminated dip-site soil samples than the control samples whilst the bacterial populations were higher in the contaminated dip-site rhizosphere soil samples than in the control samples (Zhang R, personnel communication). Edvantoro et. al. (2003) also reported increased bacterial populations in two dip sites studied. In some situations, high concentrations of soluble inorganic arsenic in soils have been observed to stimulate the development of bacteria and fungi and increase bacterial activity, yet in other situations to decrease the activity (McLaughlin et. al., 2000). The concentrations and forms of arsenic available to the soil microbial community and the interacting effects of co-contaminants are factors that complicate the response to arsenic contamination. Bizily et. al. (1999), have reported the transgenic plant *A. thaliana* engineered to express a modified organomercurial lyase grew rapidly on different concentration of highly toxic organomercurials. Further, Dhanker et. al. (2002) and Ute, 2005 have reported engineered transgenic *Arabidopsis* plants over expressing the *E. coli* ArsC with an additional transgene under a constitutive promoter results in more tolerance to AsV and Cd (Cadmium) accumulation, suggesting, engineering plants with bacterial genes seems to be a positive approach in bioremediation. Apparently, a good understanding of the bacterial arsenic resistance mechanism holds the key to our objective.

The aims of this thesis were to further identify the five CDB strains based on morphological, 16S rDNA sequencing and phylogeny, to investigate their resistance to arsenical and antimonial compounds, to carry out bioinformatic analysis of the CDB3 Ars proteins to other well known bacteria, to functionally characterize the *arsA* and *arsD*

genes of CDB3 *ars* cluster 1 and finally to investigate the expression and regulation of this cluster.

Chapter 2

Characterization and Identification of Arsenic Resistant Bacterial Strains Isolated from Cattle Dip Sites

2.1 Introduction

Arsenic and DDT were extensively used in Australia during early 1900s for cattle ticks, which have threatened the Australian beef and dairy industry. As a result, many cattle dip sites have been left heavily contaminated with arsenic (McLaren et. al., 1998). To date numerous phylogenetically distinct bacterial species have been isolated and identified from As contaminated soils, ground waters and surface waters. The contamination of these sites has become a major concern to public health (Saltikov et. al., 2003). As previously discussed in Chapter 1, the dipping of cattle in Australia to control cattle ticks has left 1000-1600 cattle tick dip sites heavily contaminated with arsenic (Smith et. al., 2003; Van Zwieten et. al., 2003). More then 1600 cattle tick dip sites were constructed in North Eastern New South Wales and along the Queensland border.

The five soil bacterial strains isolated by former members of our group from cattle dip sites in Northern NSW were previously referred to as As3, As4, As8, As9 and As10. Preliminary investigation was carried out to identify these bacterial strains based on fatty acid profile (Table 2.1) and incomplete 16S rDNA sequencing.

Table 2.1: Similarity Index of bacterial fatty acids profile (Chopra et. al., 2007)

Bacterial Strains	Similarity Index	Match bacteria
As3	0.767* 0.686 0.603	*Arthrobacter ilicis** *Paenibacillus polymyxa* *Kocuria varians*
As4	0.430 0.356	*Ochrobactrum anthropi* *Phyllobacterium myrisnacearum*
As8	0.467 0.367 0.324	*Bacillus cereus* GC subgroup A *Bacillus mycoides* GC subgroup A *Bacillus thuringlensis kurstakii*
As9	0 .549*	*Bacillus mycoides* GC subgroup A*
As10	0.686*	*Serratia marcescens* GC subgroup A*

Similarity Index is a value which expresses the FAME (fatty acid methyl ester) similarity between the unknown isolate and the database match. * indicates an excellent FAME analysis match

The fatty acid profile results showed very good matches for As3; As9 and As10 to known bacteria. Based on incomplete 16S rDNA sequencing the five bacterial strains were tentatively identified as As3- *Arthrobacter* sp., As4-*Sinorhizobium* sp., As8 and As9 as *Bacillus* spp and As10- *Serratia* sp. (Zhang R, personnel communication). Meanwhile, some preliminary arsenic resistance assays were carried out which demonstrated that these strains were highly resistant to arsenic in laboratory conditions (Xu, unpublished).Since the data obtained by these previous studies were not complete*, further study was carried out as presented in this Chapter to identify and characterize the five bacterial strains isolated from arsenic contaminated soil; to test the antibiotic resistance profile of these isolated strains and finally to elucidate the level of resistance offered by these strains to AsV, AsIII and SbIII. From here on the names of the bacterial strains were changed from As to CDB which refers to Cattle Dip Bacteria and numbered from 1 to 5 for easy identification.

* I would like to acknowledge the following previous work or assistance by antibiotic and arsenic resistances by Z. Xu; Previous incomplete 16S rDNA sequencing by Ed Moore; and assistance in re-sequencing the complete 16S rDNA by Ying Yang and Xi Luo.

2.2 Materials and Methods

2.2.1 Bacterial strains

The bacterial strains used in this study are listed in Table 2.2. *E. coli* strains were grown in LB medium at 37°C (Sambrook et. al., 1989) with ampicillin (Amp) (100 μg/mL) or kanamycin (Km) (50 μg/mL) added as required. All CDB strains were grown in LB medium at 30°C with different antibiotics due to different antibiotic resistance profiles. CDB2 and CDB5 strains were grown with any of the three antibiotics Km (50 μg/mL), Amp (100 μg/mL) and Streptomycin (Strep) (50 μg/mL) respectively. CDB1 with Km (50 μg/mL), and CDB3 and CDB4 were cultured on LB medium alone as they exhibited sensitivity to all tested antibiotics.

Table 2.2: Bacterial strains used in the study of Chapter 2

Strains and plasmid	Genotype or description	Reference or Source
E. coli DH5α	Φ80dlacZΔM15, recA1, endA1, gyrA96, thi-1, hsdR17 (r_K-,m_K+), supE44, relA1, deoR, Δ (lacZYA-argF) U169	Stratagene
E. coli AW3110	K-12 F⁻IN (*rrn*D-*rrn*E) Δ*ars*::*cam* (Cm^r, the chromosomal *arsRBC* deleted)	Gift from Dr. B.P. Rosen. (Carlin et. al., 1995)
CDB1	Km^r	This study
CDB2	Amp^r, Cm^r, $Strep^r$, Km^r	This study
CDB3		This study
CDB4		This study
CDB5	Amp^r, $Strep^r$, Km^r	This study

2.2.2 Identification of bacteria

Isolated pure bacterial strains were subjected to Gram stain, morphological observation and 16S ribosomal RNA gene sequencing.

2.2.3 16S rDNA Sequencing

2.2.3.1 Nucleic acid preparation

Overnight culture from a single colony in LB with antibiotic selection was diluted 100 times with the same medium and again incubated at 30°C with shaking until the O.D. 600_{nm} reached 0.4–0.5. Cultures were centrifuged at 5000 g for 10 min, bulk supernatant decanted and the cells resuspended by vortexing in the residual supernatant. Equal volume of 0.5% SDS, 0.1 M NaOH was added and vortexed to lyse the cells. An equal volume of TE saturated phenol/chloroform/isoamyl alcohol (25:24:1) was added and mixed thoroughly. After centrifugation at 10,000 g for 10 min, the upper aqueous layer was transferred to a new tube to which a 10% volume of 3 M Na acetate pH 5.4 was added and mixed followed by 2.5 volumes of 95% ethanol. DNA was spooled out using a sealed Pasteur pipette, washed in 70% ethanol, dissolved in sterile d. H_2O and stored at –20 °C for future use.

2.2.3.2 PCR amplification and purification of the product

Primers used for amplification

Both sets of primers were designed from fD1 and rP2 (Weisburg et. al., 1991), which are capable of amplifying 16S from a wide variety of bacterial taxa. Primers used for the amplification of 16S rDNA from all the strains and primers were designed for sequencing the middle part of 16S rDNA from CDB2 and CDB4 are shown in Table 2.3.

Table 2.3: List of primers used for PCR amplification and sequencing of 16S rDNA from different strains.

Primers	Sequences 5' to 3'	Reference
16S-Forward (fD1)	AGAGTTTGATCCTGGCTCAG	Weisburg et. al., 1991
16S-Reverse (rP2)	ACGGCTACCTTGTTACGACTT	
CDB1 16S-R1	TGGGTAAGGTTCTGCGCG	Designed by current study
CDB2 16S-R1	ACCACATGCTCCACCGCT	
CDB3 16S-F1	CCAGAGCTCAACTCTGGA	
CDB4 16S-R1	CCAGAGCTCAACTCTGGA	
CDB5 16S-R1	AGTTCTCTGGATGTCAAG	

Abbreviation: f, forward; r, reverse; D, distal; P, proximal; R1 and F1, reverse and forward primers for the middle part.

The reaction mix for each PCR amplification consisted of 10 μL PCR mix (Promega), 1 μL of each primer (20 μM), 2 μL of 1:10 diluted DNA template (CDB2, 1-3 μg/μL), and nuclease-free water to a total of 20 μL. The primers were synthesized by Sigma, and were delivered desiccated. These were diluted with TE buffer (1×) to form stock solutions (200 μM), which were further diluted to give a working concentration of 20 μM.

The PCR amplification was performed with a Gene Amp PCR system 9600 (Perkin Elmer). The program consisted of a preheating step (94°C for 3 min), followed by 30 cycles of 94°C, 30 sec. → 62°C, 30 sec. → 72°C, 1 min, the samples were then extended at 72°C for 10 min, and finally held at room temperature. 5 μL of the PCR products were loaded onto a 1.5% TAE agarose gel for checking the quality and quantity.

Purification of PCR products

PCR products were purified by G-50 sephadex column chromatography. G-50 sephadex columns were prepared by adding sephadex into a 3 mL syringe slowly avoiding the disturbed bubbles, allowing flow through for 3-4 times followed by

centrifugation at 400 g for 4 min. More sephadex was added to the column and centrifuged for another 4 min until a 1ml column formed. 60 μL of the diluted PCR samples were then applied to the column and centrifuged at 400 g for 4 min. Purified elutes were collected in a sterile eppendorf tube.

2.2.3.3 Sequencing methods

The cycle sequencing reaction mixture was set up, each containing 0.5 μL of 'Big Dye' reaction mix (Perkin-Elmer), 2 μL of purified PCR products, 2 μL of 5X buffer and 1ul primer (~3.2 pmol, 16S-F, 16S-R and 16S-R1, Sigma-Aldrich, see table 2.2). The time program used for the cycle sequencing consisted of 30 cycles of 96^0C 10 sec, 62^0C 5 sec and 60^0C 4 min. Ethanol precipitation was used to purify the extension products after cycle sequencing. After transferring all of the extension products into micro centrifuge tubes, 2 μL of 3M Na Acetate (pH 5.4) and 50 μL of 95% ethanol were added and mixed. These were left at room temperature for 10 minutes before centrifugation (20 min at 13,000 rpm). The supernatant was removed by aspiration and the DNA pellet was washed with 250 μL of 70% ethanol followed by centrifuging 20 min at 13,000 rpm. The pellet was aspirated and air-dried before sequencing. Sequencing analysis was performed using the ABI PRISM^{TM} 337 DNA Sequencer. The sequencing of the full length of the 16S rDNA of four strains were divided into three parts and assembled using the auto assembler. The primers for sequencing the middle part were designed according to the sequences of the forward and reverse sequences at both ends (Table 2.3).

2.2.3.4 Sequence analysis

The PCR product was sequenced in both directions and the sequences obtained were checked using Chromas (version 1.45). Sequences from different regions (forward,

middle and reverse) were assembled and multiple alignments were done using Bestfit (Felsenstein, 1989) and ClustalW (Felsenstein, 1989). The 16S rDNA sequences were compared with sequences deposited in the Genbank and EMBL. Searching for homologous sequences in data bank was carried out using blastn analysis (Altschul et. al., 1990; http://www.ncbi.nlm.nih.gov/blastn). The sequences used for the multiple alignments were retrieved from Genbank and EMBL. The relevant accession numbers are shown in Table 2.5.

2.2.4 Phylogenetic Analysis

The phylogenetic trees were constructed by the multiple alignments of all the strains and the related species using Neighbor-Joining/UPGMA method (version 3.573c) in PHYLIP package (Felsenstein, 1989). Bootstrap confidence values from 1000 replications of each sequence were generated and the majority rule consensus tree was derived using SEQBOOT and CONSENSUS programs in PHYLIP.

2.2.5 Taxonomy tree

In order to clarify the relationship between different species, a taxonomy tree involving related species were acquired from NCBI (www.ncbi.nlm.gov/taxonomy/ Browser / wwwtax.cgi).

2.2.6 Measurement of Minimum Inhibitory Concentration (MIC)

Overnight cultures of bacteria were spread onto LB agar plates containing different concentrations of arsenate (AsV) and arsenite (AsIII) and incubated at 30°C for three days. The minimum concentrations at which bacteria failed to grow were designated as MIC.

2.2.7 Growth inhibition

1/50 of overnight culture of bacteria was inoculated into 10 mL LB liquid medium, containing increasing amounts of arsenate, arsenite and antimonite. Cells were grown at 30°C for soil bacteria and at 37°C for *E. coli* for 4-7 hours until control (without arsenic) grew to O.D. 600_{nm} = 0.6-0.7. The assays were carried out in triplicates (technical) and error bars represent standard deviation. Each assay was repeated at least three times.

2.3 Results

2.3.1 Morphology and Gram stain of the CDB strains

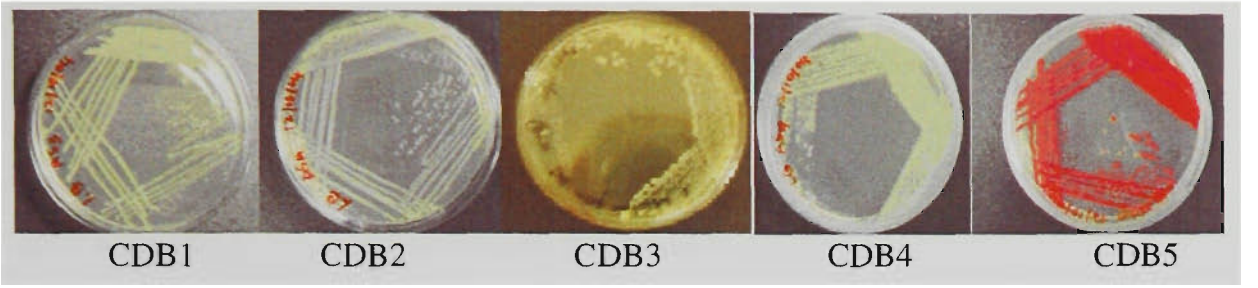


Figure 2.1: Morphologically different bacterial species isolated from contaminated cattle dip site soil.

The five strains isolated from highly arsenic contaminated cattle dip site soils were each distinctive in morphology on LB plate (Figure 2.1, Table 2.4).

Table 2.4: Colony morphological observation of different bacterial strains isolated from cattle dip site

Bacterial Strain	Colony Morphology					
	Colour	**Shape**	**Surface**	**Elevation**	**Edges**	**Opacity**
CDB1	Dark yellowish	Circular	Smooth	Flat	Entire	Transparent
CDB2	Light yellowish	Circular	Smooth	Effuse	Entire	Transparent
CDB3	White	Irregular	Smooth	Flat	Entire	Transparent
CDB4	White	Irregular	Rough	Flat	Erose/Lobate	Opaque
CDB5	Red pigmentation	Circular	Smooth	Low convex	Entire	Translucent

Smears prepared from the bacterial colonies were examined by Gram staining and were identified as shown in Figure 2.2 as Gram +ve and Gram –ve bacteria. CDB3 and CDB4 were Gram +ve rods arranged single or in chains, CDB2 was Gram –ve cocobacilli, CDB5 was Gram –ve rods and CDB1 was Gram +ve curved rods.

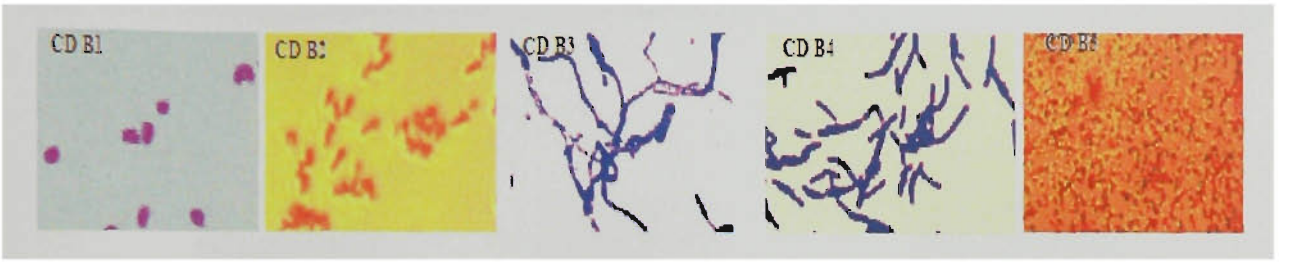


Figure 2.2: Confocal microscopic images of the bacterial strains isolated from cattle dip sites, obtained after performing Gram stain. CDB1 (Gram +ve); CDB2 (Gram -ve); CDB3 (Gram +ve), CDB4 (Gram +ve) and CDB5 (Gram -ve).

2.3.2 Extraction of genomic DNA and PCR amplification products of 16S rDNA

The genomic DNA was extracted from the bacteria and subjected to gel electrophoresis (Figure 2.3) to view quality and estimate quantity.

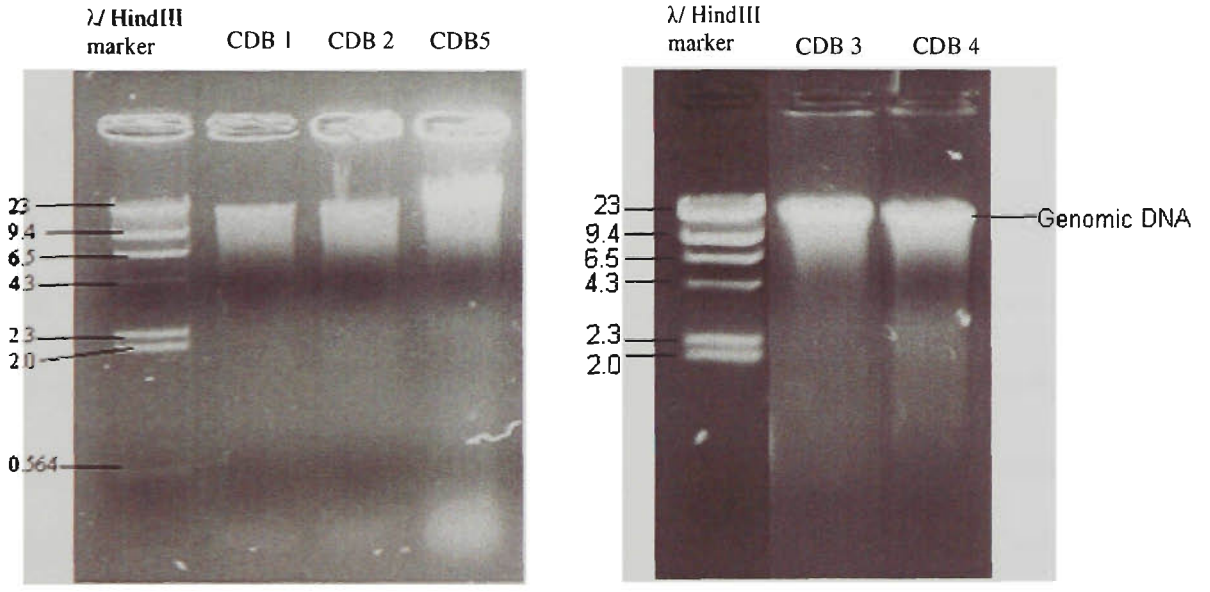


Figure 2.3: Genomic DNA isolated from CDB strains. 0.8% agarose gel along with λ/*Hind*III molecular weight markers. Sizes of markers are indicated in kb on left.

An approximately 1500 bp fragment of DNA was amplified from each CDB1, CDB2, CDB3, CDB4 and CDB5 using the primers 16S-F and 16S-R. The PCR conditions used in this study seem to give a dependable yield of full sized product (Figure 2.4).

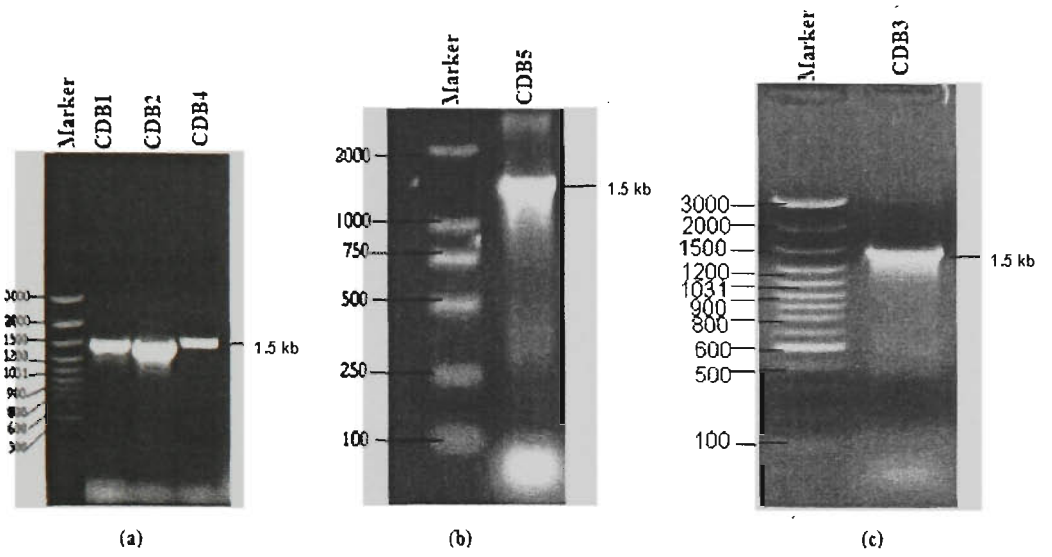


Figure 2.4: PCR amplification of the 16S rDNA of CDB1 (a), CDB2 (a), CDB3 (c), CDB4 (a) and CDB5 (b). The PCR products were run on ethidium bromide stained 1.5% agarose gels. Sizes of markers are indicated in kb on left.

2.3.3 Sequences of 16S rDNA

The amplified 16S rDNA fragments of CDB1, CDB2, CDB3, CDB4 and CDB5 were sequenced from both directions by using forward (fD1) and reserve (rP2) primers. The middle part sequences (unreached by the end primers) were acquired by using newly designed specific primers. Certain parts (towards the ends) were only sequenced in one direction, but the results presented in this portion were all based on 2 or more sequencing reactions with very clear electropherograms. The entire sequencing data obtained was auto assembled using the ABI prism autoassembler and a consensus sequence was generated for each strain which was used to perform blastn. The sequences obtained were 1433 bp (CDB1), 1341 bp (CDB2), 1463 bp (CDB3), 1422 bp (CDB4) and 1449 bp (CDB5) respectively (Figure 2.5).

CDB1:

ATATAGCCGACAGCTCACCGTAGCAAGTCGAACGATGATCCCAGCTTGCTGGGGGATTAGTGGCGAACGGGTGAGTAACACGTGAGTAACCTGCCCTTGACT
CTGGGATAAGCCTGGGAAACTGGGTCTAATACCGGATATGACTCCTCATCGCATGGTGGGGGGTGGAAAGCTTTTGTGGTTTTGGATGGACTCGCGGCCTAT
CAGCTTGTTGGTGGGGTAATGGCCTACCAAGGCGACGACGGGTAGCCGGCCTGAGAGGGTGACCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACG
GGAGGCAGCAGTGGGGAATATTGCACAATGGGCGCAAGCCTGATGCAGCGACGCCGCGTGAGGGATGACGGCCTTCGGGTTGTAAACCTCTTTCAGTAGGGA
AGAAGCGTAAGTGACGGTACCTGCAGAAGAAGCGCCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGCGCAAGCGTTATCCGGAATTATTGGGCG
TAAAGAGCTCGTAGGCGGTTTGTCGCGTCTGCTGTGAAAGACCGGGGCTCAACTCCGGTTCTGCAGTGGGTACGGGCAGACTAGAGTGCAGTAGGGGAGACT
GGAATTCCTGGTGTAGCGGTGAAATGCGCAGATATCAGGAGGAACACCGATGGCGAAGGCAGGTCTCTGGGCTGTAACTGACGCTGAGGAGCGAAAGCATGG
GGAGCGAACAGGATTAGATACCCTGGTAGTCCATGCCGTAAACGTTGGGCACTAGGTGTGGGGGACATTCCACGTTTTCCGCGCCGTAGCTAACGCATTAAG
TGCCCCGCCTGGGGAGTACGGCCGCAAGGCTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGCGGAGCATGCGGATTAATTCGATGCAACGCGAAG
AACCTTACCAAGGCTTGACATGAACCGGAAAGACCTGGAAACAGGTGCCCCGCTTGCGGTCGGTTTACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTG
AGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTCGTTCTATGTTGCCAGCGGTTCGGCCGGGGACTCATAGGAGACTGCCGGGGTCAACTCGGAGGAAG
GTGGGGACGACGTCAAATCATACATGCCCCTTATGTCTTGGGCTTCACGCATGCTACAATGGCCGGTACAAAGGGTTGCGATACTGTGAGGTGGAGCTAATC
CCAAAAAGCCGGTCTCAGTTCGGATTGGGGTCTGCAACTCGACCCCATGAAGTCGGAGTCGCTAGTAATCGCAGATCAGCAACGCTGCGGTGAATACGTTAC
CCGGGCCTTGTACACACCGCCCGTCAAGTCACGAAAGTTGGTAACACCCGAAGCCGGTAGCCTAACCCTTGTGGGGGGAGCTTAGATATAGAGCCGGAACGA
CCGAC

CDB2:

GGTAGCCGGTAAGCGGCAGACGGGTGAGTAACGCGTGGGAATCTACCTTTTGCTACGGAATAACTCAGGGAAACTTGTGCTAATACCGTATGTGCCCTTCGG
GGGAAAGATTTATCGGCAAAGGATGAGCCCGCGTTGGATTAGCTAGTTGGTGGGGTAAAGGCCCACCAAGGCGACGATCCATAGCTGGTCTGAGAGGATGAT
CAGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCCATGCCGCGTGAGT
GATGAAGGCCCTAGGGTTGTAAAGCTCTTTCACCGGTGAAGATAATGACGGTAACCGGAGAAGAAGCCCCGGCTAACTTCGTGCCAGCAGCCGCGGTAATAC
GAAGGGGGCTAGCGTTGTTCGGATTTACTGGGCGTAAAGCGCACGTAGGCGGACTTTTAAGTCAGGGGTGAAATCCCAGAGCTCAACTCCGGAACTGCCTTT
GATACTGGAAGTCTTGAGTATGGTAGAGGTGAGTGGAATTCCGAGTGTAGAGGTGAAATTCGTAGATATTCGGAGGAACACCAGTGGCGAAGGCGGCTCACT
GGACCATTACTGACGCTGAGGTGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGAATGTTAGCCGTCGGGGTGTT
TACACTTACGGTGGCGCAGCTAATCGCATTAAACATTCCGCCTGGGGAGTACGGTCGCAAGATTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGT
GGAGCATGTGGTTTAATTCGAAGCAACGCGCAGAACCTTACCAGCCCTTGACATACCGGTCGCGGACACAGAGATGTGTCTTTCAGTTCGGCTGGACCGGAT
ACAGGTGCTGCATGGCTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTCGCCTTTAGTTGCCATCATTTAGTTGGGCA
CTCTAAAGGGACTGCCAGTGATAAGCTGGAGGAAGGTGGGGATGACGTCAAGTCCCTCATGGCCCTTACGGGCTGGGCTACACACGTGCTACAATGGTGGTG
ACAGTGGGCAGCAAGCACGCGAGTGTGAGCTAATCTCCAAAAGCCATCTCAGTTCGGATTGCACTCTGCAACTCGAGTGCATGAAGTTGGAATCGCTAGTAA
TCGCGGATCAGCATGCCGCGGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCATGGGAGTTGGTTCTGCCCGAAGGCTTGAGAGTTTGATTC
CTGGCTCCGGATGAG

CDB3:

GAGTAATCGGCTGAGTATACGTAGCAAGTCGAGCGAATGGATTCGTAGCTTGCTCTTATGAAGTTAGCGGCGGACGGGTGAGTAACACGTGGGTA
ACCTGCCCATAAGACTGGGATAACTCCGGGAAACCGGGGCTAATACCGGATAATATTTTGAACTGCATGGTTCGAAATTGAAAGGCGGCTTCGGC
TGTCACTTATGGATGGACCCGCGTCGCATTAGCTAGTTGGTGAGGtAACGGCTCaCCAAGGCAACGATGCGTAGCCGACCTGAGAGGGTGATCGG
CCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTAGGGAATCTTCCGCAATGGACGAAAGTCTGACGGAGCAACGCCGCGT
GAGTGATGAAGGCTTTCGGGTCGTAAAACTCTGTTGTTAGGGAAGAACAAGTGCTAGTTGAATAAGCTGGCACCTTGACGGTACCTAACCAGAAA
GCCACGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGTGGCAAGCGTTATCCGGAATTATTGGGCGTAAAGCGCGCGCAGGTGGTTTCTT
AAGTCTGATGTGAAAGCCCACGGCTCAACCGTGGAGGGTCATTGGAAACTGGGAGACTTGAGTGCAGAAGAGGAAAGTGGAATTCCATGTGTAGC
GGTGAAATGCGTAGAGATATGGAGGAACACCAGTGGCGAAGGCGACTTTCTGGTCTGTAACTGACACTGAGGCGCGAAAGCGTGGGGAGCAAACA
GGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGAGTGCTAAGTGTTAGAGGGTTTCCGCCCTTTAGTGCTGAAGTTAACGCATTAAGCACT
CCGCCTGGGGAGTTCGGCCGCAAGGCTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGTGGAGCATGTGGTTTAATTCGAAGCAACGCG
AAGAACCTTACCAGGTCTTGACATCCTCTGAAAACCCTAGAGATAGGGCTTCTCCTTCGGGAGCAGAGTGACAGGTGGTGCATGGTTGTCGTCAG
CTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTGATCTTAGTTGCCATCATTAAGTTGGGCACTCTAAGGTGACTGCCGG
TGACAAACCGGAGGAAGGTGGGGATGACGTCAAATCATCATGCCCCTTATGACCTGGGCTACACACGTGCTACAATGGACGGTACAAAGAGCTGC
AAGACCGCGAGGTGGAGCTAATCTCATAAAACCGTTCTCAGTTCGGATTGTAGGCTGCAACTCGCCTACATGAAGCTGGAATCGCTAGTAATCGC
GGATCAGCATGCCGCGGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCACGAGAGTTTGTAACACCCGAAGTCGTTAGATTTCCA
TCCTGGATCCGGCTAGATATAGATCCGGTACATACCAG

CDB4:

ACAGGGAGCCGAGCAAGCAGCGAAGGATAAGAGCTTGCTCTTATGAAGTTAGCGGCGGACGGGTGAGTAACACGTGGGTAACCTGCCCATAAGACTGGGATA
ACTCCGGGAAACCGGGGCTAATACCGGATAATATTTTGAACTGCATGGTTCGAAATTGAAAGGCGGCTTCGGCTGTCACTTATGGATGGACCCGCGTCGCAT
TAGCTAGTTGGTGAGGTAACGGCTCACCAAGGCAACGATGCGTAGCCGACCTGAGAGGGTGATCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACG
GGAGGCAGCAGTAGGGAATCTTCCGCAATGGACGAAAGTCTGACGGAGCAACGCCGCGTGAGTGATGAAGGCTTTCGGGTCGTAAAACTCTGTTGTTAGGGA
AGAACAAGTGCTAGTTGAATAAGCTGGCACCTTGACGGTACCTAACCAGAAAGCCACGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGTGGCAAGC
GTTATCCGGAATTATTGGGCGTAAAGCGCGCGCAGGTGGTTTCTTAAGTCTGATGTGAAAGCCCACGGCTCAACCGTGGAGGGTCATTGGAAACTGGGAGAC
TTGAGTGCAGAAGAGGAAAGTGGAATTCCATGTGTAGCGGTGAAATGCGTAGAGATATGGAGGAACACCAGTGGCGAAGGCGACTTTCTGGTCTGTAACTGA
CACTGAGGCGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGAGTGCTAAGTGTTAGAGGGTTTCCGCCCTTTAGT
GCTGAAGTTAACGCATTAAGCACTCCGCCTGGGGAGTACGGCCGCAAGGCTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGTGGAGCATGTGGTT
TAATTCGAAGCAACGCGAAGAACCTTACCAGGTCTTGACATCCTCTGAAAACCCTAGAGATAGGGCTTCTCCTTCGGGAGCAGAGTGACAGGTGGTGCATGG
TTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTGATCTTAGTTGCCATCATTAAGTTGGGCACTCTAAGGTGACTGC
CGGTGACAAACCGGAGGAAGGTGGGGATGACGTCAAATCATCATGCCCCTTATGACCTGGGCTACACACGTGCTACAATGGACGGTACAAAGAGCTGCAAGA
CCGCGAGGTGGAGCTAATCTCATAAAACCGTTCTCAGTTCGGATTGTAGGCTGCAACTCGCCTACATGAAGCTGGAATCGCTAGTAATCGCGGATCAGCATG
CCGCGGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCACGAGAGTTTGTAACACCCGAAGTCGGTGGGGTACCTATGGAGCCAGCC

CDB5:

TGGCGGGCGGTAATACGTAGCAAGTCGAGCGGTAGCACAGGGGAGCTTGCTCCCTGGGTGACGAGCGGCGGACGGGTGAGTAATGTCTGGGAAACTGCCTGA
TGGAGGGGGATAACTACTGGAAACGGTAGCTAATACCGCATAACGTCGCAAGACCAAAGAGGGGGACCTTCGGGCCTCTTGCCATCAGATGTGCCCAGATGG
GATTAGCTAGTAGGTGGGGTAATGGCTCACCTAGGCGACGATCCCTAGCTGGTCTGAGAGGATGACCAGCCACACTGGAACTGAGACACGGTCCAGACTCCT
ACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGCAAGCCTGATGCAGCCATGCCGCGTGTGTGAAGAAGGCCTTCGGGTTGTAAAGCACTTTCAGCGA
GGAGGAAGGTGGTGAACTTAATACGCTCATCAATTGACGTTACTCGCAGAAGAAGCACCGGCTAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGCAA
GCGTTAATCGGAATTACTGGGCGTAAAGCGCACGCAGGCGGTTTGTTAAGTCAGATGTGAAATCCCCGGGCTCAACCTGGGAACTGCATTTGAAACTGGCAA
GCTAGAGTCTCGTAGAGGGGGGTAGAATTCCAGGTGTAgCGGTGAAATGCGTAGAGATCTGGAGGAATACCGGTGGCGAAGGCGGCCCCCTGGACGAAGACT
GACGCTCAGGTGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCTGTAAACGATGTCGATTTGGAGGTTGTGCCCTTGAGGCGTGGC
TTCCGGAGCTAACGCGTTAAATCGACCGCCTGGGGAGTCGGCCGCAAGGTTAAACTCAAATGAATTGACGGGGGCCCGCACAAGCGGTGGAGCATGTGGTTT
AATTCGATGCAACGCGAAGAACCTTACCTACTCTTGACATCCAGAGAACTTAGCAGAGATGCTTTGGTGCCTTCGGGAACTCTGAGACAGGTGCTGCATGGC
TGTCGTCAGCTCGTGTTGTGAAATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTATCCTTTGTTGCCAGCGGTTCGGCCGGGAACTCAAAGGAGACTGC
CAGTGATAAACTGGAGGAAGGTGGGGATGACGTCAAGTCATCATGGCCCTTACGAGTAGGGCTACACACGTGCTACAATGGCGTATACAAAGAGAAGCGACC
TCGCGAGAGCAAGCGGACCTCATAAAGTACGTCGTAGTCCGGATTGGAGTCTGCAACTCGACTCCATGAAGTCGGAATCGCTAGTAATCGTAGATCAGAATG
CTACGGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCATGGGAGTGGGTTGCAAAAGAAGTAGGTAGCTTAACCTTCGGGAGGGCGTTTAGA
TTTATTATACTGCTCCGCTGA

Figure 2.5: 16S rDNA sequences obtained from the five strains isolated from cattle dip-sites. The length of the sequences obtained were 1433 bp (CDB1), 1341 bp (CDB2), 1463 bp (CDB3), 1422 bp (CDB4), and 1449 bp (CDB5).

2.3.4 Homologous identity and multiple sequence alignments

A blastn search was performed for all five CDB sequences and few most related to each are listed in Table 2.5. The choice of representative 16S rDNA sequences for these species were made according to the results obtained from blastn and ClustalW.

The 26 related species were chosen according to their pairwise alignment score and identity. The species with higher score and identity with the tested strains were considered to be the possible related species and used for further analysis (Figure 2.6). CDB1, CDB2, CDB3, CDB4 and CDB5 were multiple aligned respectively with related species by using ClustalW (Figure 2.6 and Appendix 3) and compared with the result of the blastn. The homologous identity from blastn was similar to the multiple alignment result.

Table 2.5: The 16S rDNA sequence homologous identity of the tested stains with the most relevant species retrieved from the Genbank by blastn

Tested Strains	**Related bacterial species**	**Accession number**	**Length (bp)**	**Score**	**Identity (%)**	**Compared location of the subject sequence**
CDB1 (1433bp)	*Arthrobacter histidinolovorans*	AHRNA16S	1480	2662	99	23-1407
	Arthrobacter sp.	AB017649	1442	2597	99	23-1407
	Arthrobacter ilicis	X83407	1457	2407	98	51-1394
	Arthrobacter polychromogenes	AB167181	1398	2422	98	51-1394
	Arthrobacter sulfonivorans	AF235091	1457	2407	98	51-1394
CDB2 (1341bp)	*Ochrobactrum* sp.	OS16SRRA	1435	2496	99	12-1311
	Ochrobactrum sp. LMG20564	AF452128	1442	2450	99	12-1311
	Ochrobactrum grignonense	OGR242581	1383	2442	99	12-1311
	Ochrobactrum anthropi	OAN276036	1438	2357	98	12-1311
	Brucella canis	BRURGDD	1418	1606	97	12-881 898-1311
	Brucella melitensis	AY594216	1412	1606	97 97	12-881 898-1311
	Sinorhizobium sp. TB8-7-II	AY599692	1469	2026	95	4-1305
	Sinorhizobium fredii	X67231	1437	1968	95	1-1305
CDB3 (1463bp)	*Bacillus cereus* G8639	AY138273	1554	2722	100	24-1414
	Bacillus thuringiensis	AY741718	1524	2714	100	24-1414
	Bacillus anthracis	AY138383	1554	2710	98	24-1414
CDB4 (1422bp)	*Bacillus cereus* G8639	AY138273	1554	2732	100	28-1407
	Bacillus thuringiensis	AY741718	1524	2728	100	28-1407
	Bacillus anthracis	AY138383	1554	2720	99	28-1407
	Bacterium sp. Te27R	AY587813	1514	2716	99	28-1407
CDB5 (1449bp)	*Serratia marcescens*	AY514434	1424	2740	100	23-1422
	Enterobacteriaceae bacterium	AY513469	1435	1374	99	20-1415
	Pseudomonas fluorescens	ABO91837	1506	1369	99	20-1422

Multiple alignments of 16S rDNA sequences

CDB1:

```
                    1          11         21         31         41         51
A. sulfonivorans    ----------------------GGCGGCGTGCTTAACACATGCAAGTCGAACG-ATGAA
A. polychromogenes  GATCCTGGCTCAGGATGAACGCTGGCGGCGTGCTTAACACATGCAAGTCGAACG-ATGAA
A. histidinolovorans ---CCTGGCTCAGGATGAACGCTGGCGGCGTGCTTAACACATGCAAGTCGANCG-ATGAT
A. ilicis           GATCCTGGCTCAGGATGAACGCTGGCGGCGTGCTTAACACATGCAAGTCGAACG-ATGAT
A. sp.              ------------------------------TGCTTAACACATGCAAGTCGAACG-ATGAT
CDB1                ------------------ATATAGCCGACAGCTCACCGTA-GCAAGTCGAACGTATGAT
                                       *** * *  * ************ *****

                    61         71         81         91         101        111
A. sulfonivorans    CCTCACTTGTGGGGGGG-ATTAGTGGCG-AACGGGTG-AGTAACACGTGAGTAACCTGCC
A. polychromogenes  GGGGAGCTTGCTCCTGG-ATTAGTGGCG-AACGGGTG-AGTAACACGTGAGTAACCTGCC
A. histidinolovorans C-CCAGCTTNCTGGGGG-ATTAGTGGCG-AACGGGTG-AGTAACACGTGAGTAACCTGCC
A. ilicis           C-CCAGCTTGCTGGGGG-ATTAGTGGCG-AACGGGTG-AGTAACACGTGAGTAACCTGCC
A. sp.              CTCCAGCTTGCTGGGGGGATTAGTGGCG-AACGGGTG-AGTAACACGTGAGTAACCTGCC
CDB1                C-CCCGTTTGCTGGGGGTATTAGTGGCGTAACGGGTGTAGTAACACGTGAGTAACCTGCC
                          *       ** ********** ******** **********************

                    121        131        141        151        161        171
A. sulfonivorans    CTTGACTCTGGGATAAGCCTGGGAAACTGGGTCTAATACCGG-ATATG-ACGCCTCATCG
A. polychromogenes  CTTAACTCTGGGATAAGCCTGGGAAACTGGGTCTAATACCGG-ATATG-ACTCCTCATCG
A. histidinolovorans CTTGACTCTGGGATAACCCTGGGAAACTGGGTCTAATACCGG-ATATG-ACTCCTCATCG
A. ilicis           CTTGACTCTGGGATAAGCCTGGGAAACTGGGTCTAATACCGG-ATACG-ACCATCTGGCG
A. sp.              CTTGACTCTGGGATAAGCCTGGGAAACTGGGTCTAATACCGG-ATATG-ACTCCTCATCG
CDB1                CTTGACTCTGGGATAAGCCTGGGAAACTGGGTCTAATACCGGGATATGGACTCCTCATCG
                    ****************************************** ***** **       **

                    181        191        201        211        221        231
A. sulfonivorans    CATGGTGGGG-GGTGGAAAGTTTTTGTG-GTTTTGGATGGACTCGCGGCCTATCAGCTTG
A. polychromogenes  CATGGTGGGG-GGTGGAAAGCTTTATTGTGTTTTGGATGGACTCGCGGCCTATCAGCTTG
A. histidinolovorans CATGGTGGGG-GGTGGAAAGCTTTTGTG-GTTTTGGATGGACTCGCGGCCTATCAGCTTG
A. ilicis           CATGTCATGGTGGTGGAAAGCTTTTGTG-GTTTTGGATGGACTCGCGGCCTATCAGCTTG
A. sp.              CATGGTGGGG-GGTGGAAAGCTTTTGTG-GTTTTGGATGGACTCGCGGCCTATCAGCTTG
CDB1                CATGGTGGGG-GGTGGAAAGCTTTTGTG-GTTTTGGATGGACTCGCGGCCTATCAGCTTG
                    ****    ** *************  ** ********************************

                    241        251        261        271        281        291
A. sulfonivorans    TTGGTGAGGTAATGGCTCACCAAGGCGACGACGGGTAGCCGGCCTGAGAGGGTGACCGGC
A. polychromogenes  TTGGTGAGGTAATGGCTTACCAAGGCGACGACGGGTAGCCGGCCTGAGAGGGTGACCGGC
A. histidinolovorans TTGGTGGGGTAATGGCCTACCAAGGCGACGACGGGTAGCCGGCCTGAGAGGGTGACCGGC
A. ilicis           TTGGTGGGGTAATGGCCTACCAAGGCGACGACGGGTAGCCGGCCTGAGAGGGTGACCGGC
A. sp.              TTGGTGGGGTAATGGCCTACCAAGGCGACGACGGGTAGCCGGCCTGAGAGGGTGACCGGC
CDB1                TTGGTGGGGTAATGGCCTACCAAGGCGACGACGGGTAGCCGGCCTGAGAGGGTGACCGGC
                    ****** *********  ****************************************

                    301        311        321        331        341        351
A. sulfonivorans    CACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCA
A. polychromogenes  CACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCA
A. histidinolovorans CACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCA
A. ilicis           CACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCA
A. sp.              CACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCA
CDB1                CACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCA
                    ************************************************************

                    361        371        381        391        401        411
A. sulfonivorans    CAATGGGCGAAAGCCTGATGCAGCGACGCCGCGTGAGGGATGACGGCCTTCGGGTTGTAA
A. polychromogenes  CAATGGGCGCAAGCCTGATGCAGCGACGCCGCGTGAGGGATGACGGCCTTCGGGTTGTAA
A. histidinolovorans CAATGGGCGCAAGCCTGATGCAGCGACGCCGCGTGAGGGATGACGGCCTTCGGGTTGTAA
A. ilicis           CAATGGGCGCAAGCCTGATGCAGCGACGCCGCGTGAGGGATGACGGCCTTCGGGTTGTAA
A. sp.              CAATGGGCGGAAGCCTGATGCAGCGACGCCGCGTGAGGGATGACGGCCTTCGGGTTGTAA
CDB1                CAATGGGCGCAAGCCTGATGCAGCGACGCCGCGTGAGGGATGACGGCCTTCGGGTTGTAA
                    ********* ****************************************************

                    421        431        441        451        461        471
A. sulfonivorans    ACCTCTTTCAGTAGGGAAGAAGC-GAAAG-TGACGGTACCTGCAGAAGAAGCGCCGGCTA
A. polychromogenes  ACCTCTTTCAGTAGGGAAGAAGC-GAAAG-TGACGGTACCTGCAGAAGAAGCGCCGGCTA
A. histidinolovorans ACCTCTTTCAGTAGGGAAGAAGC-GTAAG-TGACGGTACCTGCAGAAGAAGCGCCGGCTA
A. ilicis           ACCTCTTTCAGTAGGGAAGAAGCCGTAAGGTGACGGTACCTGCAGAAGAAGCGCCGGCTA
A. sp.              ACCTCTTTCAGTAGGGAAGAAGCCGCAAGGTGACGGTACCTGCAGAAGAAGCGCCGGCTA
CDB1                ACCTCTTTCAGTAGGGAAGAAGC-GTAAG-TGACGGTACCTGCAGAAGAAGCGCCGGCTA
                    ********************** * *** ****************************

                    481        491        501        511        521        531
A. sulfonivorans    ACTACGTGCCA-GCAGCCGCGGTAATACGTAGGGCGCAAGCGTTATCCGGAATTATTGGG
A. polychromogenes  ACTACGTGCCA-GCAGCCGCGGTAATACGTAGGGCGCAAGCGTTATCCGGAATTATTGGG
A. histidinolovorans ACTACGTGCCA-GCAGCCGCGGTAATACGTAGGGCGCAAGCGTTATCCGGAATTATTGGG
A. ilicis           ACTACGTGCCA-GCAGCCGCGGTAATACGTAGGGCGCAAGCGTTATCCGGAATTATTGGG
A. sp.              ACTACGTGCCA-GCAGCCGCGGTAATACGTAGGGCGCAAGCGTTATCCGGAATTATTGGG
CDB1                ATTACGTGCCATGCAGCCGCGGTAATACGTAGGGCGCAAGCGTTATCCGGAATTATTGGG
                    *********** ************************************************
```

```
                    541       551       561       571       581       591
A. sulfonivorans    CGTAAAGAGCTCGTAGGCGGTTTGTCGCGTCTGCCGTGAAAGTCCGGGGCTCAACTCCGG
A. polychromogenes  CGTAAAGAGCTCGTAGGCGGTTTGTCGCGTCTGCCGTGAAAGTCCGGGGCTCAACTCCGG
A. histidinolovorans CGTAAAGAGCTCGTAGGCGGTTTGTCGCGTCTTCTGTGAAAGACCGGGGCTCAACTCCGG
A. ilicis           CGTAAAGAGCTCGTAGGCGGTTTGTCGCGTCTGCTGTGAAAGACCGGGGCTCAACTCCGG
A. sp.              CGTAAAGAGCTCGTAGGCGGTTTGTCGCGTCTGCTGTGAAAGACCGGGGCTCAACTCCGG
CDB1                CGTAAAGAGCTCGTAGGCGGTTTGTCGCGTCTGCTGTGAAAGACCGGGGCTCAACTCCGG
                    ********************************* * * ****  ******************

                    601       611       621       631       641       651
A. sulfonivorans    ATCTGCGGTGGGTACGGGCAGACTAGAGTG-ATGTAGGGGAGACTGGAATTCCTGGTGTA
A. polychromogenes  ATCTGCGGTGGGTACGGGCAGACTAGAGTG-ATGTAGGGGAGACTGGAATTCCTGGTGTA
A. histidinolovorans TTCTCCAGTGGGTACGGGCAGACTAGAGTG-CAGTAGGGGAGACTGGAATTCCTGGTGTA
A. ilicis           TTCTGCAGTGGGTACGGGCAGACTAGAGTG-CAGCAGGGGAGACTGGAATTCCTGGTGTA
A. sp.              TTCTGCAGTGGGTACGGGCAGACTAGAGTG-CAGTAGGGGAGACTGGAATTCCTGGTGTA
CDB1                TTCTGCAGTGGGTACGGGCAGACTAGAGTGCATGTAGGGGAGACTGGAATTCCTGGTGTA
                     ***** ********************    * *************************

                    661       671       681       691       701       711
A. sulfonivorans    GCGGTGAAATGCGCA-GATATCAGGAGGAACACCGATGGCGAAGGCAGGTCTCTGGGCAT
A. polychromogenes  GCGGTGAAATGCGCA-GATATCAGGAGGAACACCGATGGCGAAGGCAGGTCTCTGGGCAT
A. histidinolovorans GCGGTGAAATGCGCA-GATATCAGGAGGAACACCCATGGCGAAGGCAGGTCTCTGGGCTG
A. ilicis           GCGGCGAAATGCGCA-GATATCAGGAGGAACACCCATGGCGAAGGCAGGTCTCTGGGCTG
A. sp.              GCGGTGAAATGCGCA-GATATCAGGAGGAACACCGATGGCGAAGGCAGGTCTCTGGGCTG
CDB1                GCGGTGAAATGCGCATGATATCAGGAGGAACACCGATGGCGAAGGCAGGTCTCTGGGCTG
                    *************** ****************************************

                    721       731       741       751       761       771
A. sulfonivorans    TAACTGACGCTGAGGAGCGAAAGCATGGGGAGCGAACAGGATTAGATACCCTGGTAGTCC
A. polychromogenes  TAACTGACGCTGAGGAGCGAAAGCATGGGGAGCGAACAGGATTAGATACCCTGGTAGTCC
A. histidinolovorans TAACTGACGCTGAGGAGCGAAAGCATGGGGAGCGAACAGGATTAGATACCCTGGTAGTCC
A. ilicis           TAACTGACGCTGAGGAGCGAAAGCTAGGGGAGCGAACAGGATTAGATACCCTGGTAGTCC
A. sp.              TAACTGACGCTGAGGAGCGAAAGCATGGGGAGCGAACAGGATTAGATACCCTGGTAGTCC
CDB1                TAACTGACGCTGAGGAGCGAAAGCATGGGGAGCGAACAGGATTAGATACCCTGGTAGTCC
                    ************************************************************

                    781       791       801       811       821       831
A. sulfonivorans    ATGCCGTAAACGTTGGGCACTAGGTGTGGGGGACATTCCACGTTTTCCGCGCCGTAGCTA
A. polychromogenes  ATGCCGTAAACGTTGGGCACTAGGTGTGGGGGACATTCCACGTTTTCCGCGCCGTAGCTA
A. histidinolovorans ATGCCGTAAACGTTGGGCACTAGGTGTGGGGGACATTCCACGTTTTCCGCGCCGTAGCTA
A. ilicis           ATGCCGTAAACGTTGGGCACTAGGTGTGGGGGACATTCCACGTTTTCCGCGCCGTAGCTA
A. sp.              ATGCCGTAAACGTTGGGCACTAGGTGTGGGGGACATTCCACGTTTTCCGCGCCGTAGCTA
CDB1                ATGCCGTAAACGTTGGGCACTAGGTGTGGGGGACATTCCACGTTTTCCGCGCCGTAGCTA
                    ************************************************************

                    841       851       861       871       881       891
A. sulfonivorans    ACGCATTAAGTGCCCCGCCTGGGGAGTACGGCCGCAAGGCTAAAACTCAAAGGAATTGAC
A. polychromogenes  ACGCATTAAGTGCCCCGCCTGGGGAGTACGGCCGCAAGGCTAAAACTCAAAGGAATTGAC
A. histidinolovorans ACGCATTAAGTGCCCCGCCTGGGGAGTACGGCCGCAAGGCTAAAACTCAAAGGAATTGAC
A. ilicis           ACGCATTAAGTGCCCCGCCTGGGGAGTACGGCCGCAAGGCTAAAACTCAAAGGAATTGAC
A. sp.              ACGCATTAAGTGCCCCGCCTGGGGAGTACGGCCGCAAGGCTAAAACTCAAAGGAATTGAC
CDB1                ACGCATTAAGTGCCCCGCCTGGGGAGTACGGCCGCAAGGCTAAAACTCAAAGGAATTGAC
                    ************************************************************

                    901       911       921       931       941       951
A. sulfonivorans    GGGGGCCCGCACAAGCGGCGGAGCATGCGGATTAATTCGATGCAACGCGAAGAACCTTAC
A. polychromogenes  GGGGGCCCGCACAAGCGGCGGAGCATGCGGATTAATTCGATGCAACGCGAAGAACCTTAC
A. histidinolovorans GGGGGCCCGCACAAGCGGCGGAGCATGCGGATTAATTCGATGCAACGCGAAGAACCTTAC
A. ilicis           GGGGGCCCGCACAAGCGGCGGAGCATGCGGATTAATTCGATGCAACGCGAAGAACCTTAC
A. sp.              GGGGGCCCGCACAAGCGGCGGAGCATGCGGATTAATTCGATGCAACGCGAAGAACCTTAC
CDB1                GGGGGCCCGCACAAGCGGCGGAGCATGCGGATTAATTCGATGCAACGCGAAGAACCTTAC
                    ************************************************************

                    961       971       981       991       1001      1011
A. sulfonivorans    CAAGGCTTGACATGGACTGGAAATACCTGGAAACAGGTGCCCCCGCTTGCGGCCGGTTCA
A. polychromogenes  CAAGGCTTGACATGAACCGGTAATACCTGGAGACAGGTGCCCC-GCTTGCGGTCGGTTTA
A. histidinolovorans CAAGGCTTGACATGAACCGGAAAGACCTGGAAACAGGTGCCCC-GCTTGCGGTCGGTTTA
A. ilicis           CAAGGCTTGACATGAACCGGTAATACCTGGAAACAGGTGCCCC-GCTTGTGGTCGGTTTA
A. sp.              CAAGGCTTGACATGAACCGGAAAGACCTGGAAACAGGTGCCCC-GCTTGCGGTCGGTTTA
CDB1                CAAGGCTTGACATTAACCGGAAAGACCTGGAAACAGGTGCCCC-GCTTGCGGTCGGTTTA
                    ******************************************* ****************

                    1021      1031      1041      1051      1061      1071
A. sulfonivorans    CAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACG
A. polychromogenes  CAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACG
A. histidinolovorans CAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACG
A. ilicis           CAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACG
A. sp.              CAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACG
CDB1                CAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACG
                    ************************************************************

                    1081      1091      1101      1111      1121      1131
A. sulfonivorans    AGCGCAACCCTCGTTCTATGTTGCCAGCGCGTGATGGCGGGGACTCATAGGAGACTGCCG
A. polychromogenes  AGCGCAACCCTCGTTCTATGTTGCCAGCACGTGATGGTGGGGACTCATAGGAGACTGCCG
A. histidinolovorans AGCGCAACCCTCGTTCTATGTTGCCAGCG-GTTCGGCCGGGGACTCATAGGAGACTGCCG
A. ilicis           AGCGCAACCCTCGTTCTATGTTGCCAGCG-GTTCGGCCGGGGACTCATAGGAGACTGCCG
A. sp.              AGCGCAACCCTCGTTCTATGTTGCCAGCG-GTTCGGCCGGGGACTCATAGGAGACTGCCG
CDB1                AGCGCAACCCTCGTTCTATGTTGCCAGCG-GTTCGGCCGGGGACTCATAGGAGACTGCCG
                    ***************************  **   *  **********************
```

```
                        1141       1151       1161       1171       1181       1191
A. sulfonivorans        GGGTCAACTCGGAGGAAGGTGGGGACGACGTCAAATCAT-CATGCCCCTTATGTCTTGGG
A. polychromogenes      GGGTCAACTCGGAGGAAGGTGGGGACGACGTCAAATCAT-CATGCCCCTTATGTCTTGGG
A. histidinolovorans    GGGTCAACTCGGAGGAAGGTGGGGACGACGTCAAATCAT-CATGCCCCTTATGTCTTGGG
A. ilicis               GGGTCAACTCGGAGGAAGGTGGGGACGACGTCAAATCAT-CATGCCCCTTATGTCTTGGG
A. sp.                  GGGTCAACTCGGAGGAAGGTGGGGACGACGTCAAATCAT-CATGCCCCTTATGTCTTGGG
CDB1                    GG-TCAACTCGGAGGAAGGTGGGGACGACGTCAAATCATACATGCCCCTTATGTCTTGGG
                        ** ************************************ ********************

                        1201       1211       1221       1231       1241       1251
A. sulfonivorans        CTTCACGCATGCTACAATGGCCGGTACAAAGGGTTGCGATACTGTGAGGTGGAGCTAATC
A. polychromogenes      CTTCACGCATGCTACAATGGCCGGTACAAAGGGTTGCGATACTGTGAGGTGGAGCTAATC
A. histidinolovorans    CTTCACGCATGCTACAATGGCCGGTACAAAGGGTTGCGATACTGTGAGGTGGAGCTAATC
A. ilicis               CTTCACGCATGCTACAATGGCCGGTACAAAGGGTTGCGATACTGTGAGGTGGAGCTAATC
A. sp.                  CTTCACGCATGCTACAATGGCCGGTACAAAGGGTTGCGATACTGTGAGGTGGAGCTAATC
CDB1                    CTTCACGCATGCTACAATGGCCGGTACAAAGGGTTGCGATACTGTGAGGTGGAGCTAATC
                        ************************************************************

                        1261       1271       1281       1291       1301       1311
A. sulfonivorans        CCAAAAAGCCGGTCTCAGTTCGGATTGGGGTCTGCAACTCGACCCCATGAAGTCGGAGTC
A. polychromogenes      CCAAAAAGCCGGTCTCAGTTCGGATTGGGGTCTGCAACTCGACCCCATGAAGTCGGAGTC
A. histidinolovorans    CCAAAAAGCCGGTCTCAGTTCGGATTGGGGTCTGCAACTCGACCCCATGAAGTCGGAGTC
A. ilicis               CCAAAAAGCCGGTCTCAGTTCGGATTGGGGTCTGCAACTCGACCCCATGAAGTCGGAGTC
A. sp.                  CCAAAAAGCCGGTCTCAGTTCGGATTGGGGTCTGCAACTCGACCCCATGAAGTCGGAGTC
CDB1                    CCAAAAAGCCGGTCTCAGTTCGGATTGGGGTCTGCAACTCGACCCCATGAAGTCGGAGTC
                        ************************************************************

                        1321       1331       1341       1351       1361       1371
A. sulfonivorans        GCTAGTAATC-GCAGATCAGCAACGCTGCGGTGAATACGTT-CCCGGGCCTTGTACACAC
A. polychromogenes      GCTAGTAATC-GCAGATCAGCAACGCTGCGGTGAATACGTT-CCCGGGCCTTGTACACAC
A. histidinolovorans    GCTAGTAATC-GCAGATCAGCAACGCTGCGGTGAATACGTT-CCCGGGCCTTGTACACAC
A. ilicis               GCTAGTAATC-GCAGATCAGCAACGCTGCGGTGAATACGTT-CCCGGGCCTTGTACACAC
A. sp.                  GCTAGTAATC-GCAGATCAGCAACGCTGCGGTGAATACGTT-CCCGGGCCTTGTACACAC
CDB1                    GCTAGTAATCCGCAGATCAGCAACGCTGCGGTGAATACGTTACCCGGGCCTTGTACACAC
                        ********** ****************************** ******************

                        1381       1391       1401       1411       1421       1431
A. sulfonivorans        CGCCCGT-CAAGTCACGAAAGTTGGTAACACCCGAAGCCGGTG-GCCTAACCCCTTATGG
A. polychromogenes      CGCCCGT-CAAGTCACGAAAGTTGGTAACACCCGAAGCCGGTG-GCCTAACCCCTTGTGG
A. histidinolovorans    CGCCCGT-CAAGTCACGAAAGTTGGTAACACCCGAAGCCGGTG-GCCTAACCC-TTGTGG
A. ilicis               CGCCCGT-CAAGTCACGAAAGTTGGTAACACCCGAAGCCGGTG-GCCTAACCC-TTGTGG
A. sp.                  CGCCCGT-CAAGTCACGAAAGTTGGTAACACCCGAAGCCGGTG-GCCTAACCC-TTGTGG
CDB1                    CGCCCGTTCAAGTCACGAAAGTTGGTAACACCCGAAGCCGGTTAGCCTAACCC-TTGTGG
                        ******* ********************************** ********* ******

                        1441       1451       1461       1471       1481       1491
A. sulfonivorans        GGAGGAGCTGTCGAAGGTGGGACTGGCGATTGGGACTAAGTCGTAACAAGGTAGCCGT--
A. polychromogenes      GAGGGAGCTGTCGAAGGTGGGACTGGCGATTGGGACTAAGTCGTAACAAG----------
A. histidinolovorans    GG-GGAGCCGTCGAAGGTGGGACCGGCGATTGGGACTAAGTCGTAACAAGGTAGCCGTAC
A. ilicis               GG-GGAGCCGTCGAAGGTGGGACCGGCGATTGGGACTAAGTCGTAACAAGGTAGCCGTAC
A. sp.                  GG-GGAGCCGTCGAAGGTGGGACCGGCGATTGGGACTAAGTCGTAACAAG----------
CDB1                    GG-GGAGCT-TAGATA-TAGAGCCGGAACGACCGAC-----------------------
                        ** *****  * **   * *  ****       ***

                        1501
A. sulfonivorans        ------
A. polychromogenes      ------
A. histidinolovorans    CGGAAG
A. ilicis               CGGAAG
A. sp.                  ------
CDB1                    ------
```

CDB2:

```
                        1          11         21         31         41         51
O. sp. LMG20564         ------------------AACGAACGCTGGCGGCAGGCTTAACACATGCAAGTCGAGCGC
O. sp.                  ------------------AACGAACGCTGGCGGCAGGCTTAACACATGCAAGTCGAGCGC
CDB2                    ------------------------------------------------------------
O. grignonense          AGTTTGATCCTGGCTCAGAACGAACGCTGGCGGCAGGCTTAACACATGCAAGTCGAGCGC
B. canis                --------------TCAGAACGAACGCTGGCGGCAGGCTTAACACATGCAAGTCGAGCGC
B. melitens             --------------CTCAGAACGAACGCTGGCGGCAGGCTTAACACATGCAAGTCGAGCGC
O. anthropi             AGTTTGATCCTGGCTCAGAACGAACGCTGGCGGCAGGCTTAACACATGCAAGTCGAGCGC

                        61         71         81         91         101        111
O. sp. LMG20564         CCCGCAAGGG-AGCGGCAGACGGGTGAGTAACGCGTGGGAATCTACCTTTTGCTACGGAA
O. sp.                  CCCGCAAGGGGAGCGGCAGACGGGTGAGTAACGCGTGGGAATCTACCTTTTGCTACGGAA
CDB2                    ----------AGCGGCATACGGGTGAGTAACGCGTGGGAATCTACCTTTTGCTACGGAA
O. grignonense          CTCGCAAGAGGAGCGGCAGACGGGTGAGTAACGCGTGGGAATCTACCTTTTGCTACGGAA
B. canis                CCCGCAAGGGGAGCGGCAGACGGGTGAGTAACGCGTGGGAACGTACCATTTGCTACGGAA
B. melitens             CCCGCAAGGGGAGCGGCAGACGGGTGAGTAACGCGTGGGAACGTACCATTTGCTACGGAA
O. anthropi             CCCGCAAGGGGAGCGGCAGACGGGTGAGTAACGCGTGGGAACGTACCTTTTGCTACGGAA
                        ************************************************
```

```
                    121       131       141       151       161       171
O. sp. LMG20564     TAACTCAGGGAAACTTGTGCTAATACCGTATGTGCCCTTCGGGGG-AAAGATTTATCGGC
O. sp.              TAACTCAGGGAAACTTGTGCTAATACCGTATGTGCCCTTCGGGGG-AAAGATTTATCGGC
CDB2                TAACTCAGGGAAACTTGTGCTAATACCGTATGTGCCCTTCGGGGGGAAAGATTTATCGGC
O. grignonense      TAACTCAGGGAAACTTGTGCTAATACCGTATGTGCCCTTTTGGGG-AAAGATTTATCGGC
B. canis            TAACTCAGGGAAACTTGTGCTAATACCGTATGTGCCCTTCGGGGG-AAAGATTTATCGGC
B. melitens         TAACTCAGGGAAACTTGTGCTAATACCGTATGTGCCCTTCGGGGG-AAAGATTTATCGGC
O. anthropi         TAACTCAGGGAAACTTGTGCTAATACCGTATGTGCCCTTCGGGGG-AAAGATTTATCGGC
                    ********************************************* ************

                    181       191       201       211       221       231
O. sp. LMG20564     AAAGGATGAGCCCGCGTTGGATTAGCTAGTTGGTGGGGTAAAGGCCCACCAAGGCGACGA
O. sp.              AAAGGATGAGCCCGCGTTGGATTAGCTAGTTGGTGAGGTAAAGGCTCACCAAGGCGACGA
CDB2                AAAGGATGAGCCCGCGTTGGATTAGCTAGTTGGTGGGGTAAAGGCCCACCAAGGCGACGA
O. grignonense      AAAGGATGAGCCCGCGTTGGATTAGCTAGTTGGTAGGGTAATGGCCTACCAAGGCGACGA
B. canis            AAATGATCGGCCCGCGTTGGATTAGCTAGTTGGTGGGGTAAAGGCTCACCAAGGCGACGA
B. melitens         AAATGATCGGCCCGCGTTGGATTAGCTAGTTGGTGGGGTAAAGGCTCACCAAGGCGACGA
O. anthropi         AAAGGATCGGCCCGCGTTGGATTAGCTAGTTGGTGAGGTAAAGGCTCACCAAGGCGACGA
                    ************************************************************

                    241       251       261       271       281       291
O. sp. LMG20564     TCCATAGCTGGTCTGAGAGGATGATCAGCCACACTGGGACTGAGACACGGCCCAGACTCC
O. sp.              TCCATAGCTGGTCTGAGAGGATGATCAGCCACACTGGGACTGAGACACGGCCCAGACTCC
CDB2                TCCATAGCTGGTCTGAGAGGATGATCAGCCACACTGGGACTGAGACACGGCCCAGACTCC
O. grignonense      TCCATAGCTGGTCTGAGAGGATGATCAGCCACACTGGGACTGAGACACGGCCCAGACTCC
B. canis            TCCATAGCTGGTCTGAGAGGATGATCAGCCACACTGGGACTGAGACACGGCCCAGACTCC
B. melitens         TCCATAGCTGGTCTGAGAGGATGATCAGCCACACTGGGACTGAGACACGGCCCAGACTCC
O. anthropi         TCCATAGCTGGTCTGAGAGGATGATCAGCCACACTGGGACTGAGACACGGCCCAGACTCC
                    ************************************************************

                    301       311       321       331       341       351
O. sp. LMG20564     TACGGGAGGCAGCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCCATGCCG
O. sp.              TACGGGAGGCAGCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCCATGCCG
CDB2                TACGGGAGGCAGCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCCATGCCG
O. grignonense      TACGGGAGGCAGCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCCATGCCG
B. canis            TACGGGAGGCAGCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCCATGCCG
B. melitens         TACGGGAGGCAGCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCCATGCCG
O. anthropi         TACGGGAGGCAGCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCCATGCCG
                    ************************************************************

                    361       371       381       391       401       411
O. sp. LMG20564     CGTGAGTGATGAAGGCCCTAGGGTTGTAAAGCTCTTTCACCGGTGAAGATAATGACGGTA
O. sp.              CGTGAGTGATGAAGGTCTTAGGATTGTAAAGCTCTTTCACCGGTGAAGATAATGACGGTA
CDB2                CGTGAGTGATGAAGGCCCTAGGGTTGTAAAGCTCTTTCACCGGTGAAGATAATGACGGTA
O. grignonense      CGTGAGTGATGAAGGCCCTAGGGTTGTAAAGCTCTTTCACCGGTGAAGATAATGACGGTA
B. canis            CGTGAGTGATGAAGGCCCTAGGGTTGTAAAGCTCTTTCACCGGTGAAGATAATGACGGTA
B. melitens         CGTGAGTGATGAAGGCCCTAGGGTTGTAAAGCTCTTTCACCGGTGAAGATAATGACGGTA
O. anthropi         CGTGAGTGATGAAGGCCCTAGGGTTGTAAAGCTCTTTCACCGGTGAAGATAATGACGGTA
                    ************************************************************

                    421       431       441       451       461       471
O. sp. LMG20564     ACCGGAGAAGAAGCCCCGGCTAACTTCGTGCCAGCAGCCGCGGTAATACGAAGGGGGCTA
O. sp.              ACCGGAGAAGAAGCCCCGGCTAACTTCGTGCCAGCAGCCGCGGTAATACGAAGGGGGCTA
CDB2                ACCGGAGAAGAAGCCCCGGCTAACTTCGTGCCAGCAGCCGCGGTAATACGAAGGGGGCTA
O. grignonense      ACCGGAGAAGAAGCCCCGGCTAACTTCGTGCCAGCAGCCGCGGTAATACGAAGGGGGCTA
B. canis            ACCGGAGAAGAAGCCCCGGCTAACTTCGTGCCAGCAGCCGCGGTAATACGAAGGGGGCTA
B. melitens         ACCGGAGAAGAAGCCCCGGCTAACTTCGTGCCAGCAGCCGCGGTAATACGAAGGGGGCTA
O. anthropi         ACCGGAGAAGAAGCCCCGGCTAACTTCGTGCCAGCAGCCGCGGTAATACGAAGGGGGCTA
                    ************************************************************

                    481       491       501       511       521       531
O. sp. LMG20564     GCGTTGTTCGGATTTACTGGGCGTAAAGCGCACGTAGGCGGACTTTTAAGTCAGGGGTG-
O. sp.              GCGTTGTTCGGATTTACTGGGCGTAAAGCGCACGTAGGCGGACTTTTAAGTCAGGGGTG-
CDB2                GCGTTGTTCGGATTTACTGGGCGTAAAGCGCACGTAGGCGGACTTTTAAGTCAGGGGTGG
O. grignonense      GCGTTGTTCGGATTTACTGGGCGTAAAGCGCACGTAGGCGGACTTTTAAGTCAGGGGTG-
B. canis            GCGTTGTTCGGATTTACTGGGCGTAAAGCGCACGTAGGCGGACTTTTAAGTCAGGGGTG-
B. melitens         GCGTTGTTCGGATTTACTGGGCGTAAAGCGCACGTAGGCGGACTTTTAAGTCAGGGGTG-
O. anthropi         GCGTTGTTCGGATTTACTGGGCGTAAAGCGCACGTAGGCGGACTTTTAAGTCAGGGGTG-
                    ************************************************************

                    541       551       561       571       581       591
O. sp. LMG20564     AAATCCCAGAGCTCAACTCTGGAACTGCCTTTGATACTGGAAGTCTTGAGTATGGTAGAG
O. sp.              AAATCCCGGGGCTCAACCCCGGAACTGCCTTTGATACTGGAAGTCTTGAGTATGGTAGAG
CDB2                AAATCCCAGAGCTCAACTCTGGAACTGCCTTTGATACTGGAAGTCTTGAGTATGGTAGAG
O. grignonense      AAATCCCGGGGCTCAACCCCGGAACTGCCTTTGATACTGGAAGTCTTGAGTATGGTAGAG
B. canis            AAATCCCGGGGCTCAACCCCGGAACTGCCTTTGATACTGGAAGTCTTGAGTATGGTAGAG
B. melitens         AAATCCCGGGGCTCAACCCCGGAACTGCCTTTGATACTGGAAGTCTTGAGTATGGTAGAG
O. anthropi         AAATCCCGGGGCTCAACCCCGGAACTGCCTTTGATACTGGAAGTCTTGAGTATGGTAGAG
                    ************************************************************
```

```
                    601       611       621       631       641       651
O. sp. LMG20564     GTGAGTGGAATTCCGAGTGTAGAGGTGAAATTCGTAGATATTC-GGAGGAACACCAGTGG
O. sp.              GTGAGTGGAATTCCGAGTGTAGAGGTGAAATTCGTAGATATTC-GGAGGAACACCAGTGG
CDB2                GCGAGTGGAATTCCGAGTGTAGAGGTGAAATTCGTAGATATTCCGGAGGAACACCAGTGG
O. grignonense      GTGAGTGGAATTCCGAGTGTAGAGGTGAAATTCGTAGATATTC-GGAGGAACACCAGTGG
B. canis            GTGAGTGGAATTCCGAGTGTAGAGGTGAAATTCGTAGATATTC-GGAGGAACACCAGTGG
B. melitens         GTGAGTGGAATTCCGAGTGTAGAGGTGAAATTCGTAGATATTC-GGAGGAACACCAGTGG
O. anthropi         GTGAGTGGAATTCCGAGTGTAGAGGTGAAATTCGTAGATATTC-GGAGGAACACCAGTGG
                    ***************************************** ****************

                    661       671       681       691       701       711
O. sp. LMG20564     CGAAGGCGGCTCACTGGACCATTACT-GACGCTGAGGTGCGAAAGC-GTGGGGAGCAAAC
O. sp.              CGAAGGCGGCTCACTGGACCATTACT-GACGCTGAGGTGCGAAAGC-GTGGGGAGCAAAC
CDB2                CGAAGGCGGCTCACTGGACCATTACTCGACGCTGAGGTGCGAAAGCCGTGGGGAGCAAAC
O. grignonense      CGAAGGCGGCTCACTGGACCATTACT-GACGCTGAGGTGCGAAAGC-GTGGGGAGCAAAC
B. canis            CGAAGGCGGCTCACTGGACCATTACT-GACGCTGAGGTGCGAAAGC-GTGGGGAGCAAAC
B. melitens         CGAAGGCGGCTCACTGGACCATTACT-GACGCTGAGGTGCGAAAGC-GTGGGGAGCAAAC
O. anthropi         CGAAGGCGGCTCACTGGACCATTACT-GACGCTGAGGTGCGAAAGC-GTGGGGAGCAAAC
                    ************************** ****************** *************

                    721       731       741       751       761       771
O. sp. LMG20564     AGGATTAGATACCCTGGTAGTCCACGCCGTAAA-CGATGAATGTTAGCCGTC-GGGGTGT
O. sp.              AGGATTAGATACCCTGGTAGTCCACGCCGTAAA-CGATGAATGTTAGCCGTC-GGGGAGT
CDB2                AGGATTAGATACCCTGGTAGTCCACGCCGTAAATCGATGAATGTTAGCCGTCCGGGGTGT
O. grignonense      AGGATTAGATACCCTGGTAGTCCACGCCGTAAA-CGATGAATGTTAGCCGTC-GGGGAGT
B. canis            AGGATTAGATACCCTGGTAGTCCACGCCGTAAA-CGATGAATGTTAGCCGTC-GGGGTGT
B. melitens         AGGATTAGATACCCTGGTAGTCCACGCCGTAAA-CGATGAATGTTAGCCGTC-GGGGTGT
O. anthropi         AGGATTAGATACCCTGGTAGTCCACGCCGTAAA-CGATGAATGTTAGCCGTT-GGGGAGT
                    ********************************** ***************** **** **

                    781       791       801       811       821       831
O. sp. LMG20564     TTACACTT-CGGTGGCGCANNTAA-CGCATTAAACATTCCGCCTGGGGAGTACGGT-CGC
O. sp.              TTACTCTT-CGGTGGCGCAGCTAA-CGCATTAAACATTCCGCCTGGGGAGTACGGT-CGC
CDB2                TTACACTTACGGTGGCGCAGCTAATCGCATTAACCATTCCGCCTGGGGAGTACGGTACGC
O. grignonense      TTACTCTT-CGGTGGCGCAGCTAA-CGCATTAAACATTCCGCCTGGGGAGTACGGT-CGC
B. canis            TTACACTT-CGGTGGCGCAGCTAA-CGCATTAAACATTCCGCCTGGGGAGTACGGT-CGC
B. melitens         TTACACTT-CGGTGGCGCAGCTAA-CGCATTAAACATTCCGCCTGGGGAGTACGGT-CGC
O. anthropi         TTACTCTT-CGGTGGCGCAGCTAA-CGCATTAAACATTCCGCCTGGGGAGTACGGT-CGC
                    ******** **********  *** ******************************** ***

                    841       851       861       871       881       891
O. sp. LMG20564     AAGATTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGTGGAGCATGTGGTTTAA
O. sp.              AAGATTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGTGGAGCATGTGGTTTAA
CDB2                AAGATTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGTGGAGCATGTGGTTTAA
O. grignonense      AAGATTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGTGGAGCATGTGGTTTAA
B. canis            AAGATTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGTGGAGCATGTGGTTTAA
B. melitens         AAGATTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGTGGAGCATGTGGTTTAA
O. anthropi         AAGATTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGTGGAGCATGTGGTTTAA
                    ************************************************************

                    901       911       921       931       941       951
O. sp. LMG20564     TTCGAAGCAACGCGCAGAACCTTACCAGCCCTTGACATACCGGTCGCGGACACAG-AGAT
O. sp.              TTCGAAGCAACGCGCAGAACCTTACCAGCCCTTGACATACCGGTCGCGGACACAG-AGAT
CDB2                TT-GAAGCAACGCGCAGAACCTTACCAGCCCTTGACATACCGGTCGCGGACACAG-AGAT
O. grignonense      TTCGAAGCAACGCGCAGAACCTTACCAGCCCTTGAGATACCGGTCGCGGACACAG-AGAT
B. canis            TTCGAAGCAACGCGCAGAACCTTACCAGCCCTTGACATCCCGGTCGCGGTTAGTGGAGAC
B. melitens         TTCGAAGCAACGCGCAGAACCTTACCAGCCCTTGACATCCCGGTCGCGGTTAGTGGAGAC
O. anthropi         TTCGAAGCAACGCGCAGAACCTTACCAGCCCTTGACATACCGGTCGCGGACACAG-AGAT
                    ************************************************** ****

                    961       971       981       991       1001      1011
O. sp. LMG20564     G-TGTCTTTCAGTTCGGCTGGACCGGATACAGGTGCTGCATGGCTGTCGTCAGCTCGTGT
O. sp.              G-TGTCTTTCAGTTCGGCTGGACCGGATACAGGTGCTGCATGGCTGTCGTCAGCTCGTGT
CDB2                G-TGTCTTTCAGTTCGGCTGGACCGGATACAGGTGCTGCATGGCTGTCGTCAGCTCGTGT
O. grignonense      G-TGTCTTTCAGTTCGGCTGGACCGGATACAGGTGCTGCATGGCTGTCGTCAGCTCGTGT
B. canis            ACTATCCTTCAGTTAGGCTGGACCGGAGACAGGTGCTGCATGGCTGTCGTCAGCTCGTGT
B. melitens         ACTATCCTTCAGTTAGGCTGGACCGGAGACAGGTGCTGCATGGCTGTCGTCAGCTCGTGT
O. anthropi         G-TGTCTTTCAGTTCGGCTGGACCGGATACAGGTGCTGCATGGCTGTCGTCAGCTCGTGT
                      * ************************ ********************************

                    1021      1031      1041      1051      1061      1071
O. sp. LMG20564     CGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTCGCCTTTAGTTGCCATCATT
O. sp.              CGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTCGCCTTTAGTTGCCATCATT
CDB2                CGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTCGCCTTTAGTTGCCATCATT
O. grignonense      CGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTCGCCTTTAGTTGCCATCATT
B. canis            CGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTCGCCCTTAGTTGCCAGCATT
B. melitens         CGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTCGCCCTTAGTTGCCAGCATT
O. anthropi         CGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTCGCCCTTAGTTGCCAGCATT
                    ********************************************** ********** ****

                    1081      1091      1101      1111      1121      1131
O. sp. LMG20564     TAGTTGGGCACTCTAAAGGGACTGCCAGTGATAAGCTG-GAGGAAGGTGGGGATGACGTC
O. sp.              TAGTTGGGCACTCTAAAGGGACTGCCAGTGATAAGCTG-GAGGAAGGTGGGGATGACGTC
CDB2                TAGTTGGGCACTCTAAAGGGACTGCCAGTGATAAGCTG-GAGGAAGGTGGGGATGACGTC
O. grignonense      TAGTTGGGCACTCTAAAGGGACTGCCAGTGATAAGCTG-GAGGAAGGTGGGGATGACGTC
B. canis            CAGTTGGGCACTCTAAGGGGACTGCCGGTGATAAGCCGAGAGGAAGGTGGGGATGACGTC
B. melitens         CAGTTGGGCACTCTAAGGGGACTGCCGGTGATAAGCCGAGAGGAAGGTGGGGATGACGTC
O. anthropi         TAGTTGGGCACTCTAAGGGGACTGCCGGTGATAAGCCGAGAGGAAGGTGGGGATGACGTC
                     *************** ********* ********* * **********************
```

```
                    1141      1151      1161      1171      1181      1191
O. sp. LMG20564     AAGTCCTCATTGCCNTTACGGGCTGGGCTACACACGTGCTACAATGGTGGTGACAGTGGG
O. sp.              AAGTCCTCATGGCCCTTACGGGCTGGGCTACACACGTGCTACAATGGTGGTGACAGTGGG
CDB2                AAGTCCTCATGGCCCTTACGGGCTGGGCTACACACGTGCTACAATGGTGGTGACAGTGGG
O. grignonense      AAGTCCTCATGGCCCTTACGGGCTGGGCTACACACGTGCTACAATGGTGGTGACAGTGGG
B. canis            AAGTCCTCATGGCCCTTACGGGCTGGGCTACACACGTGCTACAATGGTGGTGACAGTGGG
B. melitens         AAGTCCTCATGGCCCTTACGGGCTGGGCTACACACGTGCTACAATGGTGGTGACAGTGGG
O. anthropi         AAGTCCTCATGGCCCTTACGGGCTGGGCTACACACGTGCTACAATGGTGGTGACAGTGGG
                    ************** ************************************************

                    1201      1211      1221      1231      1241      1251
O. sp. LMG20564     CAGCAAGCACGCGAGTGTGAGCTAATCTCCAAAAGCCATCTCAGTTCGGATTGCACTCTG
O. sp.              CAGCAAGCACGCGAGTGTGAGCTAATCTCCAAAAGCCATCTCAGTTCGGATTGCACTCTG
CDB2                CAGCAAGCACGCGAGTGTGAGCTAATCTCCAAAAGCCATCTCAGTTCGGATTGCACTCTG
O. grignonense      CAGCAAGCACGCGAGTGTGAGCTAATCTCCAAAAGCCATCTCAGTTCGGATTGCACTCTG
B. canis            CAGCGAGCACGCGAGTGTGAGCTAATCTCCAAAAGCCATCTCAGTTCGGATTGCACTCTG
B. melitens         CAGCGAGCACGCGAGTGTGAGCTAATCTCCAAAAGCCATCTCAGTTCGGATTGCACTCTG
O. anthropi         CAGCGAGCACGCGAGTGTGAGCTAATCTCCAAAAGCCATCTCAGTTCGGATTGCACTCTG
                    ************************************************************

                    1261      1271      1281      1291      1301      1311
O. sp. LMG20564     CAACTCGAGTGCATGAAGTTGGAATCGCTAGTAATCGCGGATCAGCATGCCGCGGTGAAT
O. sp.              CAACTCGAGTGCATGAAGTTGGAATCGCTAGTAATCGCGGATCAGCATGCCGCGGTGAAT
CDB2                CAACTCGAGTGCATGAAGTTGGAATCGCTAGTAATCGCGGATCAGCATGCCGCGGTGAAT
O. grignonense      CAACTCGAGTGCATGAAGTTGGAATCGCTAGTAATCGCGGATCAGCATGCCGCGGTGAAT
B. canis            CAACTCGAGTGCATGAAGTTGGAATCGCTAGTAATCGCGGATCAGCATGCCGCGGTGAAT
B. melitens         CAACTCGAGTGCATGAAGTTGGAATCGCTAGTAATCGCGGATCAGCATGCCGCGGTGAAT
O. anthropi         CAACTCGAGTGCATGAAGTTGGAATCGCTAGTAATCGCGGATCAGCATGCCGCGGTGAAT
                    ************************************************************

                    1321      1331      1341      1351      1361      1371
O. sp. LMG20564     ACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCATGGGAGTTGGTTCTGCCCGAAGG
O. sp.              ACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCATGGGAGTTGGTTCTGCCCGAAGG
CDB2                ACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCATGGGAGTTGGTTCTGCCCGAAGG
O. grignonense      ACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCATGGGAGTTGGTTCTGCCCGAAGG
B. canis            ACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCATGGGAGTTGGTTTTACCCGAAGG
B. melitens         ACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCATGGGAGTTGGTTTTACCCGAAGG
O. anthropi         ACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCATGGGAGTTGGTTTTACCCGAAGG
                    ************************************************************

                    1381      1391      1401      1411      1421      1431
O. sp. LMG20564     CACTGTGCTAACCGTAAGGAGGCAGGTGACCACGGTAGGGTCAGCGACTGGGGTGAAGTC
O. sp.              CACTGTGCTAACCGTAAGGAGGCAGGTGACCACGGTAGGGTCAGCGACTGGGGTGAAGTC
CDB2                CTT-------------------------------------------------------
O. grignonense      CACTGTGCTAACCGTA-----------------------------------------
B. canis            CGCTGTGCTAACCGCAAGGAGGCAGGCGACCACGGTAGGKTCAGCGACTGGGGTGAAGTC
B. melitens         CGCTGTGCTAACCGCAAGGAGGCAGGCGACCACGGTAGGGTCAGCGACTGGGGTG-----
O. anthropi         CGCTGTGCTAACCGCAAGGAGGCAGGCGACCACGGTAGGGTCAGCGACTGGGGTGAAG--
                    *

                    1441      1451      1461      1471
O. sp. LMG20564     GTAACAAGGTAGCCGTAGGGGAAC-TGC-----
O. sp.              GTAACAAGGTAGCCGTAGGGGAACCTGCGGTTG
CDB2                --------------------------------
O. grignonense      --------------------------------
B. canis            GC------------------------------
B. melitens         --------------------------------
O. anthropi         --------------------------------
```

CDB3:

```
                    1         11        21        31        41        51
B. thuringinesis    TTATTGGAGAGTTTGATCCTGGCTCAGGATGAACGCTGGCGGCGTGCCTAATACAT-GCA
B. anthracis        TTATTGGAGAGTTTGATCCTGGCTCAGGATGAACGCTGGCGGCGTGCCTAATACAT-GCA
B. cereus           --------------------GGCTCAGGATGAACGCTGGCGGCGTGCCTAATACAT-GCA
B. cereus G8639     TTATTGGAGAGTTTGATCCTGGCTCAGGATGAACGCTGGCGGCGTGCCTAATACAT-GCA
CDB3                ----------------------------GAGTAATC--GGCTGAGT----A-TACGTAGCA
                                                 **  **    *** * **    * *** * ***

                    61        71        81        91        101       111
B. thuringinesis    AGTCGAGCGAATGGATTAAGAGCTTGCTCTTATGAAGTTAGCGGCGGACGGGTGAGTAAC
B. anthracis        AGTCGAGCGAATGGATTAAGAGCTTGCTCTTATGAAGTTAGCGGCGGACGGGTGAGTAAC
B. cereus           AGTCGAGCGAATGGATTAAGAGCTTGCTCTTATGAAGTTAGCGGCGGACGGGTGAGTAAC
B. cereus G8639     AGTCGAGCGAATGGATTAAGAGCTTGCTCTTATGAAGTTAGCGGCGGACGGGTGAGTAAC
CDB3                AGTCGAGCGAATGGATCCGTAGCTTGCTCTTATGAAGTTAGCGGCGGACGGGTGAGTAAC
                    ************************************************************

                    121       131       141       151       161       171
B. thuringinesis    ACGTGGGTAACCTGCCCATAAGACTGGGATAACTCCGGGAAACCGGGGCTAATACCGGAT
B. anthracis        ACGTGGGTAACCTGCCCATAAGACTGGGATAACTCCGGGAAACCGGGGCTAATACCGGAT
B. cereus           ACGTGGGTAACCTGCCCATAAGACTGGGATAACTCCGGGAAACCGGGGCTAATACCGGAT
B. cereus G8639     ACGTGGGTAACCTGCCCATAAGACTGGGATAACTCCGGGAAACCGGGGCTAATACCGGAT
CDB3                ACGTGGGTAACCTGCCCATAAGACTGGGATAACTCCGGGAAACCGGGGCTAATACCGGAT
                    ************************************************************
```

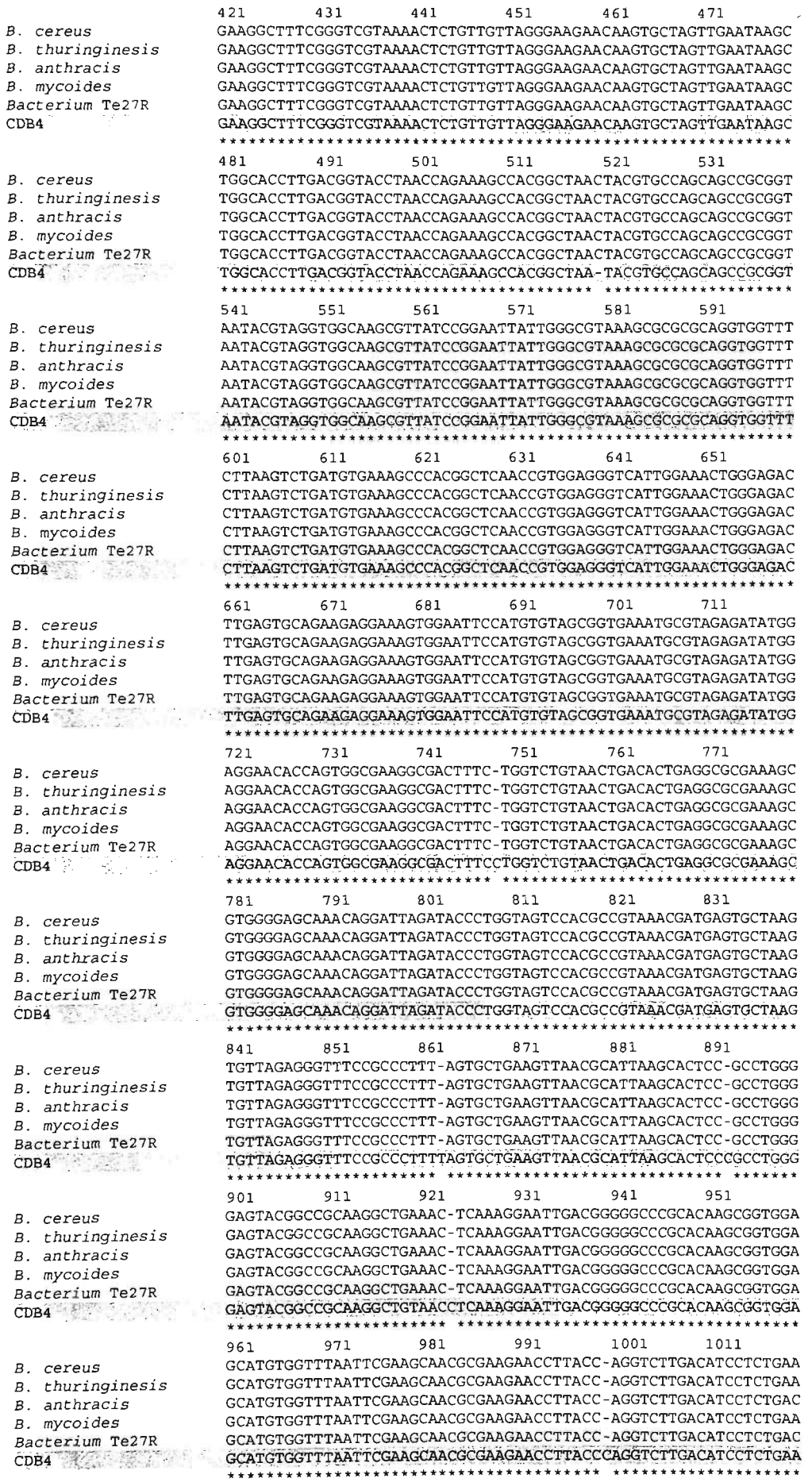
```
                    181       191       201       211       221       231
B. thuringinesis    AACATTTTGAACYGCATGGTTCGAAATTGAAAGGCGGCTTCGGCTGTCACTTATGGATGG
B. anthracis        AACATTTTGAACCGCATGGTTCGAAATTGAAAGGCGGCTTCGGCTGTCACTTATGGATGG
B. cereus           AACATTTTGAACTGCATGGTTCGAAATTGAAAGGCGGCTTCGGCTGTCACTTATGGATGG
B. cereus G8639     AAYATTTTGAACTGCATGGTTCGAAATTGAAAGGCGGCTTCGGCTGTCACTTATGGATGG
CDB3                AATATTTTGAACTGCATGGTTCGAAATTGAAAGGCGGCTTCGGCTGTCACTTATGGATGG
                    ** ************************************************************

                    241       251       261       271       281       291
B. thuringinesis    ACCCGCGTCGCATTAGCTAGTTGGTGAGGTAACGGCTCACCAAGGCAACGATGCGTAGCC
B. anthracis        ACCCGCGTCGCATTAGCTAGTTGGTGAGGTAACGGCTCACCAAGGCAACGATGCGTAGCC
B. cereus           ACCCGCGTCGCATTAGCTAGTTGGTGAGGTAACGGCTCACCAAGGCAACGATGCGTAGCC
B. cereus G8639     ACCCGCGTCGCATTAGCTAGTTGGTGAGGTAACGGCTCACCAAGGCAACGATGCGTAGCC
CDB3                ACCCGCGTCGCATTAGCTAGTTGGTGAGGTAACGGCTCACCAAGGCAACGATGCGTAGCC
                    ************************************************************

                    301       311       321       331       341       351
B. thuringinesis    GACCTGAGAGGGTGATCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGG
B. anthracis        GACCTGAGAGGGTGATCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGG
B. cereus           GACCTGAGAGGGTGATCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGG
B. cereus G8639     GACCTGAGAGGGTGATCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGG
CDB3                GACCTGAGAGGGTGATCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGG
                    ************************************************************

                    361       371       381       391       401       411
B. thuringinesis    CAGCAGTAGGGAATCTTCCGCAATGGACGAAAGTCTGACGGAGCAACGCCGCGTGAGTGA
B. anthracis        CAGCAGTAGGGAATCTTCCGCAATGGACGAAAGTCTGACGGAGCAACGCCGCGTGAGTGA
B. cereus           CAGCAGTAGGGAATCTTCCGCAATGGACGAAAGTCTGACGGAGCAACGCCGCGTGAGTGA
B. cereus G8639     CAGCAGTAGGGAATCTTCCGCAATGGACGAAAGTCTGACGGAGCAACGCCGCGTGAGTGA
CDB3                CAGCAGTAGGGAATCTTCCGCAATGGACGAAAGTCTGACGGAGCAACGCCGCGTGAGTGA
                    ************************************************************

                    421       431       441       451       461       471
B. thuringinesis    TGAAGGCTTTCGGGTCGTAAAACTCTGTTGTTAGGGAAGAACAAGTGCTAGTTGAATAAG
B. anthracis        TGAAGGCTTTCGGGTCGTAAAACTCTGTTGTTAGGGAAGAACAAGTGCTAGTTGAATAAG
B. cereus           TGAAGGCTTTCGGGTCGTAAAACTCTGTTGTTAGGGAAGAACAAGTGCTAGTTGAATAAG
B. cereus G8639     TGAAGGCTTTCGGGTCGTAAAACTCTGTTGTTAGGGAAGAACAAGTGCTAGTTGAATAAG
CDB3                TGAAGGCTTTCGGGTCGTAAAACTCTGTTGTTAGGGAAGAACAAGTGCTAGTTGAATAAG
                    ************************************************************

                    481       491       501       511       521       531
B. thuringinesis    CTGGCACCTTGACGGTACCTAACCAGAAAGCCACGGCTAACTACGTGCCAGCAGCCGCGG
B. anthracis        CTGGCACCTTGACGGTACCTAACCAGAAAGCCACGGCTAACTACGTGCCAGCAGCCGCGG
B. cereus           CTGGCACCTTGACGGTACCTAACCAGAAAGCCACGGCTAACTACGTGCCAGCAGCCGCGG
B. cereus G8639     CTGGCACCTTGACGGTACCTAACCAGAAAGCCACGGCTAACTACGTGCCAGCAGCCGCGG
CDB3                CTGGCACCTTGACGGTACCTAACCAGAAAGCCACGGCTAACTACGTGCCAGCAGCCGCGG
                    ************************************************************

                    541       551       561       571       581       591
B. thuringinesis    TAATACGTAGGTGGCAAGCGTTATCCGGAATTATTGGGCGTAAAGCGCGCGCAGGTGGTT
B. anthracis        TAATACGTAGGTGGCAAGCGTTATCCGGAATTATTGGGCGTAAAGCGCGCGCAGGTGGTT
B. cereus           TAATACGTAGGTGGCAAGCGTTATCCGGAATTATTGGGCGTAAAGCGCGCGCAGGTGGTT
B. cereus G8639     TAATACGTAGGTGGCAAGCGTTATCCGGAATTATTGGGCGTAAAGCGCGCGCAGGTGGTT
CDB3                TAATACGTAGGTGGCAAGCGTTATCCGGAATTATTGGGCGTAAAGCGCGCGCAGGTGGTT
                    ************************************************************

                    601       611       621       631       641       651
B. thuringinesis    TCTTAAGTCTGATGTGAAAGCCCACGGCTCAACCGTGGAGGGTCATTGGAAACTGGGAGA
B. anthracis        TCTTAAGTCTGATGTGAAAGCCCACGGCTCAACCGTGGAGGGTCATTGGAAACTGGGAGA
B. cereus           TCTTAAGTCTGATGTGAAAGCCCACGGCTCAACCGTGGAGGGTCATTGGAAACTGGGAGA
B. cereus G8639     TCTTAAGTCTGATGTGAAAGCCCACGGCTCAACCGTGGAGGGTCATTGGAAACTGGGAGA
CDB3                TCTTAAGTCTGATGTGAAAGCCCACGGCTCAACCGTGGAGGGTCATTGGAAACTGGGAGA
                    ************************************************************

                    661       671       681       691       701       711
B. thuringinesis    CTTGAGTGCAGAAGAGGAAAGTGGAATTCCATGTGTAGCGGTGAAATGCGTAGAGATATG
B. anthracis        CTTGAGTGCAGAAGAGGAAAGTGGAATTCCATGTGTAGCGGTGAAATGCGTAGAGATATG
B. cereus           CTTGAGTGCAGAAGAGGAAAGTGGAATTCCATGTGTAGCGGTGAAATGCGTAGAGATATG
B. cereus G8639     CTTGAGTGCAGAAGAGGAAAGTGGAATTCCATGTGTAGCGGTGAAATGCGTAGAGATATG
CDB3                CTTGAGTGCAGAAGAGGAAAGTGGAATTCCATGTGTAGCGGTGAAATGCGTAGAGATATG
                    ************************************************************

                    721       731       741       751       761       771
B. thuringinesis    GAGGAACACCAGTGGCGAAGGCGACTTTCTGGTCTGTAACTGACACTGAGGCGCGAAAGC
B. anthracis        GAGGAACACCAGTGGCGAAGGCGACTTTCTGGTCTGTAACTGACACTGAGGCGCGAAAGC
B. cereus           GAGGAACACCAGTGGCGAAGGCGACTTTCTGGTCTGTAACTGACACTGAGGCGCGAAAGC
B. cereus G8639     GAGGAACACCAGTGGCGAAGGCGACTTTCTGGTCTGTAACTGACACTGAGGCGCGAAAGC
CDB3                GAGGAACACCAGTGGCGAAGGCGACTTTCTGGTCTGTAACTGACACTGAGGCGCGAAAGC
                    ************************************************************

                    781       791       801       811       821       831
B. thuringinesis    GTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGAGTGCTAAG
B. anthracis        GTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGAGTGCTAAG
B. cereus           GTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGAGTGCTAAG
B. cereus G8639     GTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGAGTGCTAAG
CDB3                GTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGAGTGCTAAG
                    ************************************************************
```

```
                    841       851       861       871       881       891
B. thuringinesis    TGTTAGAGGGTTTCCGCCCTTTAGTGCTGAAGTTAACGCATTAAGCACTCCGCCT-GGGG
B. anthracis        TGTTAGAGGGTTTCCGCCCTTTAGTGCTGAAGTTAACGCATTAAGCACTCCGCCT-GGGG
B. cereus           TGTTAGAGGGTTTCCGCCCTTTAGTGCTGAAGTTAACGCATTAAGCACTCCGCCT-GGGG
B. cereus G8639     TGTTAGAGGGTTTCCGCCCTTTAGTGCTGAAGTTAACGCATTAAGCACTCCGCCT-GGGG
CDB3                TGTTAGAGGGTTTCCGCCCTTTAGTGCTGAAGT-AACGCATTA-GCACTCCGCCTTGGGG
                    ********************************* ********* ********** ****
                    901       911       921       931       941       951
B. thuringinesis    AGTACGGCCGCAAGGCTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGTGGAGC
B. anthracis        AGTACGGCCGCAAGGCTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGTGGAGC
B. cereus           AGTACGGCCGCAAGGCTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGTGGAGC
B. cereus G8639     AGTACGGCCGCAAGGCTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGTGGAGC
CDB3                AGTTCGGCCGCAAGGCTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGTGGAGC
                    ************************************************************
                    961       971       981       991       1001      1011
B. thuringinesis    ATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACC-AGGTC-TTGACATCCTCTGAAA
B. anthracis        ATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACC-AGGTC-TTGACATCCTCTGACA
B. cereus           ATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACC-AGGTC-TTGACATCCTCTGAAA
B. cereus G8639     ATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACC-AGGTC-TTGACATCCTCTGAAA
CDB3                ATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCCAGGTCCTTGACATCCTCTGAAA
                    ********************************** ***** ******************
                    1021      1031      1041      1051      1061      1071
B. thuringinesis    ACCCTAGAGATAGGGCTTCTCC-TTCGGGAGCAGAGTGACAGGTGGTGCATGGTTGTCGT
B. anthracis        ACCCTAGAGATAGGGCTTCTCC-TTCGGGAGCAGAGTGACAGGTGGTGCATGGTTGTCGT
B. cereus           ACCCTAGAGATAGGGCTTCTCC-TTCGGGAGCAGAGTGACAGGTGGTGCATGGTTGTCGT
B. cereus G8639     ACCCTAGAGATAGGGCTTCTCC-TTCGGGAGCAGAGTGACAGGTGGTGCATGGTTGTCGT
CDB3                ACCCTAGAGATAGGGCTTCTCCCTTCGGGAGCAGAGTGACAGGTGGTGCATGGTTGTCGT
                    ********************* ************************************
                    1081      1091      1101      1111      1121      1131
B. thuringinesis    CAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTGATCTTAGT
B. anthracis        CAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTGATCTTAGT
B. cereus           CAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTGATCTTAGT
B. cereus G8639     CAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTGATCTTAGT
CDB3                CAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTGATCTTAGT
                    ************************************************************
                    1141      1151      1161      1171      1181      1191
B. thuringinesis    TGCCATCATTAAGTTGGGCACTCTAAGGTGACTGCCGGTGACAAACCGGAGGAAGGTGGG
B. anthracis        TGCCATCATTWAGTTGGGCACTCTAAGGTGACTGCCGGTGACAAACCGGAGGAAGGTGGG
B. cereus           TGCCATCATTAAGTTGGGCACTCTAAGGTGACTGCCGGTGACAAACCGGAGGAAGGTGGG
B. cereus G8639     TGCCATCATTAAGTTGGGCACTCTAAGGTGACTGCCGGTGACAAACCGGAGGAAGGTGGG
CDB3                TGCCATCATTAAGTTGGGCACTCTAAGGTGACTGCCGGTGACAAACCGGAGGAAGGTGGG
                    ********** ************************************************
                    1201      1211      1221      1231      1241      1251
B. thuringinesis    GATGACGTCAAATCATCATGCCCCTTATGACCTGGGCTACACACGTGCTACAATGGACGG
B. anthracis        GATGACGTCAAATCATCATGCCCCTTATGACCTGGGCTACACACGTGCTACAATGGACGG
B. cereus           GATGACGTCAAATCATCATGCCCCTTATGACCTGGGCTACACACGTGCTACAATGGACGG
B. cereus G8639     GATGACGTCAAATCATCATGCCCCTTATGACCTGGGCTACACACGTGCTACAATGGACGG
CDB3                GATGACGTCAAATCATCATGCCCCTTATGACCTGGGCTACACACGTGCTACAATGGACGG
                    ************************************************************
                    1261      1271      1281      1291      1301      1311
B. thuringinesis    TACAAAGAGCTGCAAGACCGCGAGGTGGAGCTAATCTCATAAAACCGTTCTCAGTTCGGA
B. anthracis        TACAAAGAGCTGCAAGACCGCGAGGTGGAGCTAATCTCATAAAACCGTTCTCAGTTCGGA
B. cereus           TACAAAGAGCTGCAAGACCGCGAGGTGGAGCTAATCTCATAAAACCGTTCTCAGTTCGGA
B. cereus G8639     TACAAAGAGCTGCAAGACCGCGAGGTGGAGCTAATCTCATAAAACCGTTCTCAGTTCGGA
CDB3                TACAAAGAGCTGCAAGACCGCGAGGTGGAGCTAATCTCATAAAACCGTTCTCAGTTCGGA
                    ************************************************************
                    1321      1331      1341      1351      1361      1371
B. thuringinesis    TTGTAGGCTGCAACTCGCCTACATGAAGCTGGAATCGCTAGTAATCGCGGATCAGCATGC
B. anthracis        TTGTAGGCTGCAACTCGCCTACATGAAGCTGGAATCGCTAGTAATCGCGGATCAGCATGC
B. cereus           TTGTAGGCTGCAACTCGCCTACATGAAGCTGGAATCGCTAGTAATCGCGGATCAGCATGC
B. cereus G8639     TTGTAGGCTGCAACTCGCCTACATGAAGCTGGAATCGCTAGTAATCGCGGATCAGCATGC
CDB3                TTGTAGGCTGCAACTCGCCTACATGAAGCTGGAATCGCTAGTAATCGCGGATCAGCATGC
                    ************************************************************
                    1381      1391      1401      1411      1421      1431
B. thuringinesis    CGCGGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGT-CACACCACGAGAGTTTGTA
B. anthracis        CGCGGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGT-CACACCACGAGAGTTTGTA
B. cereus           CGCGGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGT-CACACCACGAGAGTTTGTA
B. cereus G8639     CGCGGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGT-CACACCACGAGAGTTTGTA
CDB3                CGCGGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTTCACACCACGAGAGTTTGTA
                    ************************************* ******************
                    1441      1451      1461      1471      1481      1491
B. thuringinesis    ACACCCGAAGTCGGTGGGGTAACCTTTTTGGAGCCAGCCGCCTA-AGGTGGGACAGATGA
B. anthracis        ACACCCGAAGTCGGTGGGGTAACCTTTTTGGAGCCAGCCGCCTA-AGGTGGGACAGATGA
B. cereus           ACACCCGAAGTCGGTGGGGTAACCTTTTTGGAGCCAGCCGCCTA-AGGTGGGACAGATGA
B. cereus G8639     ACACCCGAAGTCGGTGGGGTAACCTTTTTGGAGCCAGCCGCCTA-AGGTGGGACAGATGA
CDB3                ACACCCGAAGTCGTTAGATTT-CCATCCTGGATCCGGCTAGATATAGATCCGGTACATAC
                    ******************  ********************** ** *  *  * **
```

```
                    1501       1511       1521       1531       1541       1551
B. thuringinesis    TTGGGGTGAAGTCGTAACAAGGTAGCCGTATCGGAAGGTGCGGCTGGATCACCTCCTTTC
B. anthracis        TTGGGGTGAAGTCGTAACAAGGTAGCCGTATCGGAAGGTGCGGCTGGATCACCTCCTTTC
B. cereus           TTGGGGTGAAGTCGTAACAAGGTAGCCGTATCGGAAGGTGC------------------
B. cereus G8639     TTGGGGTGAAGTCGTAACAAGGTAGCCGTATCGGAAGGTGCGGCTGGATCACCTCCTTTC
CDB3                CAG-------------------------------------------------------
                      *
                    1561
B. thuringinesis    T
B. anthracis        T
B. cereus           -
B. cereus G8639     T
CDB3                -
```

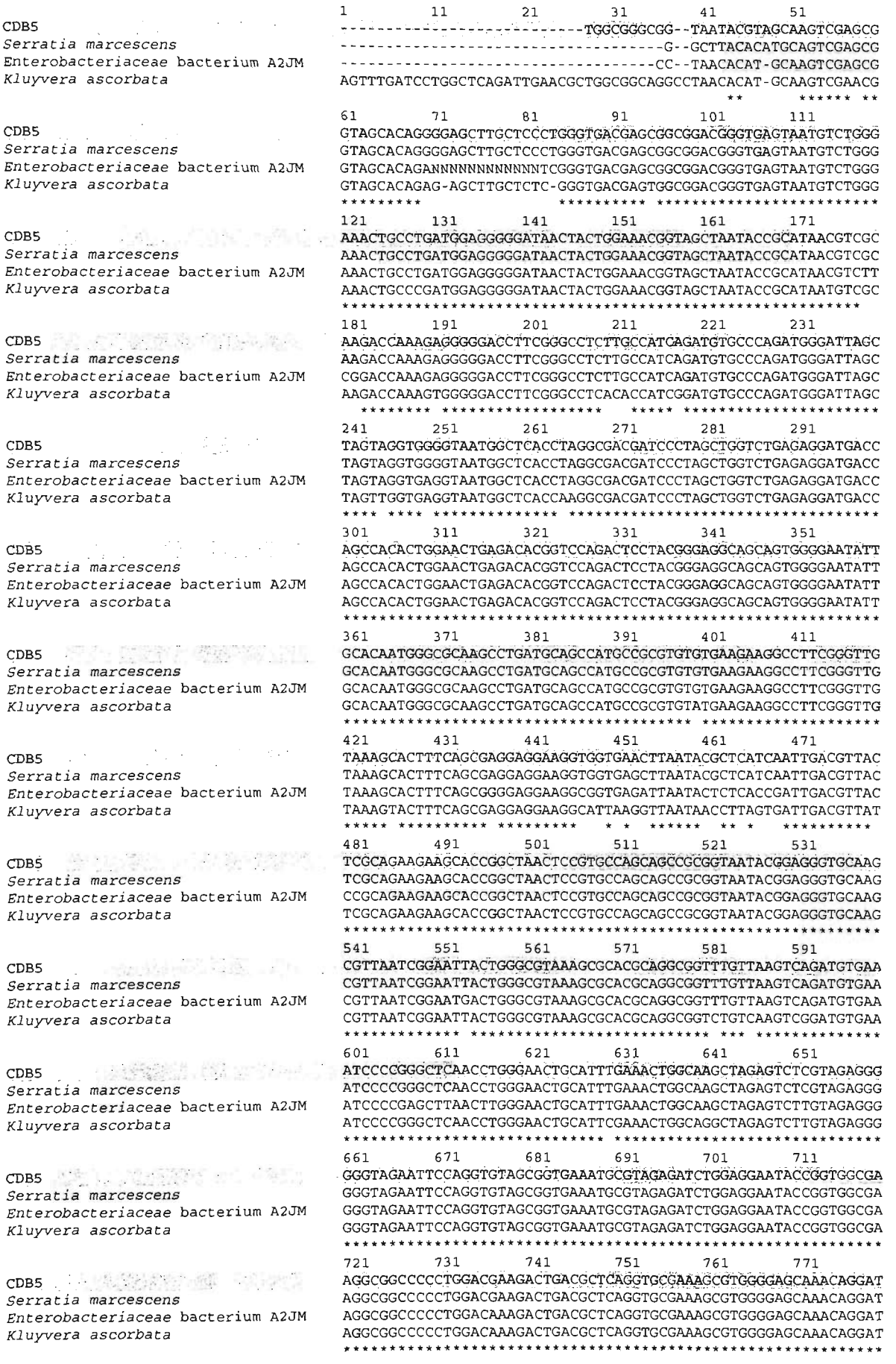
CDB4:

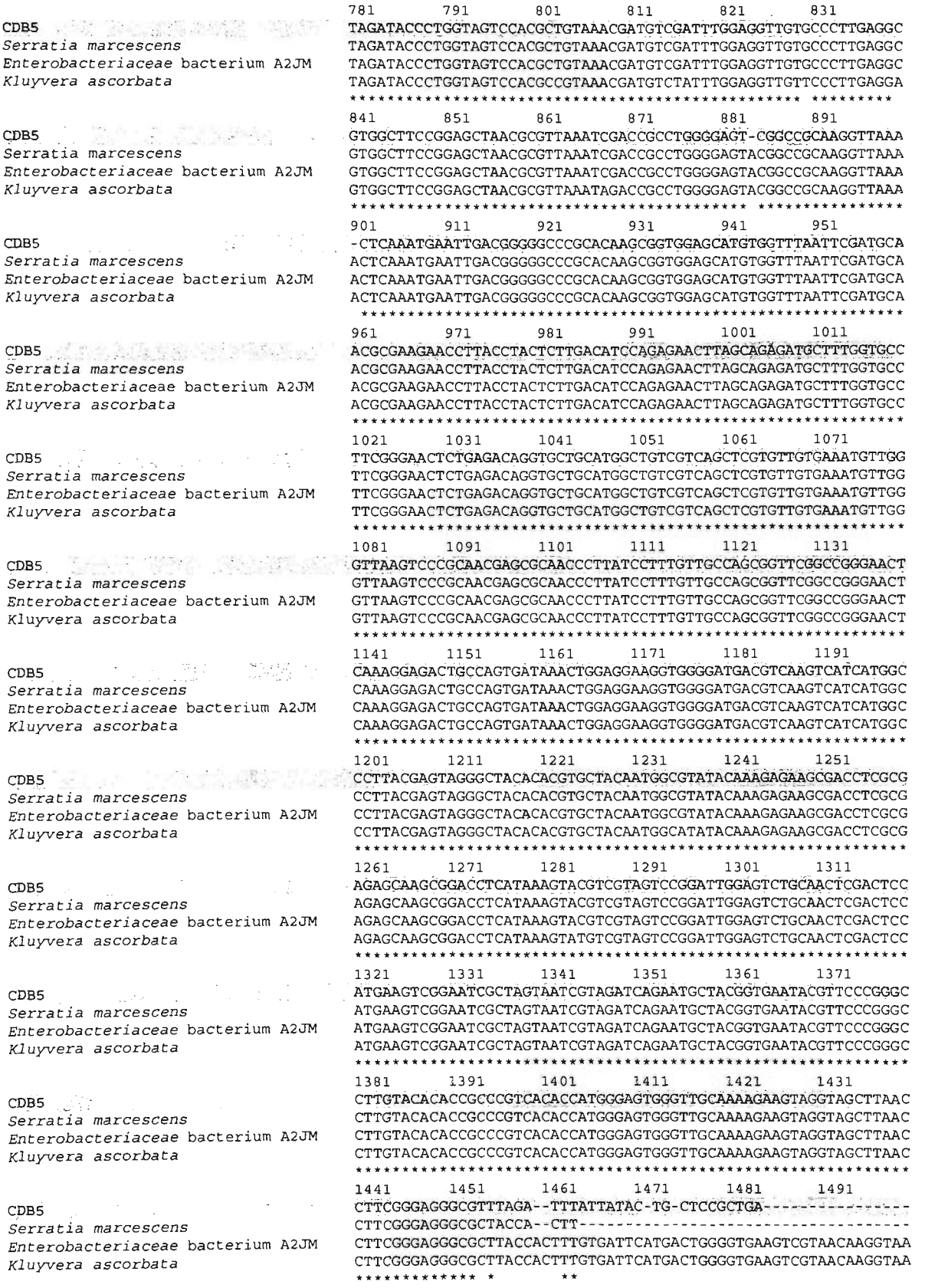
```
                    1          11         21         31         41         51
B. cereus           --------------------GGCTCAGGATGAACGCTGGCGGCGTGCCTAATACATGCAA
B. thuringinesis    --------------------CTGGCTCAGGATGAACGCTGGCGGCGTGCCTAATACATGCAA
B. anthracis        TTATTGGAGAGTTTGATCCTGGCTCAGGATGAACGCTGGCGGCGTGCCTAATACATGCAA
B. mycoides         ------------------------------------------------TAATACATGCAA
Bacterium Te27R     -----TTAGAKTTTGATCATGGCTCAGGATGAACGCTGGCGGCGTGCTTAATACATGCAA
CDB4                ------------------------------------ACAGGGAGCCGA------GCAA
                                                                 *      ****

                    61         71         81         91         101        111
B. cereus           GTCGAGCGAATGGATTAAGAGCTTGCTCTTATGAAGTTAGCGGCGGACGGGTGAGTAACA
B. thuringinesis    GTCGAGCGAATGGATTAAGAGCTTGCTCTTATGAAGTTAGCGGCGGACGGGTGAGTAACA
B. anthracis        GTCGAGCGAATGGATTAAGAGCTTGCTCTTATGAAGTTAGCGGCGGACGGGTGAGTAACA
B. mycoides         GTCGAGCGAATGGATTAAGAGCTTGCTCTTATGAAGTTAGCGGCGGACGGGTGAGTAACA
Bacterium Te27R     GTCGAGCGAATGGATTAAGAGCTTGCTCTTATGAAGTTAGCGGCGGACGGGTGAGTAACA
CDB4                G--CAGCGAA-GGAT-AAGAGCTTGCTCTTATGAAGTTAGCGGCGGACGGGTGAGTAACA
                    *   ****** **** ******************************************

                    121        131        141        151        161        171
B. cereus           CGTGGGTAACCTGCCCATAAGACTGGGATAACTCCGGGAAACCGGGGCTAATACCGGATA
B. thuringinesis    CGTGGGTAACCTGCCCATAAGACTGGGATAACTCCGGGAAACCGGGGCTAATACCGGATA
B. anthracis        CGTGGGTAACCTGCCCATAAGACTGGGATAACTCCGGGAAACCGGGGCTAATACCGGATA
B. mycoides         CGTGGGTAACCTACCCATAAGACTGGGATAACTCCGGGAAACCGGGGCTAATACCGGATG
Bacterium Te27R     CGTGGGTAACCTGCCCATAAGACTGGGATAACTCCGGGAAACCGGGGCTAATACCGGATA
CDB4                CGTGGGTAACCTGCCCATAAGACTGGGATAACTCCGGGAAACCGGGGCTAATACCGGATA
                    ************************************************************

                    181        191        201        211        221        231
B. cereus           ACATTTTGAACTGCATGGTTCGAAATTGAAAGGCGGCTTCGGCTGTCACTTATGGATGGA
B. thuringinesis    ACATTTTGAACYGCATGGTTCGAAATTGAAAGGCGGCTTCGGCTGTCACTTATGGATGGA
B. anthracis        ACATTTTGAACCGCATGGTTCGAAATTGAAAGGCGGCTTCGGCTGTCACTTATGGATGGA
B. mycoides         ATATTTTGAACTGCATAGTTCGAAATTGAAAGGCGGCTTCGGCTGTCACTTATGGATGGA
Bacterium Te27R     ACATTTTGAACYGCATGGTTCGAAATTGAAAGGCGGCTTCGGCTGTCACTTATGGATGGA
CDB4                ACATTTTGAACTGCATGGTTCGAAATTGAAAGGCGGCTTCGGCTGTCACTTATGGATGGA
                    *********** ************************************************

                    241        251        261        271        281        291
B. cereus           CCCGCGTCGCATTAGCTAGTTGGTGAGGTAACGGCTCACCAAGGCAACGATGCGTAGCCG
B. thuringinesis    CCCGCGTCGCATTAGCTAGTTGGTGAGGTAACGGCTCACCAAGGCAACGATGCGTAGCCG
B. anthracis        CCCGCGTCGCATTAGCTAGTTGGTGAGGTAACGGCTCACCAAGGCAACGATGCGTAGCCG
B. mycoides         CCCGCGTCGCATTAGCTAGTTGGTGAGGTAACGGCTCACCAAGGCGACGATGCGTAGCCG
Bacterium Te27R     CCCGCGTCGCATTAGCTAGTTGGTGAGGTAACGGCTCACCAAGGCAACGATGCGTAGCCG
CDB4                CCCGCGTCGCATTAGCTAGTTGGTGAGGTAACGGCTCACCAAGGCAACGATGCGTAGCCG
                    ************************************************************

                    301        311        321        331        341        351
B. cereus           ACCTGAGAGGGTGATCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGC
B. thuringinesis    ACCTGAGAGGGTGATCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGC
B. anthracis        ACCTGAGAGGGTGATCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGC
B. mycoides         ACCTGAGAGGGTGATCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGC
Bacterium Te27R     ACCTGAGAGGGTGATCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGC
CDB4                ACCTGAGAGGGTGATCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGC
                    ************************************************************

                    361        371        381        391        401        411
B. cereus           AGCAGTAGGGAATCTTCCGCAATGGACGAAAGTCTGACGGAGCAACGCCGCGTGAGTGAT
B. thuringinesis    AGCAGTAGGGAATCTTCCGCAATGGACGAAAGTCTGACGGAGCAACGCCGCGTGAGTGAT
B. anthracis        AGCAGTAGGGAATCTTCCGCAATGGACGAAAGTCTGACGGAGCAACGCCGCGTGAGTGAT
B. mycoides         AGCAGTAGGGAATCTTCCGCAATGGACGAAAGTCTGACGGAGCAACGCCGCGTGAGTGAT
Bacterium Te27R     AGCAGTAGGGAATCTTCCGCAATGGACGAAAGTCTGACGGAGCAACGCCGCGTGAGTGAT
CDB4                AGCAGTAGGGAATCTTCCGCAATGGACGAAAGTCTGACGGAGCAACGCCGCGTGAGTGAT
                    ************************************************************
```


421 431 441 451 461 471
B. cereus GAAGGCTTTCGGGTCGTAAAACTCTGTTGTTAGGGAAGAACAAGTGCTAGTTGAATAAGC
B. thuringinesis GAAGGCTTTCGGGTCGTAAAACTCTGTTGTTAGGGAAGAACAAGTGCTAGTTGAATAAGC
B. anthracis GAAGGCTTTCGGGTCGTAAAACTCTGTTGTTAGGGAAGAACAAGTGCTAGTTGAATAAGC
B. mycoides GAAGGCTTTCGGGTCGTAAAACTCTGTTGTTAGGGAAGAACAAGTGCTAGTTGAATAAGC
Bacterium Te27R GAAGGCTTTCGGGTCGTAAAACTCTGTTGTTAGGGAAGAACAAGTGCTAGTTGAATAAGC
CDB4 GAAGGCTTTCGGGTCGTAAAACTCTGTTGTTAGGGAAGAACAAGTGCTAGTTGAATAAGC
481 491 501 511 521 531
B. cereus TGGCACCTTGACGGTACCTAACCAGAAAGCCACGGCTAACTACGTGCCAGCAGCCGCGGT
B. thuringinesis TGGCACCTTGACGGTACCTAACCAGAAAGCCACGGCTAACTACGTGCCAGCAGCCGCGGT
B. anthracis TGGCACCTTGACGGTACCTAACCAGAAAGCCACGGCTAACTACGTGCCAGCAGCCGCGGT
B. mycoides TGGCACCTTGACGGTACCTAACCAGAAAGCCACGGCTAACTACGTGCCAGCAGCCGCGGT
Bacterium Te27R TGGCACCTTGACGGTACCTAACCAGAAAGCCACGGCTAACTACGTGCCAGCAGCCGCGGT
CDB4 TGGCACCTTGACGGTACCTAACCAGAAAGCCACGGCTAA-TACGTGCCAGCAGCCGCGGT
541 551 561 571 581 591
B. cereus AATACGTAGGTGGCAAGCGTTATCCGGAATTATTGGGCGTAAAGCGCGCGCAGGTGGTTT
B. thuringinesis AATACGTAGGTGGCAAGCGTTATCCGGAATTATTGGGCGTAAAGCGCGCGCAGGTGGTTT
B. anthracis AATACGTAGGTGGCAAGCGTTATCCGGAATTATTGGGCGTAAAGCGCGCGCAGGTGGTTT
B. mycoides AATACGTAGGTGGCAAGCGTTATCCGGAATTATTGGGCGTAAAGCGCGCGCAGGTGGTTT
Bacterium Te27R AATACGTAGGTGGCAAGCGTTATCCGGAATTATTGGGCGTAAAGCGCGCGCAGGTGGTTT
CDB4 AATACGTAGGTGGCAAGCGTTATCCGGAATTATTGGGCGTAAAGCGCGCGCAGGTGGTTT
601 611 621 631 641 651
B. cereus CTTAAGTCTGATGTGAAAGCCCACGGCTCAACCGTGGAGGGTCATTGGAAACTGGGAGAC
B. thuringinesis CTTAAGTCTGATGTGAAAGCCCACGGCTCAACCGTGGAGGGTCATTGGAAACTGGGAGAC
B. anthracis CTTAAGTCTGATGTGAAAGCCCACGGCTCAACCGTGGAGGGTCATTGGAAACTGGGAGAC
B. mycoides CTTAAGTCTGATGTGAAAGCCCACGGCTCAACCGTGGAGGGTCATTGGAAACTGGGAGAC
Bacterium Te27R CTTAAGTCTGATGTGAAAGCCCACGGCTCAACCGTGGAGGGTCATTGGAAACTGGGAGAC
CDB4 CTTAAGTCTGATGTGAAAGCCCACGGCTCAACCGTGGAGGGTCATTGGAAACTGGGAGAC
661 671 681 691 701 711
B. cereus TTGAGTGCAGAAGAGGAAAGTGGAATTCCATGTGTAGCGGTGAAATGCGTAGAGATATGG
B. thuringinesis TTGAGTGCAGAAGAGGAAAGTGGAATTCCATGTGTAGCGGTGAAATGCGTAGAGATATGG
B. anthracis TTGAGTGCAGAAGAGGAAAGTGGAATTCCATGTGTAGCGGTGAAATGCGTAGAGATATGG
B. mycoides TTGAGTGCAGAAGAGGAAAGTGGAATTCCATGTGTAGCGGTGAAATGCGTAGAGATATGG
Bacterium Te27R TTGAGTGCAGAAGAGGAAAGTGGAATTCCATGTGTAGCGGTGAAATGCGTAGAGATATGG
CDB4 TTGAGTGCAGAAGAGGAAAGTGGAATTCCATGTGTAGCGGTGAAATGCGTAGAGATATGG
721 731 741 751 761 771
B. cereus AGGAACACCAGTGGCGAAGGCGACTTTC-TGGTCTGTAACTGACACTGAGGCGCGAAAGC
B. thuringinesis AGGAACACCAGTGGCGAAGGCGACTTTC-TGGTCTGTAACTGACACTGAGGCGCGAAAGC
B. anthracis AGGAACACCAGTGGCGAAGGCGACTTTC-TGGTCTGTAACTGACACTGAGGCGCGAAAGC
B. mycoides AGGAACACCAGTGGCGAAGGCGACTTTC-TGGTCTGTAACTGACACTGAGGCGCGAAAGC
Bacterium Te27R AGGAACACCAGTGGCGAAGGCGACTTTC-TGGTCTGTAACTGACACTGAGGCGCGAAAGC
CDB4 AGGAACACCAGTGGCGAAGGCGACTTTCCTGGTCTGTAACTGACACTGAGGCGCGAAAGC
781 791 801 811 821 831
B. cereus GTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGAGTGCTAAG
B. thuringinesis GTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGAGTGCTAAG
B. anthracis GTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGAGTGCTAAG
B. mycoides GTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGAGTGCTAAG
Bacterium Te27R GTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGAGTGCTAAG
CDB4 GTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGAGTGCTAAG
841 851 861 871 881 891
B. cereus TGTTAGAGGGTTTCCGCCCTTT-AGTGCTGAAGTTAACGCATTAAGCACTCC-GCCTGGG
B. thuringinesis TGTTAGAGGGTTTCCGCCCTTT-AGTGCTGAAGTTAACGCATTAAGCACTCC-GCCTGGG
B. anthracis TGTTAGAGGGTTTCCGCCCTTT-AGTGCTGAAGTTAACGCATTAAGCACTCC-GCCTGGG
B. mycoides TGTTAGAGGGTTTCCGCCCTTT-AGTGCTGAAGTTAACGCATTAAGCACTCC-GCCTGGG
Bacterium Te27R TGTTAGAGGGTTTCCGCCCTTT-AGTGCTGAAGTTAACGCATTAAGCACTCC-GCCTGGG
CDB4 TGTTAGAGGGTTTCCGCCCTTTTAGTGCTGAAGTTAACGCATTAAGCACTCCCGCCTGGG
901 911 921 931 941 951
B. cereus GAGTACGGCCGCAAGGCTGAAAC-TCAAAGGAATTGACGGGGGCCCGCACAAGCGGTGGA
B. thuringinesis GAGTACGGCCGCAAGGCTGAAAC-TCAAAGGAATTGACGGGGGCCCGCACAAGCGGTGGA
B. anthracis GAGTACGGCCGCAAGGCTGAAAC-TCAAAGGAATTGACGGGGGCCCGCACAAGCGGTGGA
B. mycoides GAGTACGGCCGCAAGGCTGAAAC-TCAAAGGAATTGACGGGGGCCCGCACAAGCGGTGGA
Bacterium Te27R GAGTACGGCCGCAAGGCTGAAAC-TCAAAGGAATTGACGGGGGCCCGCACAAGCGGTGGA
CDB4 GAGTACGGCCGCAAGGCTGTAACCTCAAAGGAATTGACGGGGGCCCGCACAAGCGGTGGA
961 971 981 991 1001 1011
B. cereus GCATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACC-AGGTCTTGACATCCTCTGAA
B. thuringinesis GCATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACC-AGGTCTTGACATCCTCTGAA
B. anthracis GCATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACC-AGGTCTTGACATCCTCTGAC
B. mycoides GCATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACC-AGGTCTTGACATCCTCTGAA
Bacterium Te27R GCATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACC-AGGTCTTGACATCCTCTGAC
CDB4 GCATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCCAGGTCTTGACATCCTCTGAA

```
                    1021       1031       1041       1051       1061       1071
B. cereus           AACCCTAGAGATAGGGCTTCTCCTTCGGGAGCAGAGTGACAGGTGGTGCATGGTTGTCGT
B. thuringinesis    AACCCTAGAGATAGGGCTTCTCCTTCGGGAGCAGAGTGACAGGTGGTGCATGGTTGTCGT
B. anthracis        AACCCTAGAGATAGGGCTTCTCCTTCGGGAGCAGAGTGACAGGTGGTGCATGGTTGTCGT
B. mycoides         AACTCTAGAGATAGAGCTTCTCCTTCGGGAGCAGAGTGACAGGTGGTGCATGGTTGTCGT
Bacterium Te27R     AACCCTAGAGATAGGGCTTCTCCTTCGGGAGCAGAGTGACAGGTGGTGCATGGTTGTCGT
CDB4                AACCCTAGAGATAGGGCTTCTCCTTCGGGAGCAGAGTGACAGGTGGTGCATGGTTGTCGT
                    ************************************************************
                    1081       1091       1101       1111       1121       1131
B. cereus           CAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTGATCTTAGT
B. thuringinesis    CAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTGATCTTAGT
B. anthracis        CAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTGATCTTAGT
B. mycoides         CAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTGATCTTAGT
Bacterium Te27R     CAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTGATCTTAGT
CDB4                CAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTGATCTTAGT
                    ************************************************************
                    1141       1151       1161       1171       1181       1191
B. cereus           TGCCATCATTAAGTTGGGCACTCTAAGGTGACTGCCGGTGACAAACCGGAGGAAGGTGGG
B. thuringinesis    TGCCATCATTAAGTTGGGCACTCTAAGGTGACTGCCGGTGACAAACCGGAGGAAGGTGGG
B. anthracis        TGCCATCATTWAGTTGGGCACTCTAAGGTGACTGCCGGTGACAAACCGGAGGAAGGTGGG
B. mycoides         TGCCATCATTAAGTTGGGCACTCTAAGGTGACTGCCGGTGACAAACCGGAGGAAGGTGGG
Bacterium Te27R     TGCCATCATTAAGTTGGGCACTCTAAGGTGACTGCCGGTGACAAACCGGAGGAAGGTGGG
CDB4                TGCCATCATTAAGTTGGGCACTCTAAGGTGACTGCCGGTGACAAACCGGAGGAAGGTGGG
                    ************************************************************
                    1201       1211       1221       1231       1241       1251
B. cereus           GATGACGTCAAATCATCATGCCCCTTATGACCTGGGCTACACACGTGCTACAATGGA-CG
B. thuringinesis    GATGACGTCAAATCATCATGCCCCTTATGACCTGGGCTACACACGTGCTACAATGGA-CG
B. anthracis        GATGACGTCAAATCATCATGCCCCTTATGACCTGGGCTACACACGTGCTACAATGGA-CG
B. mycoides         GATGACGTCAAATCATCATGCCCCTTATGACCTGGGCTACACACGTGCTACAATGGA-CG
Bacterium Te27R     GATGACGTCAAATCATCATGCCCCTTATGACCTGGGCTACACACGTGCTACAATGGA-CG
CDB4                GATGACGTCAAATCATCATGCCCCTTATGACCTGGGCTACACACGTGCTACAATGGAGCG
                    ************************************************************ **
                    1261       1271       1281       1291       1301       1311
B. cereus           GTACAAAGAGCTGCAAGACCGCGAGGTGGAGCTAATCTCATAAAACCGTTCTCAGTTCGG
B. thuringinesis    GTACAAAGAGCTGCAAGACCGCGAGGTGGAGCTAATCTCATAAAACCGTTCTCAGTTCGG
B. anthracis        GTACAAAGAGCTGCAAGACCGCGAGGTGGAGCTAATCTCATAAAACCGTTCTCAGTTCGG
B. mycoides         GTACAAAGAGCTGCAAGACCGCGAGGTGGAGCTAATCTCATAAAACCGTTCTCAGTTCGG
Bacterium Te27R     GTACAAAGAGCTGCAAGACCGCGAGGTGGAGCTAATCTCATAAAACCGTTCTCAGTTCGG
CDB4                GTACAAAGAGCTGCAAGACCGCGAGGTGGAGCTAATCTCATAAAACCGTTCTCAGTTCGG
                    ************************************************************
                    1321       1331       1341       1351       1361       1371
B. cereus           ATTGTAGGCTGCAACTCGCCTACATGAAGCTGGAATCGCTAGTAATCGCGGATCAGCATG
B. thuringinesis    ATTGTAGGCTGCAACTCGCCTACATGAAGCTGGAATCGCTAGTAATCGCGGATCAGCATG
B. anthracis        ATTGTAGGCTGCAACTCGCCTACATGAAGCTGGAATCGCTAGTAATCGCGGATCAGCATG
B. mycoides         ATTGTAGGCTGCAACTCGCCTACATGAAGCTGGAATCGCTAGTAATCGCGGATCAGCATG
Bacterium Te27R     ATTGTAGGCTGCAACTCGCCTACATGAAGCTGGAATCGCTAGTAATCGCGGATCAGCATG
CDB4                ATTGTAGGCTGCAACTCGCCTACATGAAGCTGGAATCGCTAGTAATCGCGGATCAGCATG
                    ************************************************************
                    1381       1391       1401       1411       1421       1431
B. cereus           CCGCGGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCACGAGAGTTT-GT
B. thuringinesis    CCGCGGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCACGAGAGTTT-GT
B. anthracis        CCGCGGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCACGAGAGTTT-GT
B. mycoides         CCGCGGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCACGAGAGTTT-GT
Bacterium Te27R     CCGCGGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCACGAGAGTTT-GT
CDB4                CCGCGGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCACGAGAGTTT--G
                    ************************************************************
                    1441       1451       1461       1471       1481       1491
B. cereus           AACACCCGAAGTCGGTGGGGTAACCTTTTTGGAGCCAGCCGCCTAAGGTGGGACAGATGA
B. thuringinesis    AACACCCGAAGTCGGTGGGGTAACCTTTTTGGAGCCAGCCGCCTAAGGTGGGACAGATGA
B. anthracis        AACACCCGAAGTCGGTGGGGTAACCTTTTTGGAGCCAGCCGCCTAAGGTGGGACAGATGA
B. mycoides         AACACCCGAAGTCGGTGGGGTAACCTTTATGGAGCCAGCCGCCTAAGGTGGGACAGATGA
Bacterium Te27R     AACACCCGAAGTCGGTGGGGTAACCTTTTTGGAGCCAGCCCCCTAAGGTGGGACAGATGA
CDB4                AACACCCGAAGTCGGTGGGG-AACCT--ATGGAGCCAG---CCTAGAGTTTG--------
                    ******************* *****   *********   ********  *
                    1501       1511       1521       1531       1541       1551
B. cereus           TTGGGGTGAAGTCGTAACAAGGTAGCCGTATCGGAAGGTGC--------------------
B. thuringinesis    TTGGGGTGAAGTCGTAACAAGGTAGCCGTATCGGAAGGTGCGGCTGGAT-----------
B. anthracis        TTGGGGTGAAGTCGTAACAAGGTAGCCGTATCGGAAGGTGCGGCTGGATCACCTCCTTTC
B. mycoides         TTGGGGTGAAGTCGTAACAAGGTATCCGT--------------------------------
Bacterium Te27R     TTGGGGTGAAGTCGTAACAAGGGCCA-----------------------------------
CDB4                ------------------------------------------------------------

                    1561
B. cereus           -
B. thuringinesis    -
B. anthracis        T
B. mycoides         -
Bacterium Te27R     -
CDB4                -
```

CDB5:


1 11 21 31 41 51
CDB5 ----------TGGCGGGCGG--TAATACGTAGCAAGTCGAGCG
Serratia marcescens ----------G--GCTTACACATGCAGTCGAGCG
Enterobacteriaceae bacterium A2JM ----------CC--TAACACAT-GCAAGTCGAGCG
Kluyvera ascorbata AGTTTGATCCTGGCTCAGATTGAACGCTGGCGGCAGGCCTAACACAT-GCAAGTCGAACG
61 71 81 91 101 111
CDB5 GTAGCACAGGGGAGCTTGCTCCCTGGGTGACGAGCGGCGGACGGGTGAGTAATGTCTGGG
Serratia marcescens GTAGCACAGGGGAGCTTGCTCCCTGGGTGACGAGCGGCGGACGGGTGAGTAATGTCTGGG
Enterobacteriaceae bacterium A2JM GTAGCACAGANNNNNNNNNNNNTCGGGTGACGAGCGGCGGACGGGTGAGTAATGTCTGGG
Kluyvera ascorbata GTAGCACAGAG-AGCTTGCTCTC-GGGTGACGAGTGGCGGACGGGTGAGTAATGTCTGGG
121 131 141 151 161 171
CDB5 AAACTGCCTGATGGAGGGGGATAACTACTGGAAACGGTAGCTAATACCGCATAACGTCGC
Serratia marcescens AAACTGCCTGATGGAGGGGGATAACTACTGGAAACGGTAGCTAATACCGCATAACGTCGC
Enterobacteriaceae bacterium A2JM AAACTGCCTGATGGAGGGGGATAACTACTGGAAACGGTAGCTAATACCGCATAACGTCTT
Kluyvera ascorbata AAACTGCCCGATGGAGGGGGATAACTACTGGAAACGGTAGCTAATACCGCATAATGTCGC
181 191 201 211 221 231
CDB5 AAGACCAAAGAGGGGGACCTTCGGGCCTCTTGCCATCAGATGTGCCCAGATGGGATTAGC
Serratia marcescens AAGACCAAAGAGGGGGACCTTCGGGCCTCTTGCCATCAGATGTGCCCAGATGGGATTAGC
Enterobacteriaceae bacterium A2JM CGGACCAAAGAGGGGGACCTTCGGGCCTCTTGCCATCAGATGTGCCCAGATGGGATTAGC
Kluyvera ascorbata AAGACCAAAGTGGGGGACCTTCGGGCCTCACACCATCGGATGTGCCCAGATGGGATTAGC
241 251 261 271 281 291
CDB5 TAGTAGGTGGGGTAATGGCTCACCTAGGCGACGATCCCTAGCTGGTCTGAGAGGATGACC
Serratia marcescens TAGTAGGTGGGGTAATGGCTCACCTAGGCGACGATCCCTAGCTGGTCTGAGAGGATGACC
Enterobacteriaceae bacterium A2JM TAGTAGGTGAGGTAATGGCTCACCTAGGCGACGATCCCTAGCTGGTCTGAGAGGATGACC
Kluyvera ascorbata TAGTTGGTGAGGTAATGGCTCACCAAGGCGACGATCCCTAGCTGGTCTGAGAGGATGACC
301 311 321 331 341 351
CDB5 AGCCACACTGGAACTGAGACACGGTCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATT
Serratia marcescens AGCCACACTGGAACTGAGACACGGTCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATT
Enterobacteriaceae bacterium A2JM AGCCACACTGGAACTGAGACACGGTCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATT
Kluyvera ascorbata AGCCACACTGGAACTGAGACACGGTCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATT
361 371 381 391 401 411
CDB5 GCACAATGGGCGCAAGCCTGATGCAGCCATGCCGCGTGTGTGAAGAAGGCCTTCGGGTTG
Serratia marcescens GCACAATGGGCGCAAGCCTGATGCAGCCATGCCGCGTGTGTGAAGAAGGCCTTCGGGTTG
Enterobacteriaceae bacterium A2JM GCACAATGGGCGCAAGCCTGATGCAGCCATGCCGCGTGTGTGAAGAAGGCCTTCGGGTTG
Kluyvera ascorbata GCACAATGGGCGCAAGCCTGATGCAGCCATGCCGCGTGTATGAAGAAGGCCTTCGGGTTG
421 431 441 451 461 471
CDB5 TAAAGCACTTTCAGCGAGGAGGAAGGTGGTGAACTTAATACGCTCATCAATTGACGTTAC
Serratia marcescens TAAAGCACTTTCAGCGAGGAGGAAGGTGGTGAGCTTAATACGCTCATCAATTGACGTTAC
Enterobacteriaceae bacterium A2JM TAAAGCACTTTCAGCGGGGAGGAAGGCGGTGAGATTAATACTCTCACCGATTGACGTTAC
Kluyvera ascorbata TAAAGTACTTTCAGCGAGGAGGAAGGCATTAAGGTTAATAACCTTAGTGATTGACGTTAT
481 491 501 511 521 531
CDB5 TCGCAGAAGAAGCACCGGCTAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGCAAG
Serratia marcescens TCGCAGAAGAAGCACCGGCTAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGCAAG
Enterobacteriaceae bacterium A2JM CCGCAGAAGAAGCACCGGCTAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGCAAG
Kluyvera ascorbata TCGCAGAAGAAGCACCGGCTAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGCAAG
541 551 561 571 581 591
CDB5 CGTTAATCGGAATTACTGGGCGTAAAGCGCACGCAGGCGGTTTGTTAAGTCAGATGTGAA
Serratia marcescens CGTTAATCGGAATTACTGGGCGTAAAGCGCACGCAGGCGGTTTGTTAAGTCAGATGTGAA
Enterobacteriaceae bacterium A2JM CGTTAATCGGAATGACTGGGCGTAAAGCGCACGCAGGCGGTTTGTTAAGTCAGATGTGAA
Kluyvera ascorbata CGTTAATCGGAATTACTGGGCGTAAAGCGCACGCAGGCGGTCTGTCAAGTCGGATGTGAA
601 611 621 631 641 651
CDB5 ATCCCCGGGCTCAACCTGGGAACTGCATTTGAAACTGGCAAGCTAGAGTCTCGTAGAGGG
Serratia marcescens ATCCCCGGGCTCAACCTGGGAACTGCATTTGAAACTGGCAAGCTAGAGTCTCGTAGAGGG
Enterobacteriaceae bacterium A2JM ATCCCCGAGCTTAACTTGGGAACTGCATTTGAAACTGGCAAGCTAGAGTCTTGTAGAGGG
Kluyvera ascorbata ATCCCCGGGCTCAACCTGGGAACTGCATTCGAAACTGGCAGGCTAGAGTCTTGTAGAGGG
661 671 681 691 701 711
CDB5 GGGTAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGATCTGGAGGAATACCGGTGGCGA
Serratia marcescens GGGTAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGATCTGGAGGAATACCGGTGGCGA
Enterobacteriaceae bacterium A2JM GGGTAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGATCTGGAGGAATACCGGTGGCGA
Kluyvera ascorbata GGGTAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGATCTGGAGGAATACCGGTGGCGA
721 731 741 751 761 771
CDB5 AGGCGGCCCCCTGGACGAAGACTGACGCTCAGGTGCGAAAGCGTGGGGAGCAAACAGGAT
Serratia marcescens AGGCGGCCCCCTGGACGAAGACTGACGCTCAGGTGCGAAAGCGTGGGGAGCAAACAGGAT
Enterobacteriaceae bacterium A2JM AGGCGGCCCCCTGGACAAAGACTGACGCTCAGGTGCGAAAGCGTGGGGAGCAAACAGGAT
Kluyvera ascorbata AGGCGGCCCCCTGGACAAAGACTGACGCTCAGGTGCGAAAGCGTGGGGAGCAAACAGGAT


781 791 801 811 821 831
CDB5 TAGATACCCTGGTAGTCCACGCTGTAAACGATGTCGATTTGGAGGTTGTGCCCTTGAGGC
Serratia marcescens TAGATACCCTGGTAGTCCACGCTGTAAACGATGTCGATTTGGAGGTTGTGCCCTTGAGGC
Enterobacteriaceae bacterium A2JM TAGATACCCTGGTAGTCCACGCTGTAAACGATGTCGATTTGGAGGTTGTGCCCTTGAGGC
Kluyvera ascorbata TAGATACCCTGGTAGTCCACGCCGTAAACGATGTCTATTTGGAGGTTGTTCCCTTGAGGA
841 851 861 871 881 891
CDB5 GTGGCTTCCGGAGCTAACGCGTTAAATCGACCGCCTGGGGAGT-CGGCCGCAAGGTTAAA
Serratia marcescens GTGGCTTCCGGAGCTAACGCGTTAAATCGACCGCCTGGGGAGTACGGCCGCAAGGTTAAA
Enterobacteriaceae bacterium A2JM GTGGCTTCCGGAGCTAACGCGTTAAATCGACCGCCTGGGGAGTACGGCCGCAAGGTTAAA
Kluyvera ascorbata GTGGCTTCCGGAGCTAACGCGTTAAATAGACCGCCTGGGGAGTACGGCCGCAAGGTTAAA
901 911 921 931 941 951
CDB5 -CTCAAATGAATTGACGGGGGCCCGCACAAGCGGTGGAGCATGTGGTTTAATTCGATGCA
Serratia marcescens ACTCAAATGAATTGACGGGGGCCCGCACAAGCGGTGGAGCATGTGGTTTAATTCGATGCA
Enterobacteriaceae bacterium A2JM ACTCAAATGAATTGACGGGGGCCCGCACAAGCGGTGGAGCATGTGGTTTAATTCGATGCA
Kluyvera ascorbata ACTCAAATGAATTGACGGGGGCCCGCACAAGCGGTGGAGCATGTGGTTTAATTCGATGCA
961 971 981 991 1001 1011
CDB5 ACGCGAAGAACCTTACCTACTCTTGACATCCAGAGAACTTAGCAGAGATGCTTTGGTGCC
Serratia marcescens ACGCGAAGAACCTTACCTACTCTTGACATCCAGAGAACTTAGCAGAGATGCTTTGGTGCC
Enterobacteriaceae bacterium A2JM ACGCGAAGAACCTTACCTACTCTTGACATCCAGAGAACTTAGCAGAGATGCTTTGGTGCC
Kluyvera ascorbata ACGCGAAGAACCTTACCTACTCTTGACATCCAGAGAACTTAGCAGAGATGCTTTGGTGCC
1021 1031 1041 1051 1061 1071
CDB5 TTCGGGAACTCTGAGACAGGTGCTGCATGGCTGTCGTCAGCTCGTGTTGTGAAATGTTGG
Serratia marcescens TTCGGGAACTCTGAGACAGGTGCTGCATGGCTGTCGTCAGCTCGTGTTGTGAAATGTTGG
Enterobacteriaceae bacterium A2JM TTCGGGAACTCTGAGACAGGTGCTGCATGGCTGTCGTCAGCTCGTGTTGTGAAATGTTGG
Kluyvera ascorbata TTCGGGAACTCTGAGACAGGTGCTGCATGGCTGTCGTCAGCTCGTGTTGTGAAATGTTGG
1081 1091 1101 1111 1121 1131
CDB5 GTTAAGTCCCGCAACGAGCGCAACCCTTATCCTTTGTTGCCAGCGGTTCGGCCGGGAACT
Serratia marcescens GTTAAGTCCCGCAACGAGCGCAACCCTTATCCTTTGTTGCCAGCGGTTCGGCCGGGAACT
Enterobacteriaceae bacterium A2JM GTTAAGTCCCGCAACGAGCGCAACCCTTATCCTTTGTTGCCAGCGGTTCGGCCGGGAACT
Kluyvera ascorbata GTTAAGTCCCGCAACGAGCGCAACCCTTATCCTTTGTTGCCAGCGGTTCGGCCGGGAACT
1141 1151 1161 1171 1181 1191
CDB5 CAAAGGAGACTGCCAGTGATAAACTGGAGGAAGGTGGGGATGACGTCAAGTCATCATGGC
Serratia marcescens CAAAGGAGACTGCCAGTGATAAACTGGAGGAAGGTGGGGATGACGTCAAGTCATCATGGC
Enterobacteriaceae bacterium A2JM CAAAGGAGACTGCCAGTGATAAACTGGAGGAAGGTGGGGATGACGTCAAGTCATCATGGC
Kluyvera ascorbata CAAAGGAGACTGCCAGTGATAAACTGGAGGAAGGTGGGGATGACGTCAAGTCATCATGGC
1201 1211 1221 1231 1241 1251
CDB5 CCTTACGAGTAGGGCTACACACGTGCTACAATGGCGTATACAAAGAGAAGCGACCTCGCG
Serratia marcescens CCTTACGAGTAGGGCTACACACGTGCTACAATGGCGTATACAAAGAGAAGCGACCTCGCG
Enterobacteriaceae bacterium A2JM CCTTACGAGTAGGGCTACACACGTGCTACAATGGCGTATACAAAGAGAAGCGACCTCGCG
Kluyvera ascorbata CCTTACGAGTAGGGCTACACACGTGCTACAATGGCATATACAAAGAGAAGCGACCTCGCG
1261 1271 1281 1291 1301 1311
CDB5 AGAGCAAGCGGACCTCATAAAGTACGTCGTAGTCCGGATTGGAGTCTGCAACTCGACTCC
Serratia marcescens AGAGCAAGCGGACCTCATAAAGTACGTCGTAGTCCGGATTGGAGTCTGCAACTCGACTCC
Enterobacteriaceae bacterium A2JM AGAGCAAGCGGACCTCATAAAGTACGTCGTAGTCCGGATTGGAGTCTGCAACTCGACTCC
Kluyvera ascorbata AGAGCAAGCGGACCTCATAAAGTATGTCGTAGTCCGGATTGGAGTCTGCAACTCGACTCC
1321 1331 1341 1351 1361 1371
CDB5 ATGAAGTCGGAATCGCTAGTAATCGTAGATCAGAATGCTACGGTGAATACGTTCCCGGGC
Serratia marcescens ATGAAGTCGGAATCGCTAGTAATCGTAGATCAGAATGCTACGGTGAATACGTTCCCGGGC
Enterobacteriaceae bacterium A2JM ATGAAGTCGGAATCGCTAGTAATCGTAGATCAGAATGCTACGGTGAATACGTTCCCGGGC
Kluyvera ascorbata ATGAAGTCGGAATCGCTAGTAATCGTAGATCAGAATGCTACGGTGAATACGTTCCCGGGC
1381 1391 1401 1411 1421 1431
CDB5 CTTGTACACACCGCCCGTCACACCATGGGAGTGGGTTGCAAAAGAAGTAGGTAGCTTAAC
Serratia marcescens CTTGTACACACCGCCCGTCACACCATGGGAGTGGGTTGCAAAAGAAGTAGGTAGCTTAAC
Enterobacteriaceae bacterium A2JM CTTGTACACACCGCCCGTCACACCATGGGAGTGGGTTGCAAAAGAAGTAGGTAGCTTAAC
Kluyvera ascorbata CTTGTACACACCGCCCGTCACACCATGGGAGTGGGTTGCAAAAGAAGTAGGTAGCTTAAC
1441 1451 1461 1471 1481 1491
CDB5 CTTCGGGAGGGCGTTTAGA--TTTATTATAC-TG-CTCCGCTGA----------
Serratia marcescens CTTCGGGAGGGCGCTACCA--CTT----------
Enterobacteriaceae bacterium A2JM CTTCGGGAGGGCGCTTACCACTTTGTGATTCATGACTGGGGTGAAGTCGTAACAAGGTAA
Kluyvera ascorbata CTTCGGGAGGGCGCTTACCACTTTGTGATTCATGACTGGGGTGAAGTCGTAACAAGGTAA

```
                                    1501      1511      1521      1531      1541
CDB5                                *---------*---------*---------*---------*
Serratia marcescens                 ----------------------------------------
Enterobacteriaceae bacterium A2JM   CCGTAGGGGAACCTGCGGCTGGATCACCTCCT----------
Kluyvera ascorbata                  CCGTAGGGGAACCTGCGGCTGGATCACCTCCTTAAGCTTGGATCC
```

Figure 2.6: Multiple sequence alignment of 16S rDNA sequences from CDB strains and their homologs. An asterisk indicates all have an identical residue, a dash (-) indicates that the CDB does share the consensus residue. The alignment generated using ClustalW (Thompson et. al., 1994)

2.3.5 Phylogenetic analysis and identification of the strains

Phylogenetic trees were constructed from multiple alignments of all the CDB strains and their related species (Appendix 3). 16S rDNA sequences of the 26 DNA sequences (including CDBs) were aligned for comparison. The phylogenetic positions of all isolates within different subgroups were investigated by comparing their 16S rDNA sequences to those representatives of various genera. Four different groups can be seen from the tree: *Arthrobacter, Ochrobactrum, Bacillus* and *Serratia* (Figure 2.7). It is evident from the phylogenetic tree that CDB1 is closely related to *Arthrobacter histidinolovorans*, CDB2 to *Ochrobactrum* sp. and *Ochrobactrum grignonense*, CDB3 and CDB4 to *Bacillus* spp. and CDB5 to *Serratia marcescens*. The distance was indicated at the branches and its nodes. Neighbor – joining method was used to estimate phylogenies from nucleotide sequences and *Arthrobacter* was chosen to be the out group. 98% replicates supported the relationships exhibited.

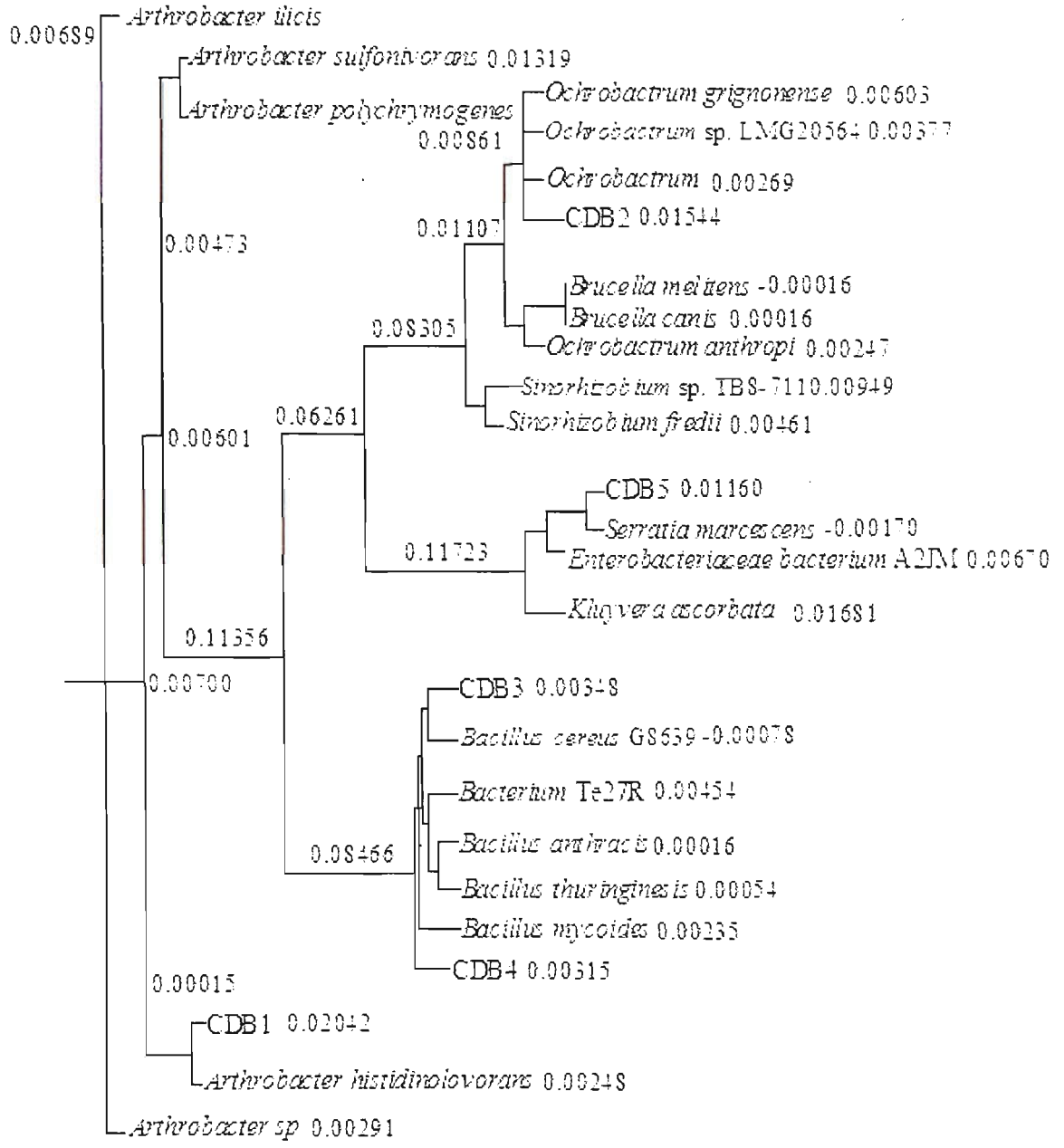


Figure 2.7: Phylogenetic tree of the CDB strains and related species retrieved from Genbank based on 16S rDNA gene homology. CDB1, CDB2, CDB3, CDB4 and CDB5 were multiple aligned by using full sequences obtained. The distances showed the differences between the different strains. The accession numbers of all strains are listed in Table 2.5.

2.3.6 Taxonomy tree

A taxonomy tree was generated through NCBI Taxonomy Browser (www.ncbi.nlm.nih.gov /taxonomy /browser) with all the CDB strains and their related species. The tree showed a similar trend to that of the phylogenetic tree. In this tree, *Ochrobactrum* belongs to the Rhizobiales group (Figure 2.8). The tree also showed that

the genera *Ochrobactrum* and *Serratia* are members of the Proteobacteria.

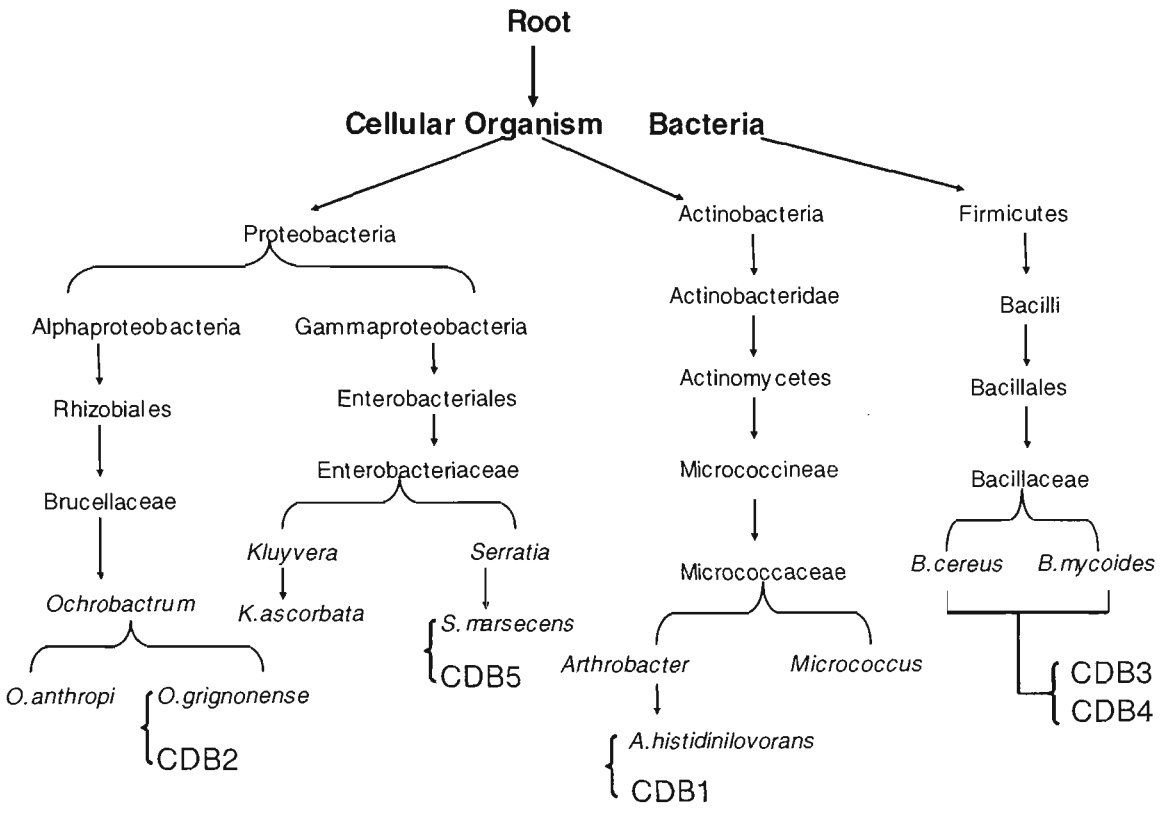
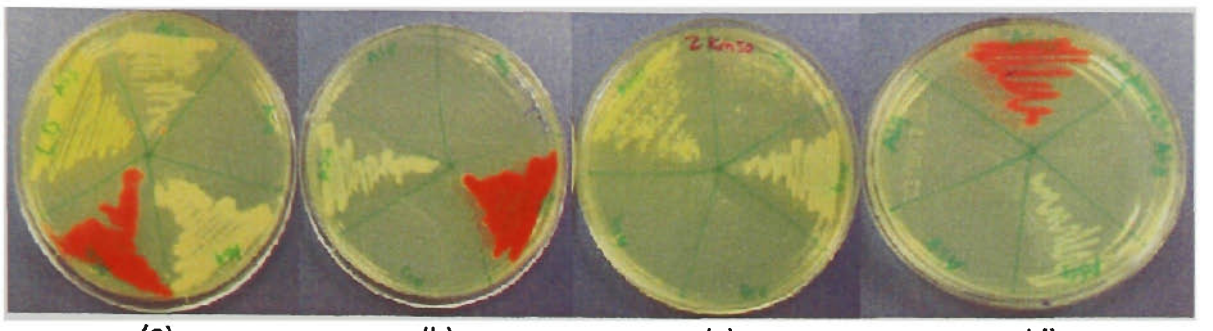


Figure 2.8: Taxonomy tree of all the related species from the taxonomy database in NCBI (drawn from www.ncbi.nlm.nih.gov /taxonomy /browser/ wwwtax.cgi)

2.3.7 Antibiotic resistance profile of the isolated bacterial strains

When tested for antibiotic resistance, CDB2 and CDB5 exhibited a wide range of resistance to antibiotics including Km (50 μg mL^{-1}), Amp (100 μg mL^{-1}) and Strep (50 μg mL^{-1}). CDB1 was resistant only to Km (50 μg mL^{-1}), while complete inhibition was observed for both CDB3 and CDB4 with all antibiotics tested (Figure 2.9).



(a) (b) (c) (d)

Figure 2.9: Growth inhibition of bacteria on different antibiotic plates. Bacterial cultures were streaked onto LB plates containing different antibiotics. Resistance to different antibiotics was observed after 24 hours of incubation at 30°C. (a) LB plate; (b) LB- Amp 100 μg/mL; (c) LB- Km 50 μg/mL; (d) LB- Strep 50 μg/mL.

2.3.8 Arsenic and antimony resistance profiles of isolated bacterial strains

The arsenic resistance of each isolated bacterial strain was tested (in triplicates and repeated three times) and compared with *E. coli* strain AW3110, in which the chromosomal *ars* operon has been deleted (Carlin et. al., 1995). Growth in liquid LB medium was more strongly inhibited by arsenite than arsenate and the *E. coli* strain AW3110 was more sensitive to both forms of arsenic (T-test indicated the significant P-value < 0.001) than any of the isolated soil bacteria (Figure 2.10). For the degree of resistance to arsenate, the order (most to least) was CDB2, CDB5, CDB4, CDB3 and CDB1 (Figure 2.10a). For resistance to arsenite, it was CDB5, CDB4, CDB3, CDB2 and CDB1 (Figure 2.10b). Strain CDB2 was exceptionally resistant to arsenate, but not to arsenite (T-test indicated the significant P-value < 0.005). When grown on LB agar plates, the isolated bacteria all showed a minimum inhibitory concentration to arsenite of 5 mM, but much higher values (more than 75 mM) to arsenate. CDB2 again was by far the most tolerant to arsenate, apparently exhibiting normal growth on agar containing 200 mM arsenate, an almost saturated concentration (Table 2.6).

Table 2.6: Minimum Inhibitory Concentration (MIC) of arsenic to CDB strains

	CDB1	CDB2	CDB3	CDB4	CDB5
arsenate (mM)	100	>200	100	75	100
arsenite (mM)	5	5	5	5	5

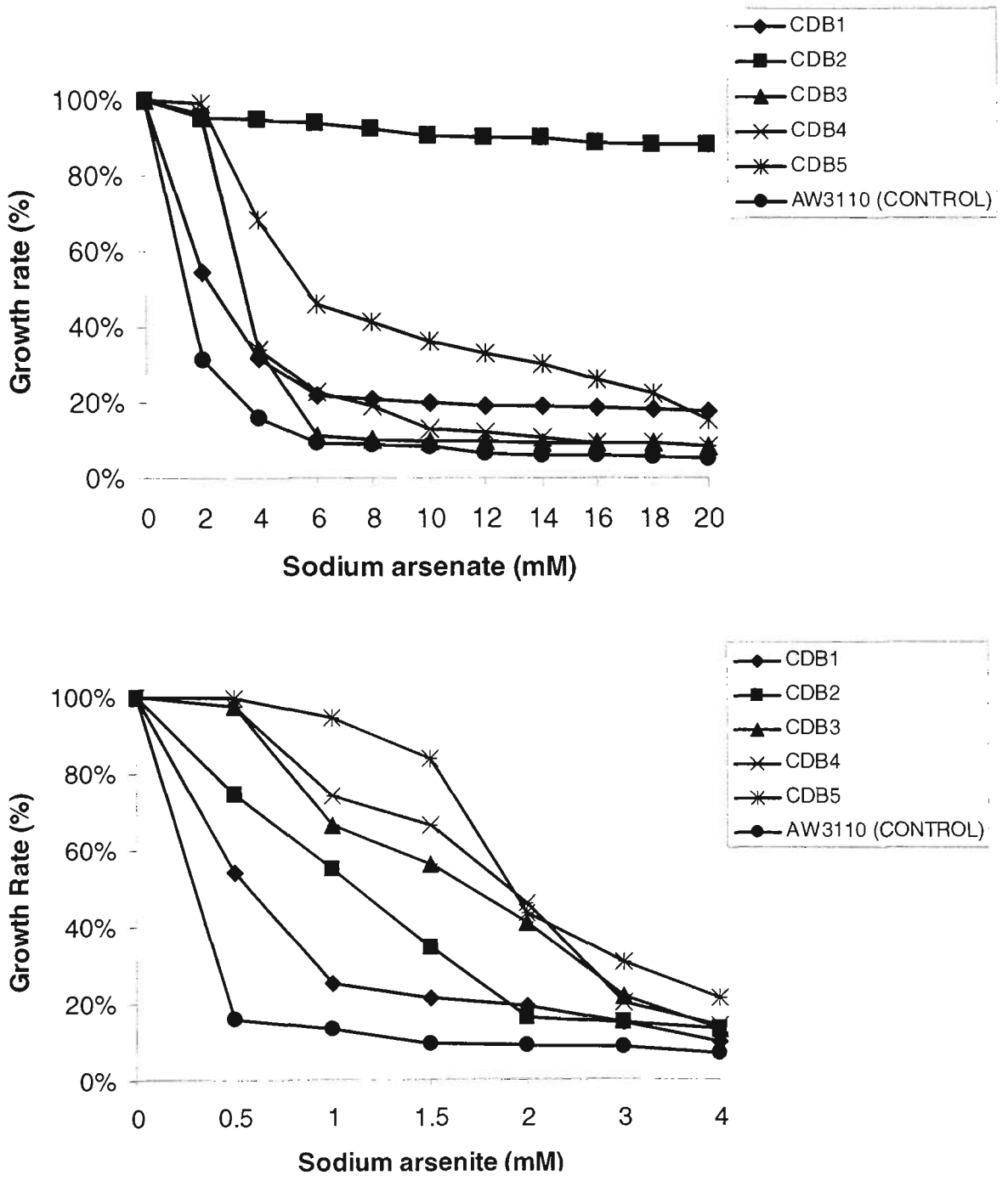


Figure 2.10: Growth inhibition of bacteria by arsenate and arsenite. Overnight cultures were diluted 50-fold into LB liquid medium containing increasing amounts of arsenate (a) and arsenite (b), incubated at 30°C for 5-7 hours. Growth rates were measured as absorbance at 600nm and indicated as percentage compared with those without arsenate and arsenite. The error bars represent SD (n=3) which are mostly small enough to be hidden behind the data points.

The antimonite resistance of these bacteria was also tested both in liquid LB medium and on LB agar plates containing different amounts of antimonite. Grown in liquid LB medium, CDB3 and CDB4 were inhibited by antimonite to a greater extent while CDB1 to a lesser extent compared with *E. coli* AW3110 (Figure 2.11) (T-test indicated the significant P-value < 0.002). *Ochrobactrum* sp. (CDB2) exhibited a higher antimonite resistance at SbIII concentration above 0.4 mM. CDB5 was the most resistant strain; 1 mM/L in LB liquid medium had no effect at all on the growth (T-test indicated the significant P-value < 0.002).

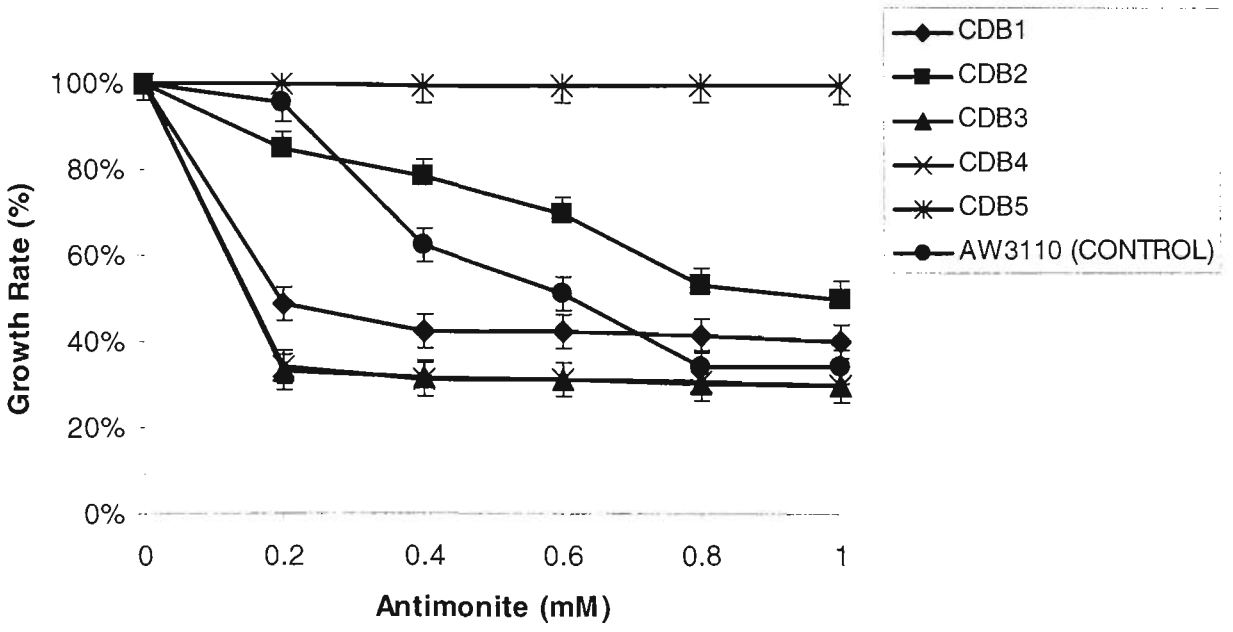


Figure 2.11: Growth inhibition of bacteria by antimonite. Overnight cultures were diluted 50-fold into LB liquid medium containing increasing amounts of antimonite, incubated at 30°C for 5-7 hours. Growth rates were measured as absorbance at 600nm and indicated as percentage compared with those without antimonite. The error bars represent SD (n=3).

2.4 Discussion

Although a large number of arsenic resistant bacteria have been reported in the literature, studies on strains existing in the cattle dip sites are very limited.

The experiments carried out in the present chapter aimed at identifying and characterizing the bacterial strains isolated from heavily contaminated cattle dip site soil in NSW. Microbial community analysis using 16S rRNA can detect and identify members and suggest phylogenetic relationships of uncultured organisms. The gene sequences encoding the 16S rRNA are highly conserved across all bacterial species (Dhruva et. al., 2003). The 16S rRNA gene is approximately 1540 bp and its sequence has been widely used as a molecular clock to estimate relationships among bacteria (phylogeny), but more recently it has also become important as a mean to identify unknown bacteria to genus or species level (Sacchi et. al., 2002). According to previous studies identification based on 16S rDNA sequencing has overcome the conventional testing of cellular fatty acid profiles and utilization of carbon (Jensen et. al., 1993; Kolbert and Persing, 1999). Therefore, sequencing of the 16S rDNA was performed to differentiate the five bacterial strains isolated for the present study. For all the five strains, using the fD1 and rP2 primers designed by Weisberg et. al., (1991) almost full length sequences were obtained; 1433 bp for CDB1, 1341 bp for CDB2, 1463 bp for CDB3, 1422 bp for CDB4 and 1449 bp for CDB5. Sequences were shorter than the complete 16S rRNA sequences. A strong and stable support can be found from the phylogenetic trees. This provided further evidence to demonstrate the effective use of 16S rDNA sequencing in the identification of different bacterial taxa.

The five isolated bacterial strains are assigned to the genera *Arthrobacter* sp. (CDB1), *Ochrobactrum* sp. (CDB2) *Bacillus* sp. (CDB3 and CDB4) and *Serratia* sp. (CDB5). This identification was based on PCR-based 16S rDNA ribotyping and similarity index of fatty acid profiles conducted by former members of our group. The two procedures concurred in identifying the most probable genus for each strain, and were consistent with the observed morphology. Ribotyping and fatty acid profiles also agreed in identifying the candidate species for two of the strains (*B. mycoides* for CDB4 & *S. marcesens* for CDB5) but conclusive identification to species level will require further research. Bhattacharya et. al. (2003) previously used PCR-based 16S rDNA ribotyping in combination with repetitive-sequence PCR fingerprinting to elucidate intraspecies diversity among strains of *Pseudomonas citronellolis* in petroleum-contaminated soils.

Phylogeny combined with taxonomy analysis included members of the alpha proteobacteria division, which reveal that the isolate CDB2 belonged to Rhizobiales group. When closely compared to our sequences, two species, *Ochrobactrum* sp. and *Ochrobactrum grignonense* (accession no. OS16SRRA and OGR242581 in the Genbank/EMBL/DDJB DNA sequence database) showed 99% identity with the 16S sequence of CDB2. Further analysis including members of gamma proteobacteria division revealed that the isolate CDB5 belonged to Enterobacteriaceae. When compared, *Serratia marcescens* (accession no. AY514434) showed 100% identity to CDB5, while, CDB1 which showed 99% identity to *Arthrobacter histidinolovorans* (accession no. AHRNA16S) assigned to Micrococcaceae. CDB3 and CDB4 showed 99% identity with *Bacillus* spp. belonging to Firmicutes. The isolates represent groups of soil bacteria

belonging to novel lineages within divisions of actinobateria, proteobacteria and firmicutes.

The five isolated rhizosphere bacterial strains showed a range of sensitivity to arsenate and arsenite and the most frequently occurring strains were not necessarily the least sensitive. In our laboratory assays, AW3110 *E. coli* were used as an *ars* operon negative control. An *ars* operon-containing *E. coli* strain DH5α showed comparable resistance to the CDB strains except CDB2 against arsenate (data not shown). Many arsenic-resistant bacterial strains have been reported belonging to the taxonomy groups identified in Figure 2.8, with the exception of *Ochrobactrum* sp., to which CDB2 belongs. One of the two common strains in the dip-site soils, *Arthrobacter* sp., was actually the most sensitive of the five identified strains to both arsenate and arsenite. Perhaps, this bacterium inhabits a niche in the soil environment where the bioavailability of arsenic compounds is particularly low. However, the other common strain, *Ochrobactrum* sp., exhibited exceptional resistance to arsenate accompanied by moderate resistance to arsenite (Figure 2.10). This suggests that the mechanism of arsenate resistance in this bacterium relies on prevention of arsenate uptake, rather than on reduction of intracellular arsenate to arsenite followed by its rapid export. In contrast CDB5 (*Serratia* sp.) and CDB4 (*Bacillus* sp.) exhibited a minimal AsV resistance but, in turn exhibited a higher level of AsIII resistance than CDB2 (Figure 2.10). This indicates the difference in regulation of different bacterial strains in response to various concentrations of AsV and AsIII. Also, CDB5 exhibited a very high resistance to antimonite whereas CDB3 and CDB4 were completely inhibited as shown in Figure 2.11. This is interesting as CDB3 and CDB4 exhibit a moderate level of AsV and AsIII

resistance but offers no resistance to antimonite (Figure 2.10 and 2.11). Further molecular studies to identify the genetic determinants in these bacteria that confer resistance to metalloids are presented in the following chapter.

Chapter 3

Genetic Characterization and Sequence Analysis of CDB3 *ars* Gene Cluster 1

3.1 Introduction

The most common source of elevated arsenite concentrations in the Australian environment is attributed to anthropogenic activities (Smith et. al., 2003). Five arsenic resistant bacterial strains were isolated and identified as discussed in Chapter 2, the next step was to confirm the genetic determinants for resistance. Two *ars* gene clusters were cloned and sequenced (Luo, 2006) from *Bacillus* sp. CDB3. CDB3 *ars* gene cluster 1 consisting of 8 full intact open reading frames, which were deduced to code for proteins homologous to ArsR, YqcL, ArsC, ArsD, ArsA, TrxB, IscA and protein phosphatase respectively, and, cluster 2 consisting of 4 full intact open reading frames encoding proteins homologous to ArsR, ORF2, YqcL and ArsC respectively (Figure 3.1).

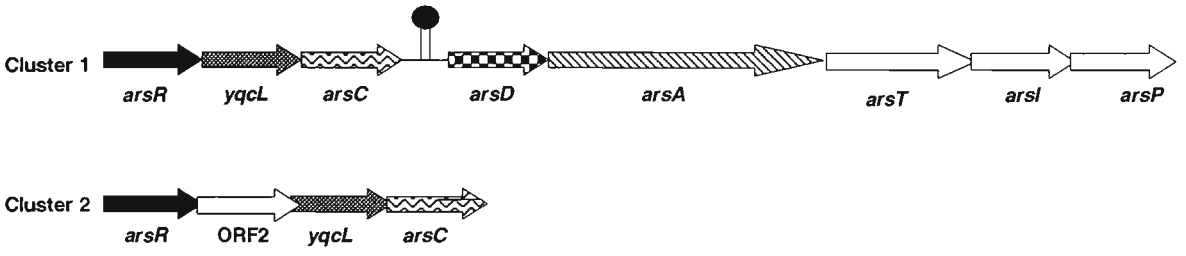
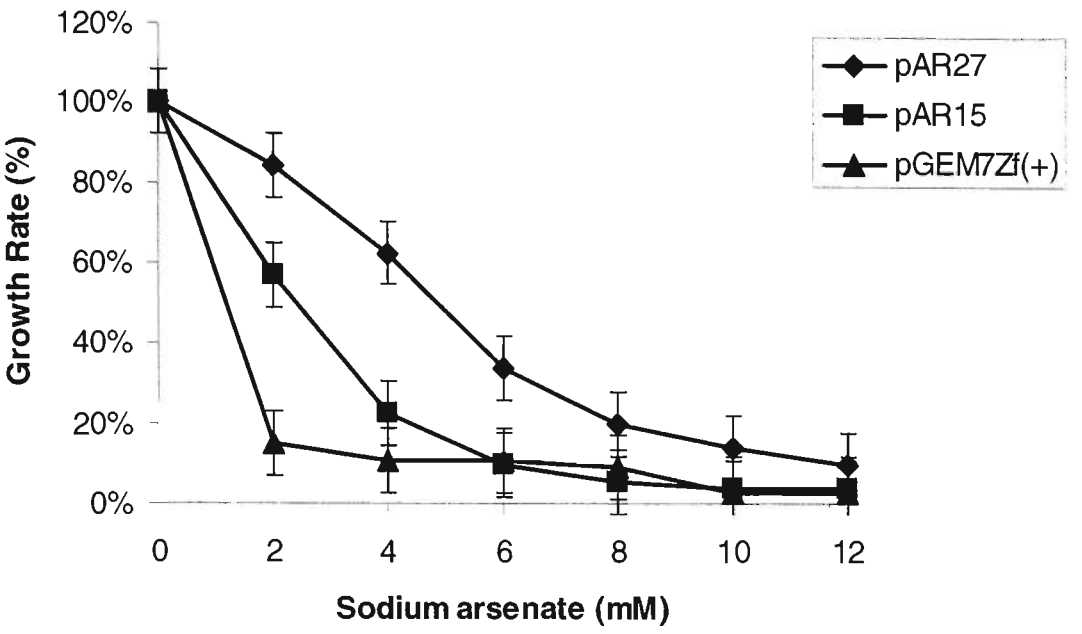
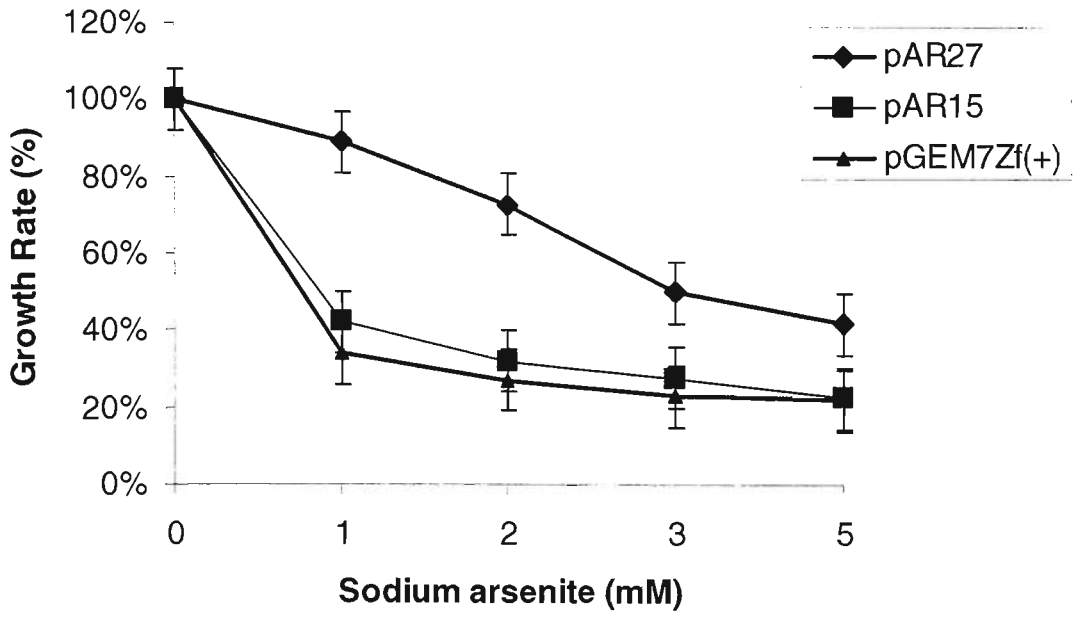


Figure 3.1: The structure of two *Bacillus* sp. CDB3 *ars* gene clusters. The last three genes of cluster 1 coding for TrxB, IscA and phosphatase, respectively, are designated *arsT*, *arsI* and *arsP*. T - stands for a potential RNA transcriptional terminator (Luo, 2006).

Once transformed into *E. coli* both clusters conferred a significant increase in arsenite and arsenate resistance to host cells (Figure 3.2).



(a)



(b)

Figure 3.2: Growth inhibition of *E. coli* AW3110 harboring plasmids pAR27 (CDB3 *ars* cluster 1), pAR15 (CDB3 *ars* cluster 2) and pGEM7Zf (+) (control) to arsenate (a) and arsenite (b). Overnight cultures were diluted 50-fold into LB liquid medium containing increasing amounts of arsenite and arsenate, incubated at 37°C for 5-7 hours. Growth rates were measured as absorbance at 600_{nm} and indicated as percentage compared with those without arsenate or arsenite (Luo, 2006).

Whilst the second cluster is very similar in structure and sequence to that identified on the *B. subtilis* skin element (Bobrowicz et. al., 1997), the first cluster shows some novel features. Firstly, the *arsD* and *arsA* are located behind *arsC*, rather than located between *arsR* and *arsB* as found in other *arsD/A* containing clusters [e.g. the plasmid-borne operon from Gram negative bacteria *E. coli* pR773 (Dey et. al., 1995)]. The intergeneic region between *arsC* and *arsD* possesses a potential transcriptional terminator and a potential promoter (Z. Xu and Q. Wang, personnel communication) suggesting that the gene cluster may contain more than one transcription unit. Twenty-three bp downstream of the *arsA* stop codon is the sixth ORF which shares homology to the thioredoxin reductase gene (*trxB*) which has also recently been identified in *ars* operon of *Streptomyces* spp. (Wang et. al., 2006). Located downstream from *trxB* are the two novel genes *arsI* and *arsP* whose products share homology with iron sulphur cluster assembly proteins and protein tyrosine phosphatases, respectively. These two ORFs (*arsI* and *P*) have not been identified in any known *ars* gene clusters until now. In addition, this is the first identified *ars* cluster in which an *arsA* gene is linked with *yqcl*.

Upon successful cloning and sequencing done by former members of our lab (Luo, 2006) further studies were needed to investigate the function and regulation of this operon in revealing the mechanism of arsenic resistance. Firstly, Southern blot analysis was performed to confirm the structure of the two clusters in the genome. Since *ars* operons have been found on either plasmids or chromosome, it was also necessary to investigate the localization of the two CDB3 *ars* clusters. In her thesis, (Luo 2006)

carried our more detailed sequence analysis on the ArsT, I and P of *ars* cluster 1, but not R, B, C, D and A. Hence, similar efforts have been made to investigate detailed sequence features of these five proteins.

3.2 Materials and Methods

3.2.1 Bacterial strains and plasmids

Bacterial strains and plasmids used in this study are listed in Table 3.1. *E. coli* strains were grown in LB medium at 37°C (Sambrook et. al., 1989) with ampicillin (100 μg ml^{-1}) or kanamycin (50 μg ml^{-1}) added as required.

Table 3.1: Bacterial strains and plasmids used in the study of Chapter 3

Strains and plasmid	**Genotype or description**	**Reference or Source**
E. coli AW3110	K-12 F⁻IN (*rrn*D-*rrn*E) Δ*ars::cam* (Cm^r, the chromosomal *arsRBC* deleted)	Gift from Dr. B.P. Rosen. (Carlin et. al., 1995)
CDB3- *Bacillus* sp.	Isolated from cattle dip-sites	(Chopra et. al., 2007)
Bacillus subtilis		Zhang lab stock
Bacillus cereus		Zhang lab stock
pGEM7Zf(+)	Cloning vector (Ap^r) 3 kb in size.	Promega
pAR27	A 7013 bp partial *Sau*3AI fragment containing CDB3 *arsRBCDATIP* cloned into pGEM7Zf(+) at *Bam*HI site	Provided by X. Luo, 2006

3.2.2 Preparation of DIG labeled DNA probe by PCR

To prepare DNA probes for Southern blot hybridization and localization studies, a PCR DNA labeling kit (Boehringer Mannheim) was used in accordance to the manufacturers protocol. The PCR primers were designed based on the CDB3 *ars* DNA / 16S rDNA sequences and synthesized by Sigma (Table 3.2). The PCR reactions were performed using a GeneAmp PCR system 9600 (Perkin Elmer). A typical reaction mixture (50 μL) consisted of, a 20 pg of DNA template (the same PCR product unlabeled), 1 μL of each primer (20 μM), 0.2 mM of DIG-dNTP mix and 2.5 units of *Taq*

DNA polymerase in the manufacturers buffer. After an initial denaturation step of 3 min at 95°C, 30 cycles of denaturation (95°C, 45 sec), annealing (54°C 30 sec) and extension (72°C, 2 min) was performed, followed by a final cooling step at 25°C, thus completing the reaction.

Table 3.2: Primers used for preparing DIG labeled DNA probes

Primer Name	**Primer Sequence 5'-3'**	**Upstream (Forward)/ Downstream (Reverse)**
As4-4	TTAACAAGAGTGTCACAG	Upstream
As4-15	TATCTAGAGCCATCTTTGATTTCTTCC	Downstream
As4-6	AGGACCAGACGTATTGC	Upstream
As4-5	ATGACCCGTTGGCGCTG	Downstream
16S fD1	AGAGTTTGATCCTGGCTCAG	Upstream
16S rP2	ACGGCTACCTTGTTACGACTT	Downstream

3.2.3 DNA extraction and restriction digestion

Extraction of DNA and subsequent restriction enzyme digestion for Southern blot analysis was performed as mentioned in Chapter 5, section 5.2.3.

3.2.4 Localization of CDB3 *ars* cluster 1

Plasmid chromosome separation was done according to the rapid method by Kado and Liu (1981). 3 mL of the overnight culture was pelleted and resuspended in 500 μL of TE buffer [40 mM Tris-acetate and 2 mM sodium EDTA pH 7.9 (adjusted with glacial acetic acid)] containing lysozyme at a final concentration of 3 mg/mL. 1 mL of the lysis buffer (50 mM Tris and 3% SDS, pH 12.6) was added and mixed. The samples were then heated at 65°C for 10 mins in a water bath and equal volume of phenol-chloroform solution (1:1 vol/vol) was added. The solution was vortexed for 30 secs and centrifuged

at 13000 g for 15 min. Avoiding the precipitate at the interface, the upper aqueous phase was transferred to a new eppendorf tube. 30-40 μL of the sample was loaded on to a 0.7% TE agarose gel and electrophoresis carried out at 100V for 2-3 hours until the tracking dye migrated to the bottom of the gel. The gel was stained with ethidium bromide and viewed over a short wave UV light source. DNA blot hybridization was then carried out as detailed in the next section (3.2.5)

3.2.5 Southern blot hybridization

After gel electrophoresis, the intact or enzyme restricted DNA was transferred from gel to membrane via the capillary transfer method (Sambrook et. al., 1989). The gel was placed upside down and blotted with a nylon membrane (Hybond^{TM}-N^{+}, Amersham) using 0.4 M NaOH as transfer solution for at least 4 hours.

The nylon membrane blots were placed in a hybridization tube containing 20 mL of DIG easy hybridization solution (Roche, 1 mL/10 cm^2). These were pre-hybridized for 2 hours at 42°C in a hybridization oven with rotation. The DIG-labeled DNA probe (about 15 μL) was heat-denatured in a boiling water bath for 10 min and chilled quickly in an ice bath to prevent re-annealing of the denatured strands. The denatured probe was diluted in an appropriate amount of the hybridization solution pre-warmed to 42°C. The pre-hybridization solution in the tubes was collected for future use and the hybridization solution containing the DIG-labeled probe was added. After overnight hybridization at 42°C, the membranes were washed twice in 2 × SSC wash solution containing 0.1% SDS (sodium dodecyl sulphate) at room temperature, (5 min for each time). The membranes

were then washed twice in 0.5 × SSC wash solution containing 0.1% SDS at 68°C, 15 min per wash. A DIG label detection kit (Boehringer Mannheim) was used to illuminate the labeled DNA on blot. The membrane was equilibrated in washing buffer for 1 min and blocked by gentle agitating in blocking solution for 45 min. The anti-DIG-AP (alkaline phosphatase) was diluted 1:10,000 in 20 mL blocking solution to form detection buffer. After blocking, the membrane was immersed in the detection buffer with gentle shaking for 30 min. The membrane was then washed twice in washing buffer (15 min per each) and equilibrated in detection buffer for 2 min. After adding sufficient amount of chemiluminescent substrate CSPD (disodium 3-(4-methoxyspiro-[1,2-dioxetane-3-2′(5′ chloro)-tricyclo[3.3.1.1^{3,7}]decane]-4-yl)phenyl phosphate), the membrane was incubated at 37°C for 15 min. The hybridization signals were visualised by exposing the membrane to an X-ray film (Kodak), in an X-ray cassette for an appropriate time. Finally, the exposed film was developed and fixed.

3.2.6 Computational sequence analysis

Searching for homologous sequences in Genbank were carried out using blastx analysis (Altschul et. al., 1990) and subsequent alignment was done as stated in Chapter 4, section 4.2.1, followed by phylogenetic analysis using Neighbor-Joining/UPGMA method (version 3.573c) in PHYLIP package (Felsenstein, 1989) was used.

3.3 Results

3.3.1 Probe preparation

Three probes were prepared, a 1.9 kb DNA fragment (probe 1) covering part of *arsR, B* and *C* genes, a 1.5 kb PCR DNA fragment (probe 2) covering the *arsDA* and 1.5 kb covering the 16S rDNA sequence of CDB3.

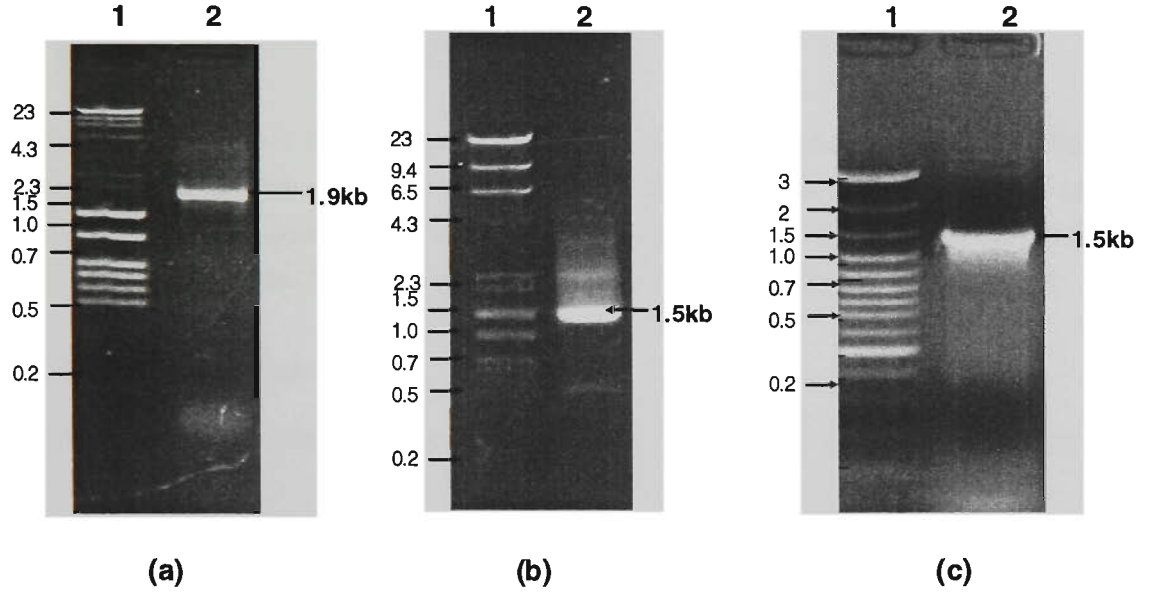


Figure 3.3: Agarose gel (2%) electrophoresis of PCR amplified DNA fragments. (a) probe1 (As4-4/15), (b) probe 2 (As4-6/5) and (c) 16S rDNA probe (16S-fD1/rP2). Each panel DNA markers are in Lane 1 and the amplified probes in Lane 2. Marker sizes in kb are indicated by arrow on left.

3.3.2 Southern blot analysis

Southern-blot analysis was performed on the extracted genomic DNA from CDB3 digested with *Eco*RI (shown in Figure 3.4). Probe 1 hybridized to three bands – 5, 1.4 and 0.35 kb and probe 2 to only the 5 kb band, exactly as expected based on the restriction map from CDB3 cluster 1 sequence (Luo, 2006). These results confirm the successful cloning of the cluster and also indicate that it has only a single copy in the CDB3 genome. Results with other probes (not presented) agreed with these.

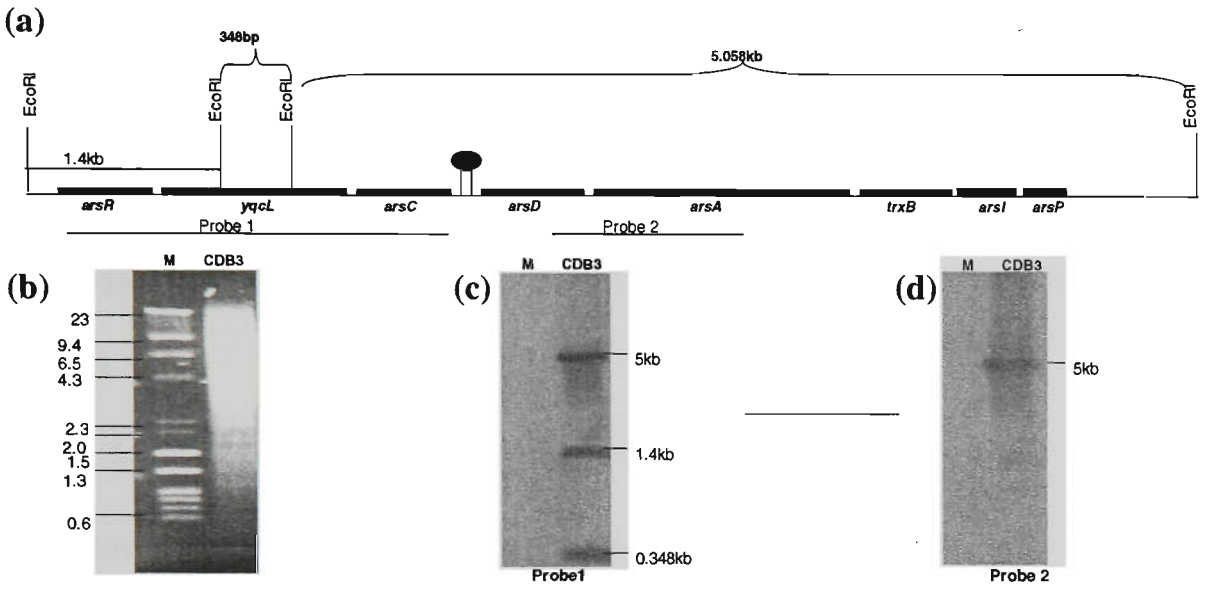


Figure 3.4: Southern analysis of CDB3 *ars* cluster 1. a - Restriction map drawn to scale showing the sizes of the fragments obtained with respect to each probe. b - Ethidium bromide stained 0.8% TAE agarose gel image showing the *Eco*RI digested CDB3 chromosomal DNA. c and d - X-ray film showing specific bands after hybridization with probes 1 and 2, respectively. Marker sizes in kb are indicated on the left.

3.3.3 Cluster localization

In order to determine the location of CDB3 *ars* cluster 1, a gel separation method as reported by Kado and Liu (1981) was employed. Initially, this method demonstrated the existence of a large plasmid in CDB3 (Figure 3.5) while no such bands could be seen on the gel for *B. subtilis* and *B. cereus* strains which are known to contain no plasmids.

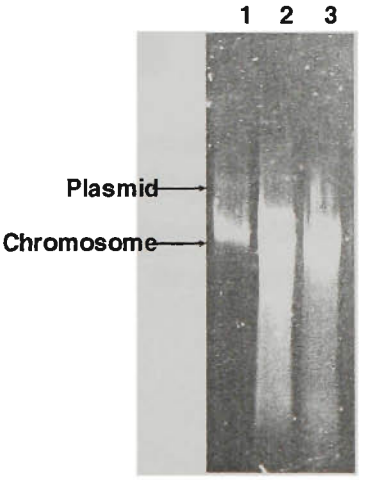


Figure 3.5: Ethidium bromide stained 0.7% TE agarose gel image showing the chromosome – plasmid separation of CDB3 (1), *B. subtilis* (2) and *B. cereus* (3) known not to contain plasmid were used as negative controls.

Initial experiments were conducted to determine the plasmid chromosome separation with known bacterial strains, such as *Agrobacterium*, *Sinorhizobium* etc. (data not shown). The gel separation system was confidently established, though the size(s) of CDB3 plasmid(s) can't be estimated. The chromosomal 16S probe control confirmed the method.

Southern blot analysis was carried out in order to localize the *ars* cluster. The DIG labeled probe 1 hybridized to the lower band as shown in Figure 3.6, indicating that the CDB3 *ars* cluster 1 is located on the chromosome. To verify this, the same DNA blot was hybridized with the CDB3 16S rDNA probe after stripping off the *ars* specific probe 1. It showed that the 16S probe hybridized only to the chromosomal band as expected, thus confirming the result.

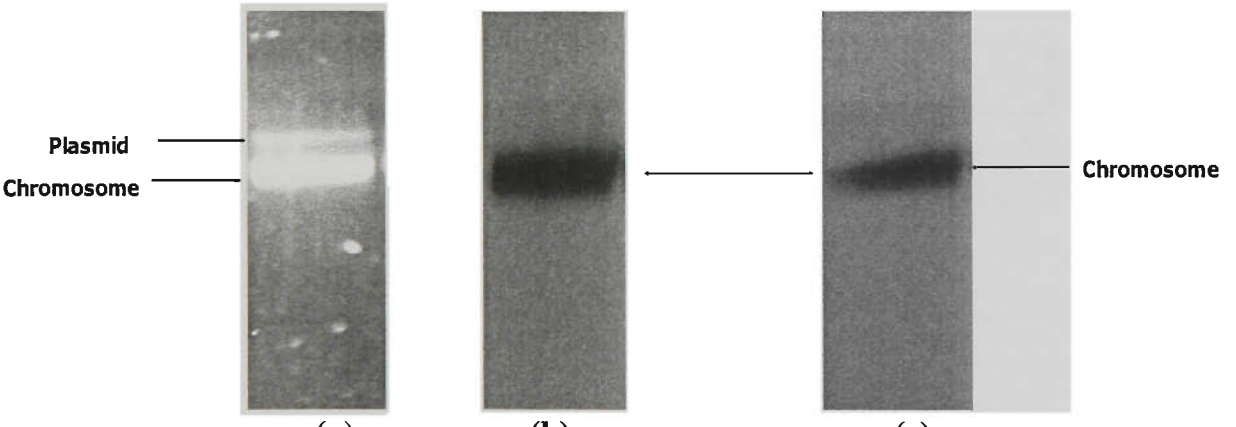


Figure 3.6: Southern blot analysis to localize CDB3 *ars* cluster 1. (a) Ethidium bromide stained 0.6% TBE agarose gel image showing the separation of the CDB3 plasmid from the chromosomal DNA. (b)- X-ray film showing specific band corresponding to the chromosomal DNA after hybridization with probe 1. (c) X-ray film showing specific band corresponding to the chromosomal DNA after hybridization with 16S probe.

3.3.4 Homology analysis of CDB3 Ars proteins

The CDB3 *ars* cluster 1 cloned in pAR27 was shown to contain putative upstream transcription elements and eight intact open reading frames (Appendix 4). Blastx and blastp searches indicated the homologies of their deduced proteins to known proteins (Table 3.3). Multiple alignments to demonstrate these homologies for the first five are presented in this section and the next chapter (Orf 2). The later three, analyzed by Luo (2006), is presented in Appendix 6.

Table 3.3: Amino acid sequence identities between the CDB3 Ars proteins and other bacterial Ars proteins

Source	ArsR	YqcL	ArsC	ArsD	ArsA	ArsT	ArsI	ArsP
Archeal Halobacterium sp. pNRC100	29.2%	-	22.9%	34.6%	34.2%	-	-	-
A. multivorum pKW301	25.2%	21.0%*	28.1%	37.6%	49.7%	-	-	-
E. coli pR773	33.0%	22.3%*	20.5%	35.0%	50.4%	-	-	-
E. coli pR46	32.1%	20.4%*	23.5%	39.5%	49.4%	-	-	-
B. subtilis Skin element	42.3%	85.6%	75.2%	-	-	-	-	-
Bacillus sp mb24**	95%	96%	96%	88%	87%	-	96%	90%
B. cereus pBC10987**	71%	98%	95%	97%	88%	-	93%	87%
L. innocua pLI100	35.5%	57%*	-	48%	56%	-	-	-
S. aureus pI258	40.6%	19.3%*	70.1%	-	-	-	-	-
Streptomyces pHZ227	50% (ArsR1) 58.9% (ArsR2)	69%*	49%	-	-	63.4%	-	-

Identities analyzed with "Best fit" program (GCG)

"-" indicates that the protein is not encoded by the corresponding *ars* cluster.

* ArsB proteins

** Operon/cluster identified from genome sequencing data; function not characterized.

3.3.4.1 ArsR

ArsR is a 13 kDa protein of 115 amino acid residues. As an arsenic-responsive repressor with high affinity for its operator site, ArsR controls the basal level of *ars* expression. An alignment of the sixteen ArsR metalloregulatory proteins encoded by the known *ars*R-bearing operons highlights the regions for metalloid-binding (Figure 3.7).

This demonstrates that all the sixteen ArsR homologs contain the vicinal cysteines (highlighted in yellow) within the metalloid binding domain (ELCVCDL), which have been demonstrated to be required for metalloid binding and induction (Shi et. al., 1994). In addition to the two cysteines conserved in the metal binding domain, the first nine and *A. multivorum* pKW301 proteins also contain another cysteine (highlighted in grey, Figure 3.7), which was revealed to be the third residue involved in metalloid binding (Shi et. al., 1994). While CDB3 ArsR lacks this cysteine residue, it contains a third cysteine residue at the 119th postion (in the alignment, shown in Figure 3.7) along with *Bacillus* sp. mb24 and *B. cereus* pBC10987 which share a similar sequence as CDB3. From the alignment, it was also found that within the putative helix-turn-helix DNA binding region (residues 41-60), the Ser-47, Ser-52, His-54 and Leu-59 residues are highly conserved (highlighted in purple), indicating that they might be important for DNA binding. Expression of the well characterized *ars* operon on pR773 is metalloregulated by ArsR, a trans-acting repressor that binds as a dimer to an operator and represses transcription (Wu and Rosen, 1993b). The relative homology shared by CDB3 ArsR to that of pR773 is quite low (33%).

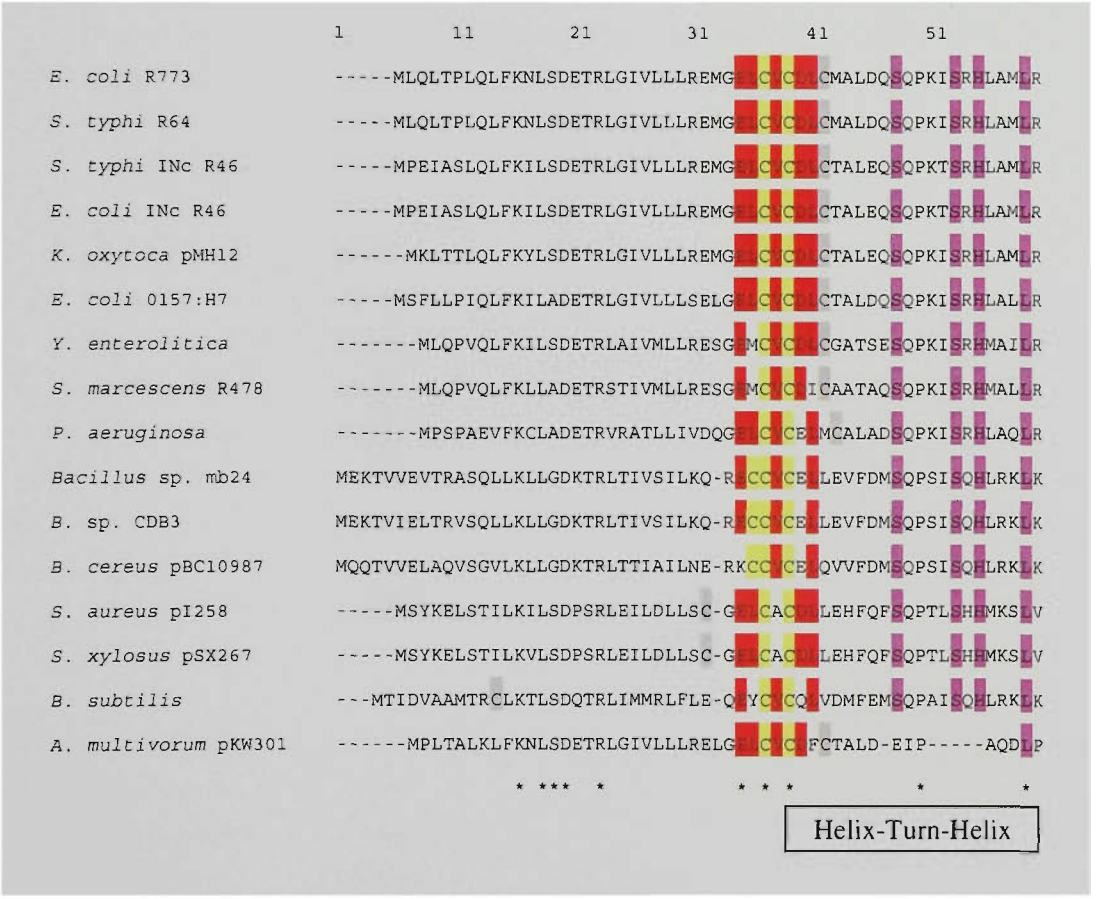
3.3.4.2 ArsD

ArsD is a novel regulatory protein and its distribution among *ars* operons appears to be limited (Li et. al., 2001). The organization of genes in those operons which encode ArsD proteins is most commonly *arsRDABC*, while the organization of the three *Bacillus* spp., CDB3 (*Bacillus* sp.), *B. cereus* pBC10987 and *Bacillus* sp. mb24 is quite different

and possess a novel gene arrangement: *arsRYCDATIP* of CDB3, and *arsRYCDAIP* in both *B. cereus* pBC10987 and *Bacillus* sp. mb24, which lack the thioredoxin reductase (*trxB*) gene. A search of available genomes returned with only ten ArsD homologues, seven of which are known to be encoded by plasmid-borne operons (Figure 3.8). The ArsD of CDB3 (*Bacillus* sp.) however is carried on the chromosome (Figure 3.6), whilst the ArsD of *Bacillus* sp. mb24 is on a transposon. Multiple sequence alignment of ten ArsD homologs, CDB3 ArsD, *E. coli* INc pR46, *A. multivorum* pKW301, *E. coli* pR773, *L. innocua* pLI100, *B. cereus* pBC10987, *Bacillus* sp. mb24, *S. typhi* R64, *S. epidermidis* RP62A, *Halobacterium* sp. pNRC100 is shown in Figure 3.9. The conserved cysteine residues are highlighted in green. Alignment of the ten ArsD homologs shows the conserved cysteine residues (Figure 3.8). It was found that CDB3 ArsD, *B. cereus* pBC10987, *Bacillus* sp. mb24 and *Halobacterium* sp. pNRC100 (Figure 3.8) only contained the first cysteine pair, Cys12-Cys13, and the cysteine residue at the 18th position; while the other four homologues have two additional cysteine pairs, Cys-112-Cy113 and Cys119-Cys120. Cys39 is present only in *E. coli* R773 and *S. typhi* R64. *E. coli* R773 ArsD was the first identified and has been demonstrated to function in the regulation of arsenical resistance (Li et. al., 2001). The alignment also shows a conserved sequence of AMCCSTGVCG denoted with asterisks and other conserved residues which may be important for metal binding recognition or in other function.

The halophilic archea *Halobacterium* sp. pNRC100 has five recognized arsenic resistance genes (Ng et. al., 1998), three of which (*arsDAR*) are organized divergently

from the downstream *arsRC* genes (Li et. al., 2001). The pNRC100 ArsD is only 34.6% homologous to CDB3 ArsD and 52% identical to R773 ArsD (Li et. al., 2001). CDB3 ArsD amino acid sequence is 39.5%, 35.0% and 37.6% identical to the ArsD proteins from *E. coli* plasmids pR46 and pR773, *Acidiphilium multivorum* AIU 301 pKW301 respectively. While, CDB3 ArsD shows 97% and 88% identity to *B. cereus* pBC10987 and *Bacillus* sp. mb24.



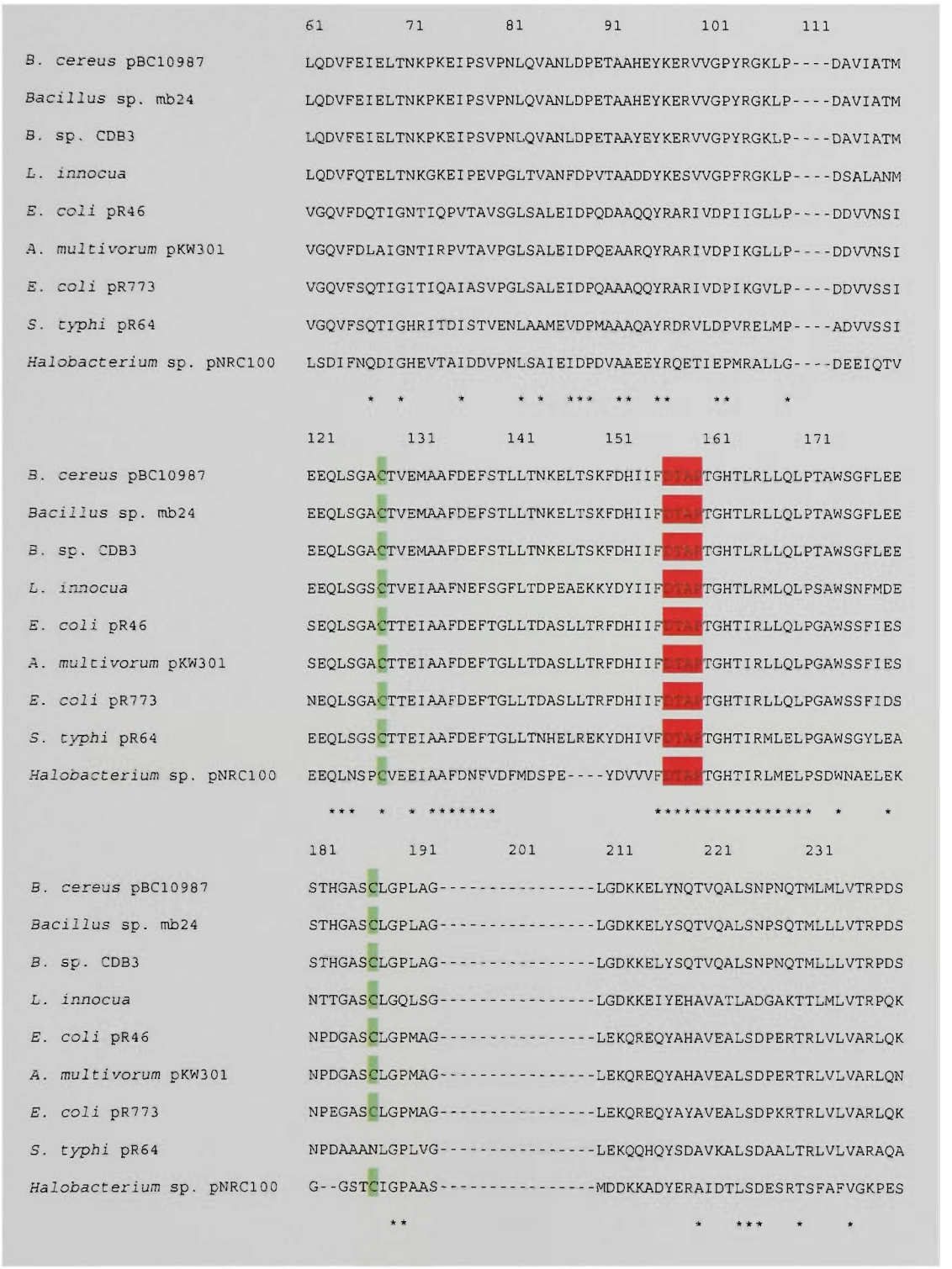
```
                          61        71        81        91        101       111
E. coli R773              ESGILLDRKQGKWVHYRLSPHIPSWAAQIIEQAWLSQQDDVQ----VIARKLASVNCSGS
S. typhi R64              ESGILLDRKQGKWVHYRLSPHIPSWAAQIIEQAWLSQQDDVQ----VIARKLASVNCSGS
S. typhi INc R46          ESGLLLDRKQGKWVHYRLSPHIPSWAALVIEQAWLSQQDDVQ----AIARKLASANCSGS
E. coli INc R46           ESGLLLDRKQGKWVHYRLSPHIPSWAALVIEQAWLSQQDDVQ----AIARKLASANCSGS
K. oxytoca pMH12          ESGLLLDPKQGKWVHYRLSPQIPSWAAQVIELAWLSQQDDVQ----AIARKLASANCSGS
E. coli 0157:H7           ESGLLLDRKQGKWVHYRLSPHIPSWAAKIIEQAWRCEQEKVQ----VIVRNLARQNCSVD
Y. enterolitica           EAELVLDRREGKWVHYRLSPHMPAWAAETITTSWHCCGKMFVSGWINQRHHPAEMNRTHS
S. marcescens R478        EAELVIDRREGKWVHYRLSPHMPAWAAGIIDTAWNCERENIR----NKLSSVASVSC---
P. aeruginosa             SAGLLLDRRQGQWVYYRLNPALPAWIHEVLQVTLRANGDWLQ----ADAARLRDMDGRPQ
Bacillus sp. mb24         DLGLVQEERRGQWIYYSLNQASDLYTLLEDILAHVPDQTEKI---KQIEKSNPTLRCGC-
B. sp. CDB3               DLGLVQEERRGQWIYYSLNQASDLYTILEDVLAHVPDQTEKI---KQIEKSNPTLRCGC-
B. cereus pBC10987        DIGLLKEERQGQWIYYSLNEESSFYPIIKDILQHVPDQKENI---QQIEKCNPTLRCKC-
S. aureus pI258           DNELVTTRKDGNKHWYQLNHAILDDIIQNLNIINTSNQ-------RCVCKNVKSGDC---
S. xylosus pSX267         DNELVTTRKNGNKHMYQLNHEFLDYINQNLDIINTSDQ-------GCACKNMKSGEC---
B. subtilis               NAGFVNEDRRGQWRYYSINGSCPEFDTLQLILHQIDQEDELLN---HIKQKKTQACCQ--
A. multivorum pKW301      SSGNAPGKRS---IAGSQAGEMGSLSLIPAYSFLGCSGD--------------------

                          121
E. coli R773              SKAVCI-
S. typhi R64              SKAVCI-
S. typhi INc R46          GKAVCI-
E. coli INc R46           GKAVCI-
K. oxytoca pMH12          GKAVCI-
E. coli 0157:H7           SKNTCS-
Y. enterolitica           FNHM---
S. marcescens R478        -------
P. aeruginosa             RASACCQ
Bacillus sp. mb24         -------
B. sp. CDB3               -------
B. cereus pBC10987        -------
S. aureus pI258           -------
S. xylosus pSX267         -------
B. subtulis               -------
A. multivorum pKW301      -------
```

Figure 3.7: A multiple sequence alignment of sixteen ArsR homologs encoded by the known *arsRBC* and *arsDA*-bearing operons. The metalloid-binding consensus region, ELCVCDL, is highlighted. The conserved cysteine residues for induction are highlighted in yellow. The other conserved cysteine residue that CDB3 does not share is highlighted in grey, while the cysteine that CDB3 has is highlighted in green. An asterisk indicates all proteins have identical residues; a blank space indicates that the CDB3 does share the consensus residue. Generated using ClustalW (Thompson et. al., 1994).

Figure 3.8: Multiple sequence alignment of ten ArsD homologs, CDB3 ArsD, *E. coli* INc pR46, *A. multivorum* pKW301, *E. coli* pR773, *L.innocua* pLI100, *B. cereus* pBC10987, *Bacillus* sp mb24, *S. typhi* R64, *S. epidermidis* RP62A, *Halobacterium* pNRC100. The conserved cystiene residues are highlighted in green. An asterisk indicates all proteins have an identical residue, blank space indicates that the CDB3 does share the consensus residue. Alignment generated using ClustalW (Thompson et. al., 1994).

3.3.4.3 ArsA

ArsA is a 63 kDa catalytic subunit that couples ATP hydrolysis to oxyanion translocation. Multiple sequence alignment was performed based on the results of a blastx search, with the highest score of ten ArsA homologs chosen, *Bacillus* sp. CDB3, *B. cereus* pBC10987, *Bacillus* sp. mb24, *E. coli* pR46, *A. multivorum* pKW301, *E. coli* pR773, *L. innocua, S. typhi* pR64, *Halobacterium* sp. pNRC100 and *B. halodurans* is shown in Figure 3.9. CDB3 ArsA shares 50% sequence identity with *E. coli* R773. Alignment of ten ArsA homologues shows that the two consensus sequences GKGGVGKT for the nucleotide binding sites, the three conserved cystiene residues cys113, cys172 and cys422, two histidine residues and one serine residue, which were demonstrated to be involved in allosteric metal binding, and the two consensus DTAP sequences [may act as a transduction domain that relays the communication between the ATPase-binding site and the allosteric AsIII-binding site] for signal transduction are highly conserved in these homologues (Zhou and Rosen, 1997).


61 71 81 91 101 111
B. cereus pBC10987 LQDVFEIELTNKPKEIPSVPNLQVANLDPETAAHEYKERVVGPYRGKLP----DAVIATM
Bacillus sp. mb24 LQDVFEIELTNKPKEIPSVPNLQVANLDPETAAHEYKERVVGPYRGKLP----DAVIATM
B. sp. CDB3 LQDVFEIELTNKPKEIPSVPNLQVANLDPETAAYEYKERVVGPYRGKLP----DAVIATM
L. innocua LQDVFQTELTNKGKEIPEVPGLTVANFDPVTAADDYKESVVGPFRGKLP----DSALANM
E. coli pR46 VGQVFDQTIGNTIQPVTAVSGLSALEIDPQDAAQQYRARIVDPIIGLLP----DDVVNSI
A. multivorum pKW301 VGQVFDLAIGNTIRPVTAVPGLSALEIDPQEAARQYRARIVDPIKGLLP----DDVVNSI
E. coli pR773 VGQVFSQTIGITIQAIASVPGLSALEIDPQAAAQQYRARIVDPIKGVLP----DDVVSSI
S. typhi pR64 VGQVFSQTIGHRITDISTVENLAAMEVDPMAAAQAYRDRVLDPVRELMP----ADVVSSI
Halobacterium sp. pNRC100 LSDIFNQDIGHEVTAIDDVPNLSAIEIDPDVAAEEYRQETIEPMRALLG----DEEIQTV
121 131 141 151 161 171
B. cereus pBC10987 EEQLSGACTVEMAAFDEFSTLLTNKELTSKFDHIIFDTAHTGHTLRLLQLPTAWSGFLEE
Bacillus sp. mb24 EEQLSGACTVEMAAFDEFSTLLTNKELTSKFDHIIFDTAHTGHTLRLLQLPTAWSGFLEE
B. sp. CDB3 EEQLSGACTVEMAAFDEFSTLLTNKELTSKFDHIIFDTAHTGHTLRLLQLPTAWSGFLEE
L. innocua EEQLSGSCTVEIAAFNEFSGFLTDPEAEKKYDYIIFDTAHTGHTLRMLQLPSAWSNFMDE
E. coli pR46 SEQLSGACTTEIAAFDEFTGLLTDASLLTRFDHIIFDTAHTGHTIRLLQLPGAWSSFIES
A. multivorum pKW301 SEQLSGACTTEIAAFDEFTGLLTDASLLTRFDHIIFDTAHTGHTIRLLQLPGAWSSFIES
E. coli pR773 NEQLSGACTTEIAAFDEFTGLLTDASLLTRFDHIIFDTAHTGHTIRLLQLPGAWSSFIDS
S. typhi pR64 EEQLSGSCTTEIAAFDEFTGLLTNHELREKYDHIVFDTAHTGHTIRMLELPGAWSGYLEA
Halobacterium sp. pNRC100 EEQLNSPCVEEIAAFDNFVDFMDSPE----YDVVVFDTAHTGHTIRLMELPSDWNAELEK
181 191 201 211 221 231
B. cereus pBC10987 STHGASCLGPLAG----------LGDKKELYNQTVQALSNPNQTMLMLVTRPDS
Bacillus sp. mb24 STHGASCLGPLAG----------LGDKKELYSQTVQALSNPSQTMLLLVTRPDS
B. sp. CDB3 STHGASCLGPLAG----------LGDKKELYSQTVQALSNPNQTMLLLVTRPDS
L. innocua NTTGASCLGQLSG----------LGDKKEIYEHAVATLADGAKTTLMLVTRPQK
E. coli pR46 NPDGASCLGPMAG----------LEKQREQYAHAVEALSDPERTRLVLVARLQK
A. multivorum pKW301 NPDGASCLGPMAG----------LEKQREQYAHAVEALSDPERTRLVLVARLQN
E. coli pR773 NPEGASCLGPMAG----------LEKQREQYAYAVEALSDPKRTRLVLVARLQK
S. typhi pR64 NPDAAANLGPLVG----------LEKQQHQYSDAVKALSDAALTRLVLVARAQA
Halobacterium sp. pNRC100 G--GSTCIGPAAS----------MDDKKADYERAIDTLSDESRTSFAFVGKPES

```
                                241       251       261       271       281       291
B. cereus pBC10987              SPLQEAERAAHELKEIGVSNQFLLVNGILKDYMQ-NDNVSNALFKRQSRALENMAEELKN
Bacillus sp. mb24               SPLQEAERAAHELKEIGVSNQFLLVNGILKDYMQ-NDNVSNALFKRQSRALENMAEELKN
B. sp. CDB3                     SPLQEAGRAAKELKEIGVNNQYLLINGVLTNYVQ-NDAISKALFTRQVRALENMSEELKG
L. innocua                      APLLEADRASKELQEIGIENQVLLVNGVLEEAT---DKVSQLIYDGQQEALAQMPDSLKA
E. coli pR46                    STLQEVARTHDELSAIGLKNQYLVINGVLPASEEKRDALAAAIWQREQEALANLPAGLSD
A. multivorum pKW301            STLQEVARTHEELAEIGLKNQYLVINGVLPEAEAEHDALAAAIWQREQEALANLPAGLSE
E. coli pR773                   STLQEVARTHLELAAIGLKNQYLVINGVLPKTEAANDTLAAAIWEREQEALANLPADLAG
S. typhi pR64                   STLKEVSHTHDELSAIGLQHQHIAINGVLPPFAGEDDPLAQSILAREEKALQAMPDNLAN
Halobacterium sp. pNRC100       SSIDEIERSASDLAELGISSQLLVVNGYLPESVC-EDPFFEGKRADEQAVIDRVESTFDQ
                                *   *       *   *        **  *               *   *

                                301       311       321       331       341       351
B. cereus pBC10987              LPTYEIPLVPFNVTGIENMRKLVQP----------MENLLISDEEANTVSIPSLQ---TL
Bacillus sp. mb24               LPTYEIPLVPFNVTGIENMRKLVQP----------MENLSISGEEANTVSIPSLQ---TL
B. sp. CDB3                     LPAYELPLVPFNVTGIENMRKLVRP----------IESLSILDEIQEEIAIPPLQ---NL
L. innocua                      FPEYSIPLRSYNVTGVENLRQLLKS----------NQGEFLT-EIVTPRAFPRLK---DI
E. coli pR46                    LPTDNLYLQPLNMVGVSALKGLLNEHA--------EITSLPEQSPQNKPENMSLS---VL
A. multivorum pKW301            LPTDTLLLQPVNMVGVSALKGLLATRS--------EALPLPVTNILYTPENLSLS---GL
E. coli pR773                   LPTDTLFLQPVNMVGVSALSRLLSTQP--------VASPSSDEYLQQRPDIPSLS---AL
S. typhi pR64                   LPRSQLYLKPFNLVGLEALRELFTESK--------AAPALPATT-LNTLDLPKLS---SL
Halobacterium sp. pNRC100       QALATYPLQPGEIAGLELLSDVGGVLYDGEEATVDVDAATRRATNEDTVDFDTFTDADAV
                                      * *       *

                                361       371       381       391       401       411
B. cereus pBC10987              ITNLSESGK-RVIFTMGKGGVGKTTVASAIAVGLAEKGHHVHLTTTDPAAHIDYVMHG--
Bacillus sp. mb24               ITNLSESGK-RVIFTMGKGGVGKTTVASAIAVGLAEKGHHVHLTTTDPAAHIDYVMHG--
B. sp. CDB3                     IADLSETGK-RVIFTMGKGGVGKTTVASAIAVGLAEKGHRVHLTTTDPAAHIDYVMHG--
L. innocua                      VNELHQSGK-KVIFTMGKGGVGKTTIAAAIATGLADKGKKVHLATTDPAAHLQFVIS---
E. coli pR46                    VDDIARSEH-GLIMLMGKGGVGKTTMAAAIAVSLADKGFNVHLTTSDPAAHLSTTLNG--
A. multivorum pKW301            VDDIARSEH-GLIMLMGKGGVGKTTMAAAIAVRLADMGFDVHLTTSDPAAHLSTTLNG--
E. coli pR773                   VDDIARNEH-GLIMLMGKGGVGKTTMAAAIAVRLADMGFDVHLTTSDPAAHLSMTLNG--
S. typhi pR64                   VDELSQTGK-GLVMTMGKGGVGKTTVAASVAVSLAKRGHKVHLTTSDPAAHLSYTLDG--
Halobacterium sp. pNRC100       AEELVPVEETRYLFFTGKGGVGKSTIASTTAVSLAEAGYETLVVTTDPAAHLADIFEQPV
                                         *********      * **         * ******
```

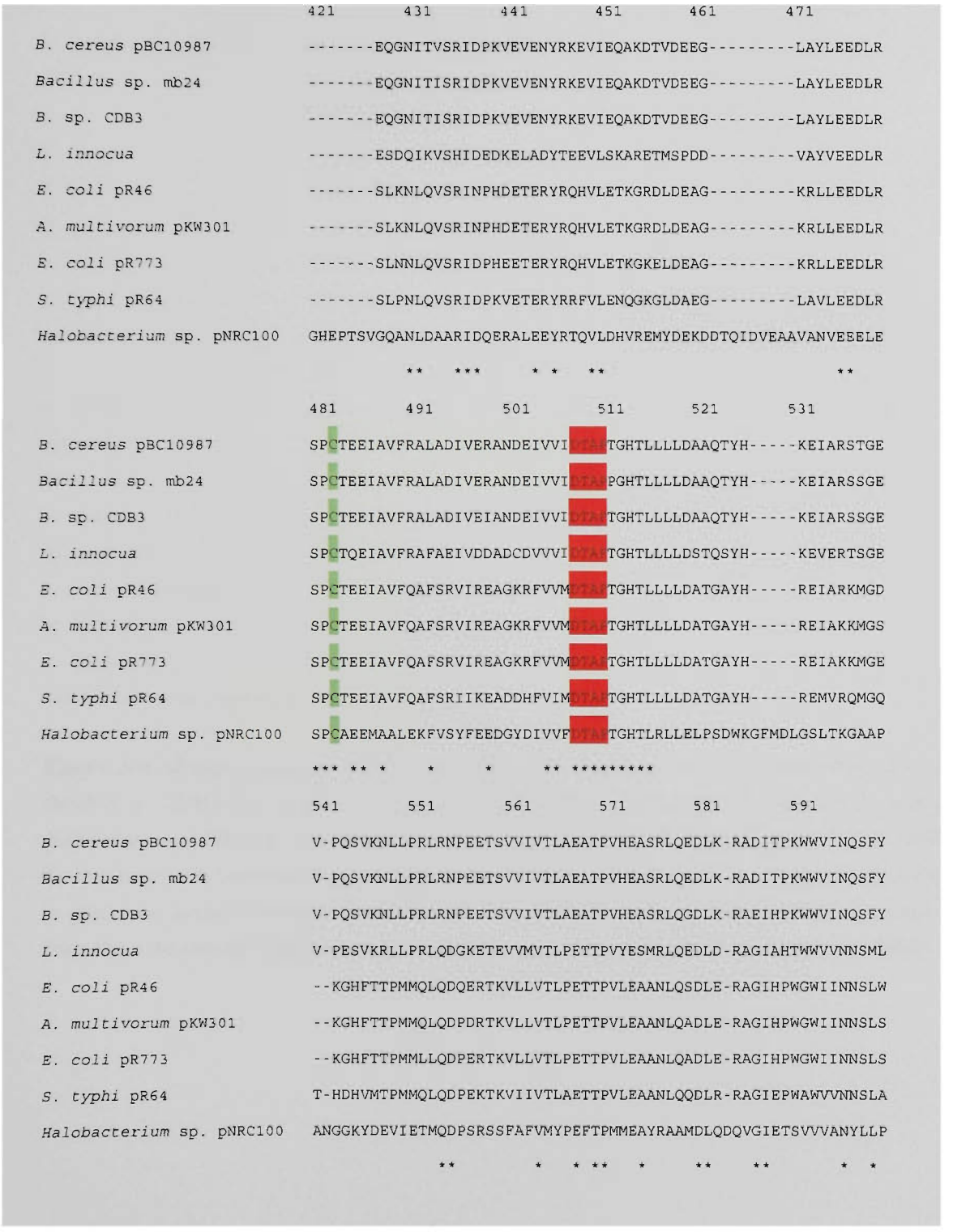

421 431 441 451 461 471
B. cereus pBC10987 -------EQGNITVSRIDPKVEVENYRKEVIEQAKDTVDEEG---------LAYLEEDLR
Bacillus sp. mb24 -------EQGNITISRIDPKVEVENYRKEVIEQAKDTVDEEG---------LAYLEEDLR
B. sp. CDB3 -------EQGNITISRIDPKVEVENYRKEVIEQAKDTVDEEG---------LAYLEEDLR
L. innocua -------ESDQIKVSHIDEDKELADYTEEVLSKARETMSPDD---------VAYVEEDLR
E. coli pR46 -------SLKNLQVSRINPHDETERYRQHVLETKGRDLDEAG---------KRLLEEDLR
A. multivorum pKW301 -------SLKNLQVSRINPHDETERYRQHVLETKGRDLDEAG---------KRLLEEDLR
E. coli pR773 -------SLNNLQVSRIDPHEETERYRQHVLETKGKELDEAG---------KRLLEEDLR
S. typhi pR64 -------SLPNLQVSRIDPKVETERYRRFVLENQGKGLDAEG---------LAVLEEDLR
Halobacterium sp. pNRC100 GHEPTSVGQANLDAARIDQERALEEYRTQVLDHVREMYDEKDDTQIDVEAAVANVEEELE
481 491 501 511 521 531
B. cereus pBC10987 SPCTEEIAVFRALADIVERANDEIVVIDTAITGHTLLLLDAAQTYH-----KEIARSTGE
Bacillus sp. mb24 SPCTEEIAVFRALADIVERANDEIVVIDTAIPGHTLLLLDAAQTYH-----KEIARSSGE
B. sp. CDB3 SPCTEEIAVFRALADIVEIANDEIVVIDTAITGHTLLLLDAAQTYH-----KEIARSSGE
L. innocua SPCTQEIAVFRAFAEIVDDADCDVVVIDTAITGHTLLLLDSTQSYH-----KEVERTSGE
E. coli pR46 SPCTEEIAVFQAFSRVIREAGKRFVVMDTAITGHTLLLLDATGAYH-----REIARKMGD
A. multivorum pKW301 SPCTEEIAVFQAFSRVIREAGKRFVVMDTAITGHTLLLLDATGAYH-----REIAKKMGS
E. coli pR773 SPCTEEIAVFQAFSRVIREAGKRFVVMDTAITGHTLLLLDATGAYH-----REIAKKMGE
S. typhi pR64 SPCTEEIAVFQAFSRIIKEADDHFVIMDTAITGHTLLLLDATGAYH-----REMVRQMGQ
Halobacterium sp. pNRC100 SPCAEEMAALEKFVSYFEEDGYDIVVFDTAITGHTLRLLELPSDWKGFMDLGSLTKGAAP
541 551 561 571 581 591
B. cereus pBC10987 V-PQSVKNLLPRLRNPEETSVVIVTLAEATPVHEASRLQEDLK-RADITPKWWVINQSFY
Bacillus sp. mb24 V-PQSVKNLLPRLRNPEETSVVIVTLAEATPVHEASRLQEDLK-RADITPKWWVINQSFY
B. sp. CDB3 V-PQSVKNLLPRLRNPEETSVVIVTLAEATPVHEASRLQGDLK-RAEIHPKWWVINQSFY
L. innocua V-PESVKRLLPRLQDGKETEVVMVTLPETTPVYESMRLQEDLD-RAGIAHTWWVVNNSML
E. coli pR46 --KGHFTTPMMQLQDQERTKVLLVTLPETTPVLEAANLQSDLE-RAGIHPWGWIINNSLW
A. multivorum pKW301 --KGHFTTPMMQLQDPDRTKVLLVTLPETTPVLEAANLQADLE-RAGIHPWGWIINNSLS
E. coli pR773 --KGHFTTPMMLLQDPERTKVLLVTLPETTPVLEAANLQADLE-RAGIHPWGWIINNSLS
S. typhi pR64 T-HDHVMTPMMQLQDPEKTKVIIVTLAETTPVLEAANLQQDLR-RAGIEPWAWVVNNSLA
Halobacterium sp. pNRC100 ANGGKYDEVIETMQDPSRSSFAFVMYPEFTPMMEAYRAAMDLQDQVGIETSVVVANYLLP

Figure 3.9: Multiple sequence alignment of nine ArsA homologs, *B. cereus* pBC10987, *Bacillus* mb24, *Bacillus* sp. CDB3, *E. coli* pR46, *A. multivorum* pKW301, *E. coli* pR773, *L. innocua, S. typhi* pR64, and *Halobacterium* pNRC100. DTAP sequences for signal transduction are highlighted in Red. GKGGVGKT for the nucleotide binding sites are highlighted in yellow. The conserved cystiene residues are highlighted in green. An asterisk indicates all proteins have an identical residue, blank space indicates that the CDB3 does share the consensus residue. Alignment generated using ClustalW (Thompson et. al., 1994).

3.3.4.4 YqcL

The *yqcl* gene of *Bacillus* sp. CDB3 is a 351 amino acid residue which encodes a membrane protein whose function is similar to that of ArsB. The *yqcl* of CDB3 shows very limited identity to the ArsB proteins of *S. aureus* pI258 (19.3%), *E. coli* pR773 (22.3%) and *E. coli* R46 (20.4%) (Table 3.2). However, it shows greater homology with *Bacillus* spp. *yqcl* gene and *ACR3* gene in *S. cerevisiae* suggesting that *Bacillus* sp. and *S. cerevisiae* possess similar proteins for extrusion of arsenite. Multiple sequence alignment (Figure 3.10) based on the results of blastx search revealed six homologs with the best match to YqcL of *B.* sp. CDB3 and ACR3 of *S. cerevisiae*.

```
                        1         11        21        31        41        51
B. cereus pBC10987      ----------------MSN-----TGKKRLSFLDRYLTLWIFLAMAVGIGVGYLSPG---
B. sp. CDB3             ----------------MSN-----TGKKRLSFLDRYLTLWIFLAMAVGIGVGYLSPG---
B. sp. mb24             ----------------MSN-----TGKKRLSFLDRYLTLWIFLAMAVGIGVGYISPG---
B. cereus ATCC10987     -------------------------MKRLSFLDRYLTLWIFLAMVVGIGLGFVFPS---
B. subtilis             -------------------------MKRLSFLDRYLTIWIFLAMALGIGLGFIFPS---
ACR3 S. cerevisiae      MSEDQKSENSVPSKVNMVNRTDILTTIKSLSWLDLMLPFTIILSIIIAVIISVYVPSSRH
                                    * *       * ** **  *   * *            *

                        61        71        81        91        101       111
B. cereus pBC10987      FVEGMNSLQVGTTSIPLAIGLILMMYPPLAKVRYEEMGRVFKD---VKVLVLSLVQNWII
B. sp. CDB3             FVEGMNSLQVGTTSIPLAIGLILMMYPPLAKVRYEEMGRVFKD---VKVLILSLVQNWII
B. sp. mb24             FVEGMNSLQVGTTSIPLAIGLILMMYPPLAKVRYEEMGRVFKD---VKVLVLSLVQNWII
B. cereus ATCC10987     VVDGLNTLQVGTTSIPLAVGLIVMMYPPLAKVRYEEMGRVFKD---VKVLVLSLVQNWII
B. subtilis             FVGGLNKLQVGTTSIPLAIGLVLMMYPPLAKVRYEEIGRVFKD---IKVLILSLVQNWII
S. cerevisiae           TFDAEGHPNLMGVSIPLTVGMIVMMIPPICKVSWESIHKYFYRSYIRKQLALSLFLNWVI
                              ****  *   ** **  **  *     *      * * ***  ** *

                        121       131       141       151       161       171
B. cereus pBC10987      GPVLMFALAVIFLPDKPEYMVGLIMIGLARCIAMVIVWNDLAKGDTEYAAGLVAFNSVFQ
B. sp. CDB3             GPVLMFALAVIFLPDKPEYMVGLIMIGLARCIAMVIVWNDLAKGDTEYAAGLVAFNSVFQ
B. sp. mb24             GPILMFALAIIFLQDKPEYMVGLIMIGLARCIAMVIVWNDLAKGDTEYAAGLVAFNSVFQ
B. cereus ATCC10987    GPVLMFVLAIIFLPDKPEYMVGLIMIGLARCIAMVIVWNDLADGDKEYAAGLVAFNSVFQ
B. subtilis             GPTLMFILAIIFLPDKPEYMIGLIMIGLARCIAMVIVWNDLSKGDTEYAAGLVAFNSIFQ
S. cerevisiae           GPLLMTALAWMALFDYKEYRQGIIMIGVARCIAMVLIWNQIAGGDNDLCVVLVITNSLLQ
                        ** **  **   * *  **  * **** *******  **    **      **  **  *
```

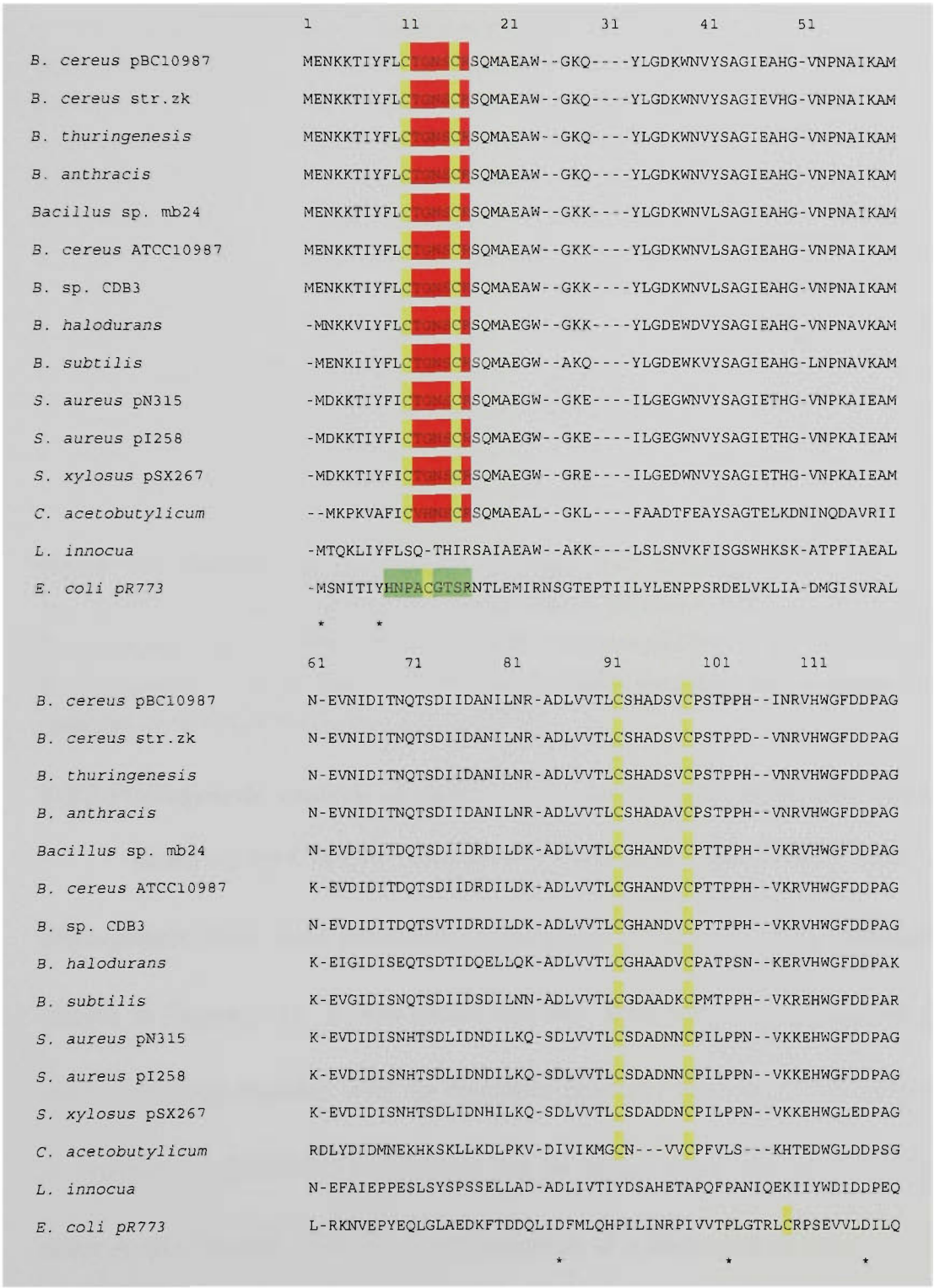
```
                        181       191       201       211       221       231
B. cereus pBC10987      MLFFSVYAYVFVTVIP-EWLGIEGAVVDITMAEVAKSVFIYLGIPFIAGMLTRLIFVKTK
B. sp. CDB3             MLFFSVYAYVFVTVIP-EWLGIEGAVVDITMAEVAKSVFIYLGIPFIAGMLTRFIFVKTK
B. sp. mb24             MLFFSVYAYVFVTVIP-EWLGIEGAVVDITMAEVAKSVFIYLGIPFIAGMLTRFIFVKTK
B. cereus ATCC10987    MLFFSVYAYVFVTVIP-EWLGIEGAIVNITMVEVAKSVCIYLGIPFIAGMLTRFVLVKLK
B. subtilis             MLFSSVYAYIFVTVIP-QWLGMEGAVVNITMAEVAKSVFIYLGVPFIAGMVTRYIFVKVK
S. cerevisiae           MVLYAPLQIFYCYVISHDHLNTSNRVL---FEEVAKSVGVFLGIPLGIGIIIRLGSLTIA
                        *           *     *           ******   ** *   *   *
                        241       251       261       271       281       291
B. cereus pBC10987      GREWYEKVFIPKISPITLIALLFTIIVMFSLKGEVIVSVPFDVVRIAIPLLIYFVIMFFV
B. sp. CDB3             GREWYEKVLIPKISPITLIALLFTIVVMFSLKGEVIVSVPFDVVRIAIPLLIYFVIMFFV
B. sp. mb24             GREWYEKVFIPKISPITLIALLFTIIVMFSLQGEAIISVPFDVVRIAIPLLIYFVVMFFV
B. cereus ATCC10987    GRQWYEKVFIPKISPLTLLALLFTIIVMFSLKGEMIVSVPLDVVRIAIPLLIYFIVMFFV
B. subtilis             GKEWYEKVFIPKISPITLIALLFTIIVMFSLKGDVIVSLPLDVVRVAIPLLIYFVLMFFV
S. cerevisiae           GKSNYEKYILRFISPWAMIGFHYTLFVIFISRGYQFIHEIGSAILCFVPLVLYFFIAWFL
                        *   ***     ***   *    *  * *   *              **  **    *
                        301       311       321       331       341       351
B. cereus pBC10987      SFFMG----------------------KKIG------ASYGVSTTLAFTAGSNNFELAIA
B. sp. CDB3             SFYMG----------------------KKVG------ASYGVSTTLAFTAGSNNFELAIA
B. sp. mb24             SFFMG----------------------KKIG------ASYGVSTTLAFTAGSNNFELAIA
B. cereus ATCC10987    SFFMG----------------------RKIG------ANYPVTTTLAFTAGSNNFELAIA
B. subtilis             SFFLG----------------------KKIG------ANYAVTTTLAFTAGSNNFELAIA
S. cerevisiae           TFALMRYLSISRSDTQRECSCDQELLLKRVWGRKSCEASFSITMTQCFTMASNNFELSLA
                         *                         *         *    * *  **  ******  *
                        361       371       381       391       401
B. cereus pBC10987      VAVGVFGIQSGAAFAAVIGPLVEVPVMIALVNVALWFQRKYFQTQPK
B. sp. CDB3             VAVGVFGIQSGAAFAAVIGPLVEVPVMIALVNVALWFQRKYFQTQPK
B. sp. mb24             VAVGVFGIQSGAAFAAVIGPLVEVPVMIALVNVALWFQRKYFQTQPK
B. cereus ATCC10987    VAVGVFGIHSGAAFAAVIGPLVEVPVMIALVNVAFWFKRKYFNDQPV
B. subtilis             VAVGVFGIHSGAAFAAVIGPLVEVPVMIALVKVALWFQRKYFGSHSM
S. cerevisiae           IAISLYGNNSKQAIAATFGPLLEVPILLILAIVARILKPYYIWNNRN
                              * *  * **  *** ***    *  **      *
```

Figure 3.10: Multiple sequence alignment of six YqcL homologs, *B.* sp. CDB3, *B. subtilis, B. cereus* pBC10987, *B. cereus*ATCC10987, *B.* sp. mb24, and *S. cerevisiae.* An asterisk indicates all proteins have an identical residue, blank space indicates that the CDB3 does share the consensus residue. Alignment generated using ClustalW (Thompson et. al., 1994)

3.3.4.5 ArsC

ArsC of *E. coli* pR773 has only one cystiene residue (Cys12) while all others except *L. innocua* have four cystiene residues (Cys10, Cys15, Cys82 and Cys89) involved in arsenate reduction. The CDB3 ArsC protein shows significantly higher homology (70%) to members of the $ArsC_{sa}$ family, than to the $ArsC_{ec}$ family (20%) (Table 3.2). The $ArsC_{sa}$ family of proteins is characterized by four conserved cystiene residues. The alignment of the fifteen ArsC protein sequences in Figure 3.11 shows these conserved cystiene residues highlighted in yellow. Arsenate reductase from Gram positive and Gram negative bacteria have been shown to share very low sequence homology (about 10% identity), but can catalyze the same chemical reaction. *E. coli* ArsC is coupled to glutaredoxin for its enzyme activity, whilst, *S. aureus* requires thioredoxin and thioredoxin reductase. The difference in redox partner explains the difference in sequence. Cys82 and Cys89 are known to form a disulfide bridge upon oxidation (Bennett et. al., 2001). *Bacillus* spp. including CDB3 share about 70% identity with the *S. aureus* pI258 reductase, with the three important cystiene residues being conserved and are distinctly related to mammalian low molecular weight protein tyrosine phosphatase (LMWPTPase) family with about 18% sequence identity (Bennett et. al., 2001), which includes a p-loop with the characteristic CX_5R sequence motif flanked by a β-strand and an α-helix (Messens et. al., 2002). The CX_5R sequence (CTGNSCR) is conserved in all the strains except *L. innocua* and *E. coli* R773 ArsC. *E. coli* R773 ArsC has been shown not to exhibit phosphatase activity (Demel et. al., 2004), but to possess a

distinct HX_3CX_3R motif (**HNPACGTSR**) (Figure 3.11) instead of the HX_3CX_5R motif, and to partially resemble crambin and glutaredoxin (Martin et. al., 2001).



```
                          121       131       141       151
B. cereus pBC10987         ------KEWSEFQRVRDEIGERIKRFSETGE------
B. cereus str.zk           ------KEWSEFQRVRDEIGERIKRFSETGE------
B. thuringenesis           ------KEWSEFQRVRDEIGERIKCFSETGE------
B. anthracis               ------KEWPEFQRVRDEIGERIKRFSETGE------
Bacillus sp. mb24          ------QEWSVFQRVRDEIGARIKKYAETGE------
B. cereus ATCC10987        ------QEWSVFQRVRDEIGARIKKYAETGE------
B. sp. CDB3                ------QEWSVFQKVRDEIGERIKKFAETGE------
B. halodurans              AEGTDEEKWAVFRRVRDEIGKRIKTFAETGK------
B. subtilis                AQGTEEEKWAFFQRVRDEIGNRLKEFAETGK------
S. aureus pN315            ------KEWSEFQRVRDEIKLAIEKFKLR--------
S. aureus pI258            ------KEWSEFQRVRDEIKLAIEKFKLR--------
S. xylosus pSX267          ------KEWSEFQRVRDEIKLAIENFKLR--------
C. acetobutylicum          ------KSDEEFIRTAKTIEEKVKDLAKRIINKEIEL
L. innocua                 -EIALPKKWASYQEVCDNIALSVKNLEHVLIEA----
E. coli pR773              ----D-AQKGAFTK-ED-GEKVVDEAGKRLK------
```

Figure 3.11: Multiple sequence alignment of fifteen ArsC homologs, CDB3 ArsC, *E. coli* pR773, *L. innocua* pLI100, *B. cereus* pBC10987 and ATCC10987, *Bacillus* sp mb24, *B. cereus* strain zk, *B. subtilis*, *B. halodurans*, *S. aureus* pN315, *S. aureus* pI258, *S. xylosus* pSX267, *C. acetobutylicum*, *B. anthracis* and *B. thuringenesis*. A blank space indicates that the CDB3 does share the consensus residue. Alignment generated using ClustalW (Thompson et. al., 1994).

3.3.5 Phylogenetic analysis of CDB3 ArsR, ArsD, YqcL, ArsC and ArsA proteins

Based on the ClustalW alignments of the ArsR, ArsD, YqcL, ArsC and ArsA, the phylogenetic trees were generated by Neighbor-joining/UPGMA (Felsenstein, 1989) as shown in Figure 3.12. It was found that the ArsR proteins identical to *E. coli* pR773 tended to group together with the distances normally less than 0.1 [except for ArsR from *A. multivorum* pKW301, which only has 84 amino acids instead of 116 amino acids as other ArsRs because of a insertional mutation of a nucleotide at position 122 of the *arsR* gene, Suzuki et. al., 1998], indicating that these *ars* operons probably have evolved from single ancestor (Figure 3.12a). The sequence identities were more than 80% identical when compared with the well-studied *E. coli* pR773 *ars* operon. However, the Ars

proteins from the other six *ars* operons seem to be distinct from the above group. The CDB3 ArsR protein seems to be more related to Gram-positive than to those of Gram-negative. Based on the phylogenetic tree (Figure 3.12b) which showed the divergence of ArsDs identical to R773 to that of ArsDs which lacked the terminal two pairs of cysteine residues Cys112-Cys113 and Cys119-Cys120, indicated that these might have evolved from single ancestor (Figure 3.12b). These organizations further illustrates that the *arsD* and *arsA* genes were laterally transmitted together there by facilitating the insertion of *arsDA* genes into more common *arsRBC* operon (Rosen, 1999).

It was found that the ArsC proteins identical to CDB3 harboring the CX_5R motif tended to group together with the distances less than 0.01 (Figure 3.12c), suggesting a common ancestor, whilst the *E. coli* R773 diverged from *S. aureus* pI258 which utilizes thioredoxin and thioredoxin reductase for its enzyme activity into a new group of enteric bacteria utilizing glutaredoxin with only 20% sequence homology to that of CDB3 and the rest of the Gram positive bacteria.

The ArsA phylogenetic tree as shown in Figure 3.12d represents a clear distinction from Gram negative to its counter part Gram positive bacteria. Although, the distances are less than 0.02 with the exception of *Halobacterium* sp. pNRC100 which branched out the others have shown to be evolved from a single ancestor. CDB3 shares only 50% homology with *E. coli* R773. However, the two consensus sequences GKGGVGKT and DTAP are conserved in all the strains.

The YqcL phylogenetic tree as shown in Figure 3.12e shows that *B.* sp. CDB3 shares a strong homology to *B. cereus* pBC10987 YqcL with a distance score of 0.01. Like *B. subtilis* skin element and ACR3, CDB3 YqcL has 10 transmembrane spanning regions (Chapter 4). A detailed comparison of phylogeny and sequence is presented further in Chapter 4, which focuses on arsenite efflux pumps similar to that of ArsB proteins, which have 12 transmembrane spanning regions.

(a)

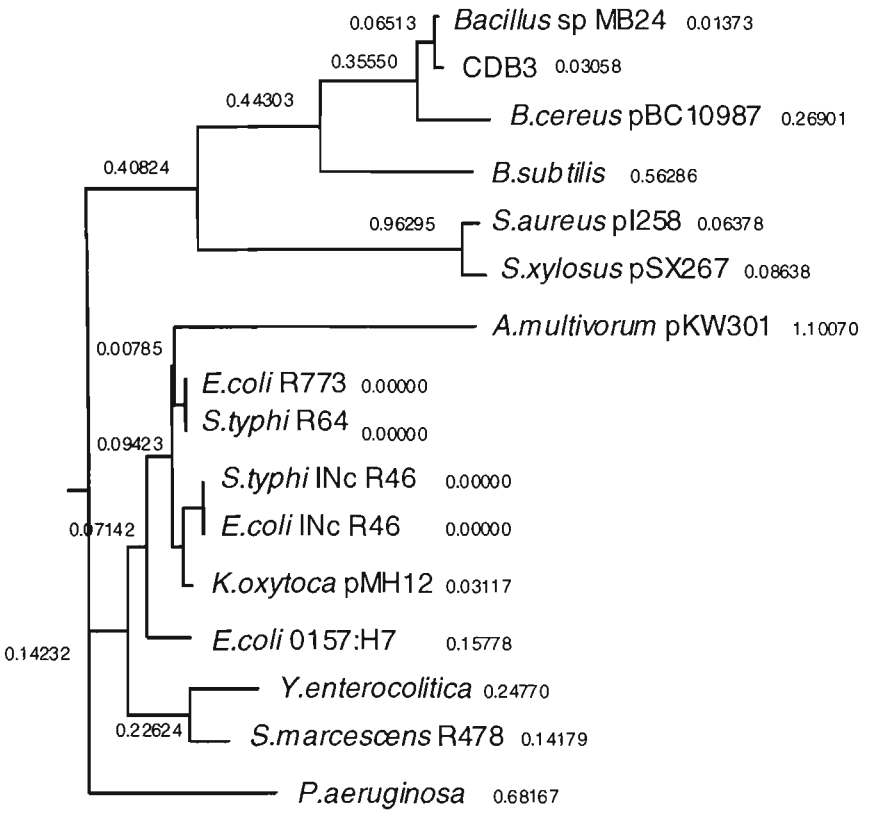


Figure 3.12a: Phylogenetic tree based on ArsR amino acid sequence. The tree was constructed by Neighbor-joining/UPGMA method. Plasmid borne *ars* genes are indicated following the organism. The figures represent the matrix distances which were based on sequence alignments.

(b)

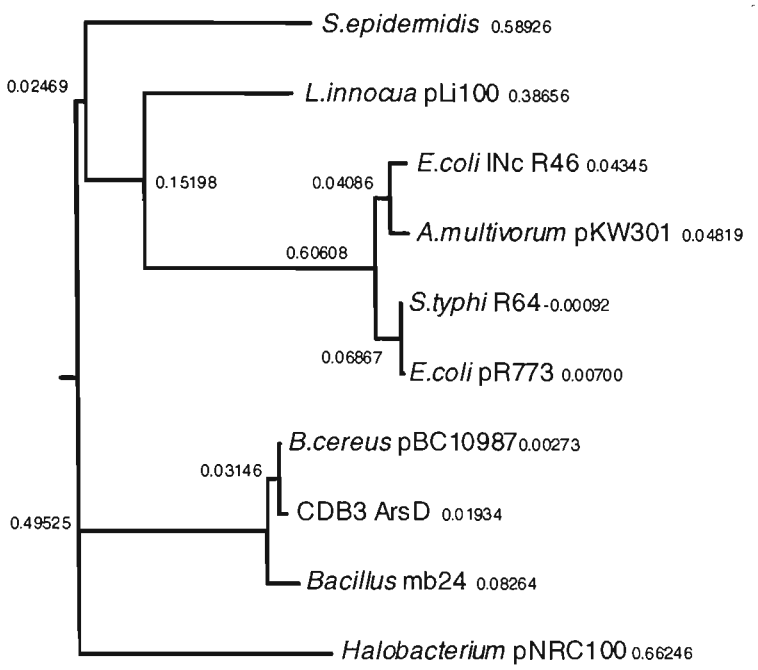


Figure 3.12b: Phylogenetic tree based on ArsD amino acid sequence. The tree was constructed by Neighbor-joining/UPGMA method. Plasmid borne *ars* genes are indicated following the organism. The figures represent the matrix distances which were based on sequence alignments.

(c)

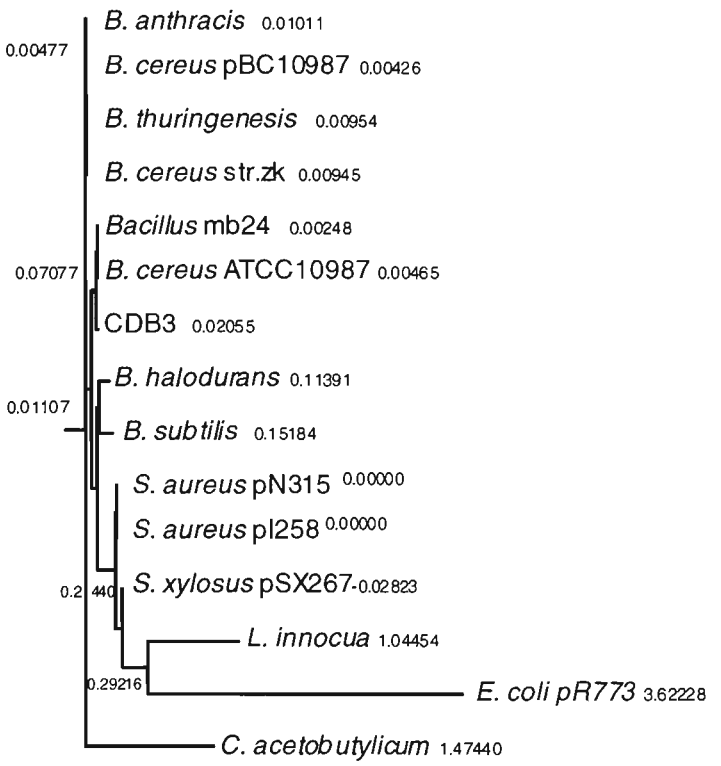


Figure 3.12c: Phylogenetic tree based on ArsC amino acid sequence. The tree was constructed by Neighbor-joining/UPGMA method. Plasmid borne *ars* genes are indicated following the organism. The figures represent the matrix distances which were based on sequence alignments.

(d)

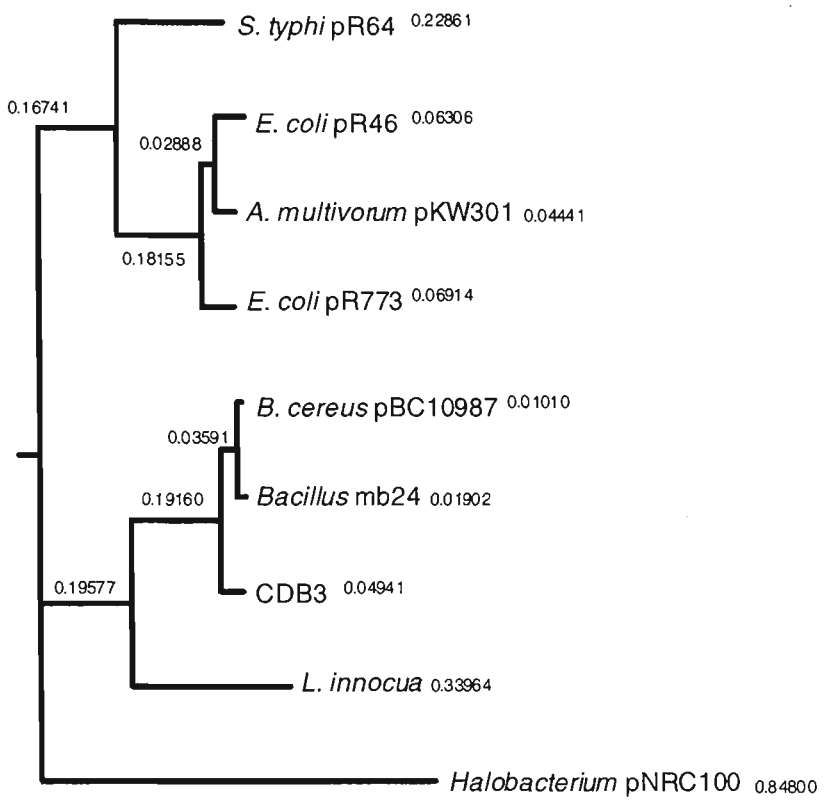


Figure 3.12d: Phylogenetic tree based on ArsA amino acid sequence. The tree was constructed by Neighbor-joining/UPGMA method. Plasmid borne *ars* genes are indicated following the organism. The figures represent the matrix distances which were based on sequence alignments.

(e)

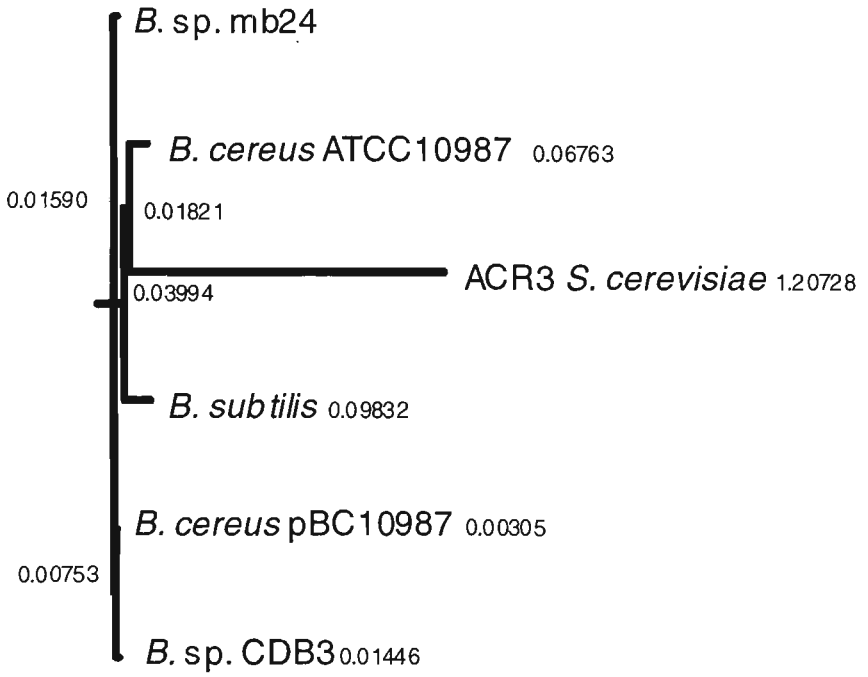


Figure 3.12e: Phylogenetic tree based on YqcL amino acid sequence. The tree was constructed by Neighbor-joining/UPGMA method. Plasmid borne *ars* genes are indicated following the organism. The figures represent the matrix distances which were based on sequence alignments.

3.4 Discussion

The CDB3 Ars proteins share a range of degrees of similarities to other bacterial Ars proteins. They show very high identity with Ars proteins from other *Bacillus* species, especially *B. cereus* pBC10987 and *B.* sp. mb24 (Table 3.3). Homologies to known characterized Ars proteins, for example: *E. coli* pR773 the CDB3 *ars* cluster was 33% (ArsR), 22.3% (YqcL) 20.5% (ArsC), 35% (ArsD) and 50.4% (ArsA) identical exhibiting a weak homology when compared to *Bacillus* spp. For example: *B. subtilis* skin element which is a Gram positive bacterium whose *ars* genes are organized as *arsRorf2BC* shares a greater homology of 42.3% (ArsR), 85.6% (YqcL) and 75.2% (ArsC) to CDB3 Ars proteins. Besides the CDB3 *ars* cluster, two more *ars* clusters from *Bacillus* spp. have recently been identified whose *ars* genes are organized as *arsRYCDA*. However, no functional characterization study has been performed and the identity is purely based on genome sequencing (Rasko et. al., 2004). The organization of these *ars* genes was different from those clusters containing *arsD* and *arsA* implying difference in gene regulation which was noticed in CDB3 as presented in Chapter 7 via northern (RNA) blot DNA-RNA hybridization analysis showing inducible synthesis of a full-length *ars* mRNA, about 7 kb in size. In the *ars* clusters characterized so far, *arsD* and *arsA* have been identified as located behind *arsC* instead of being located between *arsR* and *arsB* for example: *E. coli* pR773 *arsRDABC* cluster which is the most well studied *ars* gene cluster. The possibility of recombination resulting in artificially false gene order during cloning process was excluded based on the Southern blot analysis (Figure 3.5). Whether

arsD and *arsA* were present in the cluster from the very beginning or acquired from other sources later during the evolution processes of *ars* operons still remains unknown, but the location and gene order of *arsD* and *arsA* favor strongly the hypothesis that the *ars* operons probably acquired *arsD* and *arsA* simultaneously and more recently than other *ars* genes (Figure 3.2, Luo, 2006).

Based on the 16S rDNA sequencing, CDB3 shares approximately 99% homology to *B. cereus* (Chapter 2; Chopra et. al., 2007). The strain *B. cereus* ATCC10987 harbors two *ars* operons; one on the chromosome consisting of only three genes (*arsRYC*) with another present on pBC10987, which consists of seven genes (*arsRYCDAIP*), similar to that found in *Bacillus* sp. mb24. The CDB3 operon consists of eight ORFs, *arsRYCDATIP*. The *arsT* gene was identified and found to be a part of the cluster is absent in both *B. cereus* pBC10987 and *Bacillus* sp. mb24 *ars* clusters which makes it distinct and interesting. The order of genes appears similar but located on the plasmid pBC10987 of *B. cereus* ATCC10987 whereas, the *ars* cluster cloned from CDB3 strain is located on chromosome (Figure 3.5).

Several important features have been identified in the CDB3 Ars proteins. Firstly, YqcL, a 10 transmembrane domain (TMD) protein first identified in the *B. subtilis* skin element, endows cells with resistance to AsIII suggesting that the YqcL protein might have a similar function to that of ArsB proteins which has 12 membrane spanning regions. A detailed phylogenetic and sequence comparison on arsenite efflux pumps is presented in Chapter 4, which further details the difference in membrane topology and

sequence characters. The homology between YqcL and ArsB is quite different. The CDB3 YqcL shares weak homology with ArsBs (Table 3.2). Bobrowicz et. al. (1997) have revealed that three contiguous genes *ACR1, ACR2* and *ACR3* are involved in arsenic resistance. Interestingly, the *ACR3* gene has high similarity to the CDB3 YqcL (Figure 3.10), suggesting that *Bacillus* spp. and *S. cerevisiae* possess similar proteins for extrusion of arsenite. No functional study of the interaction of YqcL with ArsA has been reported till date. Further in Chapter 5, the functional characteristics of CDB3 ArsA and its interaction with YqcL and *E. coli* ArsB is presented which is the first experimental evidence demonstrated.

Secondly, the *arsD* from Gram negative bacteria (*E. coli* pR773 *arsRDABC* operon) probably originated from insertion of *arsDA* genes into the more common *arsRBC* cluster (Rosen, 1999; Li et. al., 2001) shows weak homology to CDB3 *arsD* (35%) where it lacks the four C-terminal cysteine residues (Figure 3.8). Yet it is still functional as a secondary regulator in controlling the basal level of gene expression. This is quite novel, as the terminal two cysteine residues (Cys119 and Cys120) where thought to be involved in induction of *arsD* expression in *E. coli* pR773 (Li et. al., 2001). A detailed study is presented in Chapter 6, which explores the diversity of the two different ArsDs characterized so far.

ars clusters can be more diversified in helping host bacteria to adapt to different living environment efficiently. The cloned *ars* gene cluster from *Bacillus* sp. CDB3 seemed a good example for further investigation with respect to the sequence analysis of

YqcL and its interaction with ArsA. The regulatory mechanism of this novel *ars* gene cluster 1 is then analyzed which is discussed in later Chapters.

Chapter 4

Phylogenetic and Protein Property Studies of Arsenite Efflux Pumps in Bacteria

4.1 Introduction

Transport across membranes is vital and membrane transporter proteins enable cells to take up essential nutrients and extrude toxic compounds. Micro organisms have acquired resistance to arsenical compounds provided by the *ars* genes located on chromosome or plasmid. An efflux pump protein, which is encoded by one of the *ars* genes, is essential for arsenite resistance exhibited by such micro organisms. Although the number of arsenite pumps characterised so far are limited, their wide occurrence in micro organisms is evident based on DNA sequence data.

Such pumps are encoded by the *arsB* or *B*-like genes. Most bacterial arsenic operons identified so far contain either three (*arsRBC*) or five genes (*arsRDABC*) (Ryan and Colleran, 2002). The *arsB* and *arsA* genes express a membrane protein and an ATPase, respectively. While ArsB alone can remove arsenite out of the cell, the ArsAB complex works more efficiently.

ArsB protein belongs to the major facilitator super family (MFS) of transporters and has 10-12 transmembrane helices. The ArsB is a 45-kDa inner membrane protein that has 12 membrane-spanning regions (Wu et. al., 1992) and is a permease which pumps arsenite out of the cell either independently as a chemiosmotic transporter or as a channel forming subunit of an ATP driven anion pump (ArsAB). Permeases can have both inside and outside polarity and function as uniporters, symporters or antiporters with a wide range of transport specificities. ArsB may pump AsO^{2-} out of the cell as a chemiosmotic transporter (Broer et. al., 1993). The molecular structure of ArsB is not known and the mechanism by which ArsB extrudes arsenite is not well understood yet. Metal efflux is coupled to a transmembrane electrochemical gradient (Dey et. al., 1995). Structurally the

ArsB arsenite pumps resemble the ABC type efflux pumps which have two groups of six TM spanning helical segments and two NBDs similar to ArsB, but there is no significant sequence similarity between the two. The function of ArsB is to act either as a potential driven secondary carrier or as a subunit of an obligatory primary ATP-coupling pump (Kuroda et. al., 1997). When there is no ArsA, ArsB transports arsenical oxyanions with energy derived from the proton-pumping respiratory chain. When an ArsB-ArsA complex is formed, the transporter complex will pump the arsenic ion out of the cell utilizing the energy of ATP. Thus, the ArsB-ArsA complex can pump arsenite through the membrane with a higher metabolic efficiency. The *yqcl* gene of *B. subtilis* skin element encodes a membrane protein whose function is similar to that of ArsB, though the two proteins appear to be only remotely related in sequence (Sato and Kobayashi, 1998). The YqcL protein has 10 membrane-spanning regions, different to most well studied ArsB proteins that have 12 such regions. While all the other reported bacterial *ars* operons also confer resistance to antimonite to their hosts (Rosen, 1999), interestingly, the wild-type strain of *B. subtilis* exhibits sensitivity to antimonite (Sato and Kobayashi, 1998), indicating that YqcL has arsenite specificity. The yeast *S. cerevisiae ACR3* gene encodes a plasma membrane arsenite-efflux transporter similar to YqcL (Bobrowicz et. al., 1997). Due to all these differences, YqcL and ACR3 have been suggested to have evolved separately from the ArsB proteins (Rosen, 1999). The arsenite pump of *Bacillus* sp. CDB3 also belongs to this group (Chapter 3).

Since arsenite pumps are widely occurring in bacteria, a computational analysis was carried out to depict the consensus sequences shared by these proteins and possibly

their evolution. In the course of analysis in my study presented in this Chapter, a possible new sub-group of arsenite pumps in proteobacteria was identified.

4.2 Materials and Methods

4.2.1 Sequence search and alignments

Searching for homologous sequences in data bank was carried out using blastx analysis (Altschul et. al., 1990; http://www.ncbi.nlm.nih.gov/blast). Sequences from different organisms were assembled and multiple alignments were done using Bestfit and ClustalW (Thompson et. al., 1994). The known ArsB sequences were compared with homologous sequences deposited in the Genbank and EMBL. A total of 26 sequences were analysed consisting of 24 bacterial, 1 archeal and yeast.

4.2.2 Sequence homology values

To evaluate the protein sequence similarity, the PRSS program (Pearson, 1988) was used. PRSS evaluates the significance of pairwise protein sequence similarity scores using Monte-Carlo analysis. The Smith-Waterman (1981) algorithm was used to calculate similarity scores between two sequences. The second sequence is shifted and compared with the first sequence 200 to 1000 times. If the similarity between the two sequences is real, the similarity score for the unshuffled sequence will be a lot higher than the shuffled score. A representative from each group was chosen and the analysis was carried out with all tested species. The relative shuffled, unshuffled and P scores obtained were then tabulated.

4.2.3 Phylogenetic analysis

Phylogenetic trees were constructed based on the multiple alignments of all the sequences using the Neighbor-Joining/UPGMA method (version 3.573c) in the PHYLIP package (Felsenstein, 1989).

4.2.4 Prediction of transmembrane domains

To predict the TMDs, hydropathy plots were constructed based on the Kyte & Doolittle (1982) hydropathy profile. Plots were generated using PepWindow program (Window size-integer 1 to 200, default is 7) available from ANGIS service (http://bioman6.angis.org.au) for all the bacterial strains used in this study.

4.2.5 Membrane topology

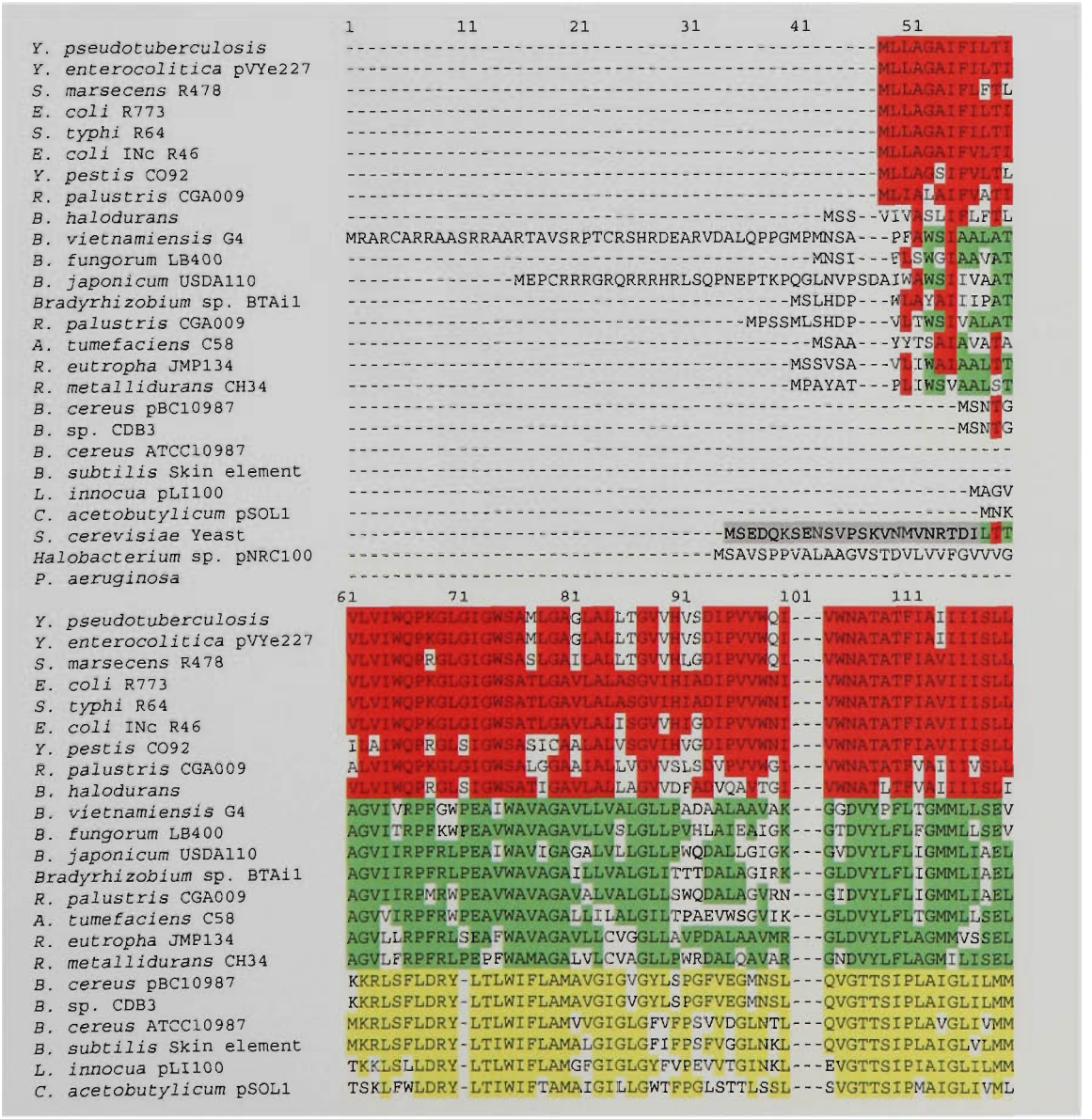
Membrane topology was predicted using the MEMSTAT3 program available from http://bioinf.cs.ucl.ac.uk/psipred/psiform.html. Protein sequences were submitted to the PSIPRED protein structure prediction server (McGuffin et. al., 2000). It predicts the secondary structure and topology of integral membrane proteins by recognizing topological models.

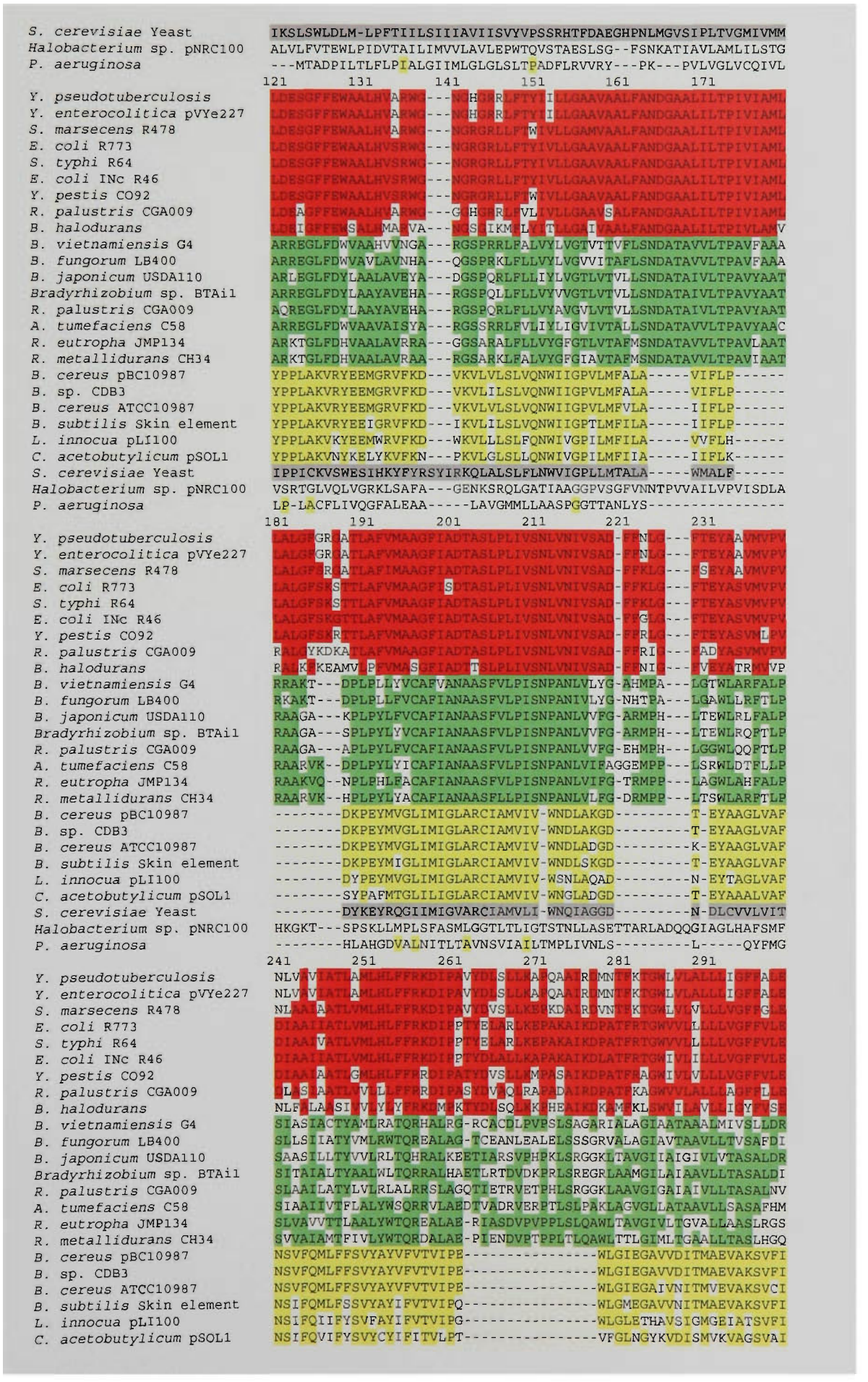
4.3 Results

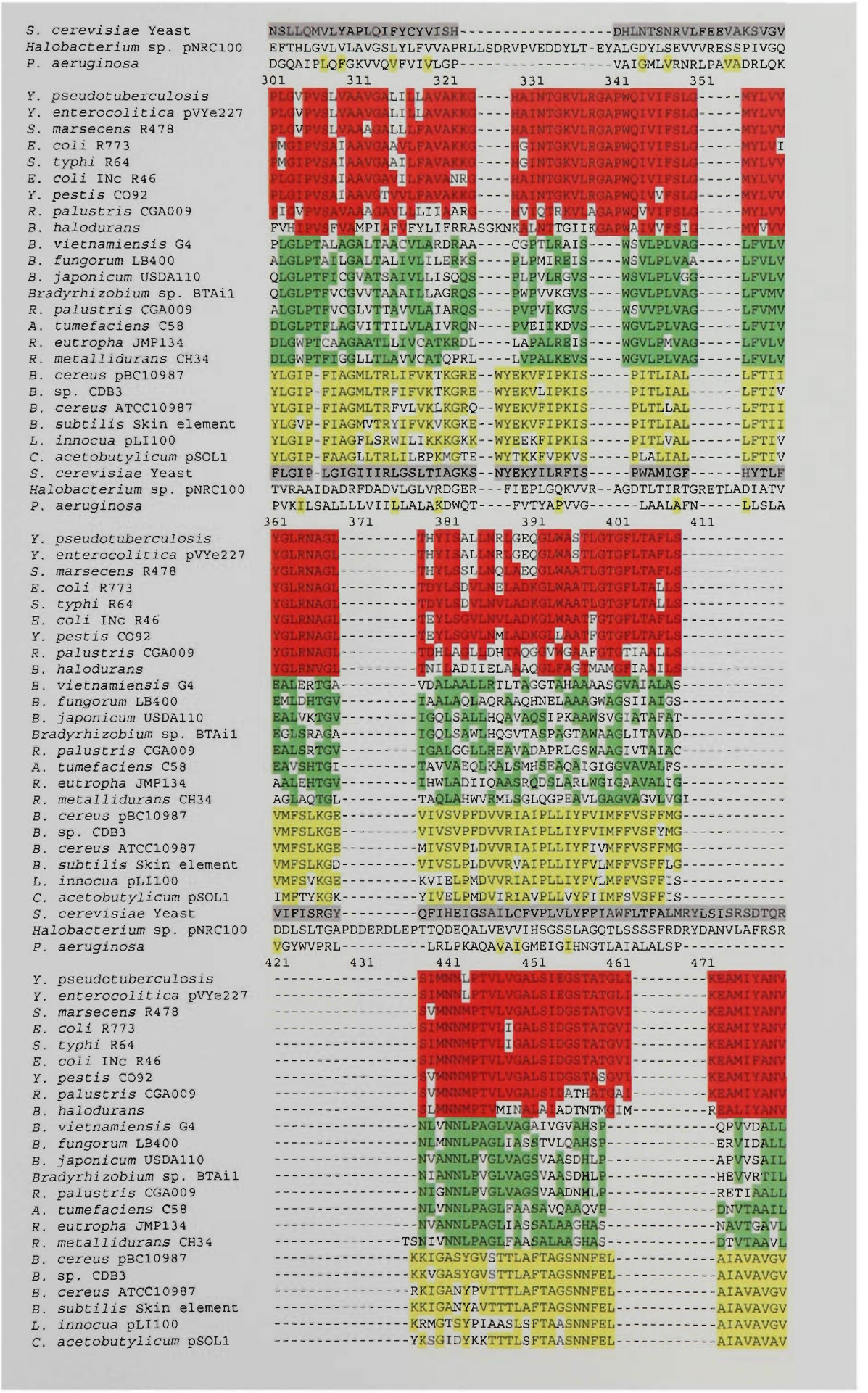
4.3.1 Multiple sequence alignments of arsenite membrane transporters and their sequence homology

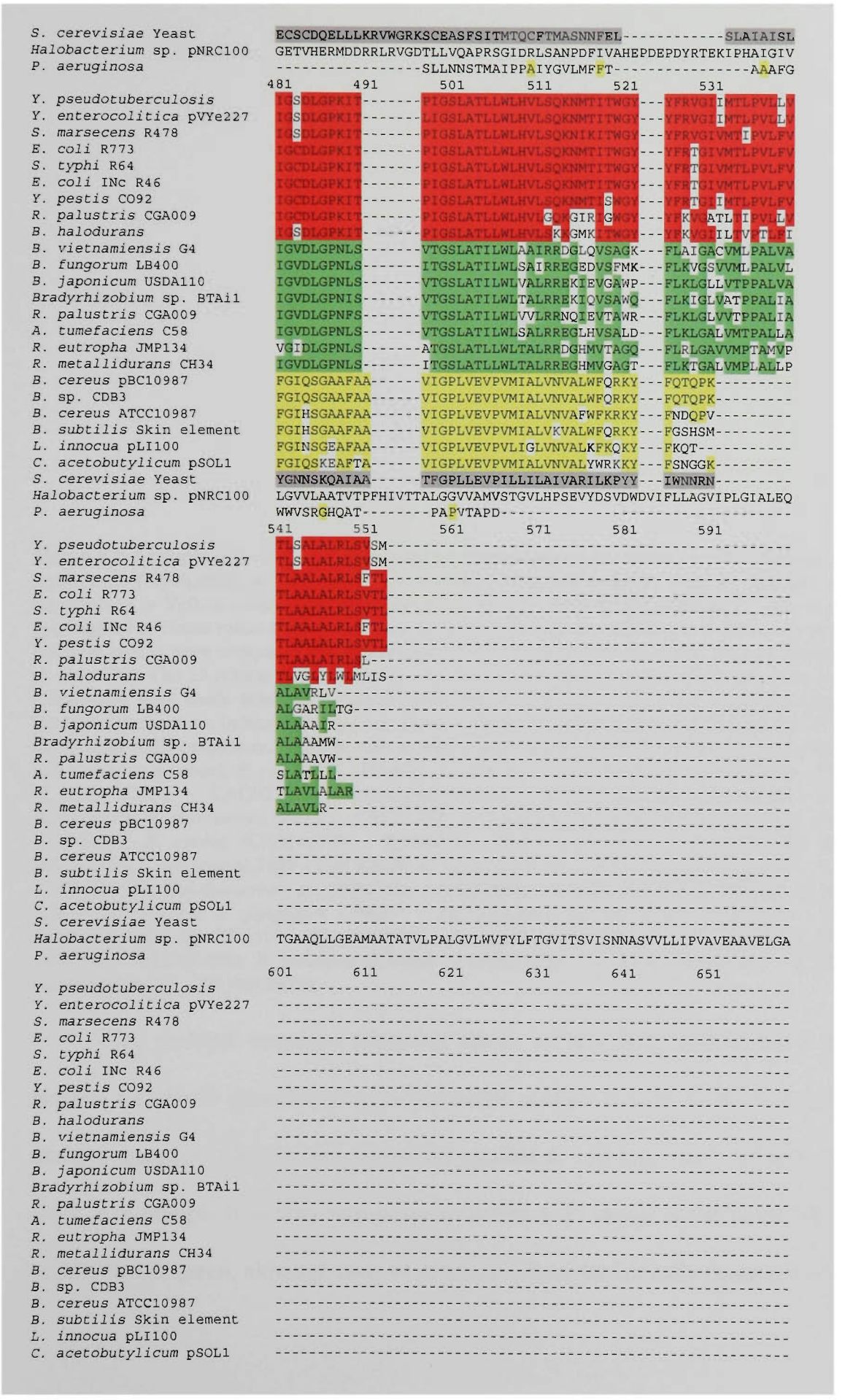
Blastx searches were performed using the *E. coli* ArsB and *Bacillus* YqcL as query sequences. 25 homologs were chosen and aligned together with the query sequences. Figure 4.1 highlights areas of importance in considering the similarity shared by different arsenite pump homologs.

The sequences of ArsB from different strains were compared with the related species acquired from Genbank through blastx (Figure 4.1).




S. cerevisiae Yeast
Halobacterium sp. pNRC100
P. aeruginosa
Y. pseudotuberculosis
Y. enterocolitica pVYe227
S. marsecens R478
E. coli R773
S. typhi R64
E. coli INc R46
Y. pestis CO92
R. palustris CGA009
B. halodurans
B. vietnamiensis G4
B. fungorum LB400
B. japonicum USDA110
Bradyrhizobium sp. BTAi1
R. palustris CGA009
A. tumefaciens C58
R. eutropha JMP134
R. metallidurans CH34
B. cereus pBC10987
B. sp. CDB3
B. cereus ATCC10987
B. subtilis Skin element
L. innocua pLI100
C. acetobutylicum pSOL1
S. cerevisiae Yeast
Halobacterium sp. pNRC100
P. aeruginosa
Y. pseudotuberculosis
Y. enterocolitica pVYe227
S. marsecens R478
E. coli R773
S. typhi R64
E. coli INc R46
Y. pestis CO92
R. palustris CGA009
B. halodurans
B. vietnamiensis G4
B. fungorum LB400
B. japonicum USDA110
Bradyrhizobium sp. BTAi1
R. palustris CGA009
A. tumefaciens C58
R. eutropha JMP134
R. metallidurans CH34
B. cereus pBC10987
B. sp. CDB3
B. cereus ATCC10987
B. subtilis Skin element
L. innocua pLI100
C. acetobutylicum pSOL1
S. cerevisiae Yeast
Halobacterium sp. pNRC100
P. aeruginosa
Y. pseudotuberculosis
Y. enterocolitica pVYe227
S. marsecens R478
E. coli R773
S. typhi R64
E. coli INc R46
Y. pestis CO92
R. palustris CGA009
B. halodurans
B. vietnamiensis G4
B. fungorum LB400
B. japonicum USDA110
Bradyrhizobium sp. BTAi1
R. palustris CGA009
A. tumefaciens C58
R. eutropha JMP134
R. metallidurans CH34
B. cereus pBC10987
B. sp. CDB3
B. cereus ATCC10987
B. subtilis Skin element
L. innocua pLI100
C. acetobutylicum pSOL1


S. cerevisiae Yeast
Halobacterium sp. pNRC100
P. aeruginosa
Y. pseudotuberculosis
Y. enterocolitica pVYe227
S. marsecens R478
E. coli R773
S. typhi R64
E. coli INc R46
Y. pestis CO92
R. palustris CGA009
B. halodurans
B. vietnamiensis G4
B. fungorum LB400
B. japonicum USDA110
Bradyrhizobium sp. BTAi1
R. palustris CGA009
A. tumefaciens C58
R. eutropha JMP134
R. metallidurans CH34
B. cereus pBC10987
B. sp. CDB3
B. cereus ATCC10987
B. subtilis Skin element
L. innocua pLI100
C. acetobutylicum pSOL1
S. cerevisiae Yeast
Halobacterium sp. pNRC100
P. aeruginosa
Y. pseudotuberculosis
Y. enterocolitica pVYe227
S. marsecens R478
E. coli R773
S. typhi R64
E. coli INc R46
Y. pestis CO92
R. palustris CGA009
B. halodurans
B. vietnamiensis G4
B. fungorum LB400
B. japonicum USDA110
Bradyrhizobium sp. BTAi1
R. palustris CGA009
A. tumefaciens C58
R. eutropha JMP134
R. metallidurans CH34
B. cereus pBC10987
B. sp. CDB3
B. cereus ATCC10987
B. subtilis Skin element
L. innocua pLI100
C. acetobutylicum pSOL1
S. cerevisiae Yeast
Halobacterium sp. pNRC100
P. aeruginosa
Y. pseudotuberculosis
Y. enterocolitica pVYe227
S. marsecens R478
E. coli R773
S. typhi R64
E. coli INc R46
Y. pestis CO92
R. palustris CGA009
B. halodurans
B. vietnamiensis G4
B. fungorum LB400
B. japonicum USDA110
Bradyrhizobium sp. BTAi1
R. palustris CGA009
A. tumefaciens C58
R. eutropha JMP134
R. metallidurans CH34
B. cereus pBC10987
B. sp. CDB3
B. cereus ATCC10987
B. subtilis Skin element
L. innocua pLI100
C. acetobutylicum pSOL1


S. cerevisiae Yeast
Halobacterium sp. pNRC100
P. aeruginosa
Y. pseudotuberculosis
Y. enterocolitica pVYe227
S. marsecens R478
E. coli R773
S. typhi R64
E. coli INc R46
Y. pestis CO92
R. palustris CGA009
B. halodurans
B. vietnamiensis G4
B. fungorum LB400
B. japonicum USDA110
Bradyrhizobium sp. BTAi1
R. palustris CGA009
A. tumefaciens C58
R. eutropha JMP134
R. metallidurans CH34
B. cereus pBC10987
B. sp. CDB3
B. cereus ATCC10987
B. subtilis Skin element
L. innocua pLI100
C. acetobutylicum pSOL1
S. cerevisiae Yeast
Halobacterium sp. pNRC100
P. aeruginosa
Y. pseudotuberculosis
Y. enterocolitica pVYe227
S. marsecens R478
E. coli R773
S. typhi R64
E. coli INc R46
Y. pestis CO92
R. palustris CGA009
B. halodurans
B. vietnamiensis G4
B. fungorum LB400
B. japonicum USDA110
Bradyrhizobium sp. BTAi1
R. palustris CGA009
A. tumefaciens C58
R. eutropha JMP134
R. metallidurans CH34
B. cereus pBC10987
B. sp. CDB3
B. cereus ATCC10987
B. subtilis Skin element
L. innocua pLI100
C. acetobutylicum pSOL1
S. cerevisiae Yeast
Halobacterium sp. pNRC100
P. aeruginosa
Y. pseudotuberculosis
Y. enterocolitica pVYe227
S. marsecens R478
E. coli R773
S. typhi R64
E. coli INc R46
Y. pestis CO92
R. palustris CGA009
B. halodurans
B. vietnamiensis G4
B. fungorum LB400
B. japonicum USDA110
Bradyrhizobium sp. BTAi1
R. palustris CGA009
A. tumefaciens C58
R. eutropha JMP134
R. metallidurans CH34
B. cereus pBC10987
B. sp. CDB3
B. cereus ATCC10987
B. subtilis Skin element
L. innocua pLI100
C. acetobutylicum pSOL1

```
S. cerevisiae Yeast          ------------------------------------------------------------
Halobacterium sp. pNRC100    NPFAFVLAVTFAASTAFMTPVGYQTNLFVYGPGGYTFSDFFRVGLPLQLLLSVVTVGGIA
P. aeruginosa                ------------------------------------------------------------
                            661
Y. pseudotuberculosis        -----
Y. enterocolitica pVYe227    -----
S. marsecens R478            -----
E. coli R773                 -----
S. typhi R64                 -----
E. coli INc R46              -----
Y. pestis CO92               -----
R. palustris CGA009          -----
B. halodurans                -----
B. vietnamiensis G4          -----
B. fungorum LB400            -----
B. japonicum USDA110         -----
Bradyrhizobium sp. BTAi1     -----
R. palustris CGA009          -----
A. tumefaciens C58           -----
R. eutropha JMP134           -----
R. metallidurans CH34        -----
B. cereus pBC10987           -----
B. sp. CDB3                  -----
B. cereus ATCC10987          -----
B. subtilis Skin element     -----
L. innocua pLI100            -----
C. acetobutylicum pSOL1      -----
S. cerevisiae Yeast          -----
Halobacterium sp. pNRC100    AFWGV
P. aeruginosa                -----
```

Figure 4.1: Multiple sequence alignment of the arsenite membrane transporter proteins belonging to *E.coli* type ArsB (highlighted in Red refers to identical conserved residues shared), *Bacillus* type YqcL (highlighted in Yellow refers to identical conserved residues shared) and predicted 3rd sub-group of ArsB (highlighted in Green refers to identical conserved residues shared). Sequences of ArsB- like proteins from different strains were compared with the related species acquired from Genbank through BlastX (Altschul et. al., 1990). The 25 related species were chosen according to their pairwise alignment score and identity. Alignments were made using the ClustalW program (Thompson et. al., 1994; http://www.angis.org.au). Protein sequences are indicated by species name. Asterisk '*' refers to strains harboring Na^+/H^+ antiporter and related AsIII permeases. NCBI protein accession numbers are listed below (indicated by species name and accession number): *E. coli* R773 – P08691; *S. typhi* R64 – BAB91586; *E. coli* INc R46 – NP_511239; *Y. pestis* CO92 – CAC92577; *Y. pseudotuberculosis* – CAH20025; *Y. enterocolitica* pVYe227 – AAD16859; *S. marcescens* R478 – CAE51702; *B. halodurans* – BAB06718; *B. cereus* pBC10987 – AAS45009; *B. cereus* ATCC10987 – AAS42118; CDB3 (*Bacillus* sp.) – AAD51846; *B.subtilis* – BAA12433; *L. innocua* pLI100 – CAC42037; *C. acetobutylicum* pSOL1 – AAK76849; ACR_3 *S. cerevisiae* – NP_015527; *Halobacterium* sp. pNRC100 – AAG20642; *P. aeruginosa* – AAG05039; *A. tumefaciens* C58 – AAL44356; *B. fungorum* LB400* - ZP_00283047; *B. japonicum* USDA110 – BAC50537; *Bradyrhizobium* sp. BTAi1* - ZP_00862573; *R. metallidurans* CH34 – ZP_00593623; *R. eutropha* JMP134 – AAZZ63030; *R. palustris* CGA009 – CAE27699; *R. palustris* CGA009* - CAE28839; *B. vietnamiensis* G4 – ZP_00426020.

The multiple sequence alignment shows a clear distinction of the 3 groups of arsenite membrane extrusion pumps. The amino acid residues highlighted in red refers to identity shared with that of the commonly existing *E. coli* type ArsB, whereas, yellow to YqcL like pumps. It is very interesting to notice a 3rd group of the pump like proteins highlighted in green, although none of these have been functionally characterized.

Further statistical analyses were performed to consolidate/verify the distinctions between the groups of protein sequences revealed by multiple alignments (Table 4.1, 4.2 and 4.3).

Table 4.1: PRSS evaluation of homology of *E. coli* pR773 ArsB to other arsenite transporter proteins

Strain	Unshuffled score	Shuffled score	P-Score
ArsB group			
S. typhi **R64**	**2693**	**66-144**	**4.9812e-102**
E. coli lNc **R46**	**2569**	**70-145**	**9.7108e-92**
Y. pestis **CO92**	**2344**	**70-131**	**1.2642e-79**
Y. pseudotuberculosis	**2358**	**69-136**	**6.6428e-75**
Y. enterocolitica **pVYe227**	**2472**	**74-155**	**3.3066e-89**
S. marsecens **R478**	**2398**	**67-140**	**6.3017e-81**
B. halodurans	**1747**	**64-121**	**1.0775e-66**
R. palustris **CGA009**	**2121**	**66-121**	**1.3816e-66**
YqcL group			
B. sp. **CDB3**	95	53-116	0.13668
B. cereus **pBC10987**	93	59-123	0.15697
B. cereus **ATCC10987**	80	59-123	0.44468
B. subtilis **"YqcL"**	71	50-125	0.6683
L. innocua **pLI100**	87	57-132	0.24437
C. acetobutylicum **pSOL1**	106	54-142	0.077319
S. cerevisiae Yeast	64	45-106	0.4999
Halobacterium **pNRC100**	132	53-110	0.0024582
P. aeruginosa	81	62-123	0.5038
3rd putative group			
B. vietnamiensis **G4**	**500**	**57-134**	**7.9101e-19**
B. japonicum **USDA110**	**458**	**58-126**	**7.5687e-16**
A. tumefaciens **C58**	**490**	**60-124**	**8.6659e-19**
R. eutropha **JMP134**	**482**	**63-123**	**1.7777e-16**
R. metallidurans **CH34**	**443**	**62-153**	**2.4532e-13**
B. fungorum **LB400***	**444**	**58-125**	**3.4568e-15**
Bradyrhirhizobium sp **BTAi1***	**479**	**57-136**	**3.0373e-17**
R. eutropha **JMP134***	**482**	**58-136**	**4.736e-16**
R. palustris **CGA009***	**500**	**63-153**	**1.2251e-17**

'*" refers to strains harboring Na^+/H^+ antiporter and related AsIII Permeases

Table 4.2: PRSS evaluation of homology of *B. subtilis* YqcL to other arsenite transporter proteins

Strain	Unshuffled score	Shuffled score	P-Score
ArsB group			
S. typhi **R64**	71	53-124	0.67409
E. coli **R773**	71	49-114	0.67355
E. coli INc **R46**	87	60-133	0.35453
Y. pestis **CO92**	91	53-124	0.25021
Y. pseudotuberculosis	88	57-123	0.23928
Y. enterocolitica **pVYe227**	75	51-126	0.65221
S. marsecens **R478**	83	58-139	0.38864
B. halodurans	78	62-127	0.69684
R. palustris **CGA009**	86	53-119	0.20476
YqcL group			
B. sp. **CDB3**	**1993**	**63-114**	**2.0801e-79**
B. cereus **pBC10987**	**2000**	**65-131**	**5.5195e-67**
B. cereus **ATCC10987**	**1990**	**66-134**	**7.8583e-77**
L. innocua **pLI100**	**1705**	**57-131**	**6.3828e-61**
C. acetobutylicum **pSOL1**	**1497**	**64-132**	**3.2664e-57**
S. cerevisiae Yeast	**669**	**48-97**	**1.1483e-29**
Halobacterium **pNRC100**	63	51-110	0.8093
P. aeruginosa	165	52-128	0.00019521
3rd putative group			
B. vietnamiensis **G4**	74	46-98	0.070849
B. japonicum **USDA110**	79	56-97	0.2269
A. tumefaciens **C58**	76	45-101	0.20295
R. eutropha **JMP134**	66	42-93	0.28216
R. metallidurans **CH34**	86	48-114	0.060893
B. fungorum **LB400***	69	48-109	0.37016
Bradyrhirhizobium sp **BTAi1***	69	46-97	0.30386
R. eutropha **JMP134***	66	47-98	0.36187
R. palustris **CGA009***	102	48-119	0.016883

"*" refers to strains harboring Na^+/H^+ antiporter and related AsIII Permeases

Table 4.3: PRSS evaluation of homology of *R. eutropha* JMP134 to other arsenite transporter proteins

Strain	Unshuffled score	Shuffled score	P-Score
ArsB group			
S. typhi **R64**	**488**	**56-131**	**3.8622e-15**
E. coli **R773**	**482**	**62-138**	**1.3219e-16**
E. coli INc **R46**	**490**	**58-153**	**3.399e-15**
Y. pestis **CO92**	**499**	**63-128**	**1.378e-14**
Y. pseudotuberculosis	**500-**	**64-148**	**4.6446e-17**
Y. enterocolitica **pVYe227**	**478**	**58-127**	**2.9809e-14**
S. marsecens **R478**	**484**	**60-127**	**1.8174e-14**
B. halodurans	**500**	**57-123**	**1.4131e-17**
R. palustris **CGA009**	**537**	**56-146**	**3.5961e-16**
YqcL group			
B. sp. **CDB3**	65	49-109	0.53518
B. cereus **pBC10987**	66	45-112	0.43706
B. cereus **ATCC10987**	70	50-89	0.28662
B. subtilis **"YqcL"**	66	48-91	0.4139
L. innocua **pLI100**	56	44-91	0.7306
C. acetobutylicum **pSOL1**	59	46-99	0.71935
S. cerevisiae Yeast	64	40-86	0.17537
Halobacterium **pNRC100**	165	56-121	0.00018294
P. aeruginosa	70	58-128	0.833
3rd putative group			
B. vietnamiensis **G4**	**1393**	**71-175**	**1.2339e-36**
B. japonicum **USDA110**	**1365**	**69-144**	**1.0905e-47**
A. tumefaciens **C58**	**1429**	**64-151**	**2.0967e-50**
R. metallidurans **CH34**	**1870**	**71-144**	**1.2947e-60**
B. fungorum **LB400***	**1358**	**64-123**	**8.9521e-48**
Bradyrhirhizobium sp **BTAi1***	**1347**	**66-141**	**1.2268e-42**
R. eutropha **JMP134***	**2646**	**80-161**	**9.1397e-78**
R. palustris **CGA009***	**1323**	**62-168**	**3.5765e-42**

'*' refers to strains harboring Na^+/H^+ antiporter and related AsIII Permeases

PRSS analysis showed significant differences between two known groups of arsenite pumps, ArsB and YqcL, and also between each of these two and the assumed third group. The similarity seen between these groups were relatively low. Individual comparisons with respect to each were made by PRSS and scores obtained show a clear distinction of the groups based on sequence homology. If the similarity between two groups of sequences is real then the similarity score for the unshuffled sequence will be much higher when compared to the shuffled score (Pearson, 1988). This was the case

with respect to each representative when compared with its own group members; whereas, the score was much lower when it was compared with members of another group. The p-score is more informative and accurate than just the similarity as it also takes the length of match and gaps inserted into account. As shown in Table 4.1, the p-scores are much higher between the *E. coli* ArsB and the pump proteins of *S. typhi* R64, *E. coli* INc R46, *Y. pestis* CO92, *Y. pseudotuberculosis*, *Y. enterocolitica* pVYe227, *S. marsecens* R478, *B. halodurans* and *R. palustris* CGA009. Effectively, this means that if one of the two sequences, used in this analysis is shuffled, there is a <0.001% chance that they could end up with the similarity as good as that between the two original sequences. Anything below 0.1% is regarded significant (Pearson, 1988); as a result there is an excellent statistical similarity between these pumps with respect to *E. coli* R773 ArsB in sequence. The YqcL group arsenite pumps show insignificant p-scores demonstrating that they are not related to the *E. coli* ArsB group. The 3rd putative group showed significant p-scores (<0.001%) when compared with the ArsB and YqcL groups, though much greater for the latter (Table 4.2 and 4.3). In this case the newly recognized 3rd putative group may be regarded as a sub group of the ArsB pumps.

Further to add to the finding of the possible 3rd sub-group of ArsB, detailed analysis was carried out with respect to protein properties, the number of TMDs, membrane topology and their respective taxonomy.

4.3.2 Phylogenetic analysis

The arsenite pump sequence similarity based phylogenetic tree (Figure 4.2) showed a clear separation of the two prokaryotic domains (bacteria and Archea). Sequence wise the eukaryotic (*S. cerevisiae*) pump is more close to the former. Minor inconsistencies were found such as *R. palustris* (α-Proteobacteria) forming a deep branch within the γ-Proteobacteria. Otherwise the major divisions and subdivisions e.g. the various Proteobacteria, Eukarya and Archaea formed expected patterns.

The separation between ArsB and YqcL groups is very distinct, supporting the hypothesis that they have been evolved separately (Rosen, 1999). However, it can be seen that there are two sub-groups within the ArsB (or ArsB-like) sequences. The second sub-group forms a branch only based on sequence similarity and not based on taxonomy. This agrees with the previous results (Figure 4.1 and Table 4.3).

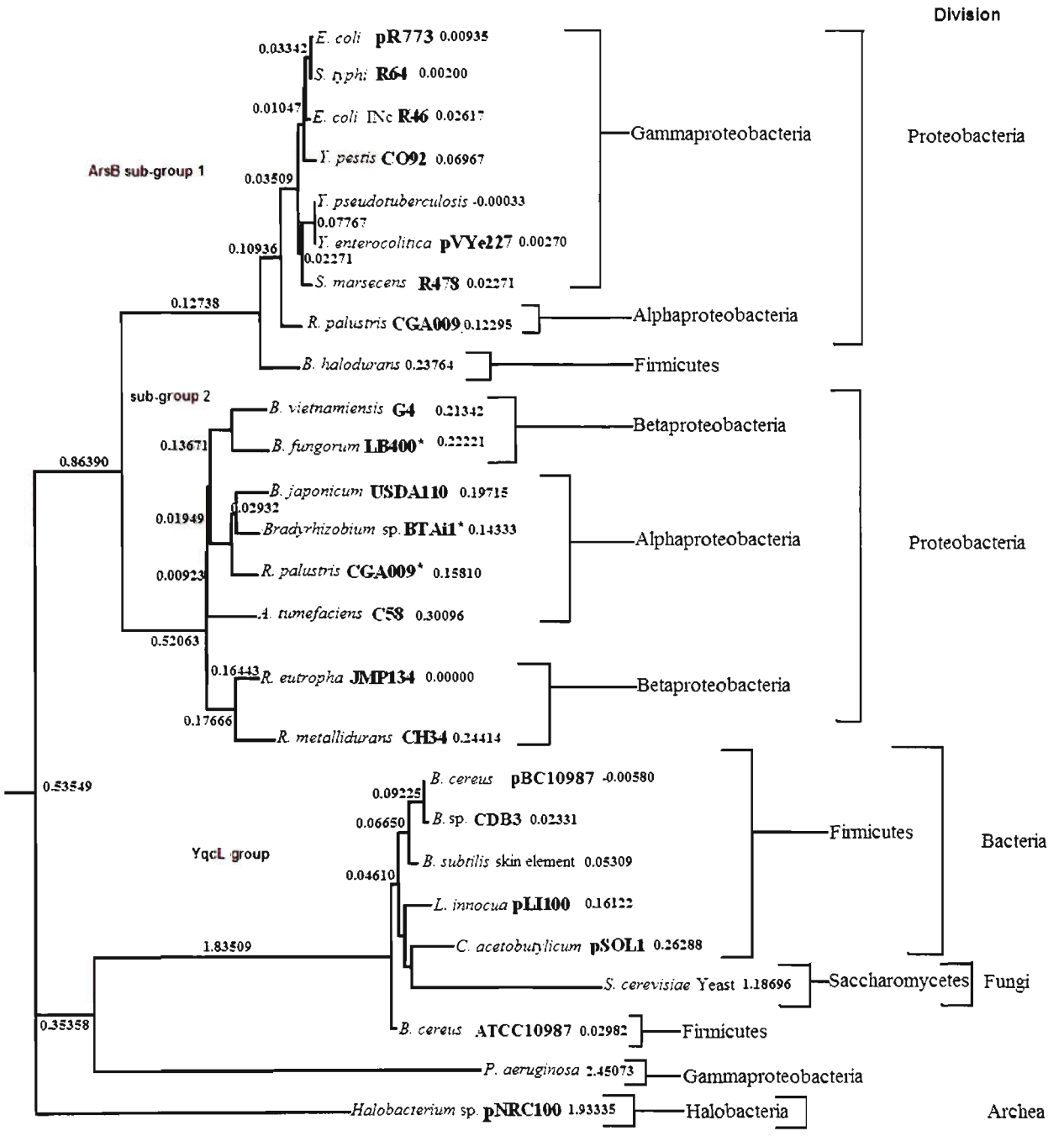


Figure 4.2: Evolutionary tree based on AsIII pump sequences. The tree was constructed by Neighbor-joining/UPGMA method. Plasmid borne pumps are indicated following the organism. The figures represent the matrix distances which were determined based on the sequence alignments. Asterisk '*' refers to strains to sub-group 3.

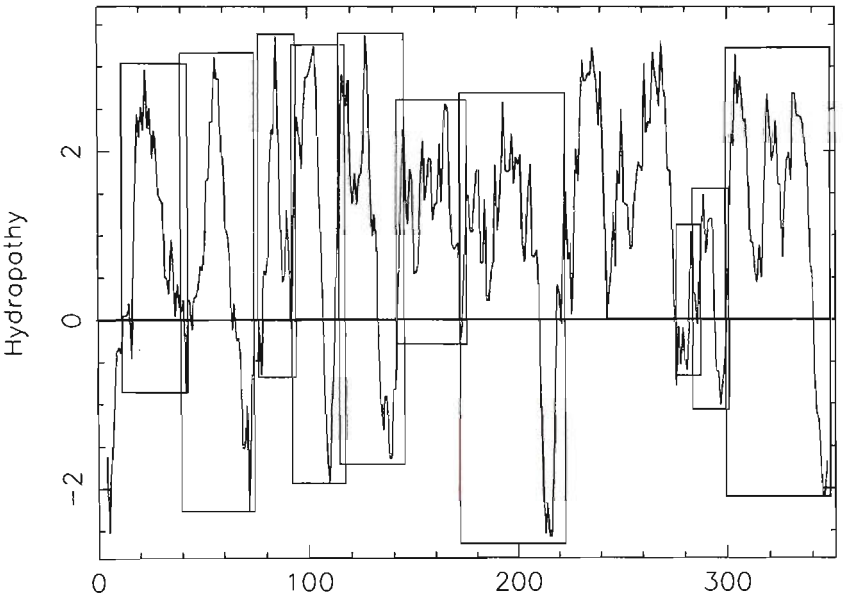
4.3.3 Other protein properties

Table 4.5 lists some properties of the 26 AsIII pump proteins. The ArsB group members are generally a bit larger in mass compared with YqcL group with the exception of *Halobacterium* sp. pNRC100 (64.9 kDa, the largest among the 26 but as an YqcL). The different groups and sub-groups vary mostly in their membrane topologies; YqcLs have 10 transmembrane domains (TMD), ArsB have 12 and ArsB like (sub-group 2) have 14 respectively. The comparison of protein properties of active arsenite transporter proteins found in all the tested bacterial species from three groups are listed in Table 4.4.

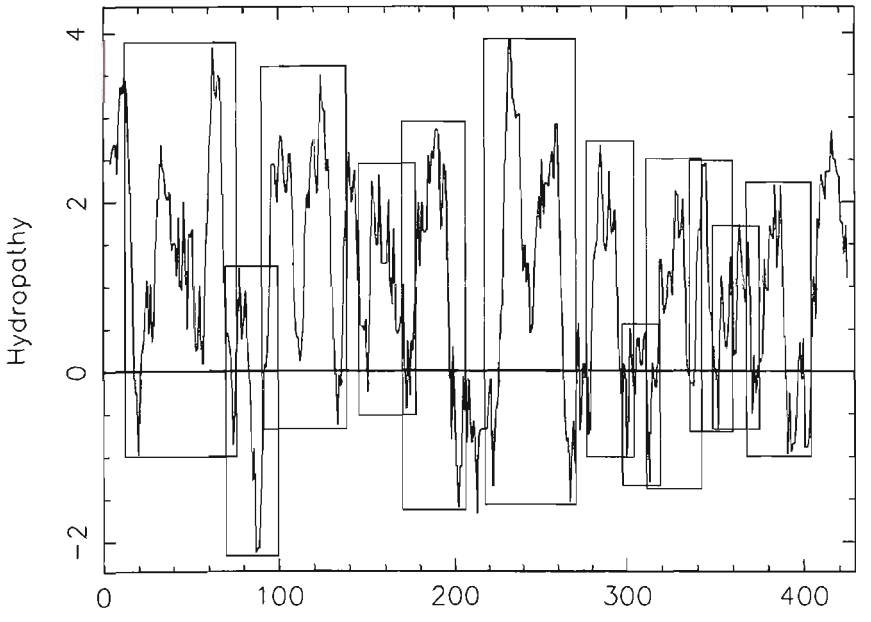
Table 4.4: Comparison of protein properties of active arsenite transporter proteins found in bacteria strains belonging to group1, 2 and the 3rd sub-group

Strain	Group	Pump group	Length (residues)	Mass (kDa)	No TMDs	pI
E. coli **pR773**	**ArsB group**	ArsB	429	45.6	12	9.56
S. typhi **R64**		ArsB	429	45.5	12	9.83
E. coli INc **R46**		ArsB	429	45.4	12	9.64
Y. pestis **CO92**		ArsB	429	45.5	12	10.28
Y. pseudotuberculosis		ArsB	429	45.5	12	10.28
Y. enterocolitica **pVYe227**		ArsB	429	45.5	12	9.96
S. marsecens **R478**		ArsB	429	45.7	12	10.24
B. halodurans		ArsB	436	47.35	12	10.23
R. palustris **CGA009**		ArsB	421	43.7	12	10.16
B. cereus **pBC10987**	**YqcL group**	YqcL	351	38.7	10	10.04
B. cereus **ATCC10987**		YqcL	346	38.3	10	9.78
B. subtilis skin element		YqcL	346	38.2	10	10.31
L. innocua **pLI100**		YqcL	348	38.8	10	10.03
C. acetobutylicum **pSOL1**		YqcL	349	38.7	10	10.07
S. cerevisiae Yeast		YqcL	404	45.8	10	10.4
Halobacterium sp. **pNRC100**		ArsB	618	64.9	12	8.66
P. aeruginosa		ArsB	297	31.3	12	4.48
B. vietnamiensis **G4**	**Sub-group 2**	ArsB	456	46.5	14	9.69
B. japonicum **USDA110**		ArsB	445	46.9	14	11.2
A. tumefaciens **C58**		ArsB	352	38.1	14	10.85
R. eutropha **JMP134**		ArsB	419	42.9	14	6.93
R. metallidurans **CH34**		ArsB	419	43.6	14	10.16
B. fungorum **LB400**		Na^+/H^+ antiporter	416	43.7	14	10.21
Bradyrhirhizobium sp. **BTAi1**		Na^+/H^+ antiporter	417	43.8	14	7.6
R. palustris **CGA009**		Na^+/H^+ antiporter	421	43.7	14	9.42

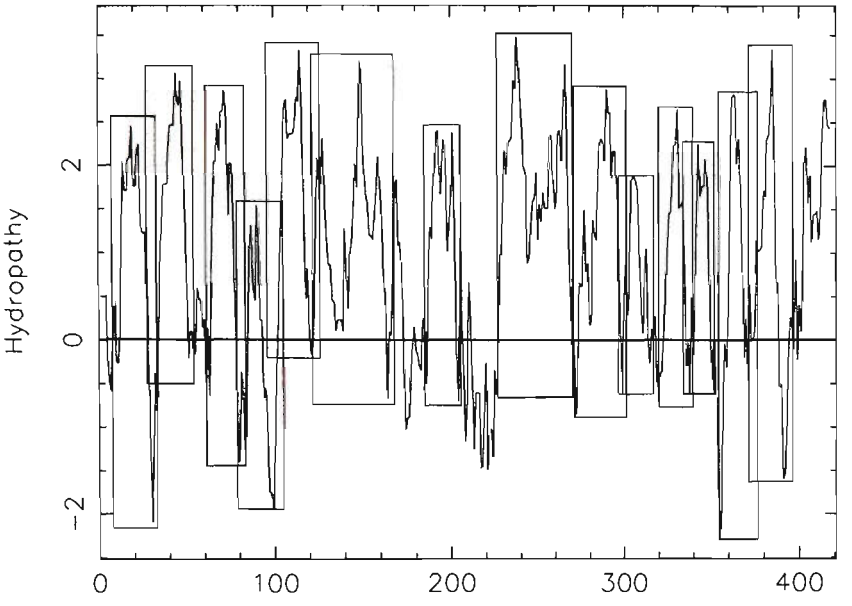
In order to predict the TMDs as given in Table 4.4 a hydropathy plot was constructed based on Kyte Doolittle hydropathy profile. Plots were generated using PepWindow program (Window size-integer 1 to 200, default is 7) available from ANGIS service for all the bacterial strains used in this study. Representatives from each of three groups distinguishing the difference in number of TMDs is presented in Figure 4.3, alternatively, hydrophobicity plots of all strains are presented in the Appendix 5.2.



(a)



(b)



(c)

Figure 4.3: Kyte Doolittle hydropathy plots generated by PepWindow showing the three representative sub-groups of arsenite membrane transporter proteins. (a) Represents the 10 TMD YqcL pump protein plot from *B.* sp. CDB3; (b) represents the 12 TMD ArsB pump protein plot from *E. coli* R773 and (c) represents the 14 TMD ArsB sub-group 2 pump protein plot from *R. eutropha* JMP134.

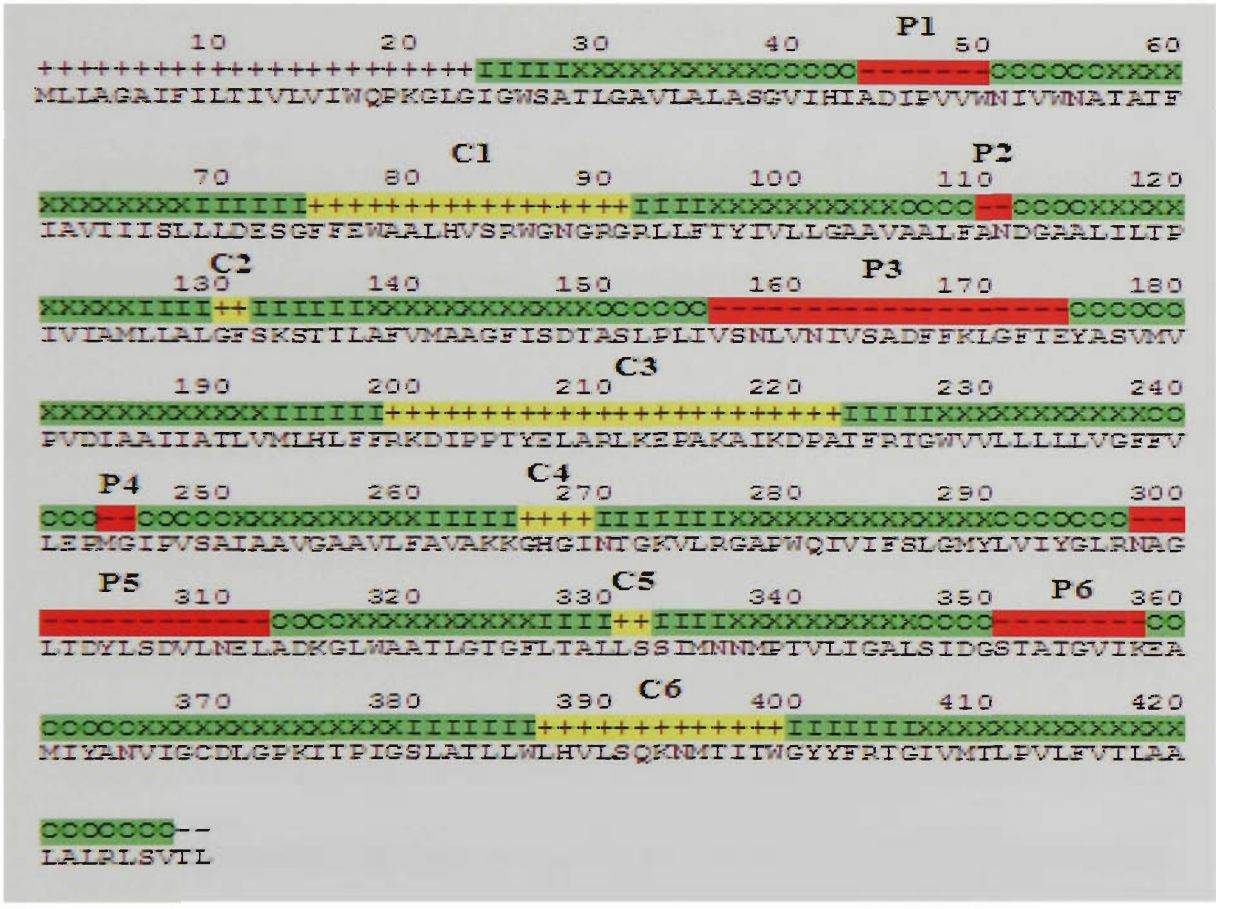
A hydropathicity profile calculation performed using the methods of Kyte and Doolittle (1982) as shown in Figures 4.3a, 4.3b, and 4.3c illustrates a clear difference between the existing groups. The TM helices were selected automatically or by user defined threshold. The program attempts to define the actual TM regions, and picks 10 or 12 or 14 regions, each of which contains the appropriate helix of interest. In Figures 4.3a and 4.3b the CDB3 YqcL and *E. coli* R773 ArsB showed the actual 10 and 12 TM spanning regions which were consistent with the previous results. As predicted the *Ralstonia eutropha* JMP134 ArsB which falls to a 3rd possible sub-group showed 14 TM spanning regions further suggesting that they are structurally homologous to the membrane protein ArsB. Although sequence homology of the *R. eutropha* JMP134 ArsB is low when compared to the well characterized 10 and 12 TM spanning proteins, the latter seems to have a similar function.

4.3.4 Membrane topology of ArsB proteins

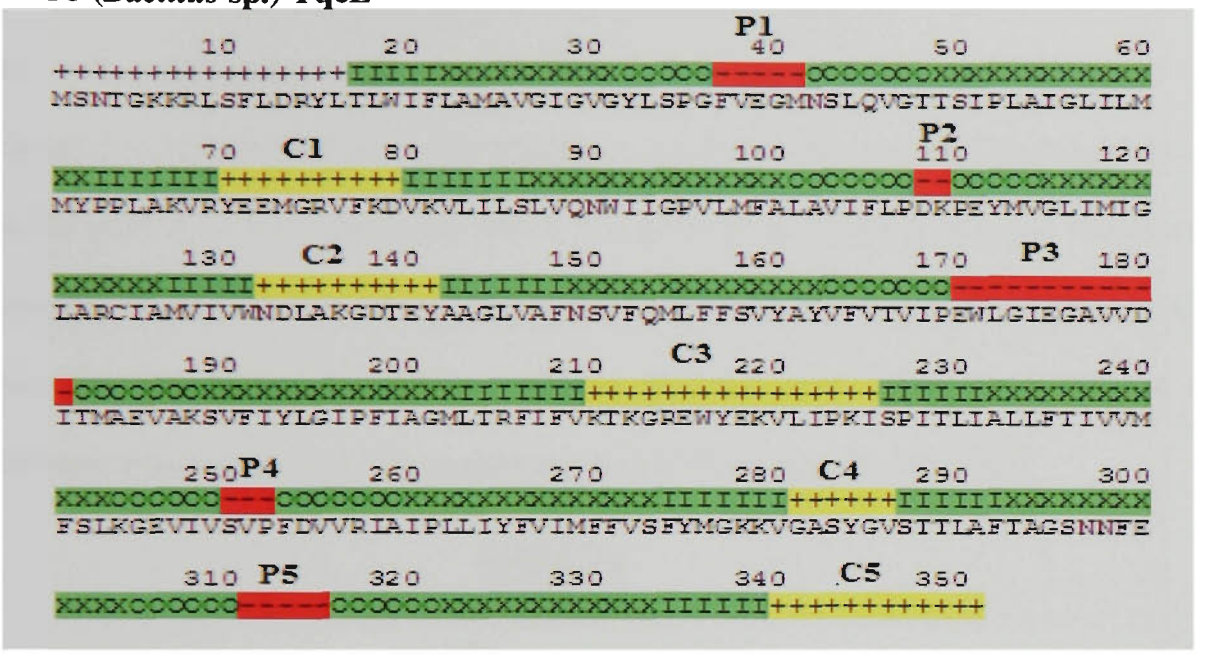
The ArsB protein is the membrane anchor of the oxyanion translocating ATPase (Tsai and Rosen, 1990). Based on the topological model of ArsB proteins, the cytoplasmic loops with charged residues can be used to predict the possible interaction with the ArsA protein (oxyanion translocating ATPase). PSIPRED (MEMSTAT3) program available from http://bioinfo.cs.ucl.ac.uk (protein prediction server) was used to predict the membrane topology of ArsB proteins (McGuffin et. al., 2000). As shown in Figure 4.5, the *E. coli* R773 shows the 6+6 arrangement (with six periplasmic and 6 cytoplasmic loops) consistent with the results obtained by Wu et. al. (1992). Based on the topological model published by Wu et. al., (1992), the cytoplasmic C3 loop has 22 residues, 10 of which are charged, and these residues would seem the more likely candidates for the interaction with ArsA protein. It is pointed out that the 6+6 arrangement is unknown; it probably represents a fundamental structural motif. It is also not clear whether the C3 loop in the ArsB protein is part of the ArsA anchoring site or fulfils some other function (Wu et. al., 1992). C1 and C4 which contain 4 and 5 charged residues are predicted to be possible candidates for ArsA interaction sites in *E. coli* R773 (Wu et. al., 1992). The data was consistent with the PSIPRED (MEMSTA3) results obtained, which predicts the membrane topology of ArsB. Further analysis was done to predict the membrane topology of the representative groups; i.e. 10 TMD and 14 TMD. The results obtained as shown in Figure 4.5, shows a 5+5 arrangement with respect to *Bacillus* sp. which possess 10 TMD. All the 5 cytoplasmic loops as predicted by MEMSTAT3 seem to be potential regions for interaction with ArsA. Due to this structural change the possibility of not extruding antimonite (SbIII) may be noted. The

ACR family of proteins seems to differ in their specificity and appear to be more specific to arsenite and not SbIII (Rosen, 1999). 6+7 arrangement with respect to *R. eutropha* JMP134 AsIII permease, which possess 13-14 TMD was seen. According to the MEMSTAT3 prediction, the C3 loop could be the possible region for ArsA interaction.

***E. coli* ArsB**



CDB3 (*Bacillus* sp.) YqcL



***R. eutropha* JMP134 (3rd putative group)**

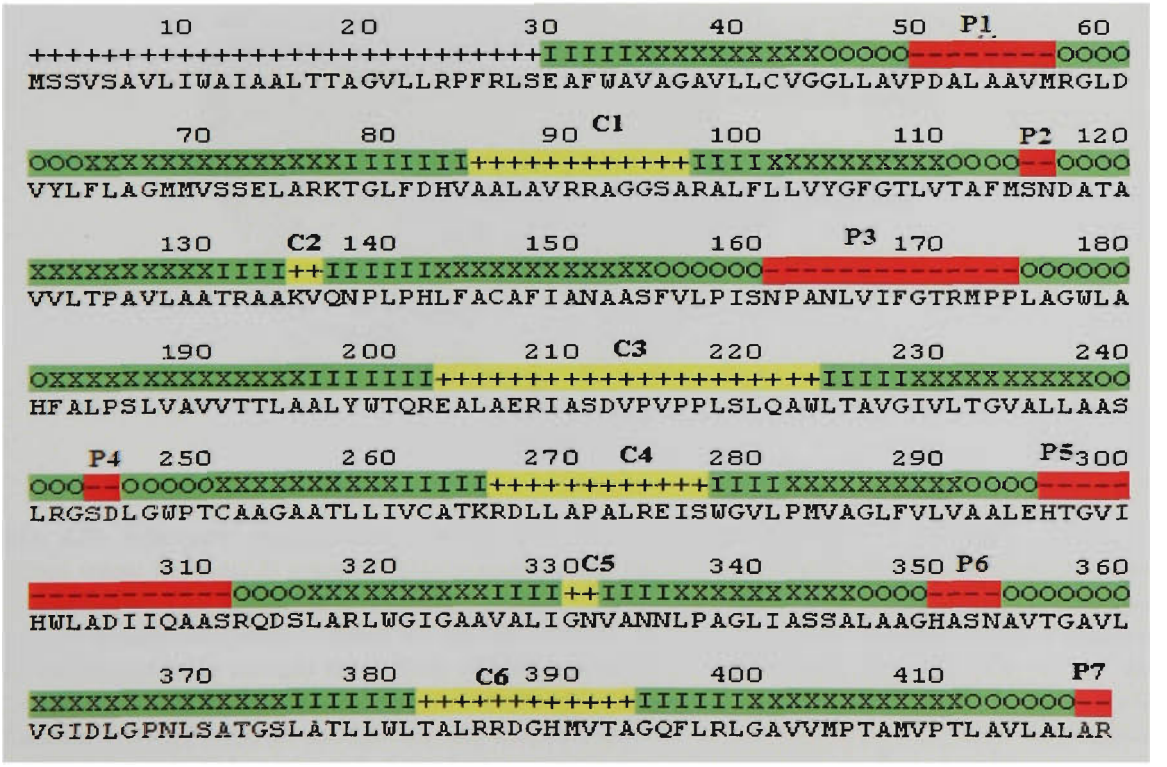


Figure 4.4: Prediction of membrane topology of ArsB protein sequences from *E. coli* R773, *Bacillus* sp. (CDB3) and *R. eutropha* JMP134 ASIII permease by MEMSTAT3, a bioinformatic analysis tool from PSIPRED (McGuffin et. al., 2000). + : Inside loop, - : Outside loop, O : Outside helix cap, X : Central transmembrane helix segment, I : Inside helix cap.

The molecular structure and the mechanism by which ArsB extrudes arsenite are not well understood yet. The driving force for arsenite pumping is thought to be the electrochemical gradient across the membrane. However, between pH 0-9 arsenite exists in the uncharged form (H_3ASO_3). Hence, it is possible that arsenite transport occurs as a symport or antiport with charged species. We propose a possible model (Figure 4.6) based on the explanation that arsenite can be transported across the membrane by different transporters based on its chemical nature (Silver and Phung, 2005).

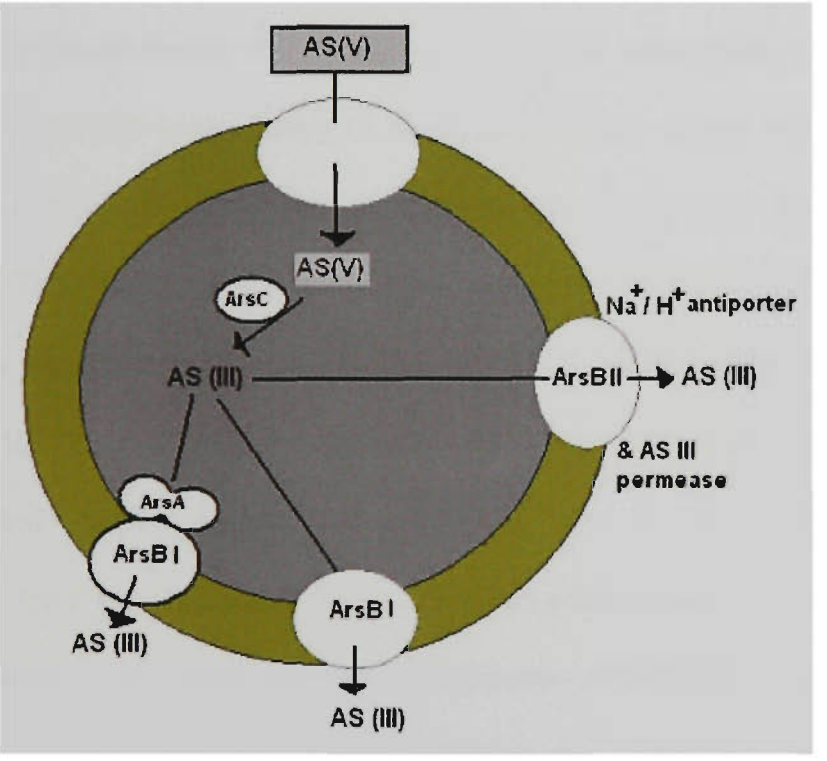


Figure 4.5: Schematic representation of the proposed model for transport of inorganic arsenic, two predicted roles; the ArsAB arsenite efflux complex and the respective homologues which catalyze NhaD Na^+/H^+ antiport. The ArsB1 is commonly existing *E. coli* type arsenite membrane protein. The ArsAB pump is a membrane pump coupled to ATPase (ArsA). The ArsAB is less common when compared to ArsB and is known to extrude more toxic arsenite out of the cell than ArsB alone. The second sub-group predicted as ArsB11 is thought to function as a Na^+/H^+ antiporter. At trace quantities of arsenite this efflux mechanism (ArsB/NhaD) is enough for the cell to resist the toxic effects, but when the concentration increases the cell would require a dedicated arsenite extrusion pump such as ArsB or ArsAB.

4.4 Discussion

ArsB proteins are members of a super family called IT (Ion-Transporter). ArsB contains 8-14 transmembrane helices and has gained the ability to function in accordance with ArsA to couple ATP hydrolysis to anion efflux (Tsai and Rosen, 1990). The ArsAB complex is similar to ATP-binding cassette transporters, but does not share any similarity in sequence (Rosen, 2002). Additional distant homologs of ArsB proteins may include members of NhaB, NhaC and NhaD (TC # 2.A.62) Na^+/H^+ antiporter families (Prakash et. al., 2003). The NhaD family is a constituent of the IT super family. ArsB proteins belong to the ArsB/NhaD super family of permeases that translocate sodium, arsenate, sulfate, and organic anions across biological membranes in all three kingdoms of life (Rosen, 2002). The NhaD Na^+/H^+ antiporters have been characterized from *Vibrio. parahemolyticus* and *Vibrio cholerae* (Herz et. al., 2003). The proteins are about 420aa long exhibiting 10-12 TMDs and have also been identified in their respective homologues and catalyze NhaD Na^+/H^+ antiport. The activity of these is seen at pH 8-9 with no activity at pH 7. Homologues are also found in proteobacteria (Herz et. al., 2003). Recent research into NhaD type antiporter from *Alkalimonas amylolytica* revealed that this antiporter helps in adaptation to alkaline environments suggesting that NhaD type antiporters might be part of regulation in adaptation to saline environments (Kurz et. al., 2006). The three NhaD homologues which have only been characterized so far from *V. parahemolyticus*, *V. cholerae* and *A. amylolytica*, exhibit Na^+/H^+ and Li^+ (Lithium)/H^+ antiport (Liu et. al., 2005).

The existence of a sub group of the ArsB-like proteins based on moderate sequence similarity with ArsB, their protein properties along with membrane topology is presented in this chapter. Two unrelated families of arsenite carriers (Figure 4.2) belonging to a 12 TMD ArsB and a 10 TMD YqcL extrusion pump proteins were identified and noticed to be quite distinct, supporting the hypothesis that they have been evolved separately (Rosen, 1999). Based on the multiple sequence alignment (Figure 4.1) the third set of sequences identified from proteobacteria, share moderate sequence similarity to that of 12 TMD pumps, whereas, showed no sequence similarity to 10 TMD extrusion pumps. However, it was shown to form a deep branch with 12 TMD ArsBs, suggesting a functional similarity based on moderate sequence similarity. Using phylogenetic approach an attempt was made in distinguishing the relationships of ArsB, ArsB-like proteins that comprise a family with the IT superfamily. Since none of the Na^+/H^+ antiporters or related AsIII permeases have been experimentally demonstrated to remove arsenite out of the cell, it is difficult to conclude whether the 3rd sub-group of ArsB-like proteins have a similar function or not. However, based on the moderate sequence homology, it can be assumed that they may possess a similar function. One possible explanation would be, at low concentrations when arsenite exists in the uncharged state (between pH 0-9), it is pumped out as a symport or antiport, however, when the concentration increases sufficiently in the cell, it requires a dedicated pump such as ArsB or ArsAB which is found in most *ars* operons.

Herz et. al., (2003) have demonstrated the roles of NhaD Na^+/H^+ antiporters in the survival of *V. cholerae* in saline environments. Most prokaryotes utilize the ubiquitous $As(OH)_3/H^+$ antiporter ArsB to extrude AsIII from cells, recently, a new pathway of

arsenic detoxification involving aquaglyceroporin channels (AqpS) has been identified in the legume symbiont *Sinorhizobium meliloti* (Yang et. al., 2005). Further functional characterization of Na^+/H^+ antiporters and related AsIII permeases in arsenic detoxification will provide a novel understanding of active efflux pump proteins.

Chapter 5

Functionality of the CDB3 *arsA* Gene

5.1 Introduction

In *E. coli* and other bacteria (Table 1), the ArsAB pump provides a greater resistance to arsenite and antimonite compared with the ArsB alone. ArsA contains two nucleotide-binding sites (NBSs) and a binding site for arsenic or antimony. The complex is similar in many ways to ATP-binding cassette (ABC) transporters, which typically have two groups of six transmembrane-spanning helical segments and two nucleotide-binding domains (NBDs). Binding of metalloids stimulates ATPase activity. The crystal structure of ArsA reveals that both NBSs and the metal-binding site are located at the interface between two homologous domains. A short stretch of residues connecting the metal-binding site to the NBSs provides a signal transduction pathway that conveys information on metal occupancy to the ATP hydrolysis sites. Based on the structural features, Zhou et. al., (2000) have proposed that the metal-binding site is involved directly in the process of vectorial translocation of arsenite or antimonite across the membrane. The relative positions of the NBS and the inferred mechanism of allosteric activation of ArsA provide a useful model for the interaction of the catalytic domains in other transport ATPases.

In the *ars* gene cluster 1 of *Bacillus* sp. CDB3, ORF 2 and 5 are deduced to specify YqcL and ATPase proteins respectively which may also form a membrane-bound pump that functions as an oxyanion-translocating ATPase to pump out the oxyanion arsenite.

Since no *arsA* gene has been identified functionally linked to the two characterized YqcL group pumps (*B. subtilis* skin element and ACR3 of *S. cerevisiae*), the functionality of CDB3 ATPase and its interaction with the YqcL pump needed to be

experimentally examined. The experimental strategy employed was to first construct a plasmid with *arsR* to monitor the basal level or background expression and/or resistance, and *arsRA* to demonstrate the interaction *arsA* with *E. coli arsB*. The constructs obtained where then transformed in *E. coli* AW3110 and *E. coli* DH5α cells. *E. coli* AW3110 was used in which the *arsRBC* operon has been deleted to demonstrate that CDB3 ArsATPase does not contribute to cellular growth, while, DH5α which harbors a chromosomal *arsRBC* operon was used to monitor the resistance profile of different constructs in increasing concentrations of arsenite to determine the interaction of CDB3 ArsA with *E. coli* chromosomal ArsB. The investigation detailed in this chapter not only confirmed these but also demonstrated that CDB3 ATPase can assist the *E. coli* 12 TMD ArsB as well.

5.2 Materials and Methods

5.2.1 Bacterial strains and plasmids

Bacterial strains and plasmids used in this study are listed in Table 5.1, with their relevant characteristics. *E. coli* strains were grown on LB agar or in LB liquid medium at 37°C. When appropriate, ampicillin or kanamycin was added at a concentration of 100 μg/mL, and 50 μg/mL respectively. Liquid cultures were shaken at 150-250 rpm in a shaking incubator (Bioline, Australia).

Table 5.1: Bacterial strains and plasmids used in the study of Chapter 5

Strain and Plasmids	Genotype or description	Reference or Source
E. coli AW3110	K-12 FIN (*rrn*D-*rrn*E) Δ*ars::cam* (Cm^r, the chromosomal *arsRBC* deleted)	Gift from Dr. B.P. Rosen. (Carlin et. al., 1995)
E. coli DH5α	Φ80dlacZΔM15, recA1, endA1, gyrA96, thi-1, hsdR17 (r_k^-,m_k^+), supE44, relA1, deoR, Δ (lacZYA-argF) U169	Stratagene, USA
pJKKmf(-)	Cloning vector (Km^r) 3.6 kb in size.	(Kirschman and Cramer, 1988)
pZX4	A 5.8 kb *Sau*3AI fragment containing *arsRYCDA* cloned into pJKKmf(-) at *Bam*HI site	Provided by Z. Xu (University of Wollongong)
pZX1	A 2160 bp *Sau*3AI fragment containing CDB3 *arsRY* and part of *arsC* cloned into pJKKmf(-) at *Bam*HI site	Provided by Z. Xu (University of Wollongong)
pR	A 1.2 kb *Eco*RV fragment consisting of *arsY* and *arsC* removed from plasmid pZX1 and religated.	Provided by Z. Xu (University of Wollongong)
pSB14 (RA)	A 1.57 kb *Eco*RV/*Cpo*I fragment consisting of *arsYC* and partial *arsD* genes removed from plasmid pZX4 and religated.	This Study

A series of plasmid constructs were made as detailed in Table 5.1. The constructs were then transformed into either *E. coli*/AW3110 or *E. coli*/DH5α and subjected to AsIII assays to confirm the functionality of *arsA*.

5.2.2 Plasmid mini-prep and restriction digestion

The bacterial colonies from the transformant plates were randomly selected and inoculated into LB medium containing appropriate antibiotics. After shaking incubation at 37°C overnight, the cell cultures were poured into the 1.5 mL eppendorf tubes and spun in a micro centrifuge for 30 sec at 13,000 rpm. The supernatant was discarded, the cell pellet was resuspended by vortexing, and 200 μL of 0.1M NaOH + 0.5% SDS was added. After gentle mixing, 200 μL of 3M sodium acetate (pH 5.4) was added and the micro centrifuge tubes were gently mixed by inversion before cooling on ice for 10 minutes. The micro centrifuge tubes were then centrifuged (10 min at 13,000 rpm) and the supernatant was transferred to a new eppendorf tube. Two volumes of 95% ethanol were added to the supernatant and the tubes were mixed by inversion before centrifugation (10 min at 13,000 rpm). The pellet was washed with 75% ethanol and then centrifuged (5 min at 13,000 rpm) again. The supernatant was removed by aspiration and the pellet containing plasmid DNA was dissolved in 30 μL of sterile distilled H_2O.

For diagnostic restriction digestion, each reaction mix consisted of 1 μL of 10X appropriate digestion buffer, 0.1 μL of RNase A (1 mg/mL) and 2 units of restriction enzymes (Promega), with the volume made up to 7.5 μL with water. The plasmid DNA sample (2.5 μL) was added and the digestion was carried out at 37°C for about 1.5 hours. The digested samples were analyzed by agarose gel electrophoresis.

5.2.3 Ligation and electro transformation

Ligation was performed by mixing vector fragment, insert (digested PCR products or plasmids), 1 μL of ligation buffer, 1 unit of T4 ligase (Promega), and H_2O up to 10 μL. DNA ligations were performed using a vector to insert molar ratio of approximately 1:3 with the vector amount normally being 50-100 ng/10 μL. The ligation mixtures were left overnight at 16°C. The DNA was precipitated out of each solution by the addition of 1 μL of 3M sodium acetate (pH 5.4) and 25 μL of 95% ethanol, followed by rapid cooling (-80°C for 20 min). The mixture was then centrifuged at room temperature (15 min at 13,000 rpm) and the pellet resuspended in 10 μL of H_2O.

Competent *E. coli* cells for electro transformation were prepared following the method of Sambrook et. al., (1989). Cells stored at -80°C were streaked on LB agar plate and incubated at 37°C overnight. A single colony from the LB agar plate was then picked and incubated in 10 mL of LB liquid medium with shaking at 37°C overnight. The cells were transferred into 1 L LB medium in a 2 L flask and cultured with 180 rpm shaking at 37°C to $O.D._{600\ nm}$ = 0.8. After briefly chilled on ice, the cells were harvested by centrifuging at 4,000 g for 10 minutes at 4°C. The pellet was resuspended in 1 L of ice-cold d. H_2O and centrifuged as before. Repeat this wash with 0.5 L of ice-cold d. H_2O. The pellet was resuspended in 20 mL of ice-cold 10% glycerol and centrifuged at 4,000 g for 10 minutes at 4°C. The final pellet was resuspended in 5 mL of 10% ice-cold glycerol and stood on ice. The cells were distributed to 1.5 mL sterile eppendorf tubes (160 μL in each) and frozen in liquid nitrogen before being stored at -80°C.

Ligated samples were used to transform competent cells by electroporation. The transformation was performed by adding 0.5-2.5 μL of the purified ligated plasmid to 40

μL of competent *E. coli* cells. The entire mixture was then placed into an ice-cold micro-cuvette (2 mm path length, Bio-Rad), placed into the Genepulser (Bio-Rad) and electroporated (2.5 kV, 25 μFD, 200 Ω). Immediately following electroporation, 960 μL of SOC medium was added to the cells. The transformed *E. coli* cells were incubated with shaking at 37°C for 1 hour and were then plated on LB agar plates containing appropriate antibiotics followed by incubation for up to 24 hours at 37°C.

5.2.4 Arsenic resistance assays

Assays performed in *E. coli* AW3110 and *E. coli* DH5α harboring plasmid constructs were carried out in LB liquid medium with 0.1 mg/ml ampicillin. Overnight cultures were diluted 100 fold into fresh LB medium containing 0.1 mg/ml ampicillin and various concentrations of sodium arsenite. Cells were incubated at 37°C for 5-7 hours and $O.D.600_{nm}$ was determined. The incubation time corresponds to the exponential growth phase of the control under the same conditions. The resistance was expressed as the percentage $O.D.600_{nm}$ compared with the control sample with no arsenic. The resistance assays were performed in triplicates, and were repeated at least four times.

5.3 Results

5.3.1 Construction of plasmid pSB14

The cloning strategy for the construction of pSB14 is shown in Figure 5.1. *Eco*RV/*Cpo*I fragment consisting of *arsYC* and *~arsD* genes (1572 bp) was removed from pZX4. The digested pZX4 consisting of *arsRA* genes was then filled-in with Klenow and ligated to make the 7.8 kb clone pSB14, and used to transform *E. coli* DH5α.

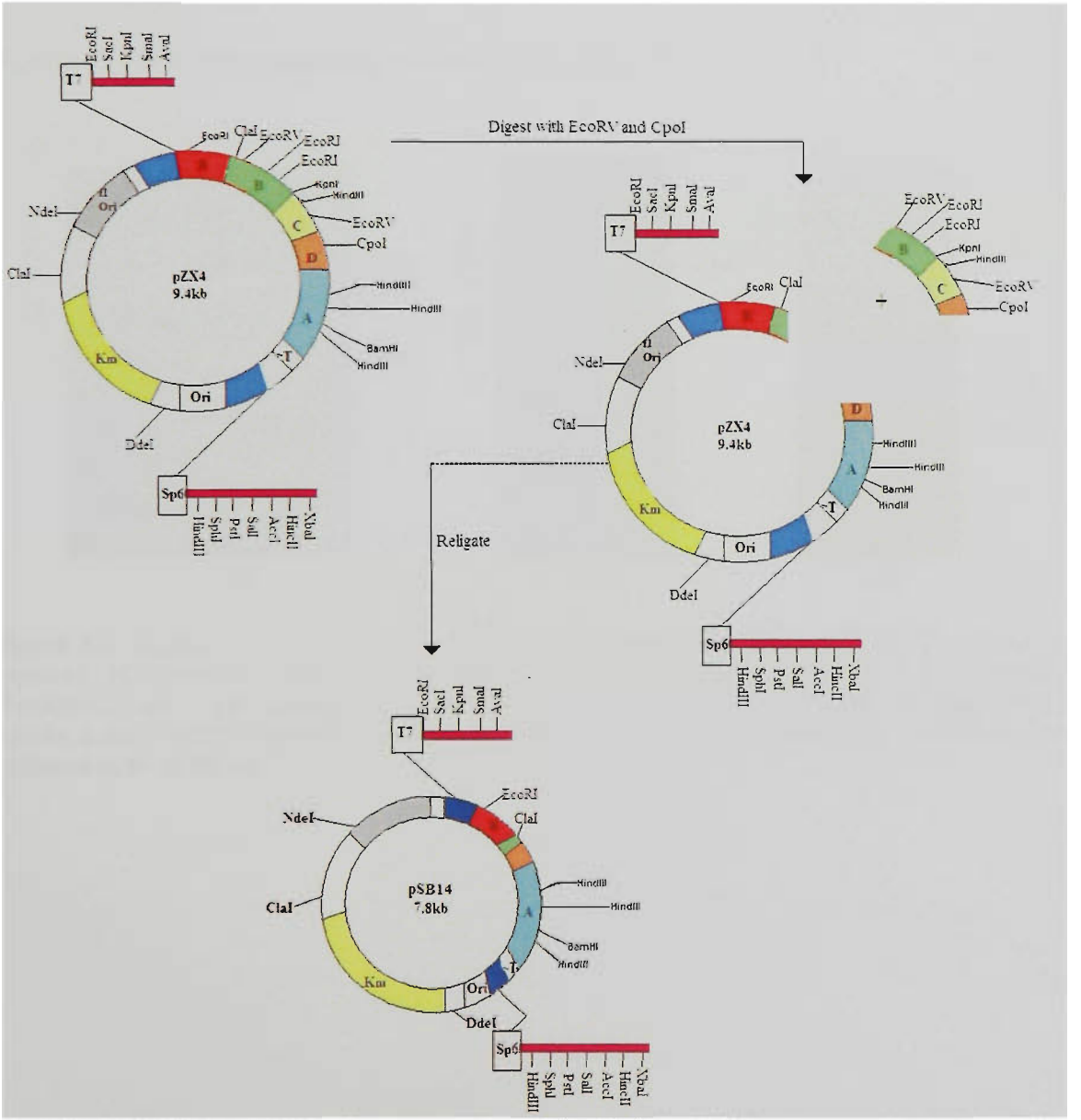


Figure 5.1: The cloning strategy for constructing pSB14. *Eco*RV/*Cpo*I fragment consisting of *arsYC* and *~arsD* genes (1572 bp) were removed from pZX4, Klenow filled in and religated.

Restriction enzyme digestion was performed to verify the constructs pR and pSB14. As shown in Figure 5.2 a, b and c, respectively. Plasmid pR (a) was digested with a combination of enzymes and showed the expected fragment sizes at 3.6 kb and ~1 kb when digested with *Sma*I and *Xba*I and *Eco*RI and *Xba*I; while 4.6 kb linear band when digested with *Xba*I alone.

pSB14 digested with *Hin*dIII and *Sac*I showed the expected fragment sizes at 5.2 kb, 1.3 kb, 830 bp, and 460 bp (Figure 5.2 c) for *Hin*dIII digestion and 7.8 kb when digested with *Sac*I respectively, as shown in Figure 5.2 b.

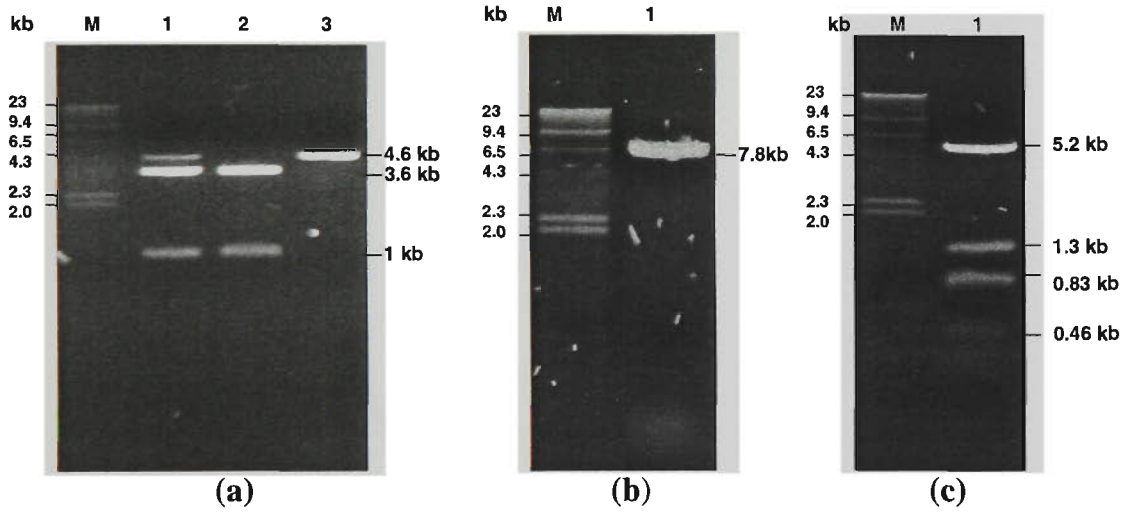


Figure 5.2: Agarose gel electrophoresis (1%) of plasmid constructs digested with different restriction enzymes. M represents λ/*Hin*dIII DNA marker. (a) Plasmid construct pR, Lane 1: pR digested with *Sma*I/*Xba*I. Lane 2: pR digested with *Eco*RI/*Xba*I. Lane 3: pR digested with *Xba*I. (b) Plasmid construct pSB14, Lane 1: pSB14 digested with *Sac*I. (c) pSB14 digested with *Hin*dIII (Lane 1). Marker sizes are indicated in kb on the left.

5.3.2 Effects of CDB3 ArsA on arsenite resistance of *arsRBC* bearing and *ars* knock-out *E. coli* cells

Prior to confirming the function and interaction of *arsA*, it was necessary to determine if ArsA itself has any effect on the cellular growth without the presence of transmembrane pump. The gene constructs (pR, pSB14) along with the vector pJkkm were transformed into *E. coli* AW3110 which is an *ars* knock-out mutant and AsIII assay was performed on increasing concentrations of sodium arsenite. The result obtained as shown in Figure 5.3, clearly indicated that this is not the case; the ATPase does not have any effect on cellular growth, as *E. coli* AW3110 cells harboring the gene constructs did not show elevated resistance to AsIII (T-test indicated the significant P-value < 0.001).

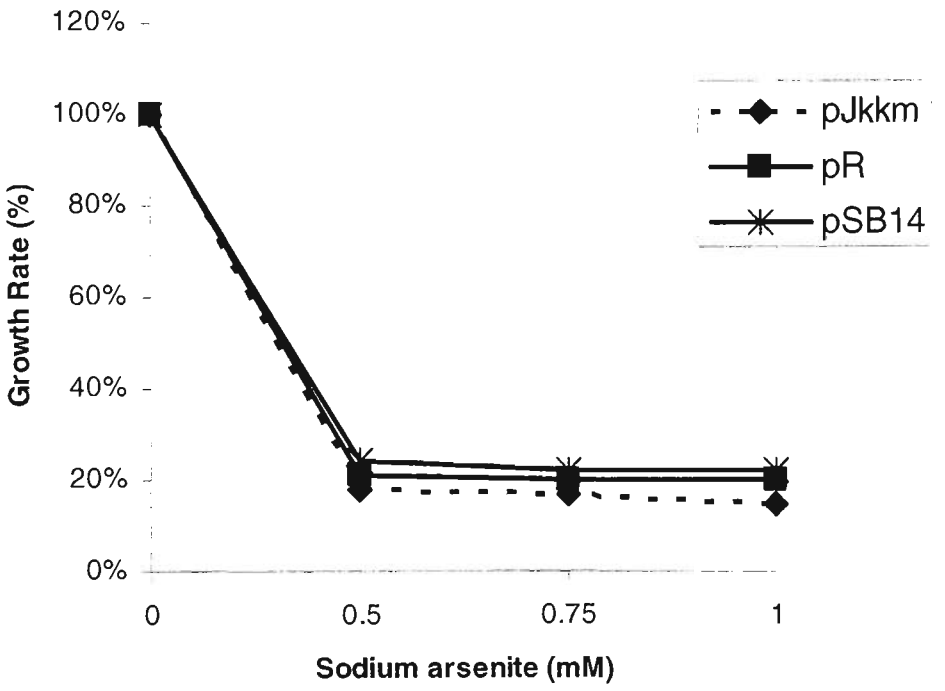


Figure 5.3: Growth inhibition by arsenite of *E. coli* AW3110 harboring different plasmid constructs. Overnight cultures were diluted 50-fold into LB liquid medium containing increasing amounts of arsenite, incubated at 37°C for 5-7 hours. Growth rates were measured as absorbance at 600nm and indicated as percentage in comparison to 0 mM.

The assay using *E. coli* AW3110 showed no resistance on increasing concentrations of AsIII since the *arsRBC* operon from AW3110 is deleted which would demonstrate why expressed sensitivity to arsenite and subsequently confirming that the ATPase itself does not contribute cellular growth, the next step was to transform these constructs into *E. coli* DH5α. The purpose of choosing DH5α as a host was due to the *arsRBC* operon located on the chromosome. The results obtained when transformed into *E. coli* DH5α as shown in Figure 5.4, indicated a significant increase in AsIII resistance with *E. coli* DH5α cells harboring pSB14, demonstrating the involvement of *arsA* gene to form an ATP driven anion pump (ArsAB complex). pSB14 harboring the *arsRA* genes exhibited a much higher resistance when compared to the control (T-test indicated the significant P-value < 0.002) indicating the interaction of CDB3 ArsA with *E. coli* ArsB in effectively removing AsIII out of the cell.

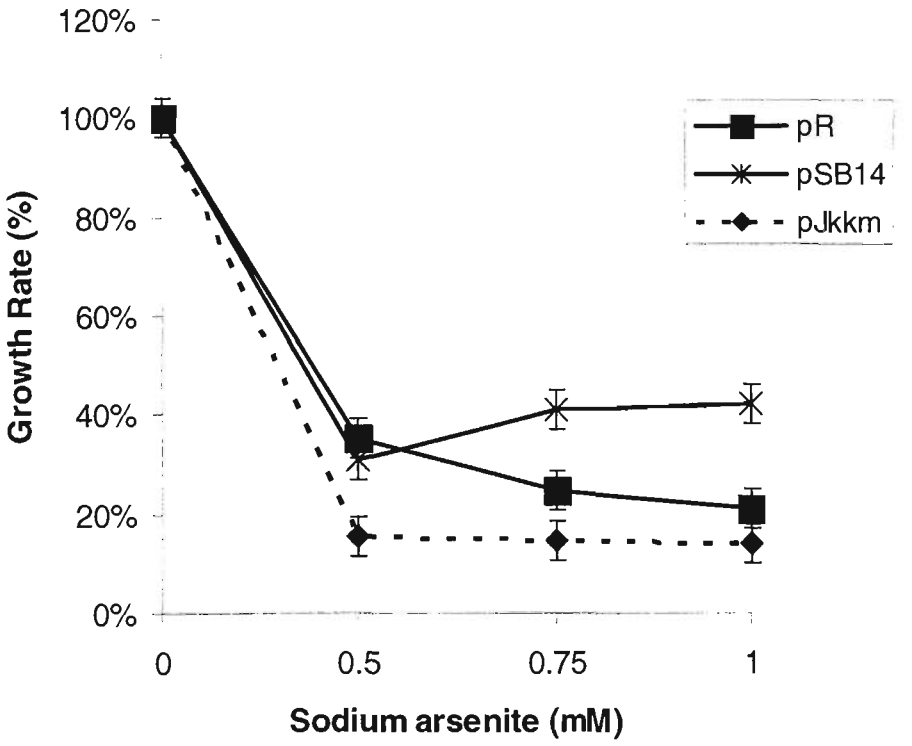


Figure 5.4: Growth inhibition by arsenite of *E. coli* DH5α harboring different plasmid constructs. Overnight cultures were diluted 50-fold into LB liquid medium containing increasing amounts of arsenite, incubated at 37°C for 5-7 hours. Growth rates were measured as absorbance at 600nm and indicated as percentage in comparison to 0 mM. The error bars represent SD (n=4).

5.4 Discussion

The ArsA ATPase is the catalytic subunit of the ArsAB pump in *E. coli* that is responsible for extruding AsIII or SbIII from inside the cell, thereby conferring higher resistance to the host than ArsB alone. Either SbIII or AsIII stimulates ArsA ATPase activity (Bhattacharjee and Rosen, 2001). The bacterial ArsA ATPase as a portion of an arsenite pump has two nucleotide binding consensus sequences in the N and C terminal halves of the protein (Walmsley et. al., 1999). Activation is associated with the dimerization of the ArsA protein (Bhattacharjee et. al., 1995). The arsenite transport system in *E. coli* pR773 exhibits a dual mode of energy coupling depending on the subunit. In the presence of both ArsA and ArsB, they form ArsAB ion-translocating ATPase which is independent of the electrochemical proton gradient. While, in the absence of ArsA, the ArsB catalyzes extrusion of AsIII coupled to electrochemical energy, suggesting ArsB is a uniporter which extrudes the AsIII anion (Chen et. al., 1996; Rosen, 2002; Meng et. al., 2004).

The CDB3 AsIII transmembrane pump protein is different from ArsB but shares greater homology to *B. subtilis* YqcL and *S. cerevisiae* ACR3 (Chapter 4). Experimental data suggests that not only CDB3 'ArsA' can interact with its own YqcL (Chapter 3, Figure 2.2) but can also interact with *E. coli* ArsB (Figure 5.4) in elevating arsenite resistance to the host indicating the function of the protein and its diversity. Much of the activity is noted when CDB3 YqcL is present, but in the absence of CDB3 YqcL the CDB3 ArsA interacted with the *E. coli* host chromosomal ArsB in removing AsIII out of the cell which was more effective than ArsB alone. This is interesting since CDB3 YqcL possess only 10 transmembrane domains with five cytoplasmic loops as potential regions

for interaction of ArsA, while *E. coli* ArsB has 12 transmembrane domains with C1, C3 and C4 loops predicted to be possible candidates for ArsA interaction (Chapter 4, Figure 4.5). To our knowledge this is the first preliminary experimental evidence that an ATPase is genetically and functionally associated with an YqcL group pump and it can also interact with an ArsB pump.

Future experiments based on the preliminary finding presented in this Chapter will contribute towards understanding the binding ability of purified ArsA protein to ArsB and/or YqcL. Previous studies conducted by Tisa and Rosen, (1990) reveal difficulties in demonstrating the direct interaction of ArsA and ArsB proteins. These difficulties were attributed to poor expression of the *arsB* gene, rapid degradation of the mRNA in the *arsB* region of *E. coli* pR773 *arsRDABC* operon (Owolabi and Rosen, 1990) and the low antigenicity of ArsB protein. Therefore, to combat these limitations, a more direct approach to confirm the interaction of these two proteins (CDB3 ArsA and *E. coli* ArsB and CDB3 YqcL-ArsA) would be worthwhile to conduct. Tisa and Rosen (1990) have effectively demonstrated the interaction of *E. coli* pR773 ArsA with ArsB by immunoblotting. A similar approach could be employed by expressing the ArsA protein and performing cellular localization studies (binding of purified ArsA protein to *E. coli* membrane) of CDB3 ArsA protein by immunoblotting using anti-ArsA antiserum with cells expressing *arsA*, *arsB*, *arsAB* and *arsYB* to detect the ArsA polypeptide.

Chapter 6

Functionality of the CDB3 *arsD* Gene

6.1 Introduction

Expression of the *ars* operon in *E. coli* plasmid R773 *ars* cluster *arsRDABC* is regulated by the first two genes, *arsR* and *arsD* which encode transacting repressors (Li et. al., 2001). Both ArsR and ArsD are 13 kDa homodimers, but share no sequence similarity. Expression of the *ars* operon is induced by arsenite and antimonite (Tsai et. al., 1997). Binding of arsenite or antimonite produces dissociation of ArsR from the operator site (Chapter 1). Increasing levels of the integral membrane ArsB protein may become toxic to the cells (Chen and Rosen, 1997) and ArsD a 120-residue repressor, regulates and controls the over expression of ArsB (Wu and Rosen, 1993). In *E. coli* R773 cells, ArsD binds to same operator site as ArsR but has a lower affinity than ArsR (Li et. al., 2001). Together ArsR and ArsD form a regulatory circuit that controls the levels of expression of the *E. coli* R773 *ars* cluster (Chen and Rosen, 1997). ArsD of pR773 has three vicinal cystiene pairs Cys12-Cys13, Cys112-Cys113 and Cys119-Cys120 (Li et. al., 2001). The role of these three vicinal cystiene pairs in metalloid sensing is detailed in Chapter 1.

The deduced ArsD protein encoded by the CDB3 *arsD* gene is structurally different to the well characterized ArsD of *E. coli* R773 since it lacks the C-terminal cystiene residues Cys112-Cys113 and Cys119-Cys120 (Figure 3.9). The ArsDs deduced from the newly genome sequenced *B. cereus* pBC10987, *Bacillus* sp. mb24 and the characterized Archeal *Halobacterium* sp. pNRC100 *ars* cluster also lack the terminal two pairs of cystiene residues while all the other bacterial strains which contain the *arsD* gene (Chapter 3) have been found to contain all three pairs of cysteine residues. Previous experiments conducted by a former member of our lab, Luo Xi, demonstrated the

contribution of *arsD* in increasing resistance to the host when compared to only 3 genes *arsRY(B)C* (Chapter 3) further illustrating that *arsD* like *arsR* regulates and controls the expression of CDB3 *ars* cluster 1.

This chapter reports my study attempting to elucidate the function of the CDB3 ArsD repressor protein with emphasis on understanding the role of ArsD in metalloid sensing and regulating the expression of CDB3 *ars* cluster 1. Deletion mutagenesis of *arsD* and DNA-protein interaction studies involving the affinity binding to *arsR* and *arsD* promoters along with AsIII responsiveness were conducted.

6.2 Materials and Methods

6.2.1 Bacterial strains and plasmids

Bacterial strains and plasmids used in this study are listed in Table 6.1, with their relevant characteristics. *E. coli* strains were grown on LB agar or in LB liquid medium at 37°C. When necessary, ampicillin was added at a concentration of 100 μg/mL.

Table 6.1: Bacterial strains and plasmids used in the study of Chapter 6

Strains and Plasmids	**Genotype or description**	**Reference or source**
E. coli JM109	*rec*A1 , *end*A1, *gyr*A96, *thi*, *hsd*R17 (r_K-, m_K+), *sup*E44, *rel*A1, Δ(*lac-pro*AB), [F', *tra*D36, proAB, *lac*IqZΔM15]	Promega
E. coli BL21 (DE3)	*hsdS gal (λcIts857 ind1 Sam7 nin5 lac*UV5-T7gene*1*)	Invitrogen
E. coli AW3110	K-12 F'IN (*rrn*D-*rrn*E) Δ*ars::cam* (Cm^r, the chromosomal *arsRBC* deleted)	Gift from Dr. B.P. Rosen. (Carlin et al., 1995)
E. coli DH5α	Φ80dlacZM15, recA1, endA1, gyrA96, thi-1, hsdR17 (r_K-,m_K+), supE44, relA1, eoR, □(lacZYA-argF) U169	Stratagene, USA
Bacillus sp CDB3	Isolated from cattle dip-sites	(Chopra et. al., 2007)
pGEM7Zf (+)	Cloning vector (Ap^r)	Promega
pAR27	A 7013 bp partial *Sau*3AI fragment containing CDB3 *arsRBCDATIP* cloned into pGEM7Zf(+) at *Bam*HI site	Provided by X. Luo, 2006
pCW12 T-easy	PCR fragment containing the putative CDB3 promoter P_{arsD} region and *arsD* ligated into pGEM-Teasy vector at *Eco*RI / *Xba*I sites (Ap^r)	Provided by C. Su, 2004
pZT	CDB3 *arsRBCDAT* cloned in pGEM7Zf(+)	Provided by X. Luo, 2006
pET11d	5.8 kb T7 expression vector (Ap^r)	Novogen
pET151/GAPDH	A 1011 bp GAPDH fragment cloned into pET151 (5.7 kb) T7 expression vector (Ap^r)	Provided by A. Cork (University of Wollongong)
pSB21	492 bp (ArsD6x(N)) PCR amplified *arsD* fragment cloned into pET11d expression vector digested with *Nco*I and *Not*I	This study
pSB1	441 bp (ArsD6x(C)) PCR amplified *arsD* fragment cloned into pET11d expression vector digested with *Nco*I and *Not*I	This study
pSB6	A 422 bp sequence covering partial *arsC*, the terminator/promoter structure and *arsD* deleted from pAR27	This study
pSB38	A 105 bp sequence of the *arsD* deleted from pAR27	This study

6.2.2 Deletion mutagenesis of *ars*D

An internal sequential deletion approach was undertaken to generate mutant *arsD* clones. Each deletion sample mix contained 0.5 µL of *Exo*III enzyme (200 u/µL- Progen) and samples of 2.5 µL each at the end of 5, 10 and 15 seconds were transferred into a eppendorf tube containing 2.4 units S1 nuclease in 7.5 µL of 1X S1 buffer kept on ice. The tubes were then placed at room temperature for 30 minutes at the end of which 1 µL of S1 stop buffer was added into each tube. The tubes were then heated at 70°C for 10 min to inactivate the S1 nuclease. The tubes were transferred to room temperature and 1 µL of Klenow mix (30 µL of Klenow buffer containing 5 u of Klenow fragment DNA polymerase - Progen) was added and the tubes were incubated for 3 min. This was followed by the addition of 1 µL of dNTP mix (0.125 mM) with 5 min incubation at 37°C. The samples were then moved to room temperature and from each, 5 µL was taken and mixed with 15 µL of ligase mix (2 µL of T4 DNA ligase buffer + 1 µL of T4 DNA ligase + 12 µL of sterile d. H_2O). After incubation for 2 hours at room temperature to facilitate ligation, the samples were transformed into *E. coli* AW3110 and spread onto LB agar plates containing 100 µg/mL of ampicillin and incubated at 37°C for up to 24 hours. Transformed colonies were picked and plasmid extraction was performed as detailed in section 5.2.2. The deletion mutants were verified by DNA sequencing.

6.2.3 Arsenic resistance assays

Assays were performed as described in sections 5.2.4, with the only difference being the use of AsV instead of AsIII. The resistance assays were performed in triplicates, and were repeated at least three times.

6.2.4 PCR amplification of the CDB3 *arsD* gene

For generating protein expression constructs, the *arsD* gene was PCR amplified from pCW12 with two sets of primers designed for N and C terminal fusions with six histidine tags. N-terminal fusion: forward primer-ArsD-6x(N) 5'A**CCATGG**CTCATCATCA **CCATCACCAT**CTCGAGTCAAAG**ATG**AAGAAGATAG3' consisting of two restriction enzyme sites *Nco*I and *Xho*I (underlined); reverse primer M13R 5'**CAGGAAACAGCTATGAC**3'. C-terminal fusion: forward primer – D-*Nco*F 5'**TACCATGG CAAAGATGAAGAAGATAGAA**3' consisting of restriction enzyme site *Nco*I (underlined); reverse primer – ArsD-6x(C) 5'GGATCC*CTA*ATGGTGATGGTGATGATGGAGCTCTTTTTTCA CGTTTAACTTTA3' consisting of two restriction enzyme sites *Bam*HI and *Xho*I (underlined). The primers (synthesized by Sigma and delivered desiccated) were diluted with TE buffer to form stock solutions (200 μM), which were further diluted to give a working concentration of 20 μM.

The reaction mix for each PCR amplification consisted of 1 unit of *Taq* DNA polymerase (Promega), 1 μL PCR buffer, 2 μL dNTP (2 mM), 1 μL of each primer (10 μM), 1 or 2 μL of DNA template (pCW12, 10 ng/μL), and nuclease-free water to a total of 10 μL. The PCR was performed with a GeneAmp PCR system 9600 (Perkin Elmer). The programme used for the amplification consisted of a preheating step (94°C for 3 min.) followed by 30 cycles of 94°C for 30 sec, 50°C for 30 sec and 72°C for 2min before the samples were finally extended at 72°C for 10 min, and then held at 25°C. 5 μL of the PCR products were loaded onto a 1.5% TAE agarose gel for checking the quality and quantity by compared with the PCR marker.

6.2.5 Plasmid construction and transformation of *E. coli*

Both the PCR products and the expression vector pET11d were digested with appropriate enzymes. The restriction digestion was carried out typically in a 50 μL restriction solution containing 10 units of appropriate restriction enzymes, 5 μL of 10X appropriate digestion buffer, 2 μg of PCR product or plasmid DNA and H_2O in an eppendorf tube. Digestion was carried out at 37°C for three hours to ensure complete digestion. The enzymes were inactivated by heating the samples at 70°C for 10 min. The samples were purified by partitioning against an equal volume of chloroform, followed by ethanol precipitation. The pellet was washed with 75% ethanol and dissolved in d. H_2O. The purified sample was run on an agarose TAE gel to check the quality and quantity. After ligation the recombinant plasmids were transformed in *E. coli* BL21*DE3 cells as described in Section 5.2.3. The transformants were verified by restriction enzyme digestion mapping.

6.2.6 DNA sequencing

In order to prepare DNA templates for sequencing, the DNA extracted was purified. Typically mini-scale prepared plasmid DNA (30 μL) was diluted to 90 μL with water containing RNase A (20 μg) and incubated at 37°C for 20 mins. A chloroform extraction was then performed: one volume of chloroform was added and the solution vortexed before centrifugation (5 min at 13,000 rpm). The aqueous upper layer was transferred to another micro centrifuge tube. DNA was precipitated out by the addition of 1/10 Volume of 3 M sodium acetate (pH 5.4) and 2.5 Volume of 95% ethanol, followed by cooling on ice for 20 min. The mixture was then centrifuged at room temperature (15

min at 13,000 rpm) and the pellet was washed with cold 75% ethanol. After another centrifugation (5 min at 13,000 rpm) the pellet was thoroughly aspirated and dissolved in 10 to 30 µL of H_2O. 1 µL was run on agarose gel electrophoresis to check for quality and quantity.

The cycle sequencing reaction and subsequent purification by ethanol precipitation of the sequencing products were followed as described in Section 2.2.3.3.

6.2.7 Expression and purification of ArsD

ArsD was purified from culture of strain *E. coli* BL21 (DE3) bearing pSB21. Cells were grown at 37°C overnight in 5 mL LB medium containing 0.1 mg/mL ampicillin. The cultures were then diluted 200 folds into 1 litre of fresh LB medium containing 0.1 mg/ml ampicillin, 1% $C_6H_{12}O_6$ (glucose) and further incubated at 37°C until OD_{600nm} reached 0.5-0.8. Production of ArsD was induced by the addition of IPTG at a final concentration of 0.1 mM for 3 hour. Induced cells were harvested by centrifugation at 5000 g at 4°C. The pelleted cells were suspended in 6 mL of Native binding buffer (250 mM NaH_2PO_4, pH 8.0; 2.5M NaCl) and disrupted by addition of lysozyme (2 mg/mL) on ice for 30 mins. Unbroken cells and membranes were removed by centrifugation at 5000 g for 15 mins. The supernatant containing the ArsD was loaded onto a Ni-NTA column (Invitrogen) pre-equilibrated with the native binding buffer. The column was washed three times with 6 mL of native wash buffer (250 mM NaH_2PO_4, pH 8.0; 2.5 M NaCl; 3 M Imidazole, pH 6.0). Fractions of 1 mL were collected by addition of native elution buffer (50 mM NaH_2PO_4, pH 8.0; 0.5 M NaCl; 3M Imidazole, pH 6.0) and analyzed by SDS-PAGE and fractions containing ArsD were pooled. The

concentration of purified ArsD was determined by Bicinchoninic acid assay (Smith et. al., 1985).

6.2.8 Preparation of DNA fragments and DIG labelling

147 bp fragment containing the *arsR* promoter region, a fragment of 70 bp containing *arsD* terminator/promoter regions and a fragment of 110 bp containing a non specific fragment was used as a negative control, were prepared from plasmid pZT using specific primers (Table 6.2).

Table 6.2: Primers used and DNA probe sizes

Primer Name	Primer Sequence 5'-3'	Corresponding position of pZT	Upstream (Forward)/ Downstream (Reverse)
Pro R-F	TTCAGTTGAATATATAAGCG	219-238 bp	Upstream
As4-13	TATCTAGAGCCATATCTATACCTCCTT	366-351 bp	Downstream
D-NcoF	TACCATGGCAAAGATGAAGAAGATAGAA	2296-2313 bp	Upstream
As4-19	TATCTAGAGTCATTGCCTTTCACTT	2681-2666 bp	Downstream
As4-18	TGAGACTGCTATGAAAG	5457-5473 bp	Upstream
As4-11	GTCTAGACTGTTTCTGTGACATC	5581-5566bp	Downstream

After PCR, fragments were purified using a G-50 sephadex spin column (Sambrook et. al., 1989). For making probes the fragments were then 3' end labeled with DIG-ddUTP using the 3' end labeling kit with DIG-ddUTP (Roche).

6.2.9 Gel mobility shift assay

To detect DNA/protein binding, gel mobility shift assays were performed after reacting 20 µL of a solution containing 10 mM Tris-CI, pH 7.6, 80 mM KCI, 0.2 mM Na_2EDTA, 0.2 mM DTT, 10% glycerol, DNA probe and various amounts of cell extracts. The binding mixtures were incubated on ice for 30 min. When required, arsenite or unlabeled DNA fragments were added individually to the binding mixture along with the

probe. Samples were loaded onto a 6% polyacrylamide gel and subjected to electrophoresis at 4°C at 120V for 1-2 hours with 45 mM Tris-Cl, 45 mM boric acid, and 1 mM Na_2EDTA, pH 8.5, as both gel and electrode buffers. Separated components were transferred onto a positively charged nylon membrane by electroblotting at 300 amps in 0.5x TBE buffer for 45 min. Detection of DIG-labeled DNA-protein complexes with alkaline phosphatase-conjugated anti-DIG antibody and CDP-star chemiluminescent alkaline phosphatase substrate was performed according to manufacturers protocol (Roche). The chemiluminescent signals were exposed on to an X-ray film.

6.3 Results

6.3.1 Functional involvement of the CDB3 *arsD* gene in arsenic resistance

To study the *in vivo* function of *arsD* gene, exonuclease III was employed to achieve deletion mutagenesis of the gene in pAR27 carrying the CDB3 *ars* gene cluster *arsRYCDATIP* which had been digested with *Cpo*I enzyme. As there is only one *Cpo*I site located near the start of *arsD* coding region, a linear band corresponding to 10 kb was visualized on the 0.8% ethidium bromide stained agarose gel (Figure 6.1b).

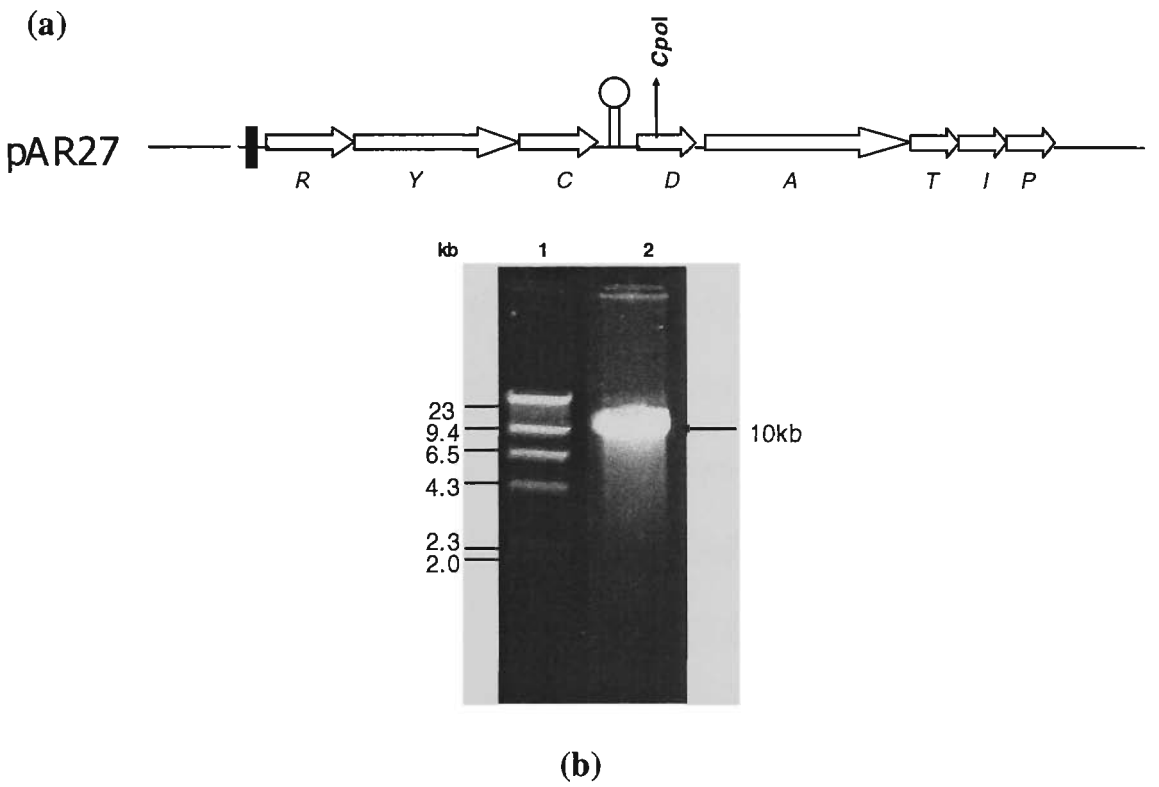


Figure 6.1: (a) Diagram to show position of single *Cpo*I site. (b) Agarose gel (0.8%) electrophoresis of *Cpo*I restriction enzyme digested plasmid pAR27. Lane-1: λ/*Hin*dIII DNA marker; Lane 2: 10 kb linear band of plasmid AR27 digested with *Cpo*I restriction enzyme. Marker sizes are indicated in kb on the left.

The exonuclease treatment was carried out at three time points (5, 15 and 25 sec). After transformation most colonies were recovered on the 5 sec plate. Plasmids isolated from 37 colonies were subjected to *Cpo*I restriction endonuclease digestion (Figure 6.2) to screen for *Cpo*I site lost mutants.

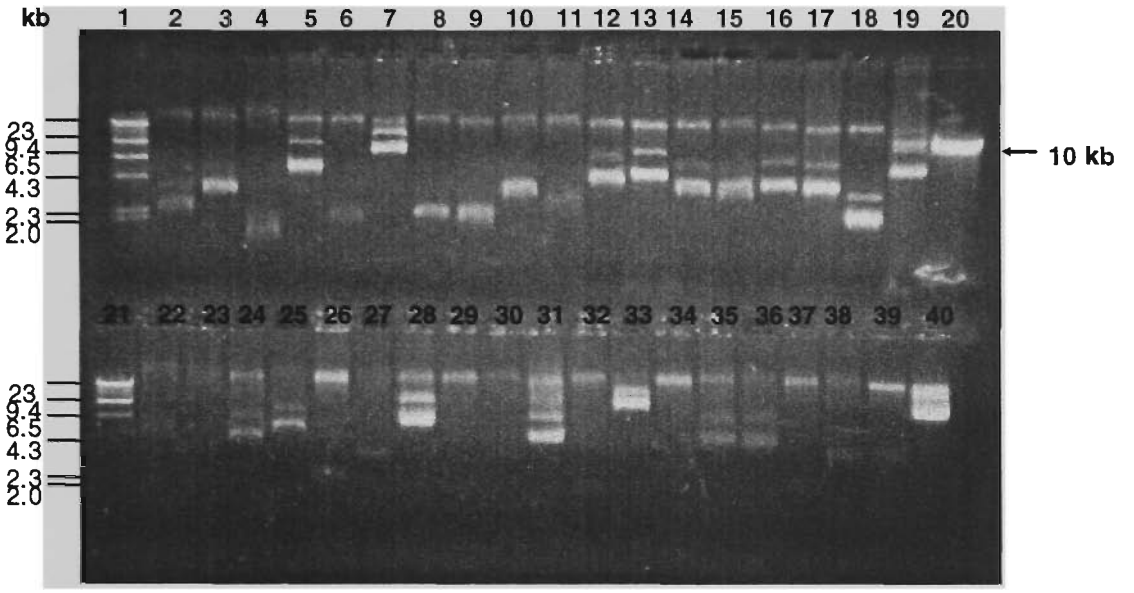
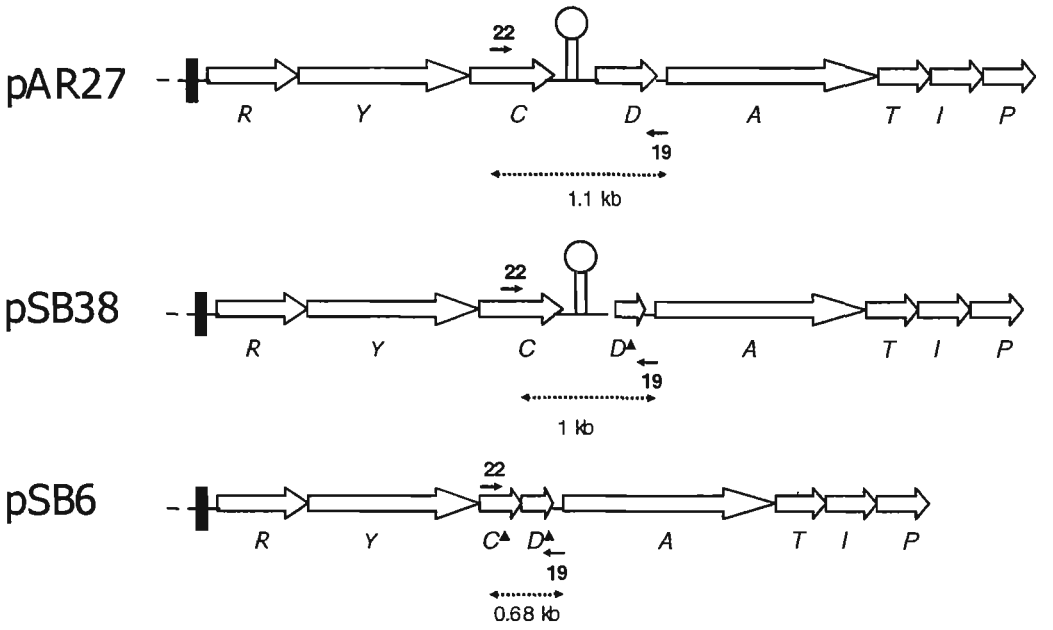
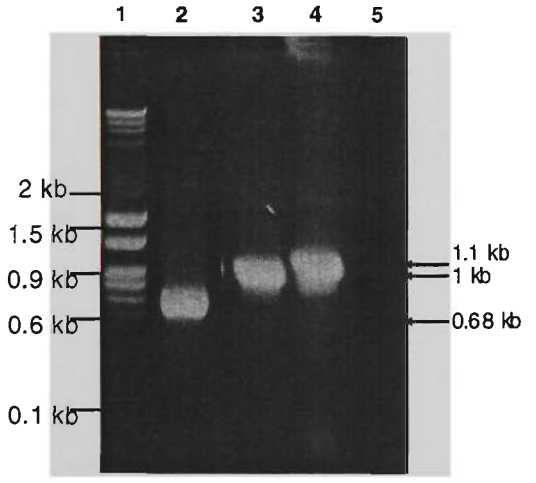


Figure 6.2: Agarose gel (0.8%) electrophoresis of *Cpo*I restriction enzyme digested plasmids. Lane-1 and 21 are λ/*Hin*dIII DNA markers; Lane 20 is pAR27 (positive control). The other lanes are plasmids from colonies resulted from exonuclease III treatment. Marker sizes are indicated in kb on the left.

The positive control pAR27 was completely digested with *Cpo*I as a single 10 kb band (lane 20), while, most of the clones screened from transformation plates seemed to be deleted for the *Cpo*I site (Figure 6.2). Five mutant clones were chosen for further confirmation by PCR using primers flanking the *arsD* gene to depict the sizes. PCR reaction was performed using primers As4-22/As4-19 (covering from the end of *arsC* to end of *arsD*) The PCR products amplified using two different sets of primers were loaded on to 1.5% agarose gel to confirm the size (Figure 6.3). As a positive control pAR27 was also used.



(a)

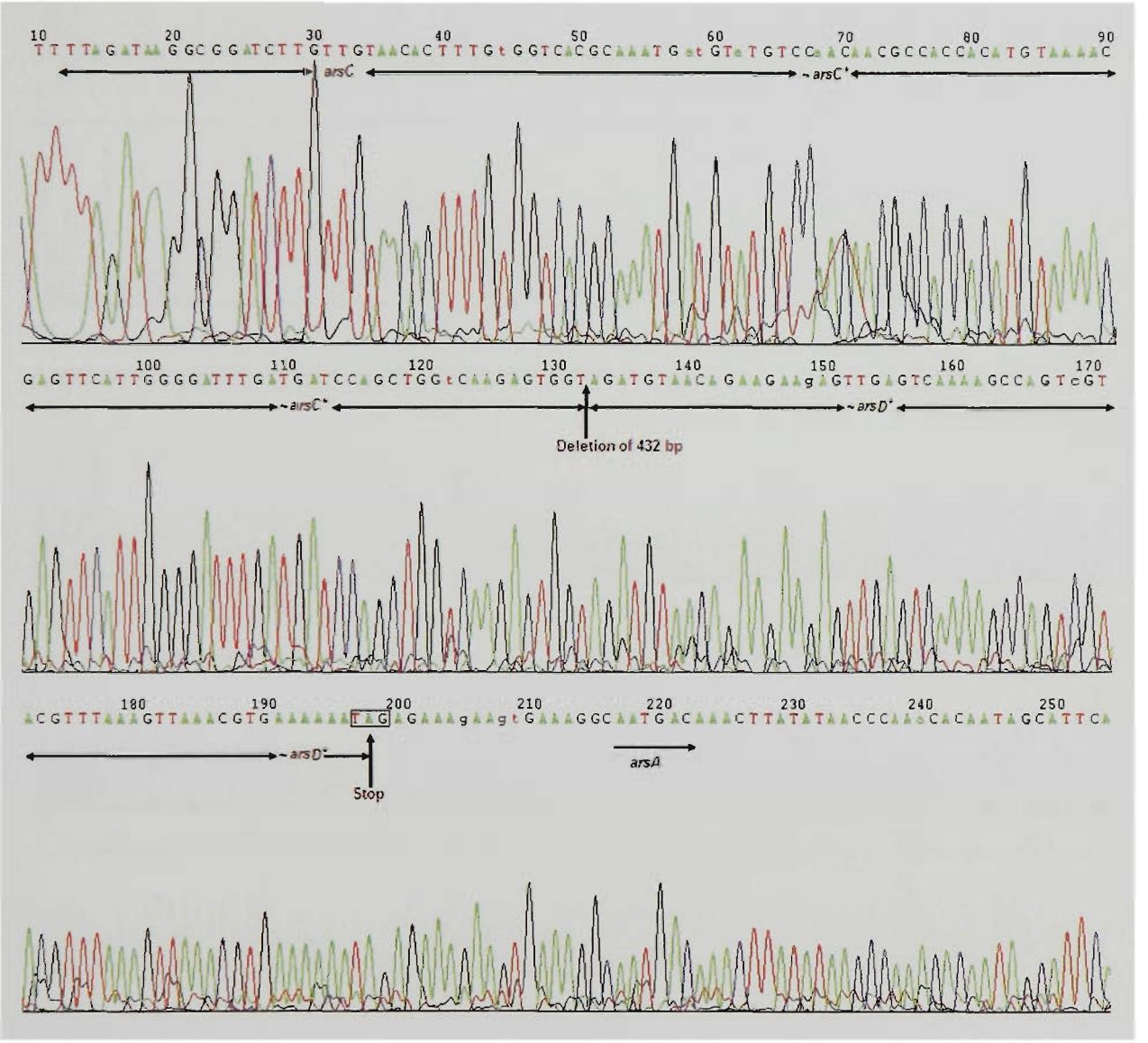


(b)

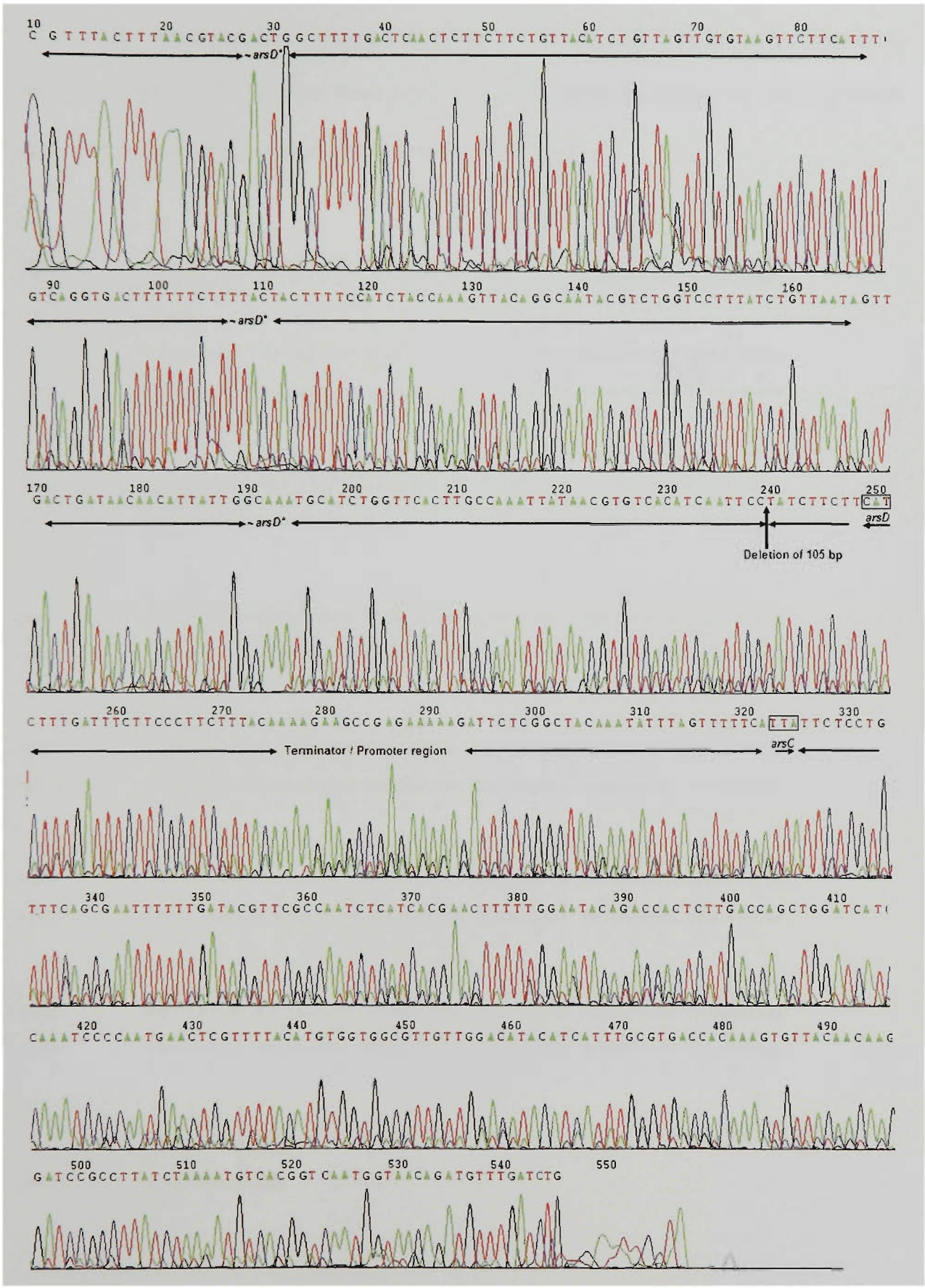
Figure 6.3: Agarose gel (1.5%) electrophoresis of PCR amplified products from *arsD* mutated clones and pAR27. (a) Expected PCR fragment sizes from all clones. (b) PCR amplified products using primers As4-22/As4-19. Lane 1: λ/*Hin*dIII 100 bp ladder DNA marker; Lane 2-5: PCR products amplified from clones two mutant clones 6 and 38, AR27 (positive control) and negative control (no template) respectively. Marker sizes are indicated in kb on the left.

The positive control pAR27 which had the intact *arsD* showed the anticipated fragment size (1120 bp) (Figure 6.3, Lane 4). No amplified product was observed in the negative control lane (Figure 6.3, Lane 5) ensuring no contamination. Clone 38 showed

slightly smaller fragment sizes (~1015 bp) compared to the positive control indicating only the disruption of *arsD* gene. The PCR products from clones 6 and 38 were subjected to sequencing to confirm the deletion using primers As4-19 and As4-22 (Figure 6.4).



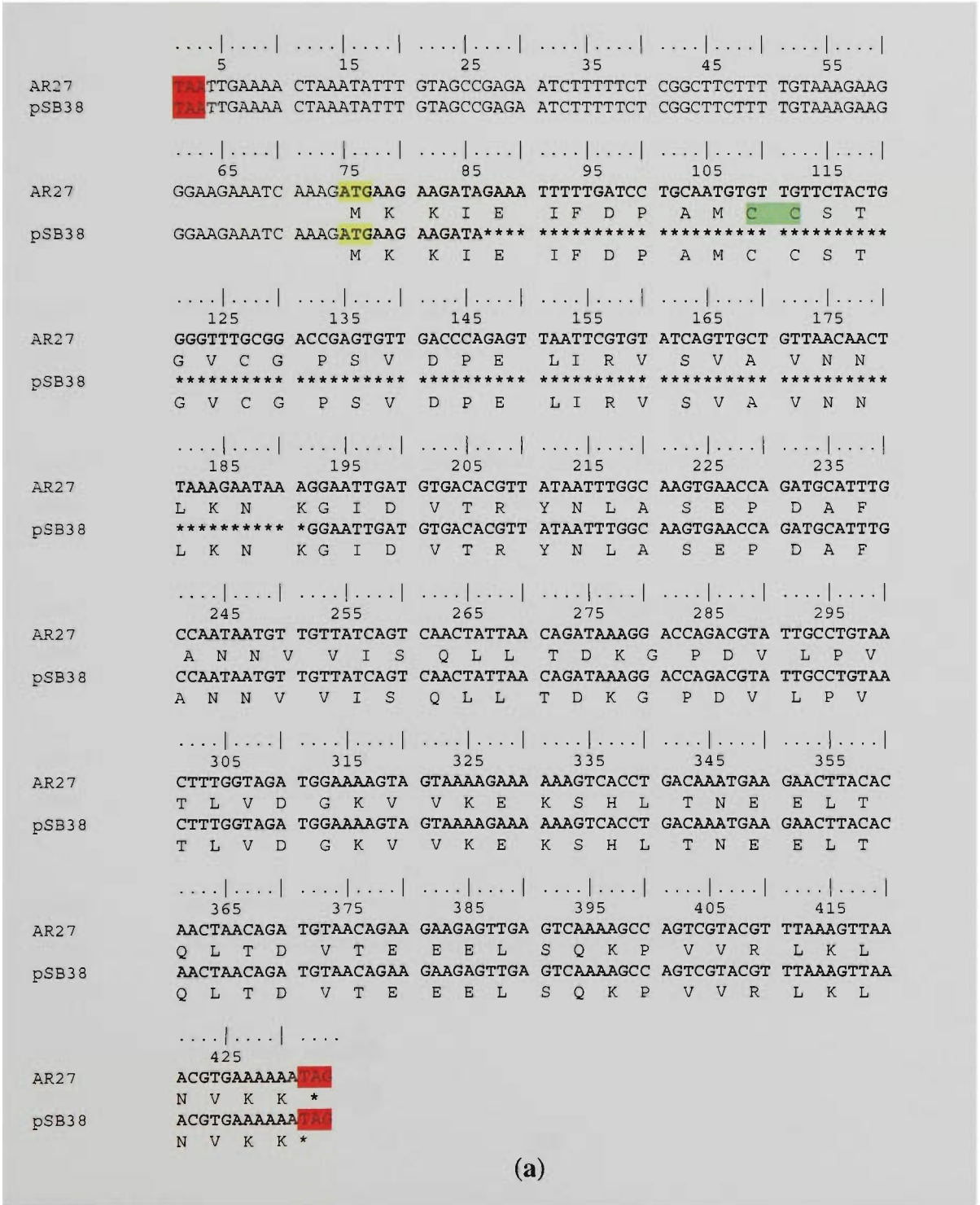
(a)



(b)

Figure 6.4: Electropherogram images of sequencing of *arsD* mutant clones 6 (a) and 38 (b). For clone 6 primer As4-22 (forward) and for clone 38 primer As4-19 (reverse) were used.

After obtaining the sequence data of the two clones, pair wise sequence alignment with wild-type *arsD* sequence from pAR27 was performed to depict the missing DNA sequence and consensus on protein sequence (Figure 6.5).



(a)

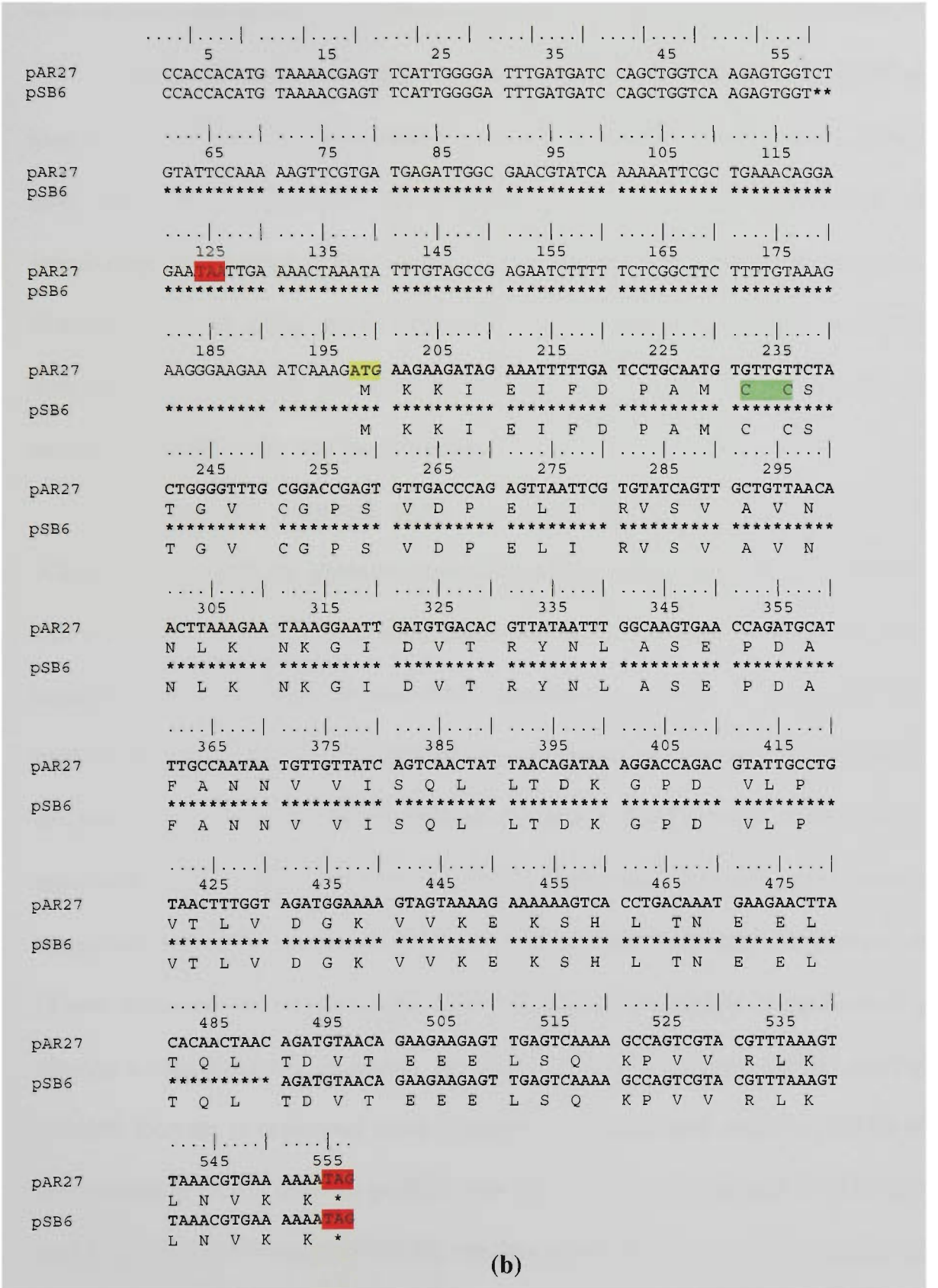
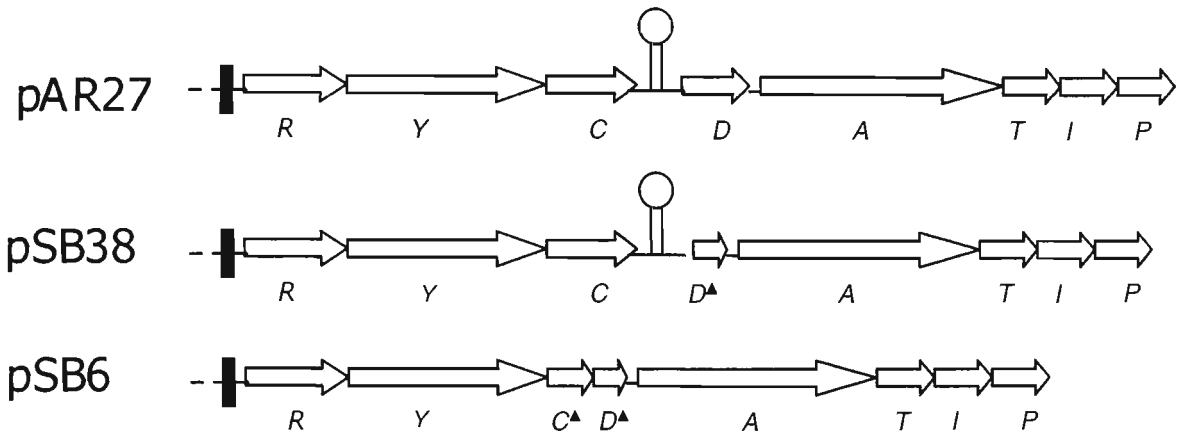


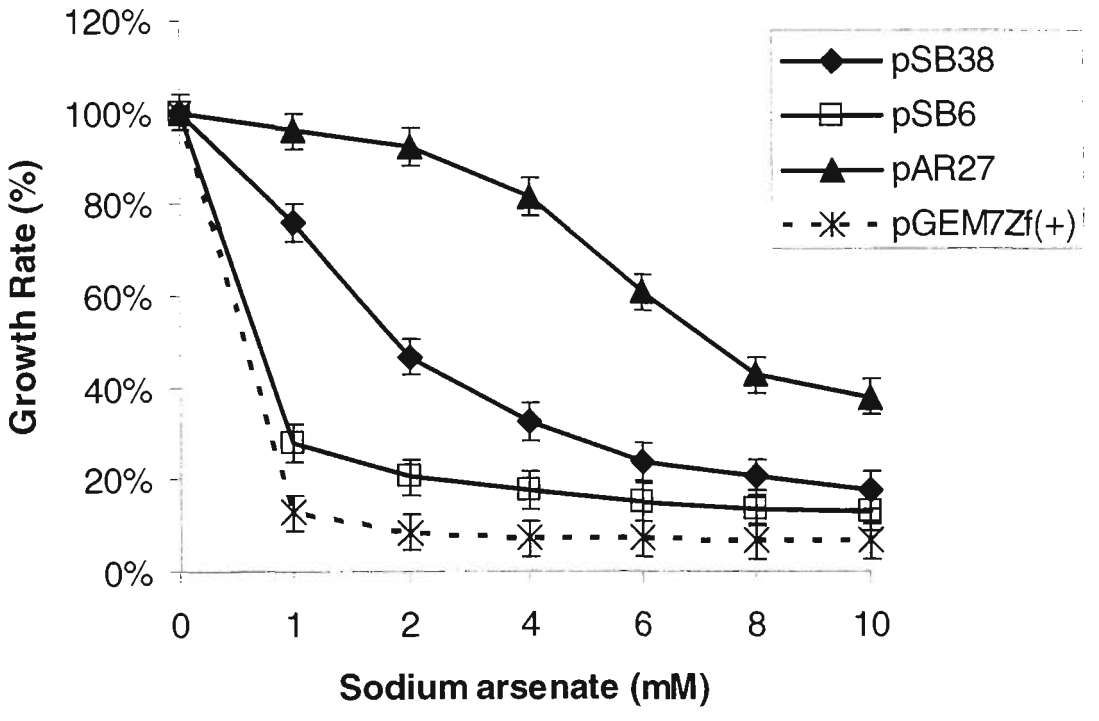
Figure 6.5: Pairwise alignments of DNA and protein sequences of *ars*D mutant clones 38 (a) and 6 (b) with wild type pAR27. The deleted bases are indicated with an asterisk '*', amino acids deleted are indicated in grey, the start codon ATG is highlighted in yellow, stop codon TGA is highlighted in red. The *arsD* gene sequence is represented in blue color; the terminator/promoter sequence upstream of *arsD* is represented in black, *arsC* gene sequence is represented in green.

105 bp of *arsD* gene were deleted from pAR27 disrupting the ORF thereby inactivating the function of the ArsD protein in pSB38. 432 bp fragment covering part of *arsC*, the terminator/promoter structure and *arsD* was deleted from pAR27 thereby inactivating the function of *arsC* and *arsD* in pSB6. In pSB38 the *arsD* mutation and/or deletion does not cause frame shift that could result in premature termination of translation. The cystiene pair Cys12-Cys13 in ArsD was deleted in both the mutants potentially affecting the metalloid binding.

The mutated constructs pSB6 and pSB38 were then transformed in *E. coli* AW3110 along with the parental plasmid pAR27 and the vector control pGEM7Zf(+). Cultures of these five strains were analysed for their ability to grow on increasing concentrations of arsenate (Figure 6.6). Compared to pAR27, *E. coli* strain harboring pSB38 showed a significant decrease in resistance to increasing concentrations of arsenate (T-test indicated the significant P-value < 0.001), while pSB6 which had a truncated *arsC*, deleted terminator/promoter region and disrupted *arsD* showed only marginally increased resistance above the control strain harboring the parental plasmid (T-test indicated the significant P-value < 0.001). These results suggest *arsD* gene is playing a significant role in arsenic resistance in the gene cluster. pAR27 *arsD* showed 25-40% increase in resistance when compared to the mutated *arsD* in pSB38, thereby, demonstrating CDB3 *arsD* in pAR27 with the missing cystiene pair Cys112-Cys113 is still functional and is essential for the regulation and control of over expression of ArsB, which itself is toxic to the cells. The decrease in resistance could be due to reduced expression of genes located downstream of *arsD* (*arsATIP*). This can be further detected by performing mRNA analysis.



(a)



(b)

Figure 6.6: (a): *ars*D mutated construct maps (pSB38 and pSB6) along with the wild type pAR27. (b): Growth inhibition of *E. coli* AW3110 harboring plasmids pSB6 pSB38, pAR27 and pGEM3Zf (+) to sodium arsenate. Overnight cultures were diluted 50-fold into LB liquid medium containing increasing amounts of sodium arsenate, incubated at 37°C for 5 hours. The error bars represent SD (n=3).

6.3.2 Construction of N and C terminal ArsD constructs

In order to understand the function/role of ArsD in metalloid sensing and regulation of CDB3 *ars* gene cluster 1, a substantial amount of purified ArsD is needed. Therefore N- and C-terminal fusions of ArsD constructs with six histidine tags (to facilitate affinity purification) were designed as shown in Figure 6.7 and 6.8.

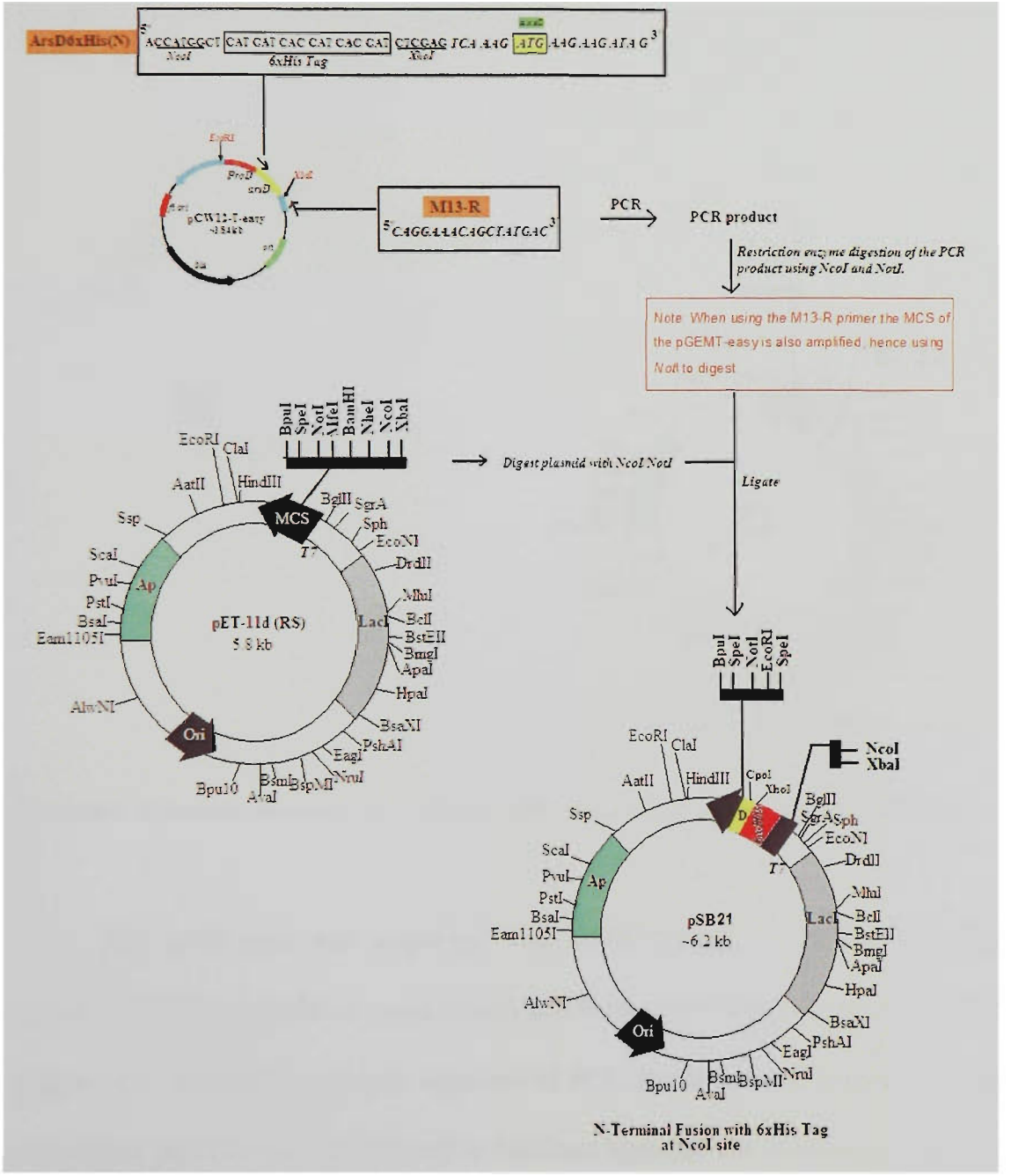


Figure 6.7: Schematic illustration of N-terminal fusion construct of *arsD* with six histidine tag.

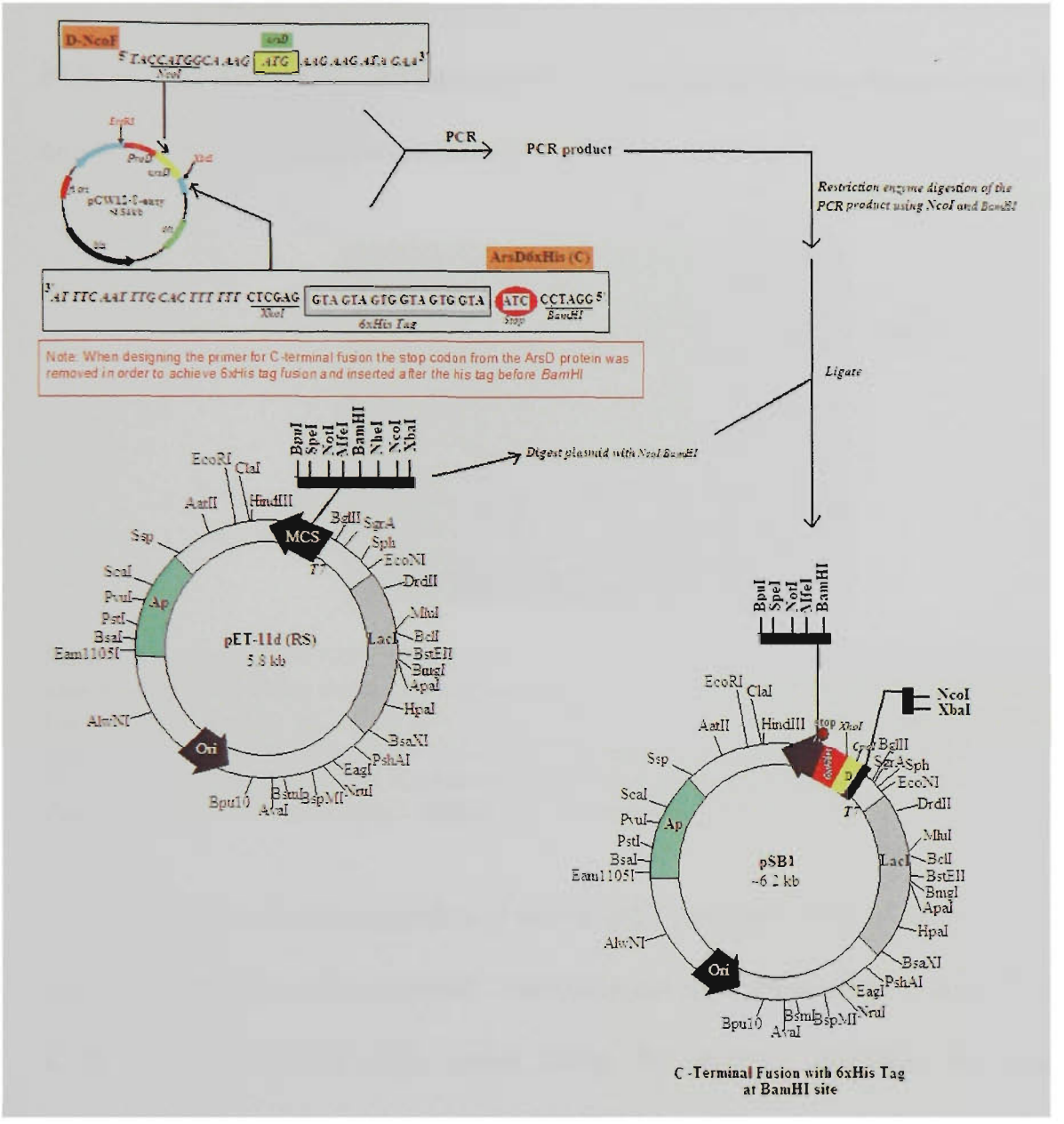


Figure 6.8: Schematic illustration of C-terminal fusion construct of *arsD* with six histidine tag.

The *arsD* gene was amplified from pCW12-Teasy by PCR using primer pair ArsD6x(N)/M13-R for N-terminal fusion and D *Nco*-F/ArsD6x(C) for C-terminal fusion (Figure 6.9). Restriction enzyme digestion of PCR amplified *arsD* fragments (Figure 6.9) and vector pET11d were performed to facilitate ligation. For N-terminal fusion both the

PCR product ArsD6x(N) and the expression vector pET11d were digested with *NcoI*/*Not*I and for C-terminal fusion with *Bam*HI/*Nco*I.

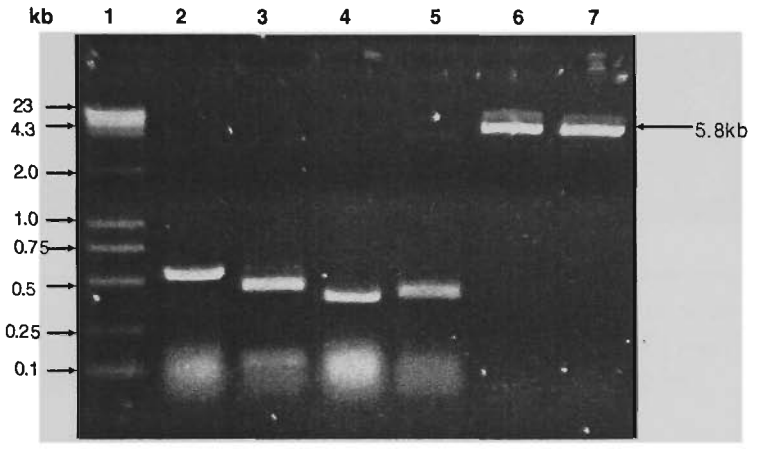


Figure 6.9: Agarose gel (2%) electrophoresis of restriction endonuclease digested PCR amplified *arsD* gene from pCW12-Teasy and expression vector pET-11d. Lane 1: λ/*Hin*dIII+100 bp ladder DNA marker; Lane 2: PCR product of ArsD6x(N); Lane 3: PCR product of ArsD6x(C); Lane 4: PCR product of ArsD6x(N) digested with *NcoI*/*Not*I, Lane 5- PCR product of ArsD6x(C) digested with *Bam*HI/*Nco*I, Lane 6- pET-11d vector digested with *NcoI*/*Not*I for N-terminal fusion and Lane 7: pET-11d vector digested with *Bam*HI/*Nco*I for C-terminal fusion. Marker sizes are indicated in kb on the left.

Upon successful digestion of vector (pET11d) and insert PCR products, ligation and transformation was performed. Transformants were screened by colony PCR (Figure 6.10) to confirm the right insert using the primers flanking the *arsD* gene, ArsD6x(N)/As4-19 for N-terminal fusion and D-Nco F/ArsD6x(C) for C-terminal fusion. A negative control (no template) was employed in the reaction to ensure no contamination of the primers nor the PCR mix or the H_2O used. The sizes of the amplified product were 425 bp [ArsD6x(N)] and 438 bp [ArsD6x(C)] as shown in Figure 6.10. No band was detected in the negative control lane, ensuring all the procedures followed were accurate and free of any contamination.

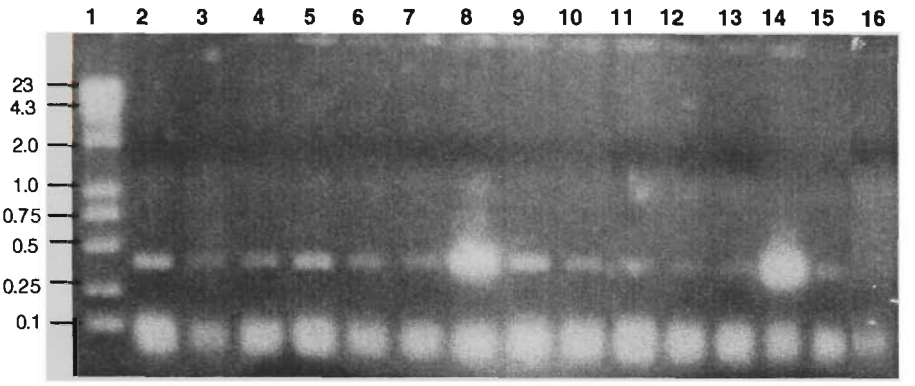
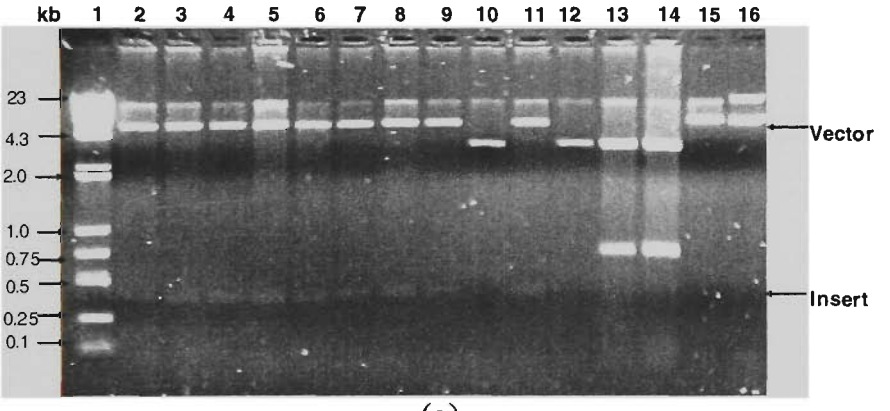
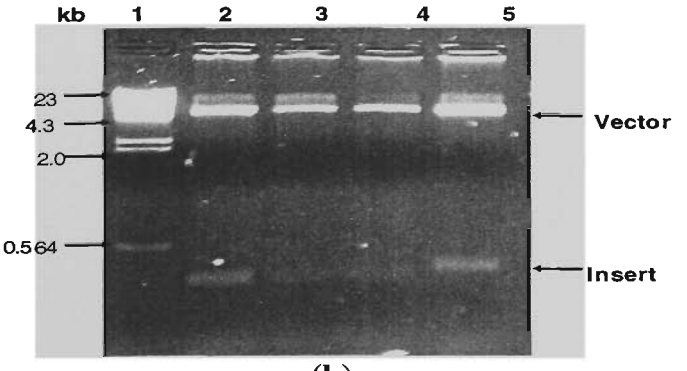


Figure 6.10: Agarose gel (1.5%) electrophoresis of colony PCR products from transformants selected on LB/Ap_{100} plates. Lane-1 is λ/*Hin*dIII+100 bp ladder DNA markers; Lanes 2 to 9 are clones selected from ArsD6x(N) transformants; Lanes 10 to 15 are clones selected from ArsD6x(C) transformants and lanes 16 is negative control (no template DNA). Marker sizes are indicated in kb on the left.

Plasmids were isolated and subjected to further restriction enzyme digestion and PCR verifications.



(a)



(b)

Figure 6.11: Agarose gel (1.5%) electrophoresis of *Bam*HI/*Xba*I, *Bam*HI/*Nco*I, *Eco*RI/*Nco*I and *Eco*RI/*Xba*I restriction enzyme digested plasmids. Gel (a) ArsD6xN/C plasmids. Lane 1: λ/*Hin*dIII 100 bp ladder DNA marker; Lanes 2 to 4: Plasmids digested with *Bam*HI/*Nco*I (ArsD6x(C)). Lane 5 to 9: Plasmids digested with *Eco*RI /*Nco*I (ArsD6x(N)). Lane 10 to 12: Plasmids digested with *Bam*HI/*Xba*I (ArsD6x(C)). Lane: 13 to 14 Plasmids digested with *Eco*RI /*Xba*I. Lane 15: pET11d digested with *Xba*I, Lane 16: pET11d digested with *Bam*HI. Gel (b) ArsD6x(N) plasmids. Lane-1: λ/*Hin*dIII 100 bp ladder DNA marker; Lane 2 to 5: Plasmids digested with *Eco*Rl /*Xba*I. Marker sizes are indicated in kb on the left.

Some combination of polylinker enzymes flanking the ArsD6xN/C were used in digestion as shown in Figure 6.11 to confirm the cloning at *Nco*I and *Bam*HI site have not resulted in deletion of polylinker sequences due to absence of sequence data covering the *Nco*I and *Bam*HI fusion sites. Three clones for each construct were finally verified to be correct in terms of restriction mapping and then subjected to DNA sequencing of the PCR product, since despite repeated efforts, sequencing of the plasmid was not achieved. A representative sequence for one of the pSB21 clones is presented in Figure 6.12 which shows no mismatches or mutations to the CDB3 *arsD* gene together with the fused N-terminal six histidine tag region.

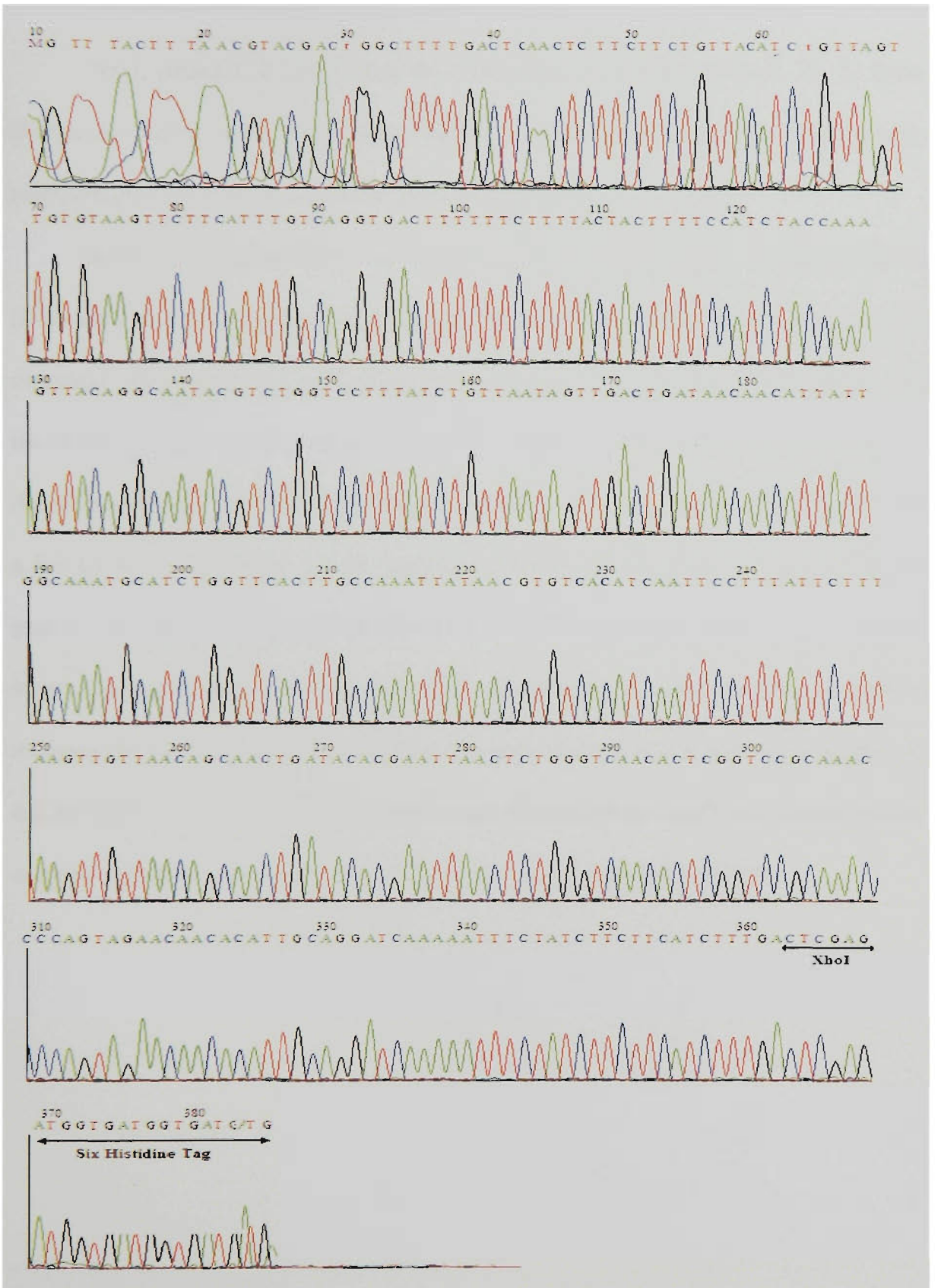


Figure 6.12: Electropherogram of the PCR sequence product of ArsD6x(N) amplified from pSB21. Six histidine tag and the *Xho*I restriction enzyme site are shown with arrows. Primer As4-19 was used and was sequenced from 3' end.

Next, clones 1, 2 and 3 from the C-terminal fusion and clones 6, 21, 22 from the N-terminal fusion were transformed into *E. coli*/BL21*(DE_3) cells for expressing the recombinant fusion protein (refer to method section 6.2.7).

Both pSB1 and pSB21 were digested with *Hin*dIII/*Xba*I, *Pst*I/*Xba*I and *Cpo*I (Figure 6.13) which showed the predicted banding pattern with respect to each enzyme digestion. The purpose of performing these digestions was to confirm the exact size of the inserted DNA, thereby eliminating any artifacts (eg. Insertion of *Nco-Not-Nco*, *Not-Nco-Not* stuffer fragments) which might result as a result of cloning. *Hin*dIII/*Pst*I showed a 0.9 kb vector band and 5.3 kb fragment covering ArsD6x(C)/(N) insert along with a part of the vector and *Pst*I/*Xba*I showed a 4.5 kb vector band and 1.7 kb covering the insert and part of the vector (Figure 6.13a). Also single enzyme digestion was performed utilizing the internal *Cpo*I site located at the start of *arsD* which revealed a single band at 6.2 kb (Figure 6.13b); all the fragment sizes obtained from single and double restriction enzyme digestion were consistent with the restriction map and position.

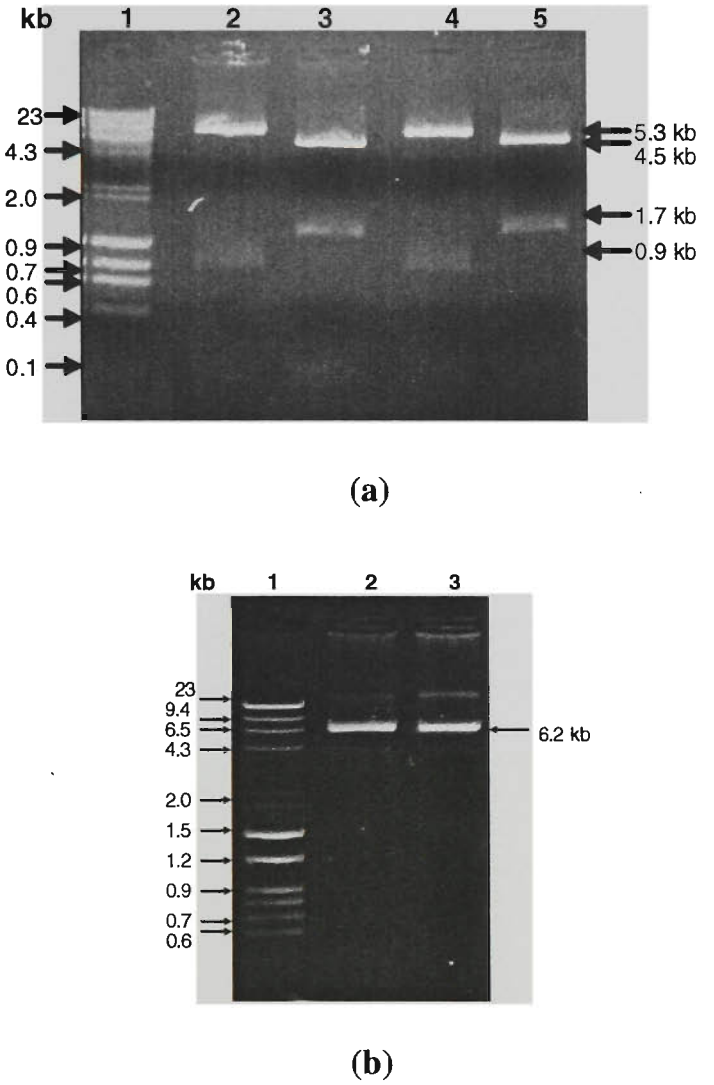
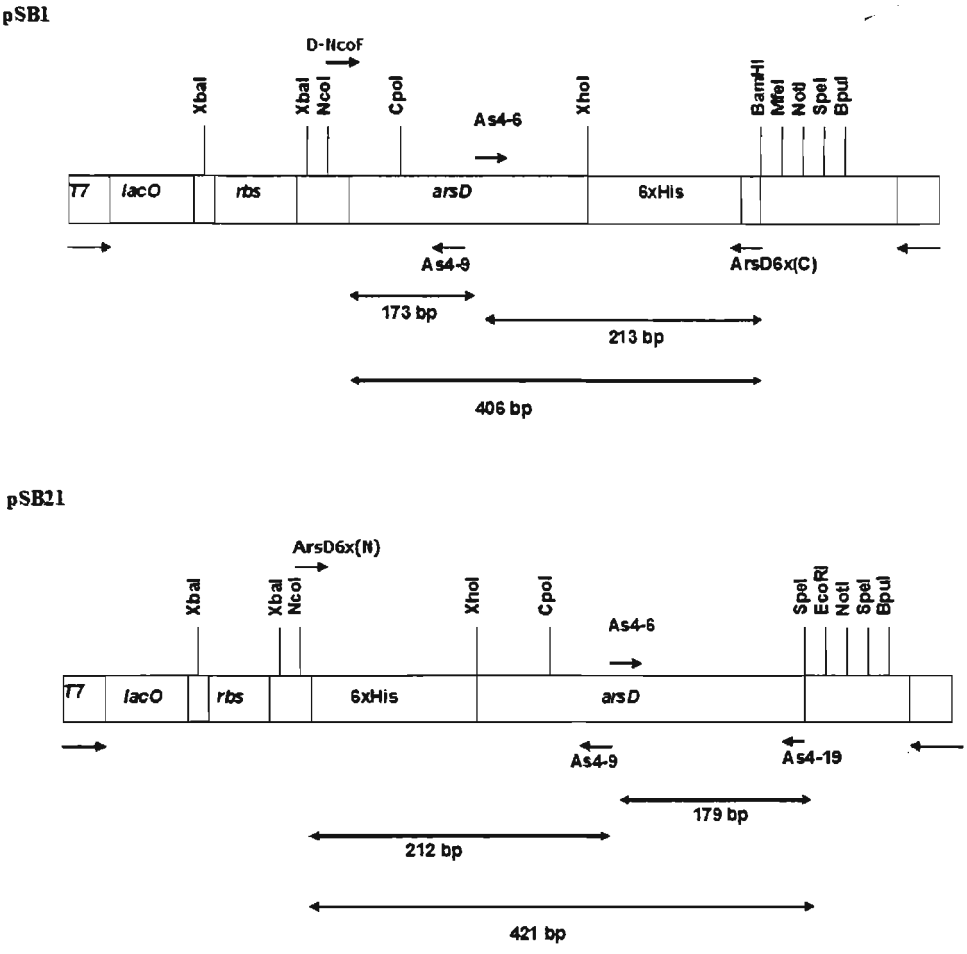
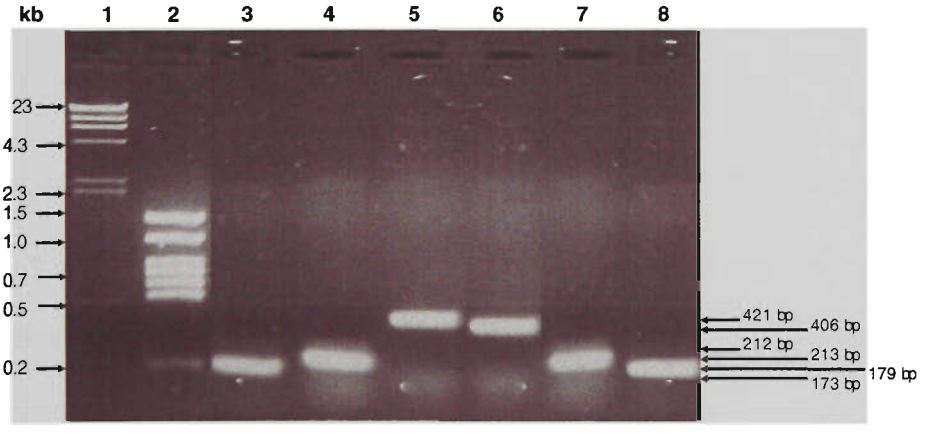


Figure 6.13: Agarose gel (1.5%) electrophoresis of *Hin*dIII/*Pst*I, *Pst*I/*Xba*I and *Cpo*I restriction enzyme digested colonies. Gel (a)- Lane-1: λ/*Hin*dIII 100 bp ladder DNA marker; Lane 2: pSB1 digested with *Hin*dIII/*Pst*I, Lane 3: pSB1 digested with *Pst*I/*Xba*I, Lane 4: pSB21 digested with *Hin*dIII/*Pst*I, Lane 5: pSB21 digested with *Pst*I/*Xba*I. Gel (b)- Lane-1: λ/*Hin*dIII 100 bp ladder DNA marker; Lane 2: pSB1 digested with *Cpo*I, Lane 3: pSB21 digested with *Cpo*I. Marker sizes are indicated in kb on the left.

Apart from restriction enzyme digestion before heading to bulk protein expression, a final confirmation of the clones were done by PCR using the internal and external primers as shown in Figure 6.14. Upon successful amplification of the PCR product from each of the two constructs, 10 µL of each sample was loaded onto a 2% agarose gel and the fragment sizes obtained for each construct were correct with respect to primer position and length of the fragment (Figure 6.14).



(a)



(b)

Figure 6.14: (a) Restriction enzyme maps of pSB1 and pSB21 showing the position of internal and external primers along with expected fragment sizes. (b) 2% agarose gel image of PCR amplified products from pSB21 and pSB1. Lane 1: λ/*Hin*dIII DNA marker, Lane 2: 200 bp ladder DNA marker, Lane 3-5: PCR product amplified from pSB21 using primers As4-6/19 (179 bp), ArsD6x(N)/As4-9 (212 bp) and ArsD6x(N)/As4-19 (421 bp), Lane 6-9: PCR product amplified from pSB1 using primers D-*Nco*F/As4-9 (173 bp), ArsD6x(C)/As4-6 (213 bp) and ArsD6x(C)/D-*Nco*F (406 bp). Marker sizes are indicated in kb on the left.

6.3.3 ArsD expression

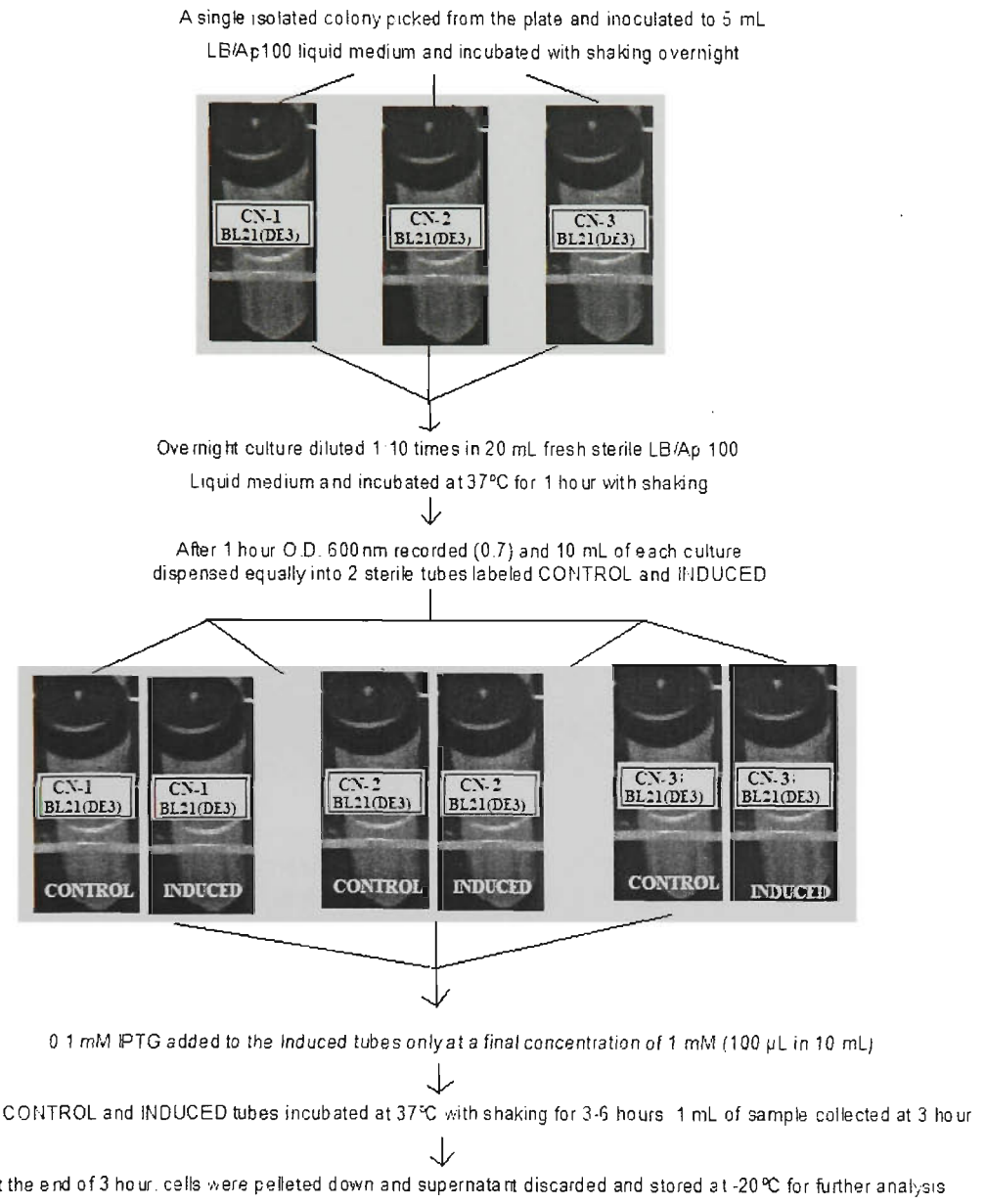


Figure 6.15: Flow chart illustrating the steps followed in expressing the ArsD6x(N)/(C) fusion protein in *E. coli* BL21*(DE_3) cells.

As illustrated in the flow chart (Figure 6.15) and in the methods section 6.2.9, overnight culture of single colony harboring either the N or C His tag construct was diluted 1:20 times in fresh LB/Ap_{100} broth containing 1% glucose and incubated at 37°C for 1 hour. When $O.D_{600nm}$ reached ~0.6 protein was expressed after adding IPTG and culturing for 3 more hours. For checking the protein expression, 1 mL of each sample

was spun down in a centrifuge, supernatant was discarded and pellet resuspended in 100 µL of 0.7% saline. After cell lysis and centrifugation, aliquots of both soluble and insoluble fractions were subjected to SDS-PAGE.

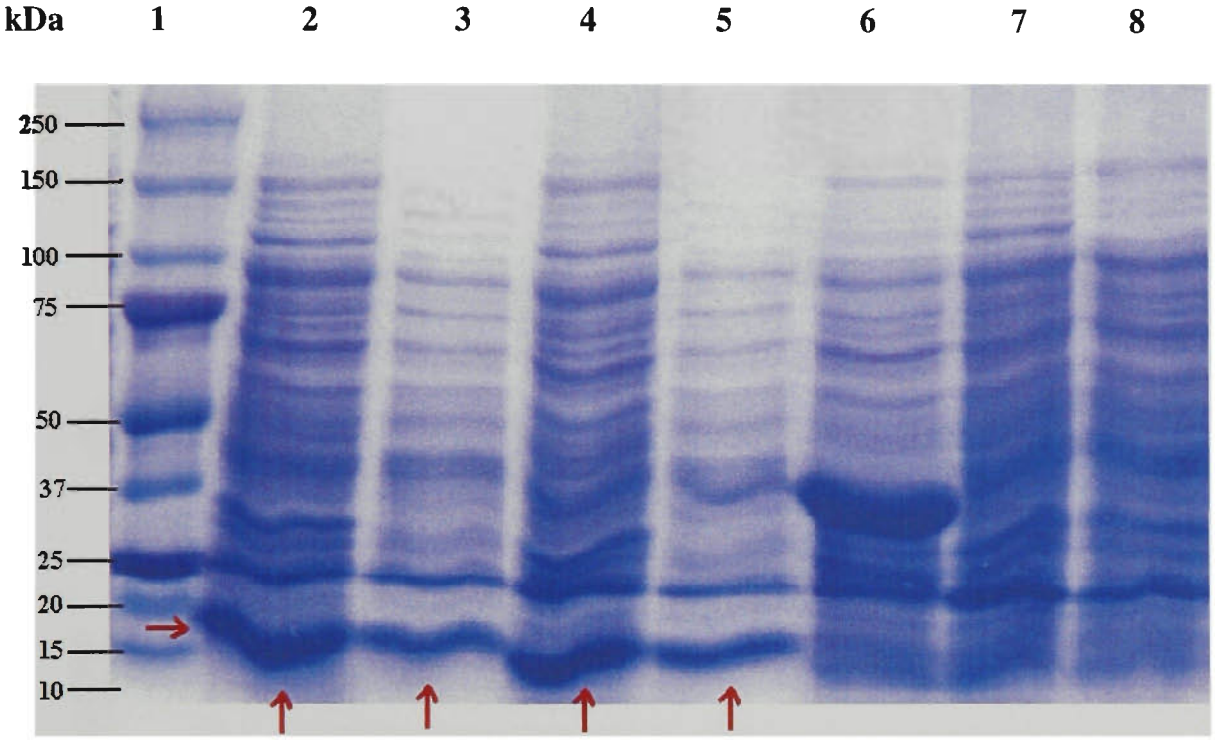


Figure 6.16: SDS-PAGE image of the expressed ArsD protein in *E.coli*BL21*(DE_3) cells. Commassive blue stained gel image of control and induced samples; Lane 1: Protein marker, Lane 2: pSB1 Induced pellet (insoluble fraction); Lane 3: pSB1 Induced supernatant (soluble fraction); Lane 4: pSB21 Induced pellet (insoluble fraction); Lane-5: pSB21 Induced supernatant (soluble fraction), Lane 6: pET151 induced supernatant (positive control), Lane 7: pSB21 control (un induced sample), Lane 8: pSB1 control (un induced sample). Arrows indicate the position of ArsD6x(N)/(C) protein.

The stained gel image as shown in Figure 6.16 demonstrated a successful protein induction and expression. A 14 kDa band corresponding to the mass of ArsD protein fused with six histidines at the N or C terminal was anticipated in the induced samples, however, the corresponding band as seen in Figure 6.16 was located between 15-20 kDa (Lane 2, 3, 4 and 5), this difference may have been due to some abnormal migration of protein bands in this gel (the exact molecular weight was visualized at 14 KDa after purification and further confirmed by Mass-spec). However, the 14 KDa band was not

apparent in the control samples showing expression of the ArsD6x(N)/(C) was under the tight control of T7 promoter. pET151-GAPDH/BL21*(DE3)construct (kindly provided by Amanda Cork, University of Wollongong) is a *Streptococcus pyogenes* M1 serotype known to express a 39.7 kDa His-tagged protein was used as a positive control (Figure 6.16, Lane-6).

6.3.4 Purification of His tagged ArsD proteins

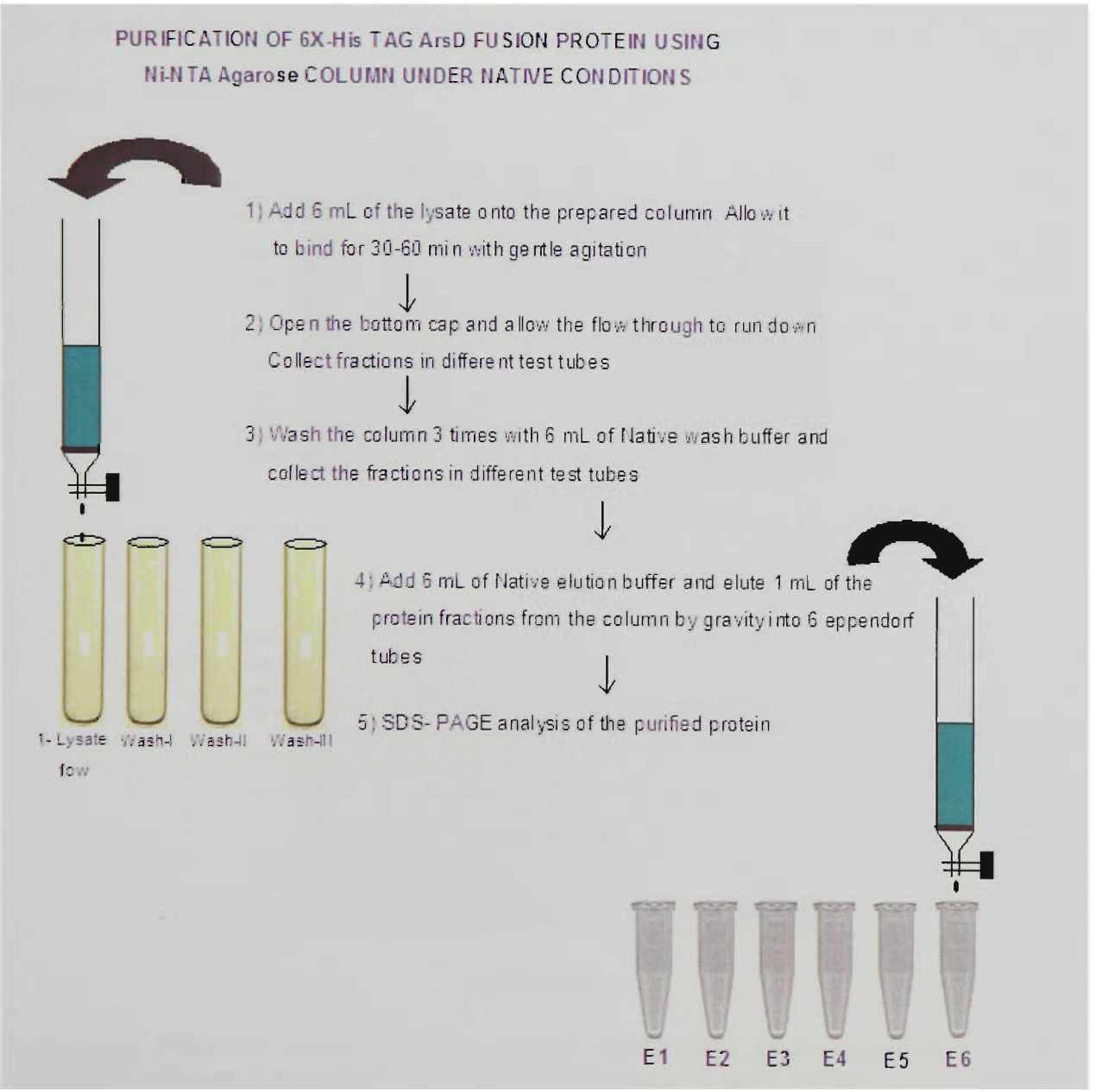


Figure 6.17: Flow chart illustrating the various steps followed in purifying the recombinant ArsD protein using a Ni-NTA agarose column under native conditions.

Bulk expression of the ArsD6x(N)/(C) were carried out and purification of the expressed proteins was achieved using a Ni-NTA agarose column. A schematic illustration of the steps followed is presented in Figure 6.17. The final eluted fractions each 1 mL was collected in eppendorf tubes and run on a 15% SDS-PAGE gel (Figure 6.18). The fractions containing the purified protein were then pooled, followed by bicinchoninic acid assay (Smith et. al., 1985) to determine the concentration of protein. The concentration was calculated using a series of standard BSA proteins to be 450 µg/mL.

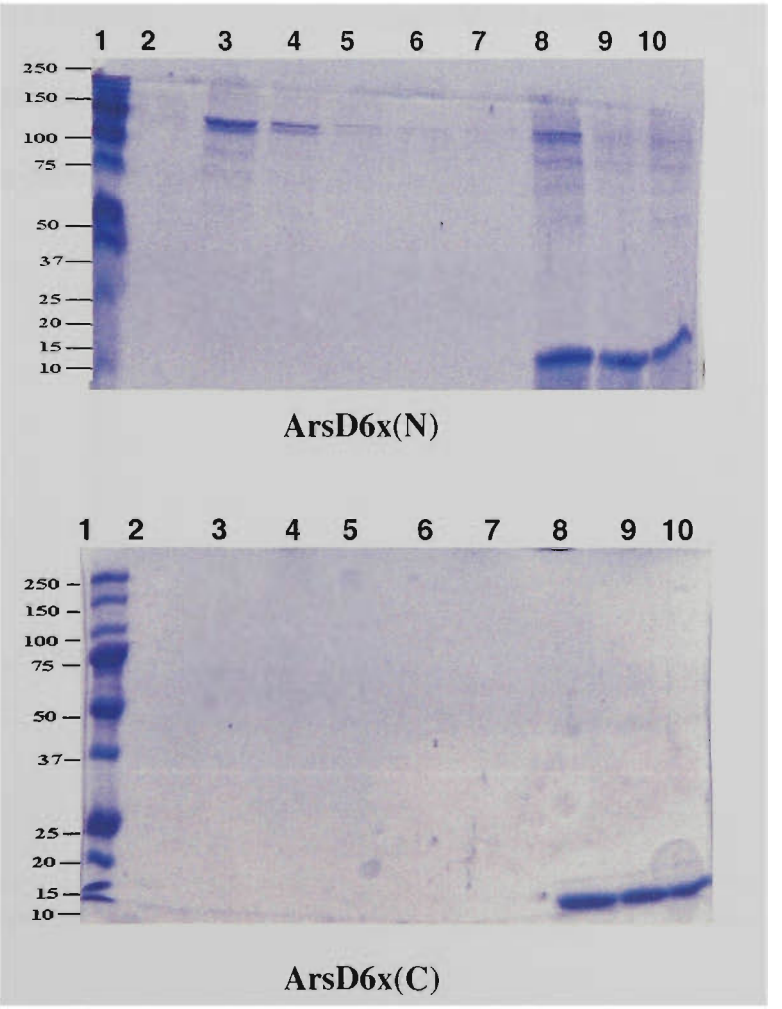


Figure 6.18: SDS-PAGE images of the two purified ArsD6x(N)/(C) proteins expressed in *E. coli* BL21*(DE_3) cells. Lane 1: protein molecular weight markers; lane 2-10: eluted fractions.

Further confirmation of the N-terminal six histidine fused protein was done by mass-spectrometry to determine the correct mass and purity. This was a collaborative work carried out by Dr. Andrew Aquilina. From the mass-spec results obtained as shown in Figure 6.19, it was evident that the protein mass matched the theoretical prediction (14467 Da) exactly confirming the identity of the recombinant protein produced being the target ArsD6x(N). There was some contamination by lysozyme due to the use of lysozyme (2 mg/mL final concentration) to break the cell. Unfortunately lysozyme has a very similar mass (14305 Da) as the recombinant ArsD protein. In an attempt to eliminate lysozyme a second round purification of the expressed protein was carried out and the contamination was reduced.

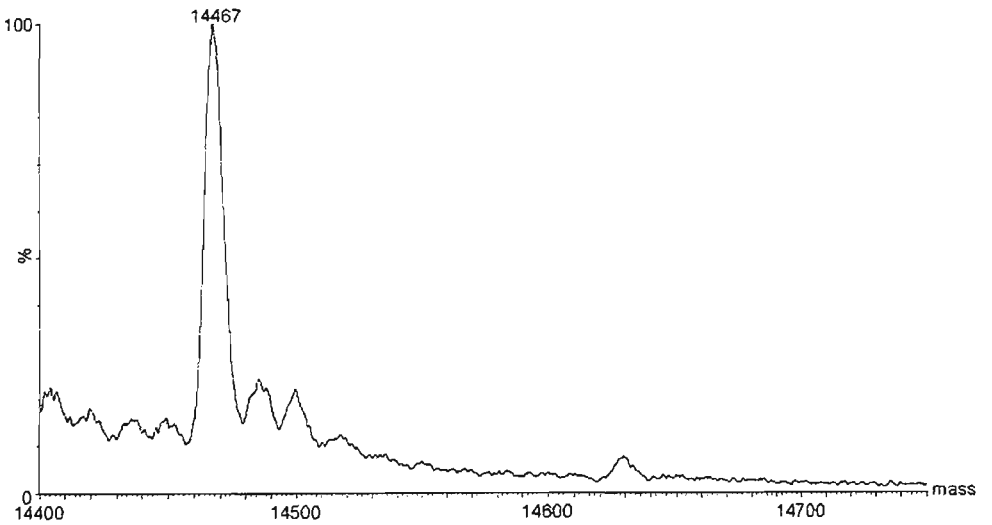


Figure 6.19: Mass spectrum of purified ArsD6x (N) protein. Purified recombinant protein was desalted with a ZipTip (Millipore) and eluted in 3 μL of 80% acetonitrile, 1% formic. The solution was analyzed using nanoelectrospray ionization mass spectrometry on a Q-ToF Ultima (Waters) using standard conditions. The raw spectrum was transformed to a mass scale to reveal a major species at 14,467 Da.

6.3.5 DNA-Protein interaction studies

To investigate the predicted repressor function of CDB3 ArsD, gel mobility shift assays were carried out to test its specific binding to both the *arsR* promoter and putative *arsD* promoter.

Gel shift assay with DIG-labeled DNA fragments

DIG-11-ddUTP DNA fragments covering the two promoters of *arsR* and *arsD* along with a nonspecific region (Figure 20a) were first obtained by PCR. The obtained sizes for the three PCR amplified fragments matched the theoretical sizes (Figure 6.20b).

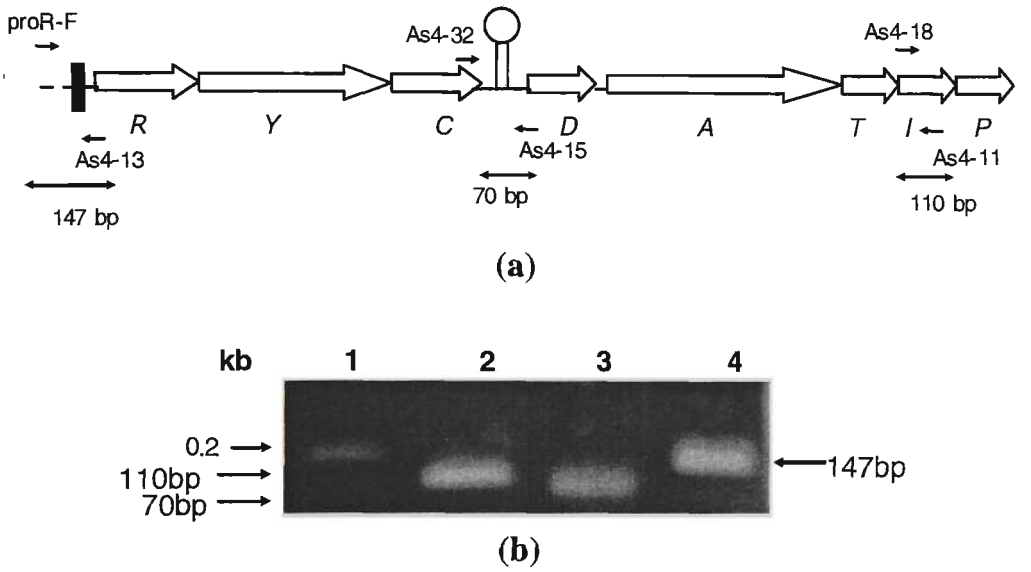


Figure 6.20: (a): Position of internal and external primers flanking the two promoters and a non specific DNA fragment along with their respective sizes. (b): Agarose gel (2%) electrophoresis of DIG-labeled PCR products. Lane 1: DNA marker; Lane 2: non-specific control fragment amplified using primers (As4-18/11); Lane 3: *arsD* promoter fragment amplified using primers (As4-32/15) and Lane 4: *arsR* promoter fragment amplified using primers (proR-F/As4-13).

Gel mobility shift assays were used to examine the DNA binding activity of ArsD. ArsD was shown to retard the migration of both *arsR* and *arsD* promoter fragments with DIG labels, no retardation of the labeled non-specific DNA fragment was detected (Figure 6.21a). Lysozyme did not have any binding (not presented).

A further assay tested the effect of arsenite on the binding. Dissociation of the ArsD protein from both promoters R and D were observed after addition of 0.1 mM

sodium arsenite (Figure 6.21b). When the reaction mixtures were mixed with cold unlabeled DNA fragments of promoter R or D respectively, the labeled binding disappeared, indicating the specific competition between the labeled and un labeled DNA fragments (Figure 6.21b).

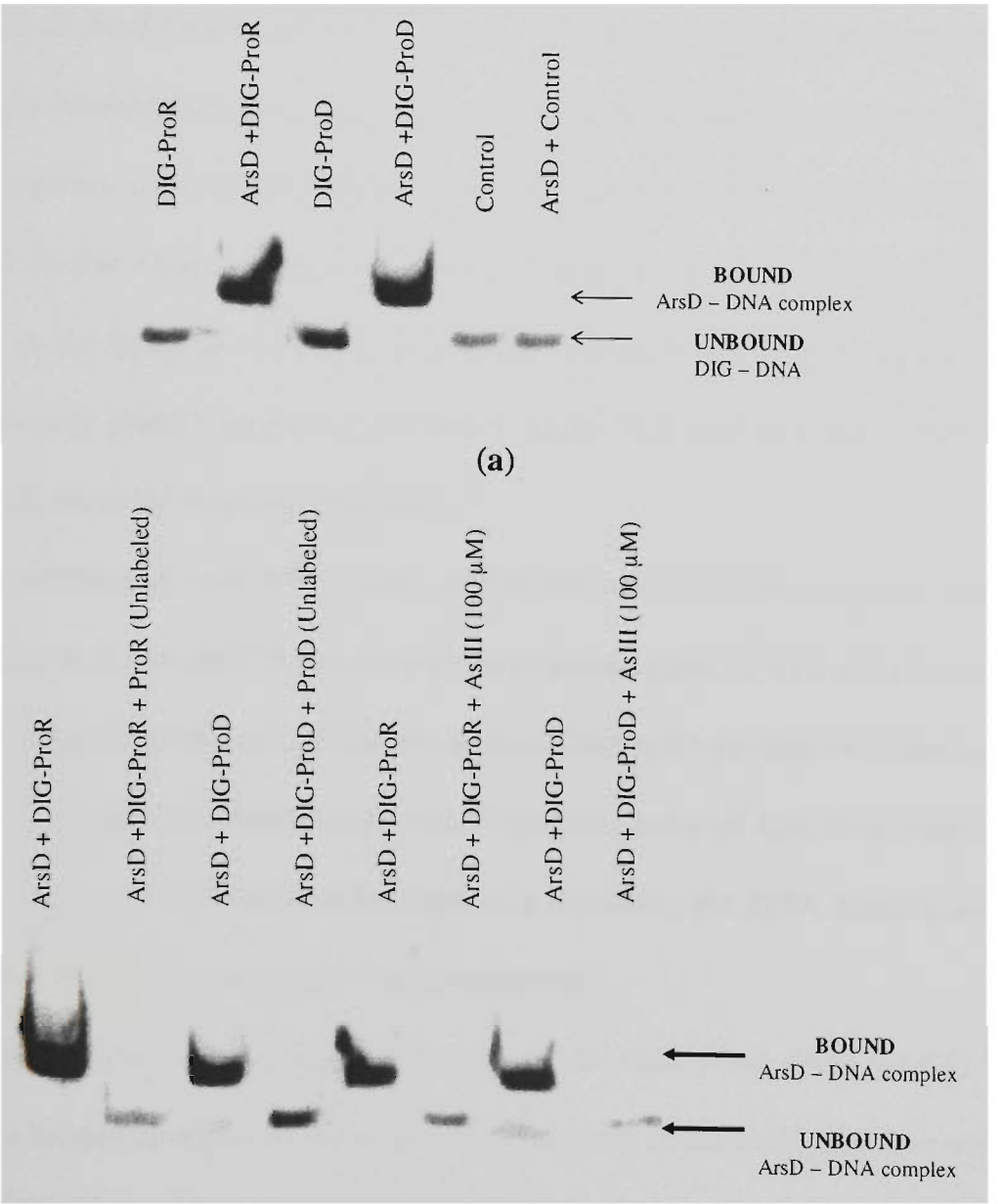


Figure 6.21: Mobility shift assays of ArsD-DNA complex formation using DIG labeled DNA fragments. DIG labeled DNA (0.40 µM) were mixed with 30 µg of purified ArsD protein; (a) Binding assay of promoter and non specific DNA fragments (control) with and without ArsD. The control DNA was amplified using primers (As4-18/11) located internally in *arsI* gene; (b): Effects of un labeled DNA fragments and sodium arsenite (0.1 mM) on the binding.

6.4 Discussion

ArsD is a novel regulatory protein and its distribution appears to be limited (Li et. al., 2001; Chapter 3). This 13 kDa protein has been shown to form homodimers to control the upper level of *ars* expression in *E. coli* R773 (Li et. al., 2001), preventing over-expression of ArsB (Wu and Rosen, 1993b). This benefits the cells as when the level of *ars* operon transcription rises, the high concentration of the integral membrane ArsB protein becomes toxic to the cell and limits the growth. This is evident by the study presented in this Chapter where by inactivation of the CDB3 ArsD protein led to a decrease in the bacterial resistance to arsenate (Figure 6.6b) when compared with the cells harboring pAR27 producing the intact ArsD. This confirms that CDB3 ArsD is functionally involved in arsenic resistance.

Apart from *E. coli* R773 study; this is the second evidence for the function of ArsD. ArsD in *E. coli* pR773 has three vicinal cysteine pairs, Cys12 and Cys13, Cys112 and Cys113 and Cys119 and Cys120, but in the CDB3 ArsD the later two pairs are absent (Chapter 3, Figure 3.9). Since these two C-terminal pairs of Cys have been reported essential for the *E. coli* ArsD to function as a regulator; the DNA binding activity of CDB3 ArsD in response to arsenic was investigated.

Additionally, a symmetrical dyad sequence with 7 bp (ATTAAAT) in each segment is located upstream of the putative –35 element of the *ars*R promoter which may be a site for repressor recognition and binding. A similar sequence was also conserved in *Bacillus subtilis* skin element (Sato and Kobayashi, 1998) while the sequence $TCATN_7TTTG$ for repressor binding conserved in *E. coli ars* operons is quite different indicating the promoter/operator upstream of *ars*R in the CDB3 *ars* gene cluster is

conserved in Gram-positive *ars* clusters and exhibits no similarity to previously identified sequences from Gram-negative *ars* operons (Xu et. al., 1996). Apart from the obvious 5' end promoter located, this DNA-protein study also included the DNA fragment located in front of *arsD* because of my speculation that a transcription regulatory (terminator or promoter) existed in this region (Wang, 2001; section 3.1 and 7.4.1).

Gel mobility shift assays have been widely used to examine DNA-protein binding properties. The assay typically involves the addition of protein to linear double stranded DNA fragments, separation of complexed and naked DNA by gel electrophoresis and visualization either by autoradiography, by staining with ethidium bromide or chemiluminescence detection is possible (Lane et. al., 1992). Assays were conducted using three DIG-labeled DNA fragments, a 147 bp 3' end labeled DNA fragment containing the CDB3 *arsR* promoter, a 70 bp 3' end labeled DNA fragment containing the CDB3 *arsD* promoter and a 110 bp 3' end labeled non-specific DNA fragment from CDB3 respectively, as target DNA. The recombinant CDB3 ArsD, was able to retard the migration of both *arsR* and *arsD* upstream fragments. No retardation was observed when the protein was incubated with the non-specific As4-18/11 fragment and neither when BSA and lysozyme were tested (result not presented), suggesting that the binding of ArsD to the *arsR* and *arsD* fragments were specific. Further study is needed using mass-spec to quantitatively study the affinity of ArsD to both R and D promoters respectively. Intrinsic protein fluorescence can be used to examine the properties of individual Cys pairs (Cys12-Cys13) as similar experiments conducted by Chen and Rosen (1997), to study the effect of metalloid binding and the resulting dissociation of the repressor from the promoter due to changes in confirmation of the ArsD protein.

The *E. coli* R773 ArsD protein has been shown to be an inducer-independent repressor protein, and the derepression could be induced by higher concentration of sodium arsenite (100 μM) compared with that of ArsR (10 μM) (Chen and Rosen, 1997). The arsenite inducible dissociation of CDB3 ArsD from *arsR* and *arsD* upstream regions was also tested and demonstrated to be true (Figure 6.21b). While, Li et. al. (2001) reported that mutation in either of the cysteine pairs Cys 12-Cys 13 and Cys 112-Cys 113 of *E. coli* ArsD eliminates derepression by the metalloid. CDB3 ArsD possesses only the N-terminal pair of Cys12-Cys13, but has no C-terminal Cys pairs. Since the CDB3 ArsD protein has only 35% identity to the *E. coli* R773 ArsD, this difference may be due to their differences in tertiary structure and binding confirmation when associated with their respective DNA partners which also vary in sequence. These warrants further investigation. One possibility would be to perform protein modeling and analyzing the role of tertiary structure and DNA binding. Many programs are available to study the putative ligand interactions to proteins. One program available from EMBL-EBI is MSDsite (www.ebi.ac.uk/msdsite), its macromolecular structure database (MSD) retrieval system allows protein structure analysis including putative bound ligands and its active site information (Golovin et. al., 2005).

Recently Lin et. al. (2006), have shown that *E. coli* ArsD can link ArsA to function as a chaperone in arsenic detoxification. Whether CDB3 ArsD has a similar function is yet to be examined.

Chapter 7

Transcriptional Regulation of CDB3 *ars* Gene Cluster 1

7.1 Introduction

Most bacterial *ars* gene clusters identified so far are organized as single operons (Chapter 1). Of all, the best characterized *arsRDABC* cluster present in *E. coli* R773 is transcribed as a single polycistronic mRNA with the two inducer-dependent trans-acting repressors ArsR and D homeostatically regulating the basal and upper levels of operon expression (Cai et. al., 1998; Li et. al., 2001). In comparison, the cloned CDB3 *ars* cluster 1 by our lab not only contains three more genes (*arsTIP*) but also exhibits a novel gene arrangement of the first five (*arsRYCDA*), where *arsDA* is located after *arsYC* instead of between *arsR* and *arsYC*.

Apart from the front promoter and operator, some other sequence features have also been found which may confer a complicated regulation mechanism for the gene expression of this cluster. As mentioned previously (Chapter 6), sequence analysis has predicted a site for repressor binding upstream of *arsR*.

Northern blot analysis was performed by a former member of our lab, Wang Q (2001), to determine whether there is a transcriptional terminator between *arsC* and *arsD*. The results showed that a 1.2 kb *arsRY* probe only detected a small RNA (1.9 kb) of about the *arsRYC* size instead of a large one corresponding to the whole cluster size indicating the possible presence of such a terminator. Another former member of our lab, Chenwei Su (2004), further performed reporter gene analysis. The results showed that at AsIII concentrations between 10 μM and 1000 μM the *arsR* 5' upstream region fused to *lacZ* exhibited an induced β-galactosidase activity demonstrating an active promoter. While no such induced β-galactosidase activity was observed by constructs harboring the

arsD 5' upstream region fused to *lac*Z. These meant that either the P_{arsD} did not exist or the promoter did exist but could not be induced by arsenite at the tested concentrations.

Therefore, this study aims to further investigate the regulatory mechanism of this *ars* gene cluster 1, attempting to reveal the transcription profile and examine the putative promoters.

7.2 Materials and Methods

7.2.1 Bacterial strains and plasmids

Bacterial strains and plasmids used in this study are listed in Table 7.1, with their relevant characteristics. *E. coli* and CDB3 (*Bacillus* sp.) strains were grown on LB agar or in LB liquid medium (Appendix 1). Ampicillin was added at a concentration of 100 μg/mL. Liquid cultures were shaken at 150-250 rpm in a shaking incubator (Bioline, Australia) at 37°C for *E. coli* and 30°C for CDB3.

Table 7.1: Bacterial strains and plasmids used in the study of Chapter 7

Bacterial Strains / Plasmid	Genotype or description	Reference or Source
E. coli AW3110	K-12 F'IN (*rrn*D-*rrn*E) Δ*ars::cam* (Cm^r, the chromosomal *arsRBC* deleted)	Gift from Dr. B.P. Rosen. (Carlin et. al., 1995)
E. coli JM109	*rec*A1 , *end*A1, *gyr*A96, *thi*, *hsd*R17 (r_K-, m_K+), *sup*E44, *rel*A1, Δ(*lac-pro*AB), [F', *tra*D36, proAB, *lac*IqZΔM15]	Promega
Bacillus sp. CDB3	Isolated from cattle dip-sites	Chopra et. al., 2007)
pAR27	A 7013 bp partial *Sau*3AI fragment containing CDB3 *arsRBCDATIP* cloned into pGEM7Zf(+) at *Bam*HI site	Provided by X. Luo, 2006
pZX11	PCR fragment of *LacZ* gene from pUJ8 fused to *arsC* of pZX1 (Km^r)	Provided by Z. Xu (University of Wollongong)
pZX12	PCR fragment of *LacZ* gene from pUJ8 fused to *trxB* of pZX4 (Km^r)	Provided by Z. Xu (University of Wollongong)
pZX13	PCR fragment of *LacZ* gene from pUJ8 fused to *trxB* of pZX10 (Km^r)	Provided by Z. Xu (University of Wollongong)
pZX14	*Sma*I- *Bam*HI fragment containing *arsRBCD* and part of *arsA* removed from pZX12 (Km^r)	Provided by Z. Xu (University of Wollongong)
pCW16	*Eco*RI-*Sac*I fragment from pCW12, containing the putative promoter P_{arsD} region, *arsD* and part of *LacZ* gene ligated into pUJ8 (Amp^r)	Provided by C. Su, 2004
pCW19	*Eco*RI-*Xba*I fragment containing the putative promoter P_{arsD} region and *arsD* deleted from pCW16 (Amp^r)	Provided by C. Su, 2004

7.2.2 RNA extraction

A single colony of CDB3 was picked up from a LB agar plate and incubated in 5 mL of LB medium containing Amp at 30°C overnight. 2.5 mL of the cells were added into 600 mL of LB medium and incubated for 5 hours with shaking until O.D.600_{nm} = 0.5. The culture was then equally divided into six flasks and incubated at 30°C. The bacteria in one flask were the control without the induction of arsenite and the cells of the other five flasks were induced by 0.1, 1, 2, 5, and 10 mM of arsenite, respectively. The cells (100 mL) were centrifuged at 5,000 g for 7 min at 4°C in a Beckman ultracentrifuge. The pellet was suspended in 5 mL of NETS buffer (0.1 M NaCl, 10 mM Tris pH8, 1 mM EDTA, 1% SDS). 5 mL of phenol: chloroform: isoamyl alcohol mixture (25:24:1) was added and the mixture was vortexed and centrifuged again for 7 min. The supernatant was transferred to a new tube. 1/10 volume of sodium acetate (3 M, pH 5.4) and 2 volumes of 95% ethanol were added into the tubes that were mixed well and stored at -80°C for 30 min. After centrifugation at 10,000 g for 20 min at 4°C, the pellet was dissolved in 150 μL of sterile d. H_2O. The quality and quantity of extracted RNA was examined on 1.5% agarose gel electrophoresis and by absorbance at 260_{nm}.

7.2.3 Preparation of DIG labeled DNA probes by PCR

To prepare DNA probes for RNA blot hybridization, a PCR DNA labeling kit (Boehringer Mannheim) was used according to the manufacturers protocol. The PCR primers were designed based on the CDB3 *ars* DNA sequence and synthesized by Sigma. The primers used and the target PCR product sizes are listed in Table 7.2. The PCR reactions were performed on a GeneAmp PCR system 9600 (Perkin Elmer). A typical reaction mixture (50 μL) consisted of, 20 pg of DNA template (the same PCR product

without label), 1 μL of each primer (20 μM), 0.2 mM of DIG-dNTP mix (Roche) and 2.5 units of *Taq* DNA polymerase in the manufacturers buffer. After an initial denaturation step consisting of 3 min at 95°C, 30 cycles of denaturation (95°C, 45 sec), annealing (54°C 30 sec) and extension (72°C, 2 min) were performed. The final step cooling to 25°C completed the reaction.

Table 7.2: Primers used and DNA probe sizes

Primer Name	**Primer Sequence 5'-3'**	**Corresponding position of pAR27**	**Upstream (Forward)/ Downstream (Reverse)**
As4-4	TTAACAAGAGTGTCACAG	386-398 bp	Upstream
As4-15	TATCTAGAGCCATCTTTGATTTCTTCC	2300-2285 bp	Downstream
As4-6	AGGACCAGACGTATTGC	2502-2518 bp	Upstream
As4-3	ATGACCCGTTGGCGCTG	3137-3121 bp	Downstream
As4-14	TAGGAACAGGACCAGC	4483-4499 bp	Upstream
CDB3 P-R	TTATACATTACAAACGTTTCAATAC	6159-6136 bp	Downstream

To test the specificity of the DNA probes, Southern-blot analysis were performed. The genomic DNA isolated from CDB3 and plasmid AR27 were subjected to *Eco*RI and *Hin*dIII restriction enzyme digestion and subsequent DNA blotting and hybridization was performed as described in section 3.2.5.

7.2.4 Northern-blot analysis

Agarose gel (1.4% final concentration) was prepared in 1 × MOPS buffer (20 mM MOPS, 5 mM sodium acetate, 1 mM EDTA, pH 7.0). The mixture was heated in the microwave to melt agarose and then after cooling the mixture to about 65°C, formaldehyde (36% v/v) was added to a final concentration of 6% before pouring. The RNA samples (5-10 μg in 10 μL) and RNA size markers (Promega), were each mixed

with 15 μL of formamide, 5.25 μL of formaldehyde, 3 μL of 10 × MOPS buffer and 1.5 μL of 1 mg/L ethidium bromide and then heated at 65°C for 5 min followed by chilling on ice. The denatured samples with 10x loading dye (3 μL) were loaded into the wells of the gel, which was then run at 60 V for 4 hours. After the electrophoresis, the formaldehyde gel was viewed under UV light and the bands of RNA markers were dotted with India ink.

RNA was transferred from gel to a nylon membrane (Hybond^{TM}-N^{+}, Amersham) by the capillary transfer method (Sambrook et. al., 1989) using 20 × SSC (175.3g NaCl and 88.2g sodium citrate in 1 L H_2O) as transfer solution. After blotting overnight, the RNA on nylon membrane was UV cross-linked in a UV Stratalinker 1800 (Stratagene) for 4 × 120 sec. The membrane was then soaked in 50 mM NaOH for 10 min followed by neutralization in 2 × SSC for 1 min. The blotted membrane was used for hybridization immediately or kept between filter papers at room temperature for later use.

The nylon membrane was placed in a hybridization bottle (with DNA side facing inward) containing pre-hybridization solution (1 mL/10 cm^2, DIG-Easy-Hyb, Roche). The membrane was pre-hybridized at 45°C for 1 hour. The probe was heat-denatured in a boiling water bath for 10 min and chilled quickly in an ice bath to prevent re-annealing of the denatured strands. The denatured probe was diluted in a small volume of hybridization solution that was just enough to run over the membrane in the bottle during rotation and was added after the pre-hybridization solution in the bottle was poured out. After at least 14 hours hybridization, the membrane was washed in 2 × SSC twice at room temperature, 15 min for each time. Then, the membrane was washed in 0.2 × SSC

containing 0.1% SDS twice at 68°C, 15 min per wash. The DIG label detection kit (Roche) was also used to illuminate the labeled RNA on blots (Section 3.2.5).

7.2.5 β-Galactosidase assay

Reporter gene analysis was carried out using *E. coli* strains harbouring several *lacZ* fusion plasmids constructed by former members of our laboratory, Z. Xu and C. Su (Su, 2004). For each strain to be assayed, a single colony was taken from a LB plate and inoculated into LB liquid medium (with appropriate antibiotic) and incubated with shaking overnight at 37°C. The overnight cultures were diluted 100 times with fresh LB medium with appropriate antibiotic and incubated until the cultures reached logarithmic growth phase ($O.D._{600}$ = 0.28-0.70). The cultures were then distributed into tubes, and divided into several groups: one group was a control with no arsenite, while the other groups were induced by different concentrations of arsenite. After incubation for an additional 1-2 hours, the cultures were placed on ice for 20 min to prevent further growth. The cell density was measured with a spectrophotometer ($O.D._{600nm}$) and then 0.1 mL of culture was mixed in a tube with 0.9 mL of Z buffer, 60 μL of chloroform and 30 μL of 0.1% SDS solution followed by vortexing for 10 seconds. After placing the tubes in a 28°C water bath for 5 min, the reaction was started by adding 0.2 mL of 4 mg/mL ONPG (*o*-nitrophenyl-β-D-galactopyranoside) to each tube and shaking the tubes for a few seconds. The tubes were incubated at 28°C until sufficient yellow color was observed. The reaction was stopped by adding 0.5 mL of a 1 M Na_2CO_3 solution and total assay time was recorded for each sample, the tubes were centrifuged at 4000 g for 5 min. Finally, the absorbance of each assay tube at both 420 nm and 550 nm was measured by using UV-spectrophotometer. The "Miller Units" of β-galactosidase activity was

calculated (Miller, 1972). The β-galactosidase assays were performed in triplicate in each experiment, and all the experiments were carried out at least four times. A schematic illustration of the various steps followed is presented in Appendix 7.

7.3 Results

7.3.1 RNA extraction

Total RNA extracted from *Bacillus* sp. CDB3 cells was first examined to check the quality and quantity. The results (Figure 7.1) on normal agarose gel showed that the quality of the RNA sample was satisfactory for Northern blot analysis.

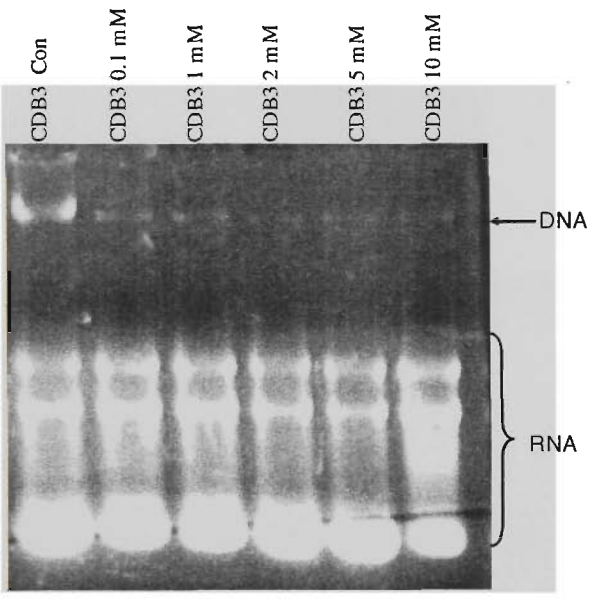


Figure 7.1: Ethidium bromide stained 1.5% agarose gel of total RNA preparations. Total RNA prepared from CDB3 cells were treated for 10 min with arsenite at different concentrations as indicated before harvesting for RNA extraction.

The extracted RNA samples were then subjected to 1.4% denaturing formaldehyde gel electrophoresis for Northern hybridization. A typical gel image is shown in Figure 7.2. Several gels were run and further blotted onto a positively charged nylon membrane for later hybridization using three different probes.

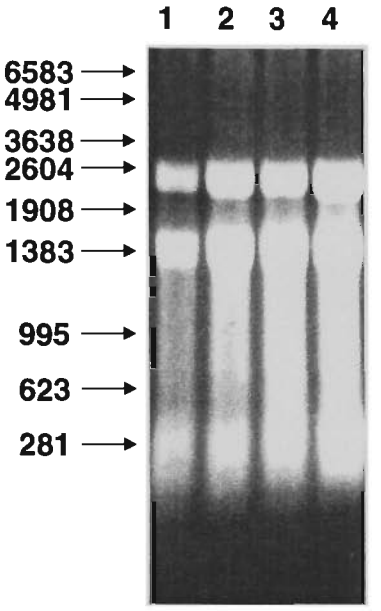


Figure 7.2: Image of an ethidium bromide stained 1.4% denaturing formaldehyde gel of total RNA extracted from *Bacillus* sp. CDB3. Lane 1: no arsenite control, Lanes 2 to 4: 0.1, 1 and 10 mM arsenite induced samples. RNA marker sizes are indicated with arrow on left.

7.3.2 Probe preparation

Three probes were selected for RNA hybridization: a 1.9 kb DNA fragment covering part of *arsRYC*, a 0.6 kb PCR DNA fragment covering the *arsDA* and a 1.7 kb PCR DNA fragment covering the *arsT* and *arsP* amplified using a DIG-UTP labeling PCR kit (Figure 7.3). These probes were designed to detect either a RNA transcript of 7 kb if the CDB3 *ars* gene cluster is transcribed together as a single transcript and/or shorter RNA transcripts to the predicted transcription stop between *arsC* and *arsD* (see above) and some other regulatory mechanisms that exist. The relative position of each of the three probes and sizes of the predicted mRNA transcript are also shown in Figure 7.3.

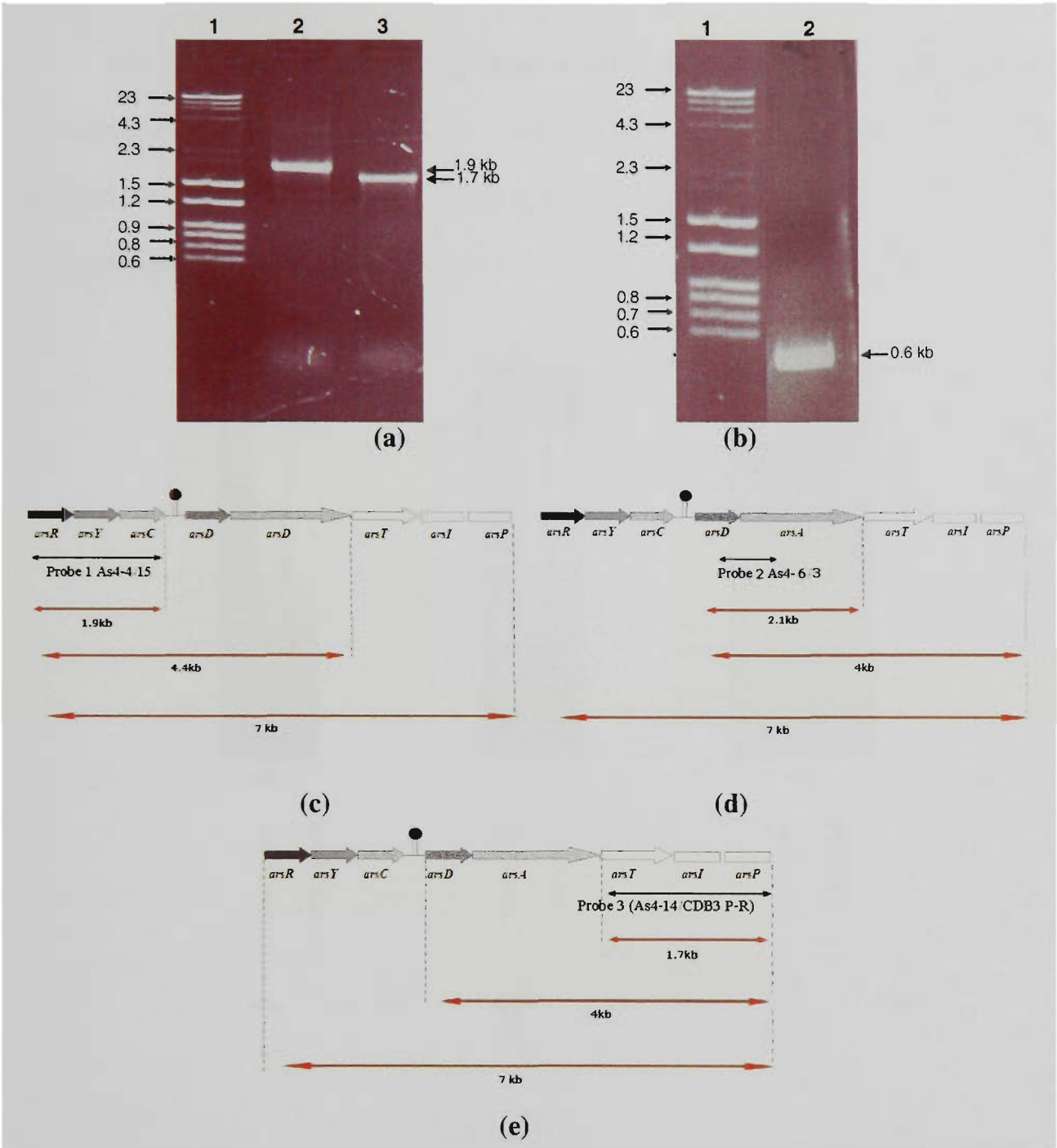


Figure 7.3: Agarose gel (2%) electrophoresis image of PCR amplified DIG labeled DNA fragments. (a): Lane 1: DNA marker; Lane 2: probe 1 (As4-4/15) showing a 1.9 kb DNA fragment covering *arsRYC* and Lane 3: probe 3 (As4-14/CDB3 P-R) showing 1.7 kb DNA fragment covering *arsTIP*. (b): Lane 1: DNA marker; Lane 2: probe 2 (As4-6/3) showing 0.636 kb DNA fragment covering *arsD* and part of *arsA*. The respective sizes are indicated by arrows. (c, d, e) Structure of *ars* cluster and the relative positions of each of the three probes used with the predicted sizes are indicated, the length of the mRNA transcript are indicated by arrows.

The specificity of each of the three probes was confirmed by Southern blot hybridization on CDB3 genomic DNA (Probe 1 and 2) and plasmid AR27 (probe 3) digested with *Eco*RI and *Hin*dIII restriction enzymes (Figure 7.4). The DNA restriction map and expected hybridization band sizes by each probe are illustrated in Figure 7.5.

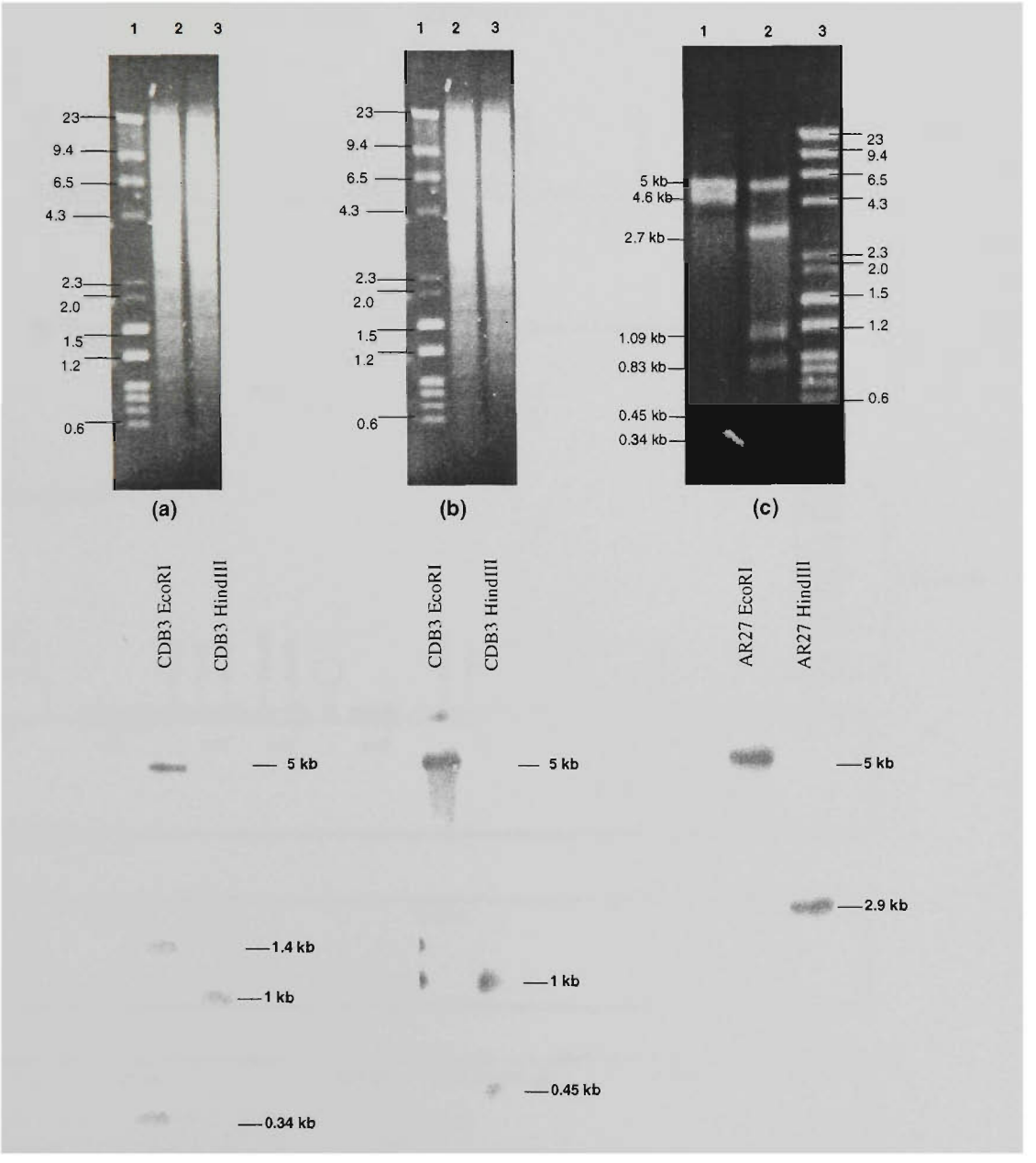


Figure 7.4: Southern blot analysis of the specificity of the three DNA probes. (a, b, c) Ethidium bromide stained 0.8 % TAE agarose gel image showing CDB3 chromosomal DNA digested with *Eco*RI and *Hin*dIII (Lane 2, 3 (a, b); (c) Plasmid AR27 digested with *Eco*RI and *Hin*dIII. Marker sizes in kb are indicated with arrow on left (Lane 1 a, b, and Lane 3-c). (d, e) X-ray film showing specific bands corresponding to CDB3 *ars* gene cluster 1 after hybridization with probes 1 (As4-4/15), probe 2 (As4-6/3) and (f) X-ray film showing specific bands corresponding to AR27 hybridized with probe 3 (CDB3 14/P-R) respectively.

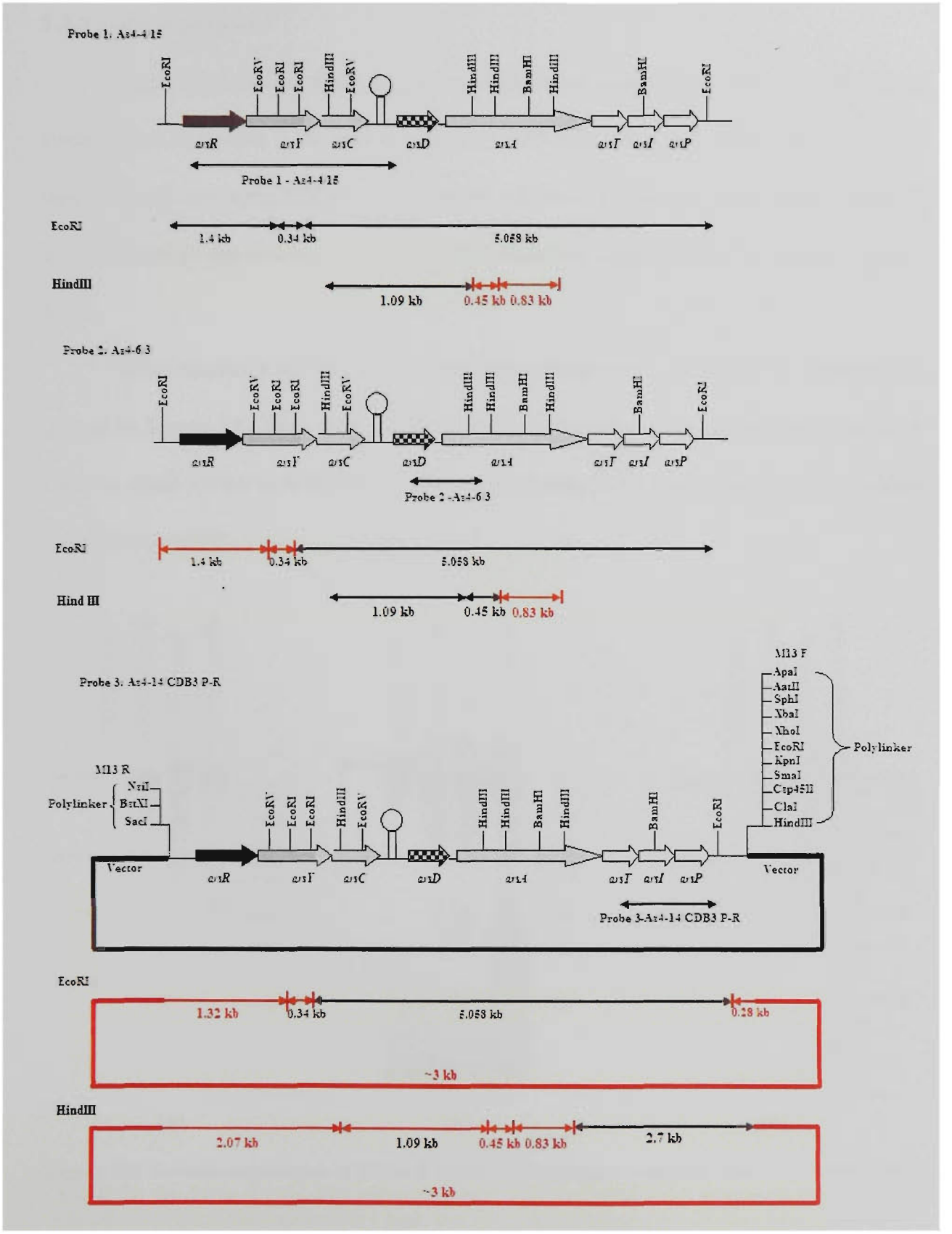


Figure 7.5: Restriction enzyme map of CDB3 *ars* gene cluster 1. The relative positions of the probes and the corresponding fragment sizes expected is shown by arrows. The arrows denoted in red refers to fragments that will not hybridize, while arrows denoted in black are the expected fragment sizes after hybridization with respective probes. The box represents the vector pGEM7Zf(+).

7.3.3 mRNA analysis

When three similar RNA blots were hybridized with three probes, (Figure 7.2), a strong band of around 7 kb was detected on each blot. This was about the size of the whole cluster and remarkably induced by the addition of arsenite to the culture medium as evidenced by the absence or very low hybridization signals in the no arsenite control lanes.

Some smaller RNA bands also appeared hybridized by the probes as indicated by arrows in Figure 7.6. They were either alternative transcripts or degradation products. A band at about 1.9 kb hybridized to probe 1 was particularly interesting since it matched the size of *arsRYC*.

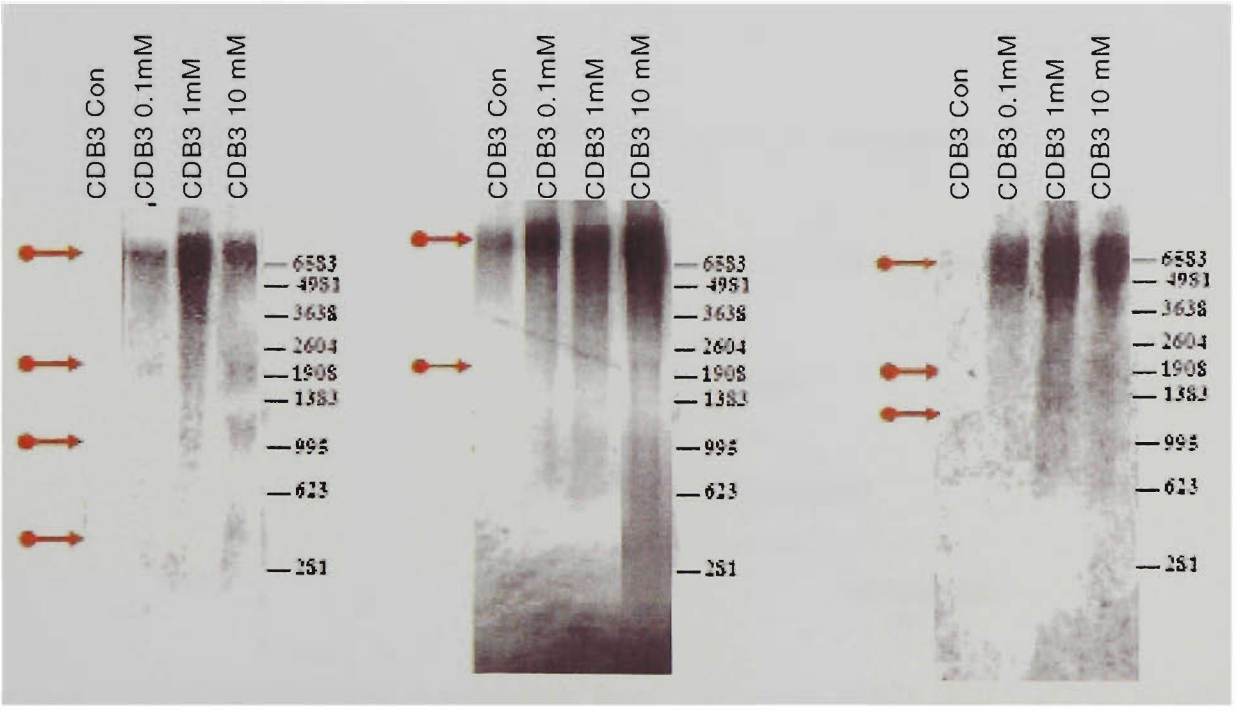


Figure 7.6: Northern blot analysis of CDB3 *ars* RNA blots hybridized with three DNA probes. (a): Probe 1 (As4-4/15). (b): Probe 2 (As4-6/3) and (c): Probe 3 (As4-14/CDB3 P-R). Hybridization signals were detected on the X-ray film exposed for 1 hour. The sizes of the transcripts are marked with arrows.

7.3.4 Reporter gene assay

To examine if other promoters are present in the *ars* gene cluster reporter gene assays were carried out with different *lacZ* fusion constructs (Figure 7.7) at arsenite concentrations higher than previously used.

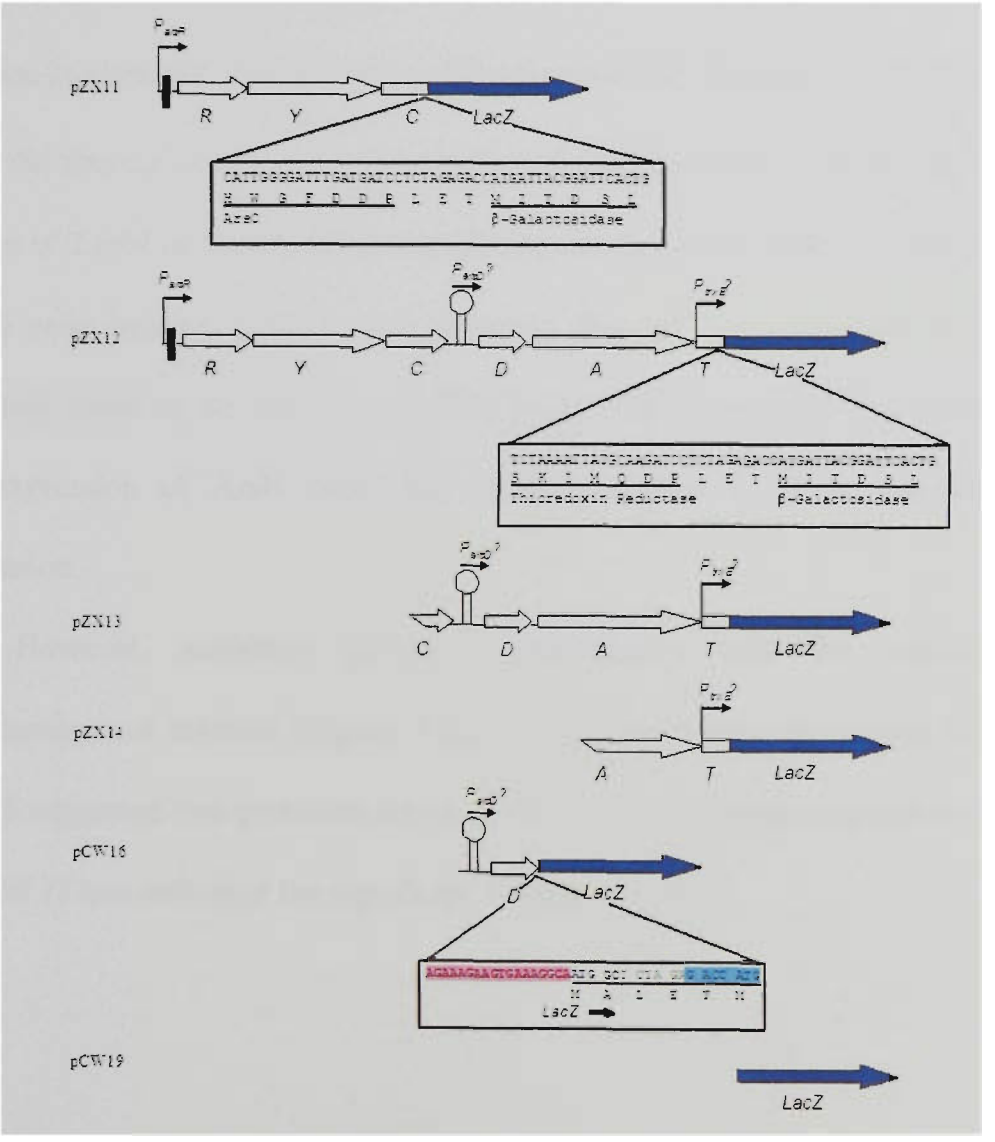
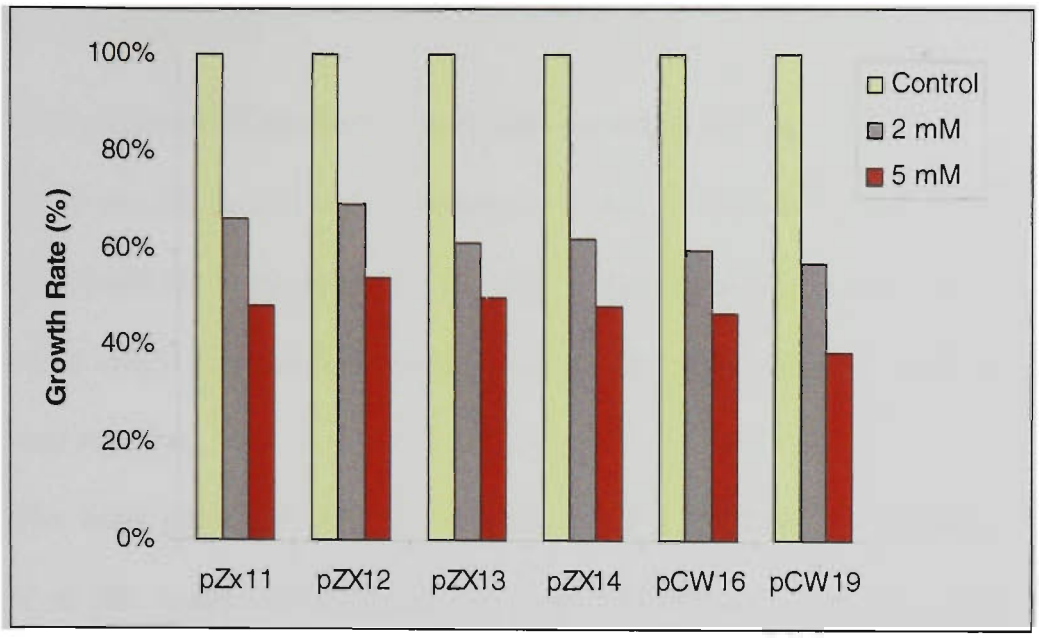


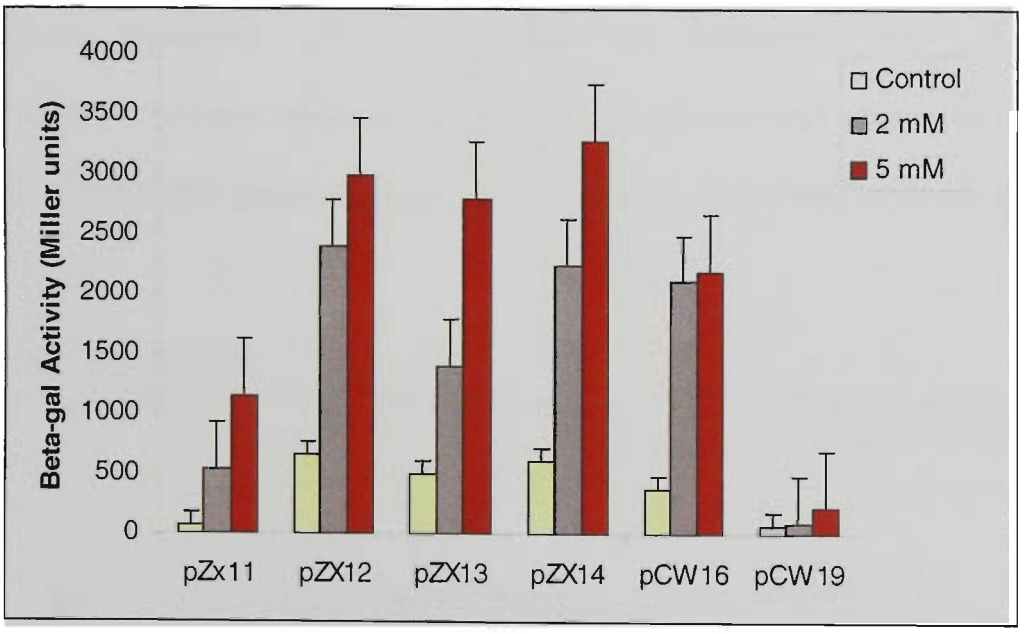
Figure 7.7: The insert contained in translational fusion constructs of pZX11, pZX12, pZX13, pZX14, pCW16 and pCW19. The putative promoters are shown by small arrows. The DNA sequences and amino acid sequences of the joint regions of the fusion proteins are indicated. The pZX11 contains intact *arsR* and *arsY* genes, and a truncated *arsC* gene, in which the *lacZ* coding region was fused in frame to the 67th codon of *ars*C. The pZX12 contains intact *arsRYCDA* genes, pZX13 contains a truncated *arsC* and intact *arsDA* genes, and pZX14 only contains part of the *arsA* gene. In these three plasmids, pZX12, pZX13 and pZX14, the *lacZ* coding region was fused, in frame, to the 165th codon of *arsT*. pCW16 contains intact *arsD* gene, the non-coding regions were indicated in red, the *lacZ* coding region were indicated in blue, while the blue ATG showed the initial code of *arsA*. Restriction mapping confirmed the four plasmids and were transformed into *E. coli* JM109, to measure their growth and the β-galactosidase activities. [From C. Su (2004)].

Figure 7.8a shows the percentage growth of *E. coli* JM109 lines transformed with different constructs. They were similar except that the pZX11 and pZX12 possessing cells showed higher arsenite resistance because of *yqcL* function. The negative control (pCW19) cells did not show significant β-galactosidase activities with or without arsenite induction as expected. The activities of β-galactosidase of all the five test constructs were low in the absence of arsenite (T-test indicated the significant P-value < 0.006). With the addition of 2 mM or 5 mM AsIII, a small increase in the activity of β-galactosidase in the *E. coli* cells bearing pZX11 was observed (Figure 7.8b). Maximal expression was previously noted by Su, (2004), with 0.01 to 0.1 mM of arsenite. This indicated that the over expression of ArsB under the strong promoter R (P_{arsR}) resulted in mRNA degradation.

However, activities increased dramatically with the addition of high concentrations of arsenite (Figure 7.8b). The induced activities shown by pZX14 and pCW16 suggested that promoter activities did exist in the upstream regions of both *arsD* and *arsT* (T-test indicated the significant P-value < 0.003).



(a)



(b)

Figure 7.8: **(a)** Cell densities of the *E. coli* JM109 bearing different gene fusion plasmids (pZX11, pZX12, pZX13, pZX14, pCW16 and pCW19) and **(b)** Activities of β-galactosidase in the *E. coli* JM109 carrying different gene fusion plasmids (pZX11, pZX12, pZX13, pZX14, pCW16 and pCW19) treated with different concentrations of arsenite for 1 hour. All assays were carried out in triplicate and repeated four times.

7.4 Discussion

7.4.1 CDB3 *ars* cluster 1 can be transcribed as a single mRNA

Three specific probes were generated and they all detected a large RNA of about 7 kb on Northern blots suggesting all the eight genes in the CDB3 *ars* cluster 1 can be expressed as single polycistronic transcript. This, to our knowledge, is the largest *ars* operon ever reported.

The work presented here also confirms that inducibilty of expression may be regulated at the transcriptional level. A 7 bp (ATTAAAT) inverted repeat located upstream of the putative –35 element of the *arsR* promoter (Figure 7.9) may be the site for repressor recognition and binding (Sato and Kobayashi, 1998). A 9 bp (AGCCGAGAA) inverted repeat is also present upstream of *arsD* which may act as a regulatory element. The putative contact points upstream of *arsR* and *arsD* are shown in Figure 7.9.

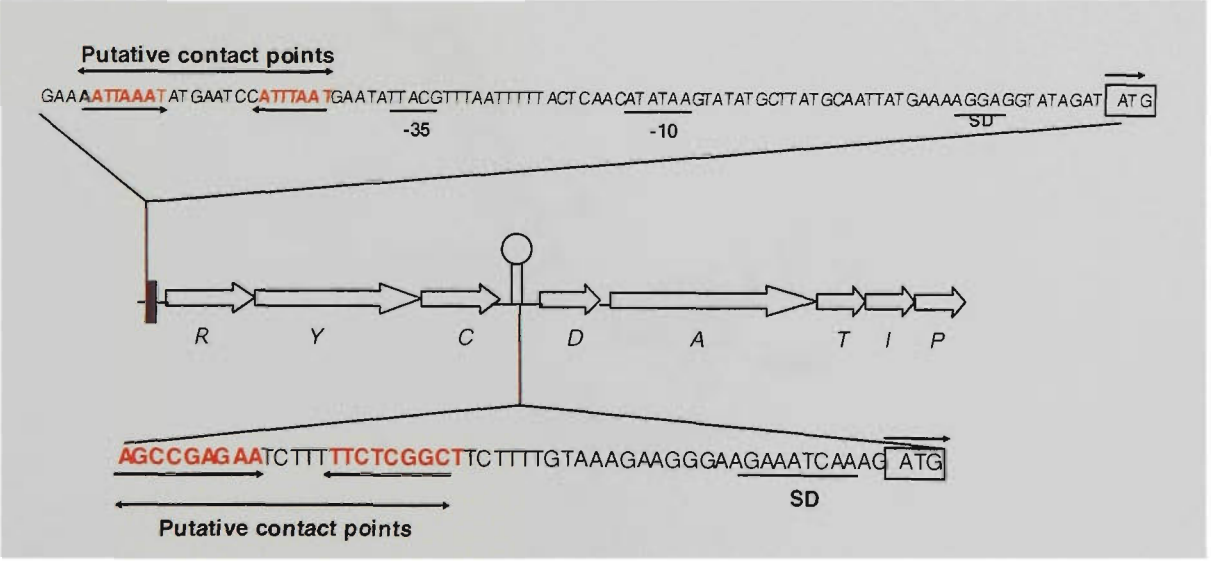


Figure 7.9: The putative contact points of *arsR* and *arsD* promoters.

There were some smaller RNA bands that hybridized to the probes. The 1.9 kb RNA detected by probe 1 matches the size of *arsRYC* supporting the prediction of a RNA terminator site located downstream of *arsC*. However, all these detected small RNA molecules might have been degradation products of the large 7 kb transcript. RNA transcript degradation was also observed in *E. coli* R773 *ars* cluster in which the *arsB* region of the polycistronic transcript decayed more rapidly than the *arsRA* and *arsC* as a result of a secondary structure at the 3' termini (Owolabi and Rosen, 1990).

From the Northern results it was not clear if any other promoter apart from the *arsR* is present. Putative transcription regulatory elements found located between *arsC* and *arsD* shows a 9 bp inverted repeat that can function as a transcription regulatory site or as a terminator site. Figure 7.10 represents the prediction of a potential mRNA secondary structure in the intergenic region between *arsC* and *arsD*.

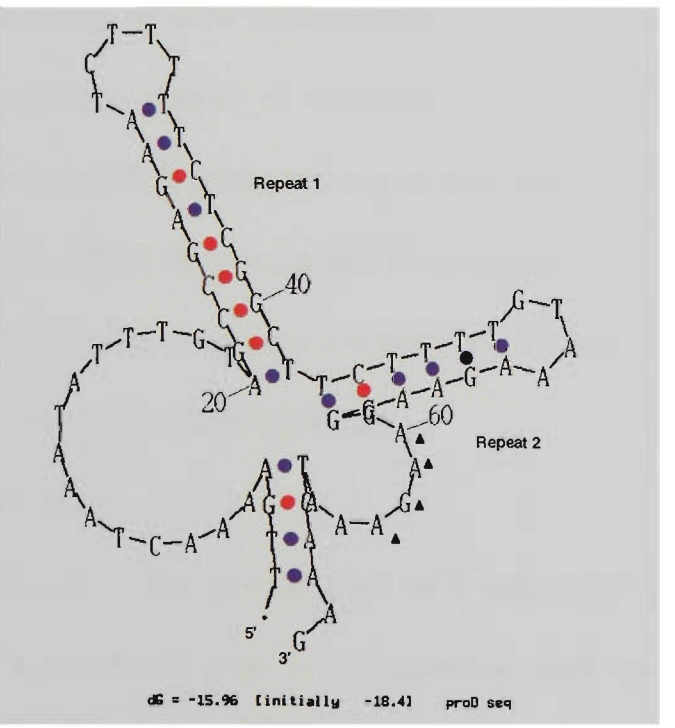


Figure 7.10: Sequence and prediction of a potential mRNA secondary structure in the intergenic region between *arsC* and *arsD*. The putative ribosome binding sequence (SD) is marked with triangles. The free energy formation of the indicated secondary structure is -15.96 kcal mol^{-1} as estimated by RNA *Mst* (Jung et. al., 2006).

The inverted repeat can be a regulatory site that is bound by regulator to inhibit the RNA polymerase to start transcription of *ars* genes. However, this 9 bp inverted repeat followed by T-rich region suggested that it could alternatively function as a termination site (Wang, 2001). When the transcription proceeds to the end of *arsC*, the secondary structure of mRNA could form a stem-loop to stop the transcription (Figure 7.10). It is interesting to find out which or both functions this inverted repeat exerts *in vivo*. Therefore to recheck, a reporter gene method was performed using high arsenite concentrations.

In prokaryotes, genes belonging to the same operon are transcribed as a single mRNA molecule. RNA polymerase binds to the promoter and synthesizes RNA chain until it reaches a transcriptional terminator. Some terminators rely on the presence of the Rho protein, whereas others function independently of Rho (Hoon et. al., 2005). Such Rho independent terminators consist of an inverted repeat followed by a stretch of thymine residues, which is seen in intergenic region between *arsC* and *arsD* of the CDB3 *ars* cluster. Hoon et. al., (2005) have suggested that unlike in *E. coli*, the Rho protein is dispensable in *B. subtilis*, which may have a limited role for Rho-dependent termination in this organism and possibly in other Firmicutes.

7.4.2 Sub operons may exist

Downstream fusion of the reporter gene *lacZ* coding for β-galactosidase to the promoter region of a structural gene is a frequently used technique to study the expression of the gene (Thomas, 2002). Previously, this method was used by C. Su. (2004) of our lab to determine whether other promoter(s) also exist in the CDB3 *ars* cluster, but results were not positive. As presented in Chapter 6 and this chapter, it is

reasonable to assume that a promoter in front of *arsD* is present. Therefore, the reporter gene assay method was employed again in my study to reexamine the hypothesis. This time, higher arsenite concentrations were used since it was assumed that the putative *arsD* promoter is activated at higher doses of inducers.

E. coli JM109 instead of AW3110 was chosen as a suitable host for performing β-galactosidase assay, based on two aspects. Firstly, the *lac-proAB* is deleted from the chromosome of *E. coli* JM109, thus leading to no background expression of *LacZ* compared to AW3110. Secondly, the *arsRBC* operon located on the chromosome of *E. coli* JM109 renders moderate resistance to arsenic, as a result cells could survive in certain concentrations of AsIII which were required for testing the ArsIII induction in my study.

The results (Figure 7.8b) showed that, when *lacZ* was under the control of putative P_{arsD} promoter (pCW16, and pZX13), the activities of β-galactosidase in the cells bearing these plasmids gradually increased in response to arsenite at high concentrations. Together with the ArsD binding feature (Chapter 6), our results clearly demonstrated the promoter activity of *arsD* upstream sequence which is induced by arsenite at high concentrations.

Another reporter gene plasmid, pZX14, was initially designed as a negative control in the assay (Figure 7.8b). Surprisingly, the induced β-galactosidase activities were also detected with this construct at high arsenite doses, suggesting that another possible sub promoter in this *ars* gene cluster 1 also exists. However, bioinformatic analysis using available tools for promoter prediction could not reveal any known promoter elements. While the existence of a promoter in front of each *arsR* and *arsD* was

evident by gel mobility shift assay (Chapter 6), further analysis is necessary to examine the binding activity of *arsT* upstream region.

Chapter 8

Concluding Remarks

Arsenic is widely distributed as a pollutant. The need to develop low cost and friendly technologies for remediation of arsenic contaminated soils and water has stimulated interest in studies on arsenic resistant organisms.

Bioremediation/phytoremediation could be a solution to the contamination problem at these sites as an alternative to currently existing physio-chemical methods. Certain bacteria and plants are capable of removing arsenic from their surroundings and could therefore be used (Takeuchi et. al., 2006). Soil contamination by arsenic and other pollutants represents a major environmental threat in Australia. Little is known about the identities of organisms and their interaction with plants growing on these sites co-contaminated with arsenic/DDT.

In addition to anthropogenic sources, arsenic occurs in soils in some regions naturally from weathering of arsenic-rich mineral deposits. Metabolism of inorganic arsenic to methylated forms by living organisms adds a range of organic arsenic compounds to many soil environments (Cullen and Reimer, 1989). Much of the extensive arsenic metabolism of bacteria and fungi (Cullen and Reimer, 1989) is associated with minimizing the concentration of arsenic within the cells, by arsenic-exporting mechanisms and by improving the specificity of phosphate uptake (Cervantes et. al., 1994). Arsenic is also utilised in metabolism as a terminal electron acceptor in dissimilatory arsenate respiration or as an electron donor in chemoautotrophic arsenite oxidation. The *arsC* gene, which codes for an arsenate reductase, transforming arsenate into arsenite, is essential for resistance by extruding arsenate from the cell (Jackson et. al., 2003). Dhankher et. al. (2002) have generated *Arabidopsis* plants that overexpresses an *E. coli* arsenate reductase ArsC, reducing AsV to AsIII using glutathione as electron

donor. When *A. thaliana* plants were transformed with *arsC* with an additional second transgene encoded by *E. coli* γ-Glutamyl cysteine synthetase (γ-ECS) under a constitutive actin (ACT2) promoter, the *A. thaliana* plant exhibited more tolerance to AsV than the wild type (Dhankher et. al., 2002). Additionally, Ute, 2005 reported that the expression of ArsC in plants also increases Cd (cadmium) tolerance and accumulation suggesting that *arsC* gene may be an ideal candidate for arsenic hyperaccumulation in plants.

The possibility of using arsenic-accumulating plants to extract arsenic from soil for remediation of contaminated land areas such as dip sites has attracted attention. The capability of the fern *Pteris vittata* to accumulate arsenic to high concentrations in its fronds has been of particular interest (Ma et. al., 2001): this capability is at least partly shared by some grasses such as *Agrostis tenuis* (Porter and Peterson, 1977). The mechanism of accumulation is poorly understood, but is mediated by rhizosphere micro organisms (Walter and Wenzel, 2002; Liu et. al., 2005). The micro organisms in soils from sites with long-term arsenic contamination may be particularly active in mediating the effect of arsenic on the plants growing on those sites.

The thesis presented is the first instance to critically examine some bacterial strains capable of resisting high concentrations of arsenic at the dip sites. Other work by our group (Chopra et. al., 2007), has illustrated that the interaction of rhizosphere microbes with plants can increase uptake of arsenic, which shows the importance of understanding plant–soil microbial interactions for developing future plant engineering strategies with bacterial *ars* genes. This phytoremediation-based approach utilizing hyperaccumulating plants to remove arsenic from soil at dip sites will be extremely beneficial as one of the mechanisms to eradicate arsenic.

Bacterial strains from cattle dip sites have provided valuable resources for research and potential application. As a result of harsh conditions at the sites and the use of so many toxicants over a long period of time, organisms that survived there might have evolved distinctive resistance mechanisms. The five isolated rhizosphere bacterial strains showed a range of resistance to arsenate and arsenite (Section 2.3.8).

After bacterial identification, further work carried out in this thesis was to study the gene function and mechanism of arsenic resistance of bacterial strain CDB3 isolated from the dip-sites; two clusters already identified makes this strain distinctive from all other bacterial strains characterized so far in this regard. The arsenical resistance cluster 1 cloned from *Bacillus* sp. CDB3 displays a number of novel features including its largest gene number of all the known *ars* operons, the unique organization of the common *arsR, B(Y), C, D and A* genes, the coupling of an ATPase with YqcL it encodes, the addition of three more genes *arsTIP* with *arsIP* being first time identified in an *ars* operon, and finally, the yet-to-be fully elucidated complicated regulatory mechanism.

Compared to other characterized bacterial *ars* clusters, CDB3 *ars* gene cluster 1 appears to have a complicated yet novel regulatory mechanism for expression. It may be expressed as a single polycistronic transcript as well as two or even three small transcripts. Apart from the upfront main promoter which exists in all the other *ars* operons, at least one other promoter (or sub-promoter) is present. The second repressor, ArsD, has been demonstrated to bind not only to the main promoter but also to another site in second promoter. While much bioinformatic and experimental evidence is yet to be obtained to elucidate the mechanism of regulation, a hypothetical model can be proposed at this stage (Figure 8.1). ArsR is the main repressor which regulates the

expression of the operon at low arsenic concentrations and ArsD controls the upper level of expression as well as the latter three genes *arsTIP*. At low concentration of arsenite, *arsRYC* is dominantly expressed and it is sufficient to resist the low-medium concentrations of arsenic. RNA polymerase will release when it is halted by ArsD. When the concentration of arsenite further increases, the ArsD repressor is released from the DNA as a result transcription resumes, probably from two positions. A surprisingly interesting result from the reporter gene analysis pointed to a third possible promoter in front of *arsT*, although no obvious promoter sequence features are identified.

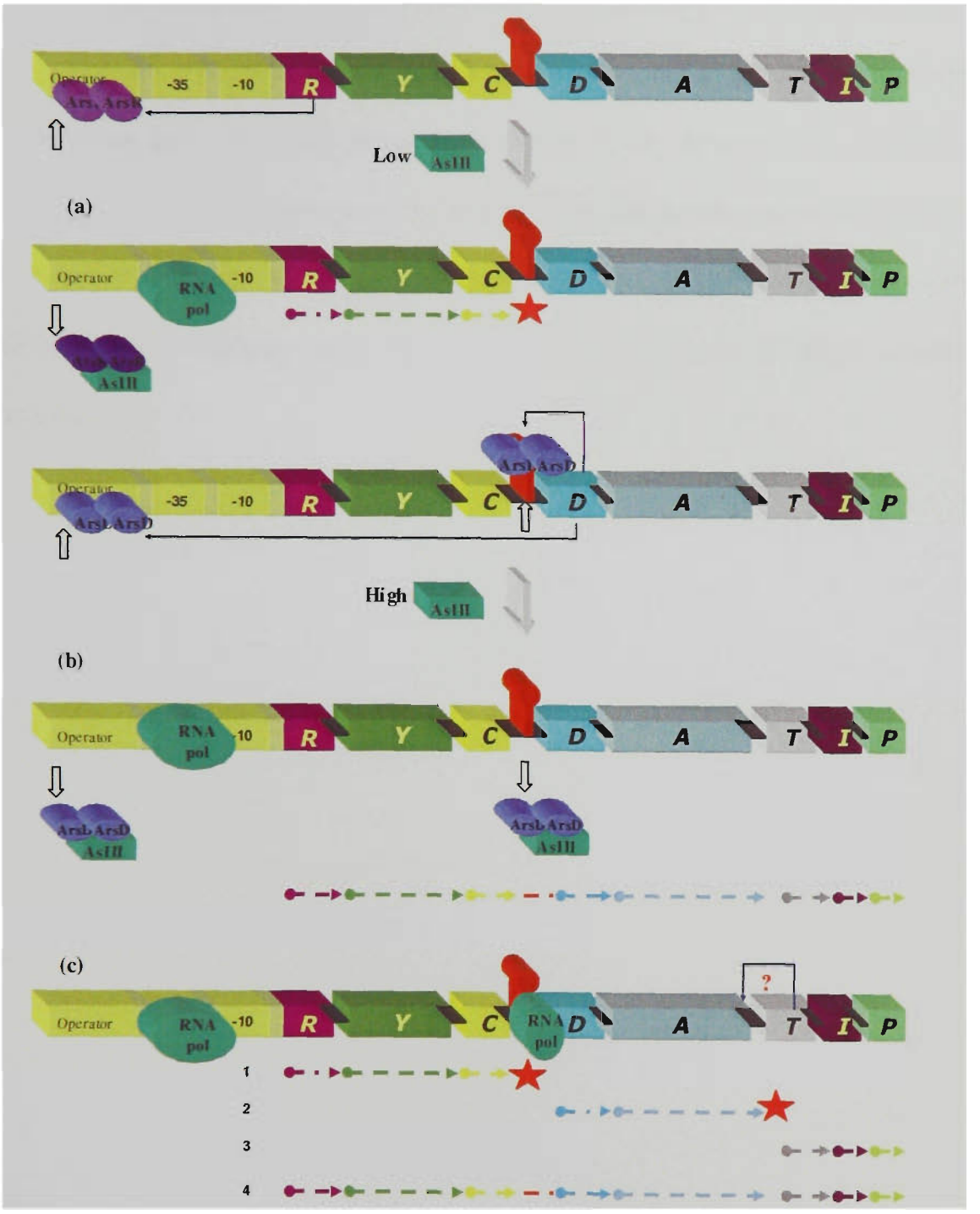


Figure 8.1: Proposed model for the regulation of CDB3 *ars* gene cluster 1. **(a)** In normal circumstances, a low level of ArsR is produced in the absence of inducers and binds to the front operator/promoter site of the *ars* operon to repress the transcription. At low concentrations of arsenite, ArsR binds to arsenite and changes the conformation of ArsR which then dissociates from the DNA. RNA polymerase binds to the promoter, resulting in transcription of the *ars* cluster. With the RNA terminator in front of *arsD*, the first three genes *arsRYC* are more abundantly expressed. **(b)** When the concentration of ArsD increases sufficiently, ArsD binds to the operator/promoter region at two positions upstream of ArsR/D and inhibits the transcription. At high concentrations of arsenite, ArsD binds to arsenite and dissociates from the DNA. RNA polymerase binds to the promoter and transcription starts again. **(c)** Possible transcripts of the CDB3 *ars* cluster 1 expression. (1): Short-transcript covering the *arsRYC*; (2): short-transcript covering the *arsDA*; (3): short-transcript covering the *arsTIP;* and (4): a single polycistronic transcript covering all eight genes. The involvement of possible *arsT* promoter is also illustrated.

Much work is still left to be carried out to further study this novel *ars* cluster. At the same time the other CDB bacteria also await to be characterized to reveal their mystery for being able to survive in the highly toxic arsenic contaminated environment. This research will lead us to new insights into the cell physiology of bacteria in combating arsenic toxicity stress and also in the development of novel remediation technologies.

BIBLIOGRAPHY

Abedin, M.J., Feldmann, J. and Meharg, A.A. 2002. Uptake kinetics of arsenic species in rice plants. Plant Physiology. 128:1120-1128.

Aggett, J. and O'Brien, G.A. 1985. Detailed model for the mobility of arsenic in lacustrine sediments based on measurements in Lake Ohakuri. Environ. Sci. Technol. 19: 231-238.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J. 1990. Basic local alignment search tool. J Mol Biol 215:403-10.

Anderson, G.L., Williams, J. and Hille, R. 1992. The purification and characterization of arsenite oxidase from *Alcaligenes faecalis*, a molybdenum-containing hydroxylase. J. Biol. Chem. 267:23674-23682.

Azcue, J and Niragu, J. 1994. Arsenic: historical perspective. In *Arsenic in the Environment.* 1. Cycling and Characterization; Niragu J. O., Ed; Wiley; New York; 15.

Bennett, M.S., Guan, Z., Laurberg, M. and Dong Su, X. 2001. *Bacillus subtilis* arsenate reductase is structurally and functionally similar to low molecular weight protein tyrosine phosphatases. PNAS. 98:13577-13582.

Bhattacharya, D., Sharma, P.M., Krishnan, S., Mishra, S. and Lal, B. 2003. Evaluation of genetic diversity among *Pseudomonas citronellolis* strains isolated from oily sludge contaminated sites, Appl Environ Microbiol. 69: 1435-1441.

Bhattacharjee, H., Li, J., Ksenzenko, M.Y. and Rosen, B.P. 1995. Role of cysteinyl residues in metalloactivation of the oxyanion-translocating ArsA ATPase. J. Biol. Chem. 270:11245-50.

Bhattacharjee, H. and Rosen, B. P. 1996. Spatial Proximity of Cys^{113}, Cys^{172}, and Cys^{422} in the Metalloactivation Domain of the ArsA ATPase. J. Biol. Chem. 271:24465-24470.

Bhattacharjee, H., Zhou, T., Li, J., Gatti, D.L., Walmsley, A.R. and Rosen, B.P. 2000. Structure-function relationships in an anion-translocating ATPase. Biochem. Soc. Trans. 28:520-6.

Bhattacharjee, H. and Rosen, B.P. 2001. Structure-function analysis of the ArsA ATPase: contribution of histidine residues. J. Bioenerg. Biomembr. 33:459-68.

Bobrowicz, P., Wysocke, R., Owsianik, G., Goffeau, A. and Ulaszewski. 1997. Isolation of three contiguous genes, ACR1, ACR2, and ACR3, involved in resistance to arsenic compounds in the yeast *Saccharomyces cerevisise*. Yeast. 13:819-828.

Booth, J. and Guidotti, G. 1997. Phosphate transport in yeast vacuoles. J. Biol. Chem. 272:20408-20413.

Bizily, S.P., Rugh, C.L., Summers, A.O. and Meagher, R.B. 1999. Phytoremediation of methylmercury population: merB expression in *Arabidopsis thaliana* confers resistance to organomercurials. Proc. Natl. Acd. Sci. USA. 96:6808-6813.

Broer, S., Ji, G., Broer, A. and Silver, S. 1993. Arsenic efflux governed by the arsenic resistance determinant of *Staphylococcus aureus* Plasmid pI258. J. Bacteriol. 175: 3480-3485.

Bruhn, D., Li, J., Silver, S., Roberto, F. and Rosen, B.P. 1996. The arsenical resistance operon of IncN plasmid R46. FEMS. Microbiol. Lett. 139:149-153.

Bull, D.C. 2000. The chemistry of chromated copper arsenate. I. Characterization of timber treatment plant chemical sludge. Wood. Sci. Technol. 34:367-376.

Bull, D.C. and Harland, P.W. 2001. The Chemistry of Chromated Copper Arsenate III. Recovery of Arsenate Content from Treatment Plant Sludge. Wood. Sci. Technol. 35: 137-141.

Butcher, B.G., Deane, S.M. and Rawlings, D.E. 2000. The chromosomal arsenic resistance genes of *Thiobacillus ferrooxidans* have an unusual arrangement and confer increased arsenic and antimony resistance to *Escherichia coli.* Appl. Envir. Microbiol. 66:1826-1833.

Butcher, B.G. and Rawlings, D.E. 2002. The divergent chromosomal *ars* operon of *Acidithiobacillus ferrooxidans* is regulated by an atypical ArsR protein. Microbiology. 148:3983-3992.

Cai, J., Salmon, K. and Dubow, M.S. 1998. A chromosomal *ars* operon homologue of *Pseudomonas aeruginosa* confers increased resistance to arsenic and antimony in *Escherichia coli.* Microbiology. 144:2705-2713.

Carlin, A., Shi, W., Dey, S. and Rosen, B.P. 1995. The *ars* operon of *Escherichia coli* confers arsenical and antimonal resistance. J. Baceriol. 177:981-986.

Cervantes, C., Ji, G., Ramirez, J. and Silver, S. 1994. Resistance to arsenic compounds in micro organisms. FEMS. Microbiol. Rev. 15:355-367.

Chen, Y., Dey, S. and Rosen, B.P. 1996. Soft metal thiol chemistry is not involved in the transport of AsIII by the Ars prump. J. Bacteriol. 178:911-913.

Chen, C.M., Misra, T.K., Silver, S. and Rosen, B.P. 1986. Nucleotide sequence of the structural genes for an anion pump. The plasmid-encoded arsenical resistance operon. J. Biol. Chem. 261:15030-15038.

Chen, Y. and Rosen, B.P. 1997. Metalloregulatory properties of the ArsD repressor. J. Biol. Chem. 272:14257-14262.

Chopra, B., Bhat, S., Mikheenko, I., Xu, Z., Yang, Y., Luo, X., Chen, H., Van Zwieten, L., Lilley, R.M. and Zhang, R. 2007. The Characteristics of Rhizosphere Microbes and Their Interaction with Plants in Arsenic-Contaminated Soils from Cattle Dip Sites. Science of Total Environment. 378:331-342.

Cullen, W.R. and Reimer, K.J. 1989. Environmental arsenic chemistry. Chem. Rev. 89: 713-764.

Das, D., Samanta, G., Mandal, B.K., Chowdhury, T.R., Chanda, C.R., Chowdhury, P.R., Basu, G.K. and Chakraborti, D. 1996. Arsenic in groundwaters in six districts of West Bengal; India. Environ. Geochem. Health. 18:5-15.

Dey, S., Dou, D. and Rosen, B.P. 1995. ATP-dependent arsenite transport in everted membrane vesicles of *Escherichia coli.* J. Biol. Chem. 269:25442-25446.

Dey, S., and Rosen, B.P. 1995. Dual mode of energy coupling by the oxanion-translocating ArsB protein. J. Bacteriol. 177:385-389.

Dey, S., Dou, D., Tisa, L.S. and Rosen, B.P. 1994. Interaction of the catalytic and the membrane subunits of an oxyanion-translocating ATPase. Arch. Biochem. Biophys. 311:418-24.

Demel, S., Shi, J., Martin, P., Rosen, B.P and Edwards, B. 2004. Arginine 60 in the ArsC arsenate reductase of *E. coli* plasmid R773 determines the chemical nature of the bound As(III) product. Protein. Sci. 13:2330-2340.

Dhanker, O.P., Li, Y., Rosen, B.P., Shi, J., Salt, D., Senecoff, J.F., Sashti, N.A. and Meagher, R.B. 2002. Engineering tolerance and hyperaccumulation of arsenic in plants by combining arsenate reductase and γ-glutamylcysteine synthetase expression. *Nature Biotech.* 20:1140-1145.

Dhruva B., Priyangshu, M.S., Krishnan, S., Sanjeet, M. and Banwari, L. 2003. Evaluation of genetic diversity among *Pseudomonas citronellolis* strains isolated from Oily sludge contaminated sites, Appl. & Environ. Microbiol. 1435-1441.

Dopson, M., Lindstrom, E. and Hallberg, K. 2001. Chromosomally encoded arsenical resistance of the moderately thermophilic acidophile *Acidithiobacillus caldus.* Extremophiles. 5:247-255.

Edvantoro, B.B., Naidu, R., Megharaj, M. and Singleton, I. 2003. Changes in microbial properties associated with long-term arsenic and DDT contaminated soils at disused cattle dip sites. Ecotoxicology and Environmental Safety. 55: 344- 351.

Edvantoro, B.B., Naidu, R., Megharaj, M., Merrington, G. and Singleton, I. 2004. Microbial formation of volatile arsenic in cattle dip site soils contaminated with arsenic and DDT. Applied Soil Ecology. 25:207- 217.

Felsenstein, J. 1989. PHYLIP -- Phylogeny Inference Package (Version 3.2). Cladistics, 5: 164-166.

Francesconi, K. A. and Kuehnelt, D. 2002. Arsenic Compounds in the Environment. Frankenberger, W. T. Jr. (ed.), In: Environmental Chemistry of Arsenic. Marcel Dekker, New York. 51-94.

Gladysheva, T.B., Oden, K.L. and Rosen, B.P. 1994. Properties of the arsenate reductase of plasmid R773. Biochemistry. 33:7288–7293.

Ghosh, M., Shen, J. and Rosen, B.P. 1999. Pathways of AsIII detoxification in *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. USA. 96:5001-5006.

Gonzalez-Chavez, C., Harris, P.J., Dodd, J. and Meharg, A.A. 2002. Arbuscular mycorrhizal fungi confer enhanced arsenate resistance on *Holcus lanatus*. New Phytologist. 155: 163-171.

Golovin, A., Dimitropoulos, D., Oldfield, T., Rachedi, A. and Henrick, K. 2005. MSDsite: A Database Search and Retrieval System for the Analysis and Viewing of Bound Ligands and Active Sites. PROTEINS: Structure, Function, and Bioinformatics 58:190-9.

Groot, D.P., Deane, S.M. and Rawlings, D.E. 2003. A transposon-located arsenic resistance mechanism from a strain of *Acidithiobacillus caldus* isolated from commercial, arsenopyrite biooxidation tanks. Hydrometallurgy. 71:115-123.

Guo, X., Li, Y., Peng, K., Hu, Y., Li, C., Xia, B. and Jin, C. 2005. Solution structures and backbone dynamics of arsenate reductase from *Bacillus subtilis*. J. Biol. Chem. 47:39601-39608.

Harrington, J.M., Laforce, M.J., Rember, W.C., Fendorf, S.E. and Rosenzweig, R.F. 1998. Phase associations and mobilization of iron and trace metals in sediments of Lake Coeur d'Alene. Environ. Sci. Technol. 32:650-666.

Hamadeh, H.K., Vargas, M., Lee, E. and Menzel, D.B. 1999. Arsenic distrupts cellular levels of p53 and mdm2: A potential mechanism of carcinogenesis. Biochemical and Biophysical research communications. 263:446-449.

Herz, K., Vimont, S., Padan, E. and Berche, P. 2003. Roles of NhaA, NhaB and NhaD Na^+/H^+ antiporters in survival of *Vibrio cholerae* in a saline environment. J. Bacteriol. 1236-1244.

Hoon, M.J.L., Makita, Y., Nakai, K. and Miyano, S. 2005. Prediction of Transcriptional Terminators in *Bacillus subtilis* and Related Species. PLoS computational Biology. 1:0210-0227.

Hsu, C.M. and Rosen, B.P. 1989. Characterization of the catalytic subunit of an anion pump. J. Biol. Chem. 264:17349-17354.

Hughes, M.P. 2002. Arsenic toxicity and potential mechanisms of action. Toxicology letters. 133:1-16.

Ilyaletdinov, A.N. and Abdrashitova, S.A. 1981. Autotrophic oxidation of arsenic by a culture of *Pseudomonas arsenitoxidans*. Microbiology. 50:197-204.

Jackson, C.R. and Dugas, S.L. 2003. Phylogenetic analysis of bacterial and archaeal *arsC* gene sequences suggests an ancient, common origin for arsenate reductase. BMC. Evol. Biol. 3:18.

Jackson, C.R., Jackson, E.F., Dugas, S.L., Gamble, K. and Williams, S.E. 2003. "Microbial transformations of arsenite and arsenate in natural environments." Recent research developments in Microbiology. 7:103-118.

Jensen, M.A., Webster, J.A. and Straus, N. 1993. Rapid identification of bacteria on the basis of PCR-amplified rDNA spacer polymorphism, Appl. & Environ. Microbiol. 62:3818-3825.

Ji, G. and Silver, S. 1992a. Reduction of arsenate to arsenite by the ArsC protein of the arsenic resistance operon of *Staphylococcus aureus* plasmid pI258. PNAS. 89:9474-9478.

Ji, G. and Silver, S. 1992b. Regulation and expression of the arsenic resistance operon from *Staphylococcus aureus* plasmid pI258. J. Bacteriol. 174:3684-3694.

Ji, G., Garber, E.A.E., Armes, L.G., Chen, C.M., Fuchs, J.A., and Silver, S. 1994. Arsenate reductase of *Staphylococcus aureus* plasmid pI258. Biochemistry. 33:7294–7299.

Jung-Da Rd., Chung-li. and Tao-yuan. 2006. RNA_{Mst}, *An efficient and flexible search tool for RNA*. Department of Computer Science and Information Engineering , National Central University, Taiwan. 300.

Kado, C.I. and Liu, S.T. 1981. Rapid procedure for detection and isolation of large and small plasmids. J. Bacteriol. 145:1365-1373.

Kaur, P. and Rosen, B.P. 1992. Plasmid encoded resistance to arsenic and antimony. Plasmid. 27:29-40.

Karkaria, C.E., Chen, C.M. and Rosen, B.P. 1990. Mutagenesis of a nucleotide-binding site of an anion-translocating ATPase. J. Biol. Chem. 265:7832–7836.

Koch, I., Feldmann, J., Wang, L., Andrewes, P., Reimer, K.J., and Cullen, W.R. 1999. Arsenic in the Meager Creek hot springs environment, British Columbia, Canada. Sci. Total. Environ. 236:101.

Kolbert, P.C. and Persing, D.H. 1999. Ribosomal DNA sequencing as a tool for identification of bacterial pathogens. Current Opinion in Microbiology. 59:945-952.

Kirschman, J. A. and J. H. Cramer. 1988. Two new tools: multi-purpose cloning vectors that carry kanamycin or spectinomycin/streptomycin resistance markers. Gene **68:**163-165.

Kumaresan, M. and Riyazuddin, P. 2001. Overview of Speciation Chemistry of Arsenic (Review), Current Science. 80:837-846.

Kuroda, M., Dey, S., Sanders, O.I. and Rosen, B.P. 1997. Alternate energy coupling of ArsB, the membrane subunit of the *ars* anion-translocating ATPase. J. Biol. Chem. 272:326-331.

Kurz, M., Brunig, A.N.S. and Galinski, E.A. 2006. NhaD type sodium/proton antiporter of *Halomonas elongata*: a salt stress response mechanism in marine habitats. Saline systems. 2:10.

Kyte, J. and Doolittle, R.F. 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. 157:105-132.

Laverman, A.M., Switzer Blum, J., Schaefer, J.K., Phillips, E.J.P., Lovley, D.R. and Oremland, R.S. 1995. Growth of strain SE-3 with arsenate and other diverse electron acceptors. Appl. Environ. Microbiol. 61:3556-3561.

Lane, D.P., Prentki, M. and Chandler. 1992. Use of gel retardation to analyze protein-nucleic acid interactions, Microbiological Reviews. 56:509-528.

Leyval, C., Turn, K., Haselwandter, K. 1997. Effect of Heavy Metal Pollution on Mycorrhizal Colonization and Function: Physiological, Ecological and Applied Aspects. Mycorrhiza. 7: 139-153.

Li, J., Liu, S. and Rosen, B.P. 1996. Interaction of ATP binding sites in the ArsA ATPase, the catalytic subunit of the Ars pump. J. Biol. Chem. 271:25247-52.

Li, S., Chen, Y. and Rosen, B.P. 2001. Role of vicinal cysteine pairs in metalloid sensing by ArsD As(III)-responsive repressor. Mol. Microbiol. 41:687-696.

Li, S., Rosen, B.P., Walmsley, M.I.B. and Walmsley. 2002. Evidence for co-operativity between the four binding sites of dimeric ArsD, and AsIII-responsive transcriptional regulator. J. Biol. Chem. 277:25992-26002.

Lin. Y.F., Walmsley, A.R. and Rosen, B.P. 2006. An arsenic metallochaperone for an arsenic detoxification pump. PNAS. 103:15617-15622.

Liu, J. and Rosen, B.P. 1997. Ligand Interactions of the ArsC Arsenate Reductase. J. Biol. Chem. 272:21084-21089.

Liu, J., Xue, Y., Wang, Q., Wei, Y., Swartz, T.H., Hicks, D.B., Ito, M., Ma, Y. and Krulwich, T.A. 2005. The activity profile of the NhaD-type Na^+ (Li^+)/H^+ antiporter from the soda lake haloalkaliphile *Alkalimonas amylocytica* is adaptive for the extreme environment. J. Bacteriol. 187:7589-7595.

Luo, X. 2006. Molecular cloning and characterization of arsenic resistance gene clusters of *Bacillus* sp. PhD thesis. *Institute of Medicinal Biotechnology, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100050. China.*

Martin, P., Demel, S., Shi, J., Gladysheva, T., Gatti, D.L., Rosen, B.P. and Edwards, F.P. 2001. Insights into the structure, salvation and mechanism of ArsC arsenate reductase, a novel arsenic detoxification enzyme. Structure. 9:1071-1081.

Ma, L.Q., Komar, K.M., Tu, C., Zhang, W., Cai, Y. and Kennelley, E.D. 2001. A fern that hyperaccumulates arsenic. Nature. 409: 579

McGuffin, L.J., Bryson, K. and Jones, D.T. 2000. The PSIPRED protein structure prediction server. Bioinformatics. 16:404-405.

McLaughlin, M. J., Hamon, R. E., McLaren, R. G., Speir, T. W. and Rogers, S. L. 2000. A bioavailability-based rationale for controlling metal and metalloid contamination of agricultural land in Australia and New Zealand. Aust J Soil Res. 38: 1037–86.

McLaren, R.G., Naidu, R., Smith, J. and Tiller, K.G. 1998. "Fractionation and distribution of arsenic in soils contaminated by cattle dip." Journal of Environmental Quality. 27:248-354.

Meng, Y.L., Liu, Z. and Rosen, B.P. 2004. Arsenite and antimonite uptake by GlpF and efflux by ArsB in *E. coli*. J. Bacteriol. 279:18334-18341.

Messens, J., Martin, J.C., Van Belle, K., Brosens, E., Desmyter, A., De Gieter, M., Wieruszeski, J.M., Willem, R., Wyns, L. and Zegers, I. 2002. All intermediates of the arsenate reductase mechanism, including an intramolecular dynamic disulfide cascade. PNAS. 99:8506-8511.

Miller, J. 1972. A short course in bacterial genetics : a laboratory manual and handbook for *Escherichia coli* and related bacteria, vol. 1. Cold Spring Harbor Laboratory Press, Plainview, N.Y.

Morris, A. 1998. Study of arsenic resistant plants grown in the soil of cattle dip sites. B. Environ. Sci. Thesis.

Mukhopadhyay, R., Shi, J. and Rosen, B.P. 2000. Purification and characterization of ACR2p, the *Saccharomyces cerevisiae* arsenate reductase. J. Biol. Chem. 275:21149–21157.

Mukhopadhyay, R., Rosen, B.P., Phung, L.T. and Silver, S. 2002. Microbial arsenic: from geocycles to genes and enzymes. FEMS. Microbiol. Rev. 26:311-25.

Mukhopadhyay, R., Zhou, Y. and Rosen, B.P. 2003. Directed evolution of a yeast arsenate reductase into a protein-tyrosine phosphatase. J. Biol. Chem. 27:24476-24480.

Neyt, C., Iriarte, M., Thi, V.H. and Cornelis, G.R. 1997. Virulence and *ars*enic resistance in Yersiniae. J. Bacteriol. 179:612-619.

Ng, W.V., Ciufo, S.A., Smith, T.M., Bumgarner, R.E., Baskin, D., Faust, J. et. al. 1998. Snapshot of a large dynamic replicon in a halophilic archeaon: megaplasmid or minichromosome? Genome. Res. 8:1131-1141.

Nickson, R., McArthur, J. and Burgess, W. 1998. Arsenic poisoning of Bangladesh ground water. Nature. 395:338.

Nimick, D.A., Moore, J.N., Dalby, C.E. and Savka, M.W. 1998. The fate of geothermal arsenic in the Madison and Missouri Rivers, Montana and Wyoming. Water. Resour. Res. 11:3051.

Nordstrom, D.K. 2002. Public health – worldwide occurrences of arsenic in ground water. Science. 296:2143-2145.

Nriagu, J. O. 1984. Arsenic in the Environment - Part I: Cycling and Characterisation., vol. 26. John Wiley & Sons. New York.

Nriagu, J. O. 2002. Environmental Chemistry of Arsenic,W. T. Frankenberger Jr., Ed. (Dekker, New York,). pp. 1–26.

Owolabi, J.B. and Rosen, B.P. 1990. Differential mRNA stability controls relative gene expression within the plkasmid encoded arsenical resistance operon. J. Bacteriol. 172:2367-2371.

Oremland, R.S. and Stolz, J.F. 2003. The ecology of arsenic. Science. 300:939-944.

Pearson, W.R. and Lipman, D.J. 1988. Improved tools for biological sequence analysis. Proc.Nat.Acad.Sci. USA. 85:2444-2448.

Pieper, D. H. and Reineke, W. 2000. Engineering bacteria for bioremediation. *Current* Opinion in Biotechnology. 11:262-270.

Porter, E.K. and Peterson, P.J. 1977. Arsenic tolerance in grasses growing on mine waste. Environmental Pollution. 14:255-265.

Roy, P. and Saha, A. 2002. Metabolism and toxicity of arsenic; a human carcinogen, *Current science*. Volume 82:38-45.

Prithivirajsingh, S., Mishra, S. and Mahadevan, A. 2001. "Detection and analysis of chromosomal arsenic resistance in Pseudomonas fluorescens strain MSP3." Biochem. Biophys. Res. Commun. 280:1393-401.

Prakash, S., Cooper, G., Singhi, S. and Saier, M.H. 2003. The ion transporter superfamily. Biochim. Biophys. Acta. 1618:79-92.

Ramaswamy, S. and Kaur, P. 1998. Nucleotide Binding to the C-terminal Nucleotide Binding Domain of ArsA. Studies with an ATP analogue, 5'-*p*-fluorosulfonylbenzoyladenosine (FSBA). J. Biol. Chem. 273:9243-9248.

Rasko, D.A., Ravel, J., Ekstad, O.A., Helgason, E., Cer, R.Z., Jiang, L., Shores, K.A., Fouts, D.E., Tourasse, N.J., Angiuoli, S.V., Kolonay, J., Nelson, W.C., Kolstc, A.B., Fraser, C.M. and Read, T.D. 2004. The genome sequence of *Bacillus cereus* ATCC10987 reveals metabolic adaptations and a large plasmid related to *Bacillus anthracis* pXO1. Nucleic Acids Research. 32:977- 988.

Rensing, C., Ghosh, M. and Rosen, B.P. 1999. Families of Soft-Metal-Ion-Transporting ATPases J. Bacteriol. 181:5891 – 5897.

Reynolds, C. 1999. A natural poison for redevelopment, VHB/Vanasse Hangen Brustlin' Inc. 2:5.

Rosen, B. P. 1999. Families of arsenic transporters. Trends. Microbiol. 7:207-212.

Rosen, B. P. 2002. Biochemistry of arsenic detoxification. FEBS. Lett. 529:86-92.

Rosen, B.P., Bhattacharjee, H., Zhou, T. and Walmsley, A.R. 1999. Mechanism of the ArsA ATPase. Biochim. Biophys. Acta. 1461:207-15.

Rosenstein, R., Peschel, P., Wieland, B. and Gotz. F. 1992. Expression and regulation of the *Staphylococcus xylosus* antimonite, arsenite, and arsenate resistance operon. J. Bacteriol. 174:3676-3683.

Rosenstein, R., Nikoleit, K. and Gotz, F. 1994. Binding of ArsR, the repressor of the *Staphylococcus xylosus* (pSX267) arsenic resistance operon to a sequence with dyad symmetry within the *ars* promoter. Mol. Gen. Genet. 242:566-572.

Rosenberg, H.R., Gerdes, R.G. and Chegwidden, K. 1977. Two systems for the uptake of phosphate in *E. coli.* J. Bacteriol. 131:505-511.

Ryan, D. and Colleran, E. 2002. Arsenical resistance in the IncHI2 plasmids. Plasmid. 47:234-240.

Sacchi, C., Whitney, A.M., Mayer, l.W., Morey, R., Steigerwalt, A., Boras, A., Weyant, R.S. and Popovic, T. 2002. Sequencing of 16S rRNA gene: A rapid tool for identification of Bacillus anthracis. Emerging Infectious Diseases. 8:1117-1123.

Saltikov, C.W., Cifuentes, A., Venkateswaran, K. and Newman, D.K. 2003. The *ars* detoxification system is advantageous but not required for AS (V) respiration by the genetically tractable *Shewanella* species Strain ANA-3. Appl. Environ. Microbiol. p2800-2809.

Sambrook, J., Fritsch, E.F. and Maniatis, T. 1989. Molecular cloning: a laboratory manual, Second ed. Cold Spring Harbour Laboratory Press, Cold Spring Harbour. New York.

Sato, T. and Kobayashi, Y. 1998. The *ars* operon in the skin element of *Bacillus subtilis* confers resistance to arsenate and arsenite. J. Bacteriol. 180:1655-1661.

San Francisco, M., L. Tisa and B. P. Rosen. 1989. Identification of the membrane component of the anion pump encoded by the arsenical resistance operon of R-factor R773. Mol Microbiol **3:**15-21.

San Francisco, M., Hope, C.L., Owolabi, J.B., Tisa, L.S. and Rosen, B.P. 1990. Identification of the metalloregulatory element of the plasmid-encoded arsenical resistance operon. Nucleic Acids Res. 18:619-624.

Sharples JM, Meharg AA, Chambers SM, Cairney JW. 2000. Mechanism of arsenate resistance in the ericoid mycorrhizal fungus *Hymenoscyphus ericae.* Plant Physiol. 124: 1327-1334.

Shen, Z.X., Chen, G.Q., Ni, J.H., Li, X.S., Xiong, S.M., Qiu, Q.Y., Zhu, J., Tang, W., Sun, G.L., Yang, K.Q, Chen, Y., Zhou, L., Fang, Z.W., Wang, Y.T., Ma, J., Zhang, P., Zhang, T.D., Chen, S.J., Chen, Z. and Wang, Z.Y. 1997. Use of arsenic trioxide in the treatment of acute promyleocytic leukemia (APL): Clinical efficacy and pharmacokinetics in relapsed patients, Blood. J. 89:3354-60.

Shi, W., Wu, J. and Rosen, B.P. 1994. Identification of a putative metal binding site in a new family of metalloregulatory proteins. J. Biol. Chem. 269:19826-19829.

Shi, W., Dong, J., Scott, R.A., Ksenzenko, M.Y. and Rosen, B.P. 1996. The role of arsenic-thiol interactions in metalloregulation of the *ars* operon. J. Biol. Chem. 271:9291-9297.

Silver, S. and Phung, L.T. 2005. Genes and enzymes involved in bacterial oxidation and reduction of inorganic arsenic. Appl. Environ. Microbiol. p 599-608.

Silver, S., Budd, K., Leahy, K., Shaw, W., Hammond, D., Novick, R., Willsky, G., Malamy, M. and Rosenberg, H. 1981. Inducible plasmid-determined resistance to arsenate, arsenite, and antimony (III) in *Escherichia coli* and *Staphylococcus aureus.* J. Bacteriol. 146:983-996.

Smedley, P. L. and Kinniburgh, D.G. 2002. A review of the source, behavior and distribution of arsenic in natural waters. Appl. Geochem. 17:517-568.

Smith, T. F. and Waterman, M. S. 1981. Identification of common molecular subsequences. J. Mol. Biol. 147:195-197.

Smith, P.K., Krohn, R.I., Hermanson, G.T., Mallia, A.K., Gartner, F.H., Provenzano, M.D., Fujimoto, E.K., Goeke, N.M., Olson, B.J. and Klenk, D.C. 1985. Measurement of protein using Bicinchoninic acid. Anal. Biochem. 150:76-85.

Smith, E., Naidu, R. and Alston, A.M. 1998. Arsenic in the soil environment: A review. Adv. Agronomy. 57:149-195.

Smith, E., Smith, J., Smith, L., Biswas, T., Correll, R. and Naidu, R. 2003. "Arsenic in Australian environment: an overview". J. Environ. Sci. Health. Part A. Tox. Hazard. Subst. Environ. Eng. 38:223-39.

Stauffer, R.E. and Thompson, J.M. 1984. Arsenic and antimony in geothermal waters of Yellowstone National Park, Wyoming, USA. Geochim. Cosmochim. Acta. 48:2547-2561.

Suzuki, K., Wakao, N., Kimura, T., Sakka, K. and Ohmiya, K. 1998. Expression and Regulation of the arsenic resistance operon of *Acidiphilium multivorum* AIU 301 plasmid pKW301 in Escherichia coli. Appl. & Environ. Microbiol. 64:411-418.

Su, C. 2004. Regulation of a novel *ars* gene cluster in *Sinorhizobium* sp. MSc research thesis. School of Biological Sciences, University of Wollongong. Australia.

Tamaki, S. and Frankenberger, W. 1992. Enviromental biochemistry of arsenic. Rev Environmental contamination & Toxicology. 124:79-110.

Tanaka, T. 1990. Arsenic in the natural environment. Part II: arsenic concentrations in thermal waters from Japan. Appl. Organomet. Chem. 4:197-203.

Takeuchi, M., Kawahata, H., Gupta, L. P., Kita, N., Morishita, Y., Ono, Y. And Komai, T. 2006. Arsenic resistance and removal by marine and non-marine bacteria. J. Biotechnol. 127:434-442.

Thompson, J.D., Higgins, D.G. and Gibson, T.J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research. 22: 4673-4680.

Thomas, C.M. 2002. *Lac*Z promoter fusions the effect of growth. Microbiology. 148: 1241-1245.

Tsai, K.J., Hsu, C.M. and Rosen B.P. 1997. Efflux mechanisms of resistance to cadmium, arsenic and antimony in prokaryotes and eukaryotes. Zoological studies. 36: 1-16.

Tisa, L.S. and Rosen, B.P. 1990. Molecular characterization of an anion pump. The ArsB protein is the membrane anchor for the ArsA protein. J. Biol. Chem. 265:190-4.

Ute, K. 2005. Phytoremediation: Novel approaches to cleaning up polluted soils. Current Opinion in Biotechnology. 16:133-141.

Van Zwieten, L., Ayres, M.R. and Morris, S.G. 2003. Influence of arsenic co-contamination on DDT breakdown and microbial activity. Environmental Pollution. 124: 331-339.

Wang, L., Chen, S., Xiao, X., Huang, X., You, D., Zhou, X. and Deng, Z. 2006. *arsRBOCT* arsenic resistance system encoded by linear plasmid pHZ227 in *Streptomyces* sp. Strain FR-008. Appl. & Environ. Microbiol. 72:3738-3742.

Wang, Q. 2001. Characterisation of an arsenic resistance gene cluster in *Sinorhizobium* sp. Biol911 Report. School of Biological Sciences, University of Wollongong. Australia.

Walmsley, A.R., Zhou, T., Borges-Walmsley, M.I. and Rosen, B.P. 1999. The ATPase mechanism of ArsA, the catalytic subunit of the arsenite pump. J. Biol. Chem. 274:16153-61.

Walter, F.J. and Wenzel, W.W. 2002. Arsenic transformations in the soil rhizosphere plant system: fundamentals and potential application to phytoremediation. J Biotech. 99: 259- 278.

Weisburg, W.G., Barns, S.M., Pelletier, D.A. and Lane, D.J. 1991. 16S Ribosomal DNA Amplification for Phylogenetic Study. J. Bacteriol. 173:697-703.

Welch, H., Westjohn, D.B., Helsel, D.R. and Wanty, R.B. 2000. Arsenic in **ground water** of the united states- occurrence and geochemistry. Groundwater. 38:589-604.

Wilkie, J.A. and Hering, J.G. 1998. Rapid Oxidation of Geothermal Arsenic(III) in Streamwaters of the Eastern Sierra Nevada". Environ. Sci. Technol. 32:657-662.

Wu, J. and Rosen, B.P. 1991. The ArsR protein is a trans-acting regulatory protein. Mol. Microbiol. 5:1331-1336.

Wu, J., Tisa, L.S. and Rosen, B.P. 1992. Membrane topology of the ArsB protein, the membrane subunit of an anion-translocating ATPase. J. Biol. Chem. 267:12570-12576.

Wu, J.H. and Rosen, B.P. 1993a. The *ars*D gene encodes a second transacting regulatory protein of the plasmid-encode arsenical resistance operon. Mol. Microbiol. 8:615-623.

Wu, J. and Rosen, B.P. 1993b. Metalloregulated expression of the a*rs* operon. J. Biol. Chem. 268:52-58.

Xu, C., Shi, W. and Rosen, B.P. 1996. The chromosomal *arsR* gene of *Escherichia coli* encodes a trans-acting metalloregulatory protein. J. Biol. Chem. 271:2427-2432.

Xu, C. and Rosen, B.P. 1997. Dimerization Is Essential for DNA Binding and Repression by the ArsR Metalloregulatory Protein of *Escherichia coli*. J. Biol. Chem. 272:15734-15738.

Yang, H.C., Cheng, J., Finan, T.M., Rosen, B.P. and Bhattacharjee, H. 2005. Novel pathway of arsenic detoxification in the legume symbiont *Sinorhizobium meliloti*. J. Bacteriol. 6991-6997.

Zhou, T. and Rosen, B.P. 1997. Tryptophan fluorescence reports nucleotide-induced conformational changes in a domain of the ArsA ATPase. J. Biol. Chem. 272:19731–19737.

Zhou, T., Radaev, S., Rosen, B.P. and Gatti, D.L. 2000, Structure of the ArsA ATPase: the catalytic subunit of a heavy metal resistance pump. EMBO. J. 19:4838-45.

APPENDICES

APPENDIX 1: BACTERIAL MEDIA

LB Broth Medium (1 litre):

Bacto-tryptone	10 g
Yeast extract	5 g
NaCl	10 g
pH 7.5	

LB Agar Medium(1 litre):

Bacto-tryptone	10 g
Yeast extract	5 g
NaCl	10 g
Agar	15 g
pH 7.5	

SOB Medium (1 litre):

Tryptone	20 g
Yeast extract	5 g
NaCl (1 M)	10 mL
KCl (1 M)	2.5 mL
Autoclave and add $MgSO_4$	10 mM

SOC Medium (100 ml):

SOB medium	100 mL
Glucose	20 mM

2YT Medium (1 litre):

Bacto-tryptone	16 g
Yeast extract	10 g
NaCl	5 g
pH 7.0	

APPENDIX 2: BUFFERS AND SOLUTIONS

0.8% TAE/TBE agarose gel + ethidium bromide (100 mL):

Agarose	0.8 g
1X TAE/TBE	100 mL
EB (1 mg/mL)	15 µL

Loading dye (10 mL):

Bromophenol blue	0.025 g
Xylene cyanol FF	0.025 g
Ficoll	1.5 g

NETS buffer (1 litre):

NaCl	0.1 M
Tris-HCl (pH 8.0)	10 mM
EDTA (pH 8.0)	1 mM
SDS	1% (W/V)

Before use, add β-Mercaptoethanol to 0.1% (V/V)

TAE buffer 50 X (1 litre):

Tris base	242 g
Glacial acetic acid	57.1 mL
EDTA (0.5M, pH 8.0)	100 mL

pH 7.0-7.5

TBE buffer 1 X (1 litre):

Tris base	10.8 g
Boric acid	5.5 g
EDTA (0.5 M, pH 8.0)	4 mL

pH 7.0-7.5

Z buffer (1 litre):

$Na_2HPO_4.7H_2O$	60 mM
$NaH_2PO_4.H_2O$	40 mM
KCl	10 mM
$MgSO_4.\ 7H_2O$	1 mM

Adjust pH to 7.0

Before use, add β-Mercaptoethanol to 50 mM

10 X MOPS (3-[N-Morpholine] propane sulfonic acid) (1 litre):

0.2 M MOPS	46.2 g
0.05 M sodium acetate	6.8 g
0.01 M EDTA	3.27 g

pH to 7.0

Binding Buffer (Gel Shift Assay):

Tris-Cl (pH 7.6)	10 mM
KCl	80 mM
Na_2EDTA	0.2 mM
DTT	0.2 mM
Glycerol	10 %

6% Non-denaturing acrylamide gel (TBE gel) (10 mL):

H_2O	7.3 mL
40% Acrylamide/bis	1.5 mL
5x TBE	0.6 mL
10% APS	0.234 mL
TEMED	0.006 mL
80% Glycerol	0.360 mL

Protein gel loading buffer:

125 mM Tris-Cl (pH 6.8)
5% SDS
25% glycerol
0.25 Bromophenol Blue
The above mixture is 2.5X stored at room temperature.
Mix with 1M DTT at 4:1 ratio, which then must be refrigerated.
Mix this with sample 1:1 ratio.

SDS-PAGE gel 15% (10 mL):

Separating gel (5 mL):

H_2O	1.1 mL
30% Acrylamide/bis	2.5 mL
1.5 M Tris (pH 8.8)	1.3 mL
10% SDS	0.05 mL
10% APS	0.05 mL
TEMED	0.002 mL

Stacking gel (5 mL):

H_2O	2.8 mL
30% Acrylamide/bis	0.83 mL
0.5 M Tris (pH 6.8)	1.26 mL
10% SDS	0.05 mL
10% APS	0.05 mL
TEMED	0.005 mL

SDS-PAGE Protein staining solution:

Comassive brilliant blue stain (1litre):

Comassive Blue R-250	0.2%	2 g
Methanol	40%	400 mL
Glacial Acetic Acid	10%	100 mL
H_2O	50%	500 mL

Rapid de-stain (1litre):

Methanol	40%	400 mL
Glacial Acetic Acid	10%	100 mL
H_2O	50%	500 mL

Final de-stain (1litre):

Glacial Acetic Acid	10%	100 mL
Glycerol	4%	40 mL
H_2O	86%	860 mL

SDS-PAGE running buffer (1litre):

	5x	**1x**
25mM Tris	15.1 g	3.028 g
250mM Glycine (pH-8.3)	94 g	18.8 g
0.1% SDS	50 mL (10%w/v)	10 mL (10%w/v)
D-H_2O	950 mL	990 mL

Ni – NTA Purification

Buffer stock solution A (10 x) (1litre):

200 mM Sodium Phosphate (monobasic) (NaH_2PO_4)	27.6 g
5 M Nacl	292.9 g
Deionized H_2O	1000 mL

Buffer stock solution B (10 x) (1litre):

200 mM Sodium Phosphate (dibasic) (NaH_2PO_4)	28.4 g
5 M Nacl	292.9 g
Deionized H_2O	1000 mL

Note: Store solutions at Room Temperature.

5 x Native purification buffer:
(250 mM NaH_2PO_4, pH 8.0; 2.5 M Nacl)

Sodium Phosphate (monobasic)	7 g
NaCl	29.2 g
H_2O	200 mL

Mix well and adjust the pH to 8.0 with NaOH and store at R.T

3 M Imidazole pH 6.0:
(3 M Imidazole; 500 mM Nacl; 20 mM Sodium phosphate buffer, pH 6.0)

Imidazole	20.6 g
Stock solution A (10 x)	8.77 mL
Stock solution B (10 x)	1.23 mL
Deionized H_2O	100 mL

Mix well and adjust the pH to 6.0 with HCl or NaOH and store at R.T

Native purification buffers:

1 x Native purification buffer (100 mL):

5 x Native Purification buffer	20 mL
Sterile d.H_2O	80 mL

Mix well and adjust the pH to 8.0

Native binding buffer:
Use 30ml of 1x native Purification buffer as the Native Binding Buffer

Native wash buffer:

1 x Native purification buffer	50 mL
3 M Imidazole, pH 6.0	335 µL

Mix well and adjust the pH to 8.0

Native elution buffer (15 mL):

1 x Native Purification buffer	13.75 mL
3 M Imidazole, pH 6.0	1.25 mL

Mix well and adjust the pH to 8.0

APPENDIX 3: Multiple sequence alignments of the 16S rDNA sequence of CDB1, CDB2, CDB3, CDB4 and CDB5 with their related bacterial species. The 26 related species were chosen according to their pairwise alignment score and identity. Multiple alignments with related species were performed using ClustalW.

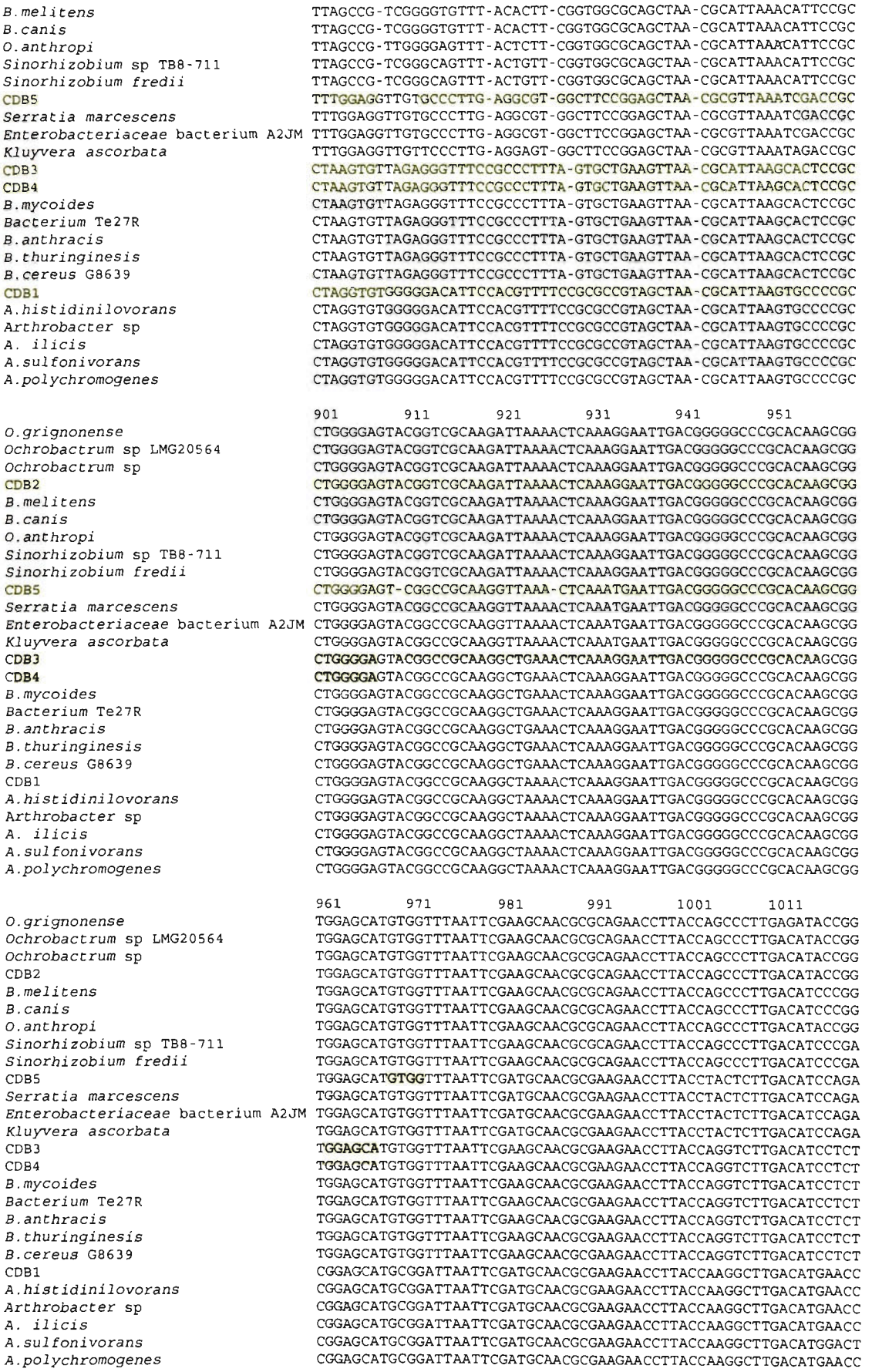
```
                                  1         11        21        31        41        51
O.grignonense                     ---------------------------------AGTTTGATCCTGGCTCAGAACGAACGC
Ochrobactrum sp LMG20564          -------------------------------------------------AACGAACGC
Ochrobactrum sp                   -------------------------------------------------AACGAACGC
CDB2                              ------------------------------------------------------------
B.melitens                        ---------------------------------------------CTCAGAACGAACGC
B.canis                           ----------------------------------------------TCAGAACGAACGC
O.anthropi                        ---------------------------------AGTTTGATCCTGGCTCAGAACGAACGC
Sinorhizobium sp TB8-711          -------------------------------------------------ACGAACGC
Sinorhizobium fredii              -------------------------------------------------AACGAACGC
CDB5                              ----------------------------------TGGCGGGCGGTAATACGTAGCA
Serratia marcescens               ---------------------------------------------GGCTTACACATGC
Enterobacteriaceae bacterium A2JM ---------------------------------------------CCTAACACAT-GCA
Kluyvera ascorbata                ---------AGTTTGATCCTGGCTCAGATTGAACGCTGGCGGCAGGCCTAACACAT-GCA
CDB3                              --------------------------GATGAACGCTGGCGGCGTGCCTAATACAT-GCA
CDB4                              ----------------------------------ACAGGGAGCCGA-------GCA
B.mycoides                        ------------------------------------------------TAATACAT-GCA
Bacterium Te27R                   -----TTAGAKTTTGATCATGGCTCAGGATGAACGCTGGCGGCGTGCTTAATACAT-GCA
B.anthracis                       TTATTGGAGAGTTTGATCCTGGCTCAGGATGAACGCTGGCGGCGTGCCTAATACAT-GCA
B.thuringinesis                   ------------------CTGGCTCAGGATGAACGCTGGCGGCGTGCCTAATACAT-GCA
B.cereus G8639                    TTATTGGAGAGTTTGATCCTGGCTCAGGATGAACGCTGGCGGCGTGCCTAATACAT-GCA
CDB1                              ---------------------------------ATATAGCCGACAGCTCACCGTA--GCA
A.histidinilovorans               -----------------CCTGGCTCAGGATGAACGCTGGCGGCGTGCTTAACACAT-GCA
Arthrobacter sp                   --------------------------GATGAACGCTGGCGGCGTGCTTAACACAT-GCA
A. ilicis                         -------------GATCCTGGCTCAGGATGAACGCTGGCGGCGTGCTTAACACAT-GCA
A.sulfonivorans                   ------------------------------------GGCGGCGTGCTTAACACAT-GCA
A.polychromogenes                 -------------GATCCTGGCTCAGGATGAACGCTGGCGGCGTGCTTAACACAT-GCA

                                  61        71        81        91        101       111
O.grignonense                     TGGCGGCAGGCTTAACACATGCAAGTCGAGCGCCTCGCAAGAGGAGCGGCAGACGGGTGA
Ochrobactrum sp LMG20564          TGGCGGCAGGCTTAACACATGCAAGTCGAGCGCCCCGCAAGGGGAGCGGCAGACGGGTGA
Ochrobactrum sp                   TGGCGGCAGGCTTAACACATGCAAGTCGAGCGCCCCGCAAG-GGAGCGGCAGACGGGTGA
CDB2                              ---------------------------------GGTAGCCGG-TAAGCGGCAGACGGGTGA
B.melitens                        TGGCGGCAGGCTTAACACATGCAAGTCGAGCGCCCCGCAAGGGGAGCGGCAGACGGGTGA
B.canis                           TGGCGGCAGGCTTAACACATGCAAGTCGAGCGCCCCGCAAGGGGAGCGGCAGACGGGTGA
O.anthropi                        TGGCGGCAGGCTTAACACATGCAAGTCGAGCGCCCCGCAAGGGGAGCGGCAGACGGGTGA
Sinorhizobium sp TB8-711          TGGCGGCAGGCTTAACACATGCAAGTCGAGCGCTC-GCAACGGGAGTGGCAGACGGGTGA
Sinorhizobium fredii              TGGCGGCAGGCTTAACACATGCAAGTCGAGCGCCCCGCAAGGGGAGCGGCAGACGGGTGA
CDB5                              AGTCGAGCGG--TAGCACAGGGGAG-CTTGCTCCCTGGGTGACGAGCGGCGGACGGGTGA
Serratia marcescens               AGTCGAGCGG--TAGCACAGGGGAG-CTTGCTCCCTGGGTGACGAGCGGCGGACGGGTGA
Enterobacteriaceae bacterium A2JM AGTCGAGCGG--TAGCACAGANNNN-NNNNNNNNTCGGGTGACGAGCGGCGGACGGGTGA
Kluyvera ascorbata                AGTCGAACGG--TAGCACAGAG-AG-CTTGCTCTC-GGGTGACGAGTGGCGGACGGGTGA
CDB3                              AGTCGAGCG----AATGGGTTGAGAGCTTGCTCTC-AAGAAGTTAGCGGCGGACGGGTGA
CDB4                              AG--CAGCG----AA--GGATAAGAGCTTGCTCTT-ATGAAGTTAGCGGCGGACGGGTGA
B.mycoides                        AGTCGAGCG----AATGGATTAAGAGCTTGCTCTT-ATGAAGTTAGCGGCGGACGGGTGA
Bacterium Te27R                   AGTCGAGCG----AATGGATTAAGAGCTTGCTCTT-ATGAAGTTAGCGGCGGACGGGTGA
B.anthracis                       AGTCGAGCG----AATGGATTAAGAGCTTGCTCTT-ATGAAGTTAGCGGCGGACGGGTGA
B.thuringinesis                   AGTCGAGCG----AATGGATTAAGAGCTTGCTCTT-ATGAAGTTAGCGGCGGACGGGTGA
B.cereus G8639                    AGTCGAGCG----AATGGATTGAGAGCTTGCTCTC-AAGAAGTTAGCGGCGGACGGGTGA
CDB1                              AGTCGAACG----AT--GATCC-CAGCTTGCTGG---GG-GATTAGTGGCGAACGGGTGA
A.histidinilovorans               AGTCGANCG----AT--GATCC-CAGCTTNCTGG---GG-GATTAGTGGCGAACGGGTGA
Arthrobacter sp                   AGTCGAACG----AT--GATCC-CAGCTTGCTGG---GG-GATTAGTGGCGAACGGGTGA
A. ilicis                         AGTCGAACG----AT--GATCC-CAGCTTGCTGG---GG-GATTAGTGGCGAACGGGTGA
A.sulfonivorans                   AGTCGAACG----AT--GAACCTCACTTGTGGGG---GG-GATTAGTGGCGAACGGGTGA
A.polychromogenes                 AGTCGAACG----AT--GAAGGGGAGCTTGCTCC---TG-GATTAGTGGCGAACGGGTGA

                                  121       131       141       151       161       171
O.grignonense                     GTAACGCGTGGG-AATCTACCTTTTGCTACGGAATAACTCAGGGAAACTTGTGCTAATAC
Ochrobactrum sp LMG20564          GTAACGCGTGGG-AATCTACCTTTTGCTACGGAATAACTCAGGGAAACTTGTGCTAATAC
Ochrobactrum sp                   GTAACGCGTGGG-AATCTACCTTTTGCTACGGAATAACTCAGGGAAACTTGTGCTAATAC
CDB2                              GTAACGCGTGGG-AATCTACCTTTTGCTACGGAATAACTCAGGGAAACTTGTGCTAATAC
B.melitens                        GTAACGCGTGGG-AACGTACCATTTGCTACGGAATAACTCAGGGAAACTTGTGCTAATAC
B.canis                           GTAACGCGTGGG-AACGTACCATTTGCTACGGAATAACTCAGGGAAACTTGTGCTAATAC
O.anthropi                        GTAACGCGTGGG-AACGTACCTTTTGCTACGGAATAACTCAGGGAAACTTGTGCTAATAC
Sinorhizobium sp TB8-711          GTAACGCGTGGG-AATCTACCCTTTTCTACGGAATAACGCATGGAAACTTGTGCTAATAC
Sinorhizobium fredii              GTAACGCGTGGG-AATCTACCCTTTTCTACGGAATAACGCAGGGAAACTTGTGCTAATAC
CDB5                              GTAATGTCTGGG-AAACTGCCTGATGGAGGGGGATAACTACTGGAAACGGTAGCTAATAC
Serratia marcescens               GTAATGTCTGGG-AAACTGCCTGATGGAGGGGGATAACTACTGGAAACGGTAGCTAATAC
Enterobacteriaceae bacterium A2JM GTAATGTCTGGG-AAACTGCCTGATGGAGGGGGATAACTACTGGAAACGGTAGCTAATAC
Kluyvera ascorbata                GTAATGTCTGGG-AAACTGCCCGATGGAGGGGGATAACTACTGGAAACGGTAGCTAATAC
CDB3                              GTAACACGTGGGTAACCTGCCCATAAGACTGGGATAACTCCGGGAAACCGGGGCTAATAC
CDB4                              GTAACACGTGGGTAACCTGCCCATAAGACTGGGATAACTCCGGGAAACCGGGGCTAATAC
```

```
B.mycoides                          GTAACACGTGGGTAACCTACCCATAAGACTGGGATAACTCCGGGAAACCGGGGCTAATAC
Bacterium Te27R                     GTAACACGTGGGTAACCTGCCCATAAGACTGGGATAACTCCGGGAAACCGGGGCTAATAC
B.anthracis                         GTAACACGTGGGTAACCTGCCCATAAGACTGGGATAACTCCGGGAAACCGGGGCTAATAC
B.thuringinesis                     GTAACACGTGGGTAACCTGCCCATAAGACTGGGATAACTCCGGGAAACCGGGGCTAATAC
B.cereus G8639                      GTAACACGTGGGTAACCTGCCCATAAGACTGGGATAACTCCGGGAAACCGGGGCTAATAC
CDB1                                GTAACACGTGAGTAACCTGCCCTTGACTCTGGGATAAGCCTGGGAAACTGGGTCTAATAC
A.histidinilovorans                 GTAACACGTGAGTAACCTGCCCTTGACTCTGGGATAACCCTGGGAAACTGGGTCTAATAC
Arthrobacter sp                     GTAACACGTGAGTAACCTGCCCTTGACTCTGGGATAAGCCTGGGAAACTGGGTCTAATAC
A. ilicis                           GTAACACGTGAGTAACCTGCCCTTGACTCTGGGATAAGCCTGGGAAACTGGGTCTAATAC
A.sulfonivorans                     GTAACACGTGAGTAACCTGCCCTTGACTCTGGGATAAGCCTGGGAAACTGGGTCTAATAC
A.polychromogenes                   GTAACACGTGAGTAACCTGCCCTTAACTCTGGGATAAGCCTGGGAAACTGGGTCTAATAC
                                    181       191       201       211       221       231
O.grignonense                       CGTATGT--------GCCCTTTTGGGGAAAGATT---------------TATCGGCAAAG
Ochrobactrum sp LMG20564            CGTATGT--------GCCCTTCGGGGGAAAGATT---------------TATCGGCAAAG
Ochrobactrum sp                     CGTATGT--------GCCCTTCGGGGGAAAGATT---------------TATCGGCAAAG
CDB2                                CGTATGT--------GCCCTTCGGGGGAAAGATT---------------TATCGGCAAAG
B.melitens                          CGTATGT--------GCCCTTCGGGGGAAAGATT---------------TATCGGCAAAT
B.canis                             CGTATGT--------GCCCTTCGGGGGAAAGATT---------------TATCGGCAAAT
O.anthropi                          CGTATGT--------GCCCTTCGGGGGAAAGATT---------------TATCGGCAAAG
Sinorhizobium sp TB8-711            CGTATGA--------GCCCTTCGGGGGAAAGATT---------------TATCGGGAAAG
Sinorhizobium fredii                CGTATGA--------GCCCTTCGGGGGAAAGATT---------------TATCGGGAAAG
CDB5                                CGCATAACGTCGCAAGACCAAAGAGGGGGACCTTCGGGC---------CTCTTGCCATCA
Serratia marcescens                 CGCATAACGTCGCAAGACCAAAGAGGGGGACCTTCGGGC---------CTCTTGCCATCA
Enterobacteriaceae bacterium A2JM   CGCATAACGTCTTCGGACCAAAGAGGGGGACCTTCGGGC---------CTCTTGCCATCA
Kluyvera ascorbata                  CGCATAATGTCGCAAGACCAAAGTGGGGGACCTTCGGGC---------CTCACACCATCG
CDB3                                CGGATAATATTTTGAACTGCATGGTTCGAAATTGAAAGGCGGCTTCGGCTGTCACTTATG
CDB4                                CGGATAATATTTTGAACTGCATGGTTCGAAATTGAAAGGCGGCTTCGGCTGTCACTTATG
B.mycoides                          CGGATGATATTTTGAACTGCATAGTTCGAAATTGAAAGGCGGCTTCGGCTGTCACTTATG
Bacterium Te27R                     CGGATAACATTTTGAACYGCATGGTTCGAAATTGAAAGGCGGCTTCGGCTGTCACTTATG
B.anthracis                         CGGATAACATTTTGAACCGCATGGTTCGAAATTGAAAGGCGGCTTCGGCTGTCACTTATG
B.thuringinesis                     CGGATAACATTTTGAACYGCATGGTTCGAAATTGAAAGGCGGCTTCGGCTGTCACTTATG
B.cereus G8639                      CGGATAAYATTTTGAACTGCATGGTTCGAAATTGAAAGGCGGCTTCGGCTGTCACTTATG
CDB1                                CGGATATGACTCCTCATCGCATGGTGGGG-GGTGGAAA---GCTTT---TGTG-GTTTTG
A.histidinilovorans                 CGGATATGACTCCTCATCGCATGGTGGGG-GGTGGAAA---GCTTT---TGTG-GTTTTG
Arthrobacter sp                     CGGATACGACCATTCCACGCATGTGGTGGTGGTGGAAA---GCTTT---TGTG-GTTTTG
A. ilicis                           CGGATACGACCATCTGGCGCATGTCATGGTGGTGGAAA---GCTTT---TGTG-GTTTTG
A.sulfonivorans                     CGGATATGACGCCTCATCGCATGGTGGGG-GGTGGAAA---GTTTT---TGTG-GTTTTG
A.polychromogenes                   CGGATATGACTCCTCATCGCATGGTGGGG-GGTGGAAA---GCTTT---ATTGTGTTTTG

                                    241       251       261       271       281       291
O.grignonense                       GATGAGCCCGCGTTGGATTAGCTAGTTGGTAGGGTAATGGCCTACCAAGGCGACGATCCA
Ochrobactrum sp LMG20564            GATGAGCCCGCGTTGGATTAGCTAGTTGGTGAGGTAAAGGCTCACCAAGGCGACGATCCA
Ochrobactrum sp                     GATGAGCCCGCGTTGGATTAGCTAGTTGGTGGGGTAAAGGCCCACCAAGGCGACGATCCA
CDB2                                GATGAGCCCGCGTTGGATTAGCTAGTTGGTGGGGTAAAGGCCCACCAAGGCGACGATCCA
B.melitens                          GATCGGCCCGCGTTGGATTAGCTAGTTGGTGGGGTAAAGGCTCACCAAGGCGACGATCCA
B.canis                             GATCGGCCCGCGTTGGATTAGCTAGTTGGTGGGGTAAAGGCTCACCAAGGCGACGATCCA
O.anthropi                          GATCGGCCCGCGTTGGATTAGCTAGTTGGTGAGGTAAAGGCTCACCAAGGCGACGATCCA
Sinorhizobium sp TB8-711            GATGAGCCCGCGTTGGATTAGCTAGTTGGTGGGGTAAAGGCCTACCAAGGCGACGATCCA
Sinorhizobium fredii                GATGAGCCCGCGTTGGATTAGCTAGTTGGTGGGGTAAAGGCCTACCAAGGCGACGATCCA
CDB5                                GATGTGCCCAGATGGGATTAGCTAGTAGGTGGGGTAATGGCTCACCTAGGCGACGATCCC
Serratia marcescens                 GATGTGCCCAGATGGGATTAGCTAGTAGGTGGGGTAATGGCTCACCTAGGCGACGATCCC
Enterobacteriaceae bacterium A2JM   GATGTGCCCAGATGGGATTAGCTAGTAGGTGAGGTAATGGCTCACCTAGGCGACGATCCC
Kluyvera ascorbata                  GATGTGCCCAGATGGGATTAGCTAGTTGGTGAGGTAATGGCTCACCAAGGCGACGATCCC
CDB3                                GATGGACCCGCGTCGCATTAGCTAGTTGGTGAGGTAACGGCTCACCAAGGCAACGATGCG
CDB4                                GATGGACCCGCGTCGCATTAGCTAGTTGGTGAGGTAACGGCTCACCAAGGCAACGATGCG
B.mycoides                          GATGGACCCGCGTCGCATTAGCTAGTTGGTGAGGTAACGGCTCACCAAGGCGACGATGCG
Bacterium Te27R                     GATGGACCCGCGTCGCATTAGCTAGTTGGTGAGGTAACGGCTCACCAAGGCAACGATGCG
B.anthracis                         GATGGACCCGCGTCGCATTAGCTAGTTGGTGAGGTAACGGCTCACCAAGGCAACGATGCG
B.thuringinesis                     GATGGACCCGCGTCGCATTAGCTAGTTGGTGAGGTAACGGCTCACCAAGGCAACGATGCG
B.cereus G8639                      GATGGACCCGCGTCGCATTAGCTAGTTGGTGAGGTAACGGCTCACCAAGGCAACGATGCG
CDB1                                GATGGACTCGCGGCCTATCAGCTTGTTGGTGGGGTAATGGCCTACCAAGGCGACGACGGG
A.histidinilovorans                 GATGGACTCGCGGCCTATCAGCTTGTTGGTGGGGTAATGGCCTACCAAGGCGACGACGGG
Arthrobacter sp                     GATGGACTCGCGGCCTATCAGCTTGTTGGTGGGGTAATGGCCTACCAAGGCGACGACGGG
A. ilicis                           GATGGACTCGCGGCCTATCAGCTTGTTGGTGGGGTAATGGCCTACCAAGGCGACGACGGG
A.sulfonivorans                     GATGGACTCGCGGCCTATCAGCTTGTTGGTGAGGTAATGGCTCACCAAGGCGACGACGGG
A.polychromogenes                   GATGGACTCGCGGCCTATCAGCTTGTTGGTGAGGTAATGGCTTACCAAGGCGACGACGGG

                                    301       311       321       331       341       351
O.grignonense                       TAGCTGGTCTGAGAGGATGATCAGCCACACTGGGACTGAGACACGGCCCAGACTCCTACG
Ochrobactrum sp LMG20564            TAGCTGGTCTGAGAGGATGATCAGCCACACTGGGACTGAGACACGGCCCAGACTCCTACG
Ochrobactrum sp                     TAGCTGGTCTGAGAGGATGATCAGCCACACTGGGACTGAGACACGGCCCAGACTCCTACG
CDB2                                TAGCTGGTCTGAGAGGATGATCAGCCACACTGGGACTGAGACACGGCCCAGACTCCTACG
B.melitens                          TAGCTGGTCTGAGAGGATGATCAGCCACACTGGGACTGAGACACGGCCCAGACTCCTACG
B.canis                             TAGCTGGTCTGAGAGGATGATCAGCCACACTGGGACTGAGACACGGCCCAGACTCCTACG
O.anthropi                          TAGCTGGTCTGAGAGGATGATCAGCCACACTGGGACTGAGACACGGCCCAGACTCCTACG
Sinorhizobium sp TB8-711            TAGCTGGTCTGAGAGGATGATCAGCCACACTGGGACTGAGACACGGCCCAAACTCCTACG
Sinorhizobium fredii                TAGCTGGTCTGAGAGGATGATCAGCCACATTGGGACTGAGACACGGCCCAAACTCCTACG
CDB5                                TAGCTGGTCTGAGAGGATGACCAGCCACACTGGAACTGAGACACGGTCCAGACTCCTACG
Serratia marcescens                 TAGCTGGTCTGAGAGGATGACCAGCCACACTGGAACTGAGACACGGTCCAGACTCCTACG
Enterobacteriaceae bacterium A2JM   TAGCTGGTCTGAGAGGATGACCAGCCACACTGGAACTGAGACACGGTCCAGACTCCTACG
Kluyvera ascorbata                  TAGCTGGTCTGAGAGGATGACCAGCCACACTGGAACTGAGACACGGTCCAGACTCCTACG
```

```
CDB3                                TAGCCGACCTGAGAGGGTGATCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACG
CDB4                                TAGCCGACCTGAGAGGGTGATCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACG
B.mycoides                          TAGCCGACCTGAGAGGGTGATCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACG
Bacterium Te27R                     TAGCCGACCTGAGAGGGTGATCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACG
B.anthracis                         TAGCCGACCTGAGAGGGTGATCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACG
B.thuringinesis                     TAGCCGACCTGAGAGGGTGATCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACG
B.cereus G8639                      TAGCCGACCTGAGAGGGTGATCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACG
CDB1                                TAGCCGGCCTGAGAGGGTGACCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACG
A.histidinilovorans                 TAGCCGGCCTGAGAGGGTGACCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACG
Arthrobacter sp                     TAGCCGGCCTGAGAGGGTGACCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACG
A. ilicis                           TAGCCGGCCTGAGAGGGTGACCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACG
A.sulfonivorans                     TAGCCGGCCTGAGAGGGTGACCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACG
A.polychromogenes                   TAGCCGGCCTGAGAGGGTGACCGGCCACACTGGGACTGAGACACGGCCCAGACTCCTACG

                                    361        371        381        391        401        411
O.grignonense                       GGAGGCAGCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCCATGCCGCGTG
Ochrobactrum sp LMG20564            GGAGGCAGCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCCATGCCGCGTG
Ochrobactrum sp                     GGAGGCAGCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCCATGCCGCGTG
CDB2                                GGAGGCAGCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCCATGCCGCGTG
B.melitens                          GGAGGCAGCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCCATGCCGCGTG
B.canis                             GGAGGCAGCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCCATGCCGCGTG
O.anthropi                          GGAGGCAGCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCCATGCCGCGTG
Sinorhizobium sp TB8-711            GGAGGCAGCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCCATGCCGCGTG
Sinorhizobium fredii                GGAGGCAGCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCCATGCCGCGTG
CDB5                                GGAGGCAGCAGTGGGGAATATTGCACAATGGGCGCAAGCCTGATGCAGCCATGCCGCGTG
Serratia marcescens                 GGAGGCAGCAGTGGGGAATATTGCACAATGGGCGCAAGCCTGATGCAGCCATGCCGCGTG
Enterobacteriaceae bacterium A2JM   GGAGGCAGCAGTGGGGAATATTGCACAATGGGCGCAAGCCTGATGCAGCCATGCCGCGTG
Kluyvera ascorbata                  GGAGGCAGCAGTGGGGAATATTGCACAATGGGCGCAAGCCTGATGCAGCCATGCCGCGTG
CDB3                                GGAGGCAGCAGTAGGGAATCTTCCGCAATGGACGAAAGTCTGACGGAGCAACGCCGCGTG
CDB4                                GGAGGCAGCAGTAGGGAATCTTCCGCAATGGACGAAAGTCTGACGGAGCAACGCCGCGTG
B.mycoides                          GGAGGCAGCAGTAGGGAATCTTCCGCAATGGACGAAAGTCTGACGGAGCAACGCCGCGTG
Bacterium Te27R                     GGAGGCAGCAGTAGGGAATCTTCCGCAATGGACGAAAGTCTGACGGAGCAACGCCGCGTG
B.anthracis                         GGAGGCAGCAGTAGGGAATCTTCCGCAATGGACGAAAGTCTGACGGAGCAACGCCGCGTG
B.thuringinesis                     GGAGGCAGCAGTAGGGAATCTTCCGCAATGGACGAAAGTCTGACGGAGCAACGCCGCGTG
B.cereus G8639                      GGAGGCAGCAGTAGGGAATCTTCCGCAATGGACGAAAGTCTGACGGAGCAACGCCGCGTG
CDB1                                GGAGGCAGCAGTGGGGAATATTGCACAATGGGCGCAAGCCTGATGCAGCGACGCCGCGTG
A.histidinilovorans                 GGAGGCAGCAGTGGGGAATATTGCACAATGGGCGCAAGCCTGATGCAGCGACGCCGCGTG
Arthrobacter sp                     GGAGGCAGCAGTGGGGAATATTGCACAATGGGCGCAAGCCTGATGCAGCGACGCCGCGTG
A. ilicis                           GGAGGCAGCAGTGGGGAATATTGCACAATGGGCGCAAGCCTGATGCAGCGACGCCGCGTG
A.sulfonivorans                     GGAGGCAGCAGTGGGGAATATTGCACAATGGGCGAAAGCCTGATGCAGCGACGCCGCGTG
A.polychromogenes                   GGAGGCAGCAGTGGGGAATATTGCACAATGGGCGCAAGCCTGATGCAGCGACGCCGCGTG

                                    421        431        441        451        461        471
O.grignonense                       AGTGATGAAGGCCCTAGGGTTGTAAAGCTCTTTCACCGGTGAAGA-------------TA
Ochrobactrum sp LMG20564            AGTGATGAAGGTCTTAGGATTGTAAAGCTCTTTCACCGGTGAAGA-------------TA
Ochrobactrum sp                     AGTGATGAAGGCCCTAGGGTTGTAAAGCTCTTTCACCGGTGAAGA-------------TA
CDB2                                AGTGATGAAGGCCCTAGGGTTGTAAAGCTCTTTCACCGGTGAAGA-------------TA
B.melitens                          AGTGATGAAGGCCCTAGGGTTGTAAAGCTCTTTCACCGGTGAAGA-------------TA
B.canis                             AGTGATGAAGGCCCTAGGGTTGTAAAGCTCTTTCACCGGTGAAGA-------------TA
O.anthropi                          AGTGATGAAGGCCCTAGGGTTGTAAAGCTCTTTCACCGGTGAAGA-------------TA
Sinorhizobium sp TB8-711            AGTGATGAAGGCCCTAGGGTTGTAAAGCTCTTTCACCGGTGAAGA-------------TA
Sinorhizobium fredii                AGTGATGAAGGCCCTAGGGTTGTAAAGCTCTTTCACCGGTGAAGA-------------TA
CDB5                                TGTGAAGAAGGCCTTCGGGTTGTAAAGCACTTTCAGCGAGGAGGAAGGTG-GTGAACTTA
Serratia marcescens                 TGTGAAGAAGGCCTTCGGGTTGTAAAGCACTTTCAGCGAGGAGGAAGGTG-GTGAGCTTA
Enterobacteriaceae bacterium A2JM   TGTGAAGAAGGCCTTCGGGTTGTAAAGCACTTTCAGCGGGGAGGAAGGCG-GTGAGATTA
Kluyvera ascorbata                  TATGAAGAAGGCCTTCGGGTTGTAAAGTACTTTCAGCGAGGAGGAAGGCA-TTAAGGTTA
CDB3                                AGTGATGAAGGCTTTCGGGTCGTAAAACTCTGTTGTTAGGGAAGAACAAGTGCTAGTTGA
CDB4                                AGTGATGAAGGCTTTCGGGTCGTAAAACTCTGTTGTTAGGGAAGAACAAGTGCTAGTTGA
B.mycoides                          AGTGATGAAGGCTTTCGGGTCGTAAAACTCTGTTGTTAGGGAAGAACAAGTGCTAGTTGA
Bacterium Te27R                     AGTGATGAAGGCTTTCGGGTCGTAAAACTCTGTTGTTAGGGAAGAACAAGTGCTAGTTGA
B.anthracis                         AGTGATGAAGGCTTTCGGGTCGTAAAACTCTGTTGTTAGGGAAGAACAAGTGCTAGTTGA
B.thuringinesis                     AGTGATGAAGGCTTTCGGGTCGTAAAACTCTGTTGTTAGGGAAGAACAAGTGCTAGTTGA
B.cereus G8639                      AGTGATGAAGGCTTTCGGGTCGTAAAACTCTGTTGTTAGGGAAGAACAAGTGCTAGTTGA
CDB1                                AGGGATGACGGCCTTCGGGTTGTAAACCTCTTTCAGTAGGGAAGAA-----GC-----G-
A.histidinilovorans                 AGGGATGACGGCCTTCGGGTTGTAAACCTCTTTCAGTAGGGAAGAA-----GC-----G-
Arthrobacter sp                     AGGGATGACGGCCTTCGGGTTGTAAACCTCTTTCAGTAGGGAAGAA-----GC-----G-
A. ilicis                           AGGGATGACGGCCTTCGGGTTGTAAACCTCTTTCAGTAGGGAAGAA-----GC-----CG
A.sulfonivorans                     AGGGATGACGGCCTTCGGGTTGTAAACCTCTTTCAGTAGGGAAGAA-----GC-----G-
A.polychromogenes                   AGGGATGACGGCCTTCGGGTTGTAAACCTCTTTCAGTAGGGAAGAA-----GC-----G-

                                    481        491        501        511        521        531
O.grignonense                       AT-------------GACGGTAACCGGAGAAGAAGCCCCGGCTAACTTCGTGCCAGCAGC
Ochrobactrum sp LMG20564            AT-------------GACGGTAACCGGAGAAGAAGCCCCGGCTAACTTCGTGCCAGCAGC
Ochrobactrum sp                     AT-------------GACGGTAACCGGAGAAGAAGCCCCGGCTAACTTCGTGCCAGCAGC
CDB2                                AT-------------GACGGTAACCGGAGAAGAAGCCCCGGCTAACTTCGTGCCAGCAGC
B.melitens                          AT-------------GACGGTAACCGGAGAAGAAGCCCCGGCTAACTTCGTGCCAGCAGC
B.canis                             AT-------------GACGGTAACCGGAGAAGAAGCCCCGGCTAACTTCGTGCCAGCAGC
O.anthropi                          AT-------------GACGGTAACCGGAGAAGAAGCCCCGGCTAACTTCGTGCCAGCAGC
Sinorhizobium sp TB8-711            AT-------------GACGGTAACCGGAGAAGAAGCCCCGGCTAACTTCGTGCCAGCAGC
Sinorhizobium fredii                AT-------------GACGGTAACCGGAGAAGAAGCCCCGGCTAACTTCGTGCCAGCAGC
CDB5                                ATACGCTCATCAATTGACGTTACTCGCAGAAGAAGCACCGGCTAACTCCGTGCCAGCAGC
```

```
Serratia marcescens                ATACGCTCATCAATTGACGTTACTCGCAGAAGAAGCACCGGCTAACTCCGTGCCAGCAGC
Enterobacteriaceae bacterium A2JM  ATACTCTCACCGATTGACGTTACCCGCAGAAGAAGCACCGGCTAACTCCGTGCCAGCAGC
Kluyvera ascorbata                 ATAACCTTAGTGATTGACGTTATTCGCAGAAGAAGCACCGGCTAACTCCGTGCCAGCAGC
CDB3                               ATAAGCTGGCACCTTGACGGTACCTAACCAGAAAGCCACGGCTAACTACGTGCCAGCAGC
CDB4                               ATAAGCTGGCACCTTGACGGTACCTAACCAGAAAGCCACGGCTAACTACGTGCCAGCAGC
B.mycoides                         ATAAGCTGGCACCTTGACGGTACCTAACCAGAAAGCCACGGCTAACTACGTGCCAGCAGC
Bacterium Te27R                    ATAAGCTGGCACCTTGACGGTACCTAACCAGAAAGCCACGGCTAACTACGTGCCAGCAGC
B.anthracis                        ATAAGCTGGCACCTTGACGGTACCTAACCAGAAAGCCACGGCTAACTACGTGCCAGCAGC
B.thuringinesis                    ATAAGCTGGCACCTTGACGGTACCTAACCAGAAAGCCACGGCTAACTACGTGCCAGCAGC
B.cereus G8639                     ATAAGCTGGCACCTTGACGGTACCTAACCAGAAAGCCACGGCTAACTACGTGCCAGCAGC
CDB1                               -TAAG---------TGACGGTACCTGCAGAAGAAGCGCCGGCTAACTACGTGCCAGCAGC
A.histidinilovorans                -TAAG---------TGACGGTACCTGCAGAAGAAGCGCCGGCTAACTACGTGCCAGCAGC
Arthrobacter sp                    -AAAG---------TGACGGTACCTGCAGAAGAAGCGCCGGCTAACTACGTGCCAGCAGC
A. ilicis                          TAAGG---------TGACGGTACCTGCAGAAGAAGCGCCGGCTAACTACGTGCCAGCAGC
A.sulfonivorans                    -AAAG---------TGACGGTACCTGCAGAAGAAGCGCCGGCTAACTACGTGCCAGCAGC
A.polychromogenes                  -AAAG---------TGACGGTACCTGCAGAAGAAGCGCCGGCTAACTACGTGCCAGCAGC

                                   541       551       561       571       581       591
O.grignonense                      CGCGGTAATACGAAGGGGGCTAGCGTTGTTCGGATTTACTGGGCGTAAAGCGCACGTAGG
Ochrobactrum sp LMG20564           CGCGGTAATACGAAGGGGGCTAGCGTTGTTCGGATTTACTGGGCGTAAAGCGCACGTAGG
Ochrobactrum sp                    CGCGGTAATACGAAGGGGGCTAGCGTTGTTCGGATTTACTGGGCGTAAAGCGCACGTAGG
CDB2                               CGCGGTAATACGAAGGGGGCTAGCGTTGTTCGGATTTACTGGGCGTAAAGCGCACGTAGG
B.melitens                         CGCGGTAATACGAAGGGGGCTAGCGTTGTTCGGATTTACTGGGCGTAAAGCGCACGTAGG
B.canis                            CGCGGTAATACGAAGGGGGCTAGCGTTGTTCGGATTTACTGGGCGTAAAGCGCACGTAGG
O.anthropi                         CGCGGTAATACGAAGGGGGCTAGCGTTGTTCGGATTTACTGGGCGTAAAGCGCACGTAGG
Sinorhizobium sp TB8-711           CGCGGTAATACGAAGGGGGCTAGCGTTGTTCGGAATTACTGGGCGTAAAGCGCACGTAGG
Sinorhizobium fredii               CGCGGTAATACGAAGGGGGCTAGCGTTGTTCGGAATTACTGGGCGTAAAGCGCACGTAGG
CDB5                               CGCGGTAATACGGAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGCGCACGCAGG
Serratia marcescens                CGCGGTAATACGGAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGCGCACGCAGG
Enterobacteriaceae bacterium A2JM  CGCGGTAATACGGAGGGTGCAAGCGTTAATCGGAATGACTGGGCGTAAAGCGCACGCAGG
Kluyvera ascorbata                 CGCGGTAATACGGAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGCGCACGCAGG
CDB3                               CGCGGTAATACGTAGGTGGCAAGCGTTATCCGGAATTATTGGGCGTAAAGCGCGCGCAGG
CDB4                               CGCGGTAATACGTAGGTGGCAAGCGTTATCCGGAATTATTGGGCGTAAAGCGCGCGCAGG
B.mycoides                         CGCGGTAATACGTAGGTGGCAAGCGTTATCCGGAATTATTGGGCGTAAAGCGCGCGCAGG
Bacterium Te27R                    CGCGGTAATACGTAGGTGGCAAGCGTTATCCGGAATTATTGGGCGTAAAGCGCGCGCAGG
B.anthracis                        CGCGGTAATACGTAGGTGGCAAGCGTTATCCGGAATTATTGGGCGTAAAGCGCGCGCAGG
B.thuringinesis                    CGCGGTAATACGTAGGTGGCAAGCGTTATCCGGAATTATTGGGCGTAAAGCGCGCGCAGG
B.cereus G8639                     CGCGGTAATACGTAGGTGGCAAGCGTTATCCGGAATTATTGGGCGTAAAGCGCGCGCAGG
CDB1                               CGCGGTAATACGTAGGGCGCAAGCGTTATCCGGAATTATTGGGCGTAAAGAGCTCGTAGG
A.histidinolovorans                CGCGGTAATACGTAGGGCGCAAGCGTTATCCGGAATTATTGGGCGTAAAGAGCTCGTAGG
Arthrobacter sp                    CGCGGTAATACGTAGGGCGCAAGCGTTATCCGGAATTATTGGGCGTAAAGAGCTCGTAGG
A. ilicis                          CGCGGTAATACGTAGGGCGCAAGCGTTATCCGGAATTATTGGGCGTAAAGAGCTCGTAGG
A.sulfonivorans                    CGCGGTAATACGTAGGGCGCAAGCGTTATCCGGAATTATTGGGCGTAAAGAGCTCGTAGG
A.polychromogenes                  CGCGGTAATACGTAGGGCGCAAGCGTTATCCGGAATTATTGGGCGTAAAGAGCTCGTAGG

                                   601       611       621       631       641       651
O.grignonense                      CGGACTTTTAAGTCAGGGGTGAAATCCCGGGGCTCAACCCCGGAACTGCCTTTGATACTG
Ochrobactrum sp LMG20564           CGGACTTTTAAGTCAGGGGTGAAATCCCGGGGCTCAACCCCGGAACTGCCTTTGATACTG
Ochrobactrum sp                    CGGACTTTTAAGTCAGGGGTGAAATCCCAGAGCTCAACTCTGGAACTGCCTTTGATACTG
CDB2                               CGGACTTTTAAGTCAGGGGTGAAATCCCAGAGCTCAACTCCGGAACTGCCTTTGATACTG
B.melitens                         CGGACTTTTAAGTCAGGGGTGAAATCCCGGGGCTCAACCCCGGAACTGCCTTTGATACTG
B.canis                            CGGACTTTTAAGTCAGGGGTGAAATCCCGGGGCTCAACCCCGGAACTGCCTTTGATACTG
O.anthropi                         CGGACTTTTAAGTCAGGGGTGAAATCCCGGGGCTCAACCCCGGAACTGCCTTTGATACTG
Sinorhizobium sp TB8-711           CGGACATTTAAGTCAGGGGTGAAATCCCAGAGCTCAACTCTGGAACTGCCTTTGATACTG
Sinorhizobium fredii               CGGACATTTAAGTCAGGGGTGAAATCCCGGGGCTCAACCCCGGAACTGCCTTTGATACTG
CDB5                               CGGTTTGTTAAGTCAGATGTGAAATCCCCGGGCTCAACCTGGGAACTGCATTTGAAACTG
Serratia marcescens                CGGTTTGTTAAGTCAGATGTGAAATCCCCGGGCTCAACCTGGGAACTGCATTTGAAACTG
Enterobacteriaceae bacterium A2JM  CGGTTTGTTAAGTCAGATGTGAAATCCCCGAGCTTAACTTGGGAACTGCATTTGAAACTG
Kluyvera ascorbata                 CGGTCTGTCAAGTCGGATGTGAAATCCCCGGGCTCAACCTGGGAACTGCATTCGAAACTG
CDB3                               TGGTTTCTTAAGTCTGATGTGAAAGCCCACGGCTCAACCGTGGAGGGTCATTGGAAACTG
CDB4                               TGGTTTCTTAAGTCTGATGTGAAAGCCCACGGCTCAACCGTGGAGGGTCATTGGAAACTG
B.mycoides                         TGGTTTCTTAAGTCTGATGTGAAAGCCCACGGCTCAACCGTGGAGGGTCATTGGAAACTG
Bacterium Te27R                    TGGTTTCTTAAGTCTGATGTGAAAGCCCACGGCTCAACCGTGGAGGGTCATTGGAAACTG
B.anthracis                        TGGTTTCTTAAGTCTGATGTGAAAGCCCACGGCTCAACCGTGGAGGGTCATTGGAAACTG
B.thuringinesis                    TGGTTTCTTAAGTCTGATGTGAAAGCCCACGGCTCAACCGTGGAGGGTCATTGGAAACTG
B.cereus G8639                     TGGTTTCTTAAGTCTGATGTGAAAGCCCACGGCTCAACCGTGGAGGGTCATTGGAAACTG
CDB1                               CGGTTTGTCGCGTCTGCTGTGAAAGACCGGGGCTCAACTCCGGTTCTGCAGTGGGTACGG
A.histidinilovorans                CGGTTTGTCGCGTCTTCTGTGAAAGACCGGGGCTCAACTCCGGTTCTCCAGTGGGTACGG
Arthrobacter sp                    CGGTTTGTCGCGTCTGCTGTGAAAGACCGGGGCTCAACTCCGGTTCTGCAGTGGGTACGG
A. ilicis                          CGGTTTGTCGCGTCTGCTGTGAAAGACCGGGGCTCAACTCCGGTTCTGCAGTGGGTACGG
A.sulfonivorans                    CGGTTTGTCGCGTCTGCCGTGAAAGTCCGGGGCTCAACTCCGGATCTGCGGTGGGTACGG
A.polychromogenes                  CGGTTTGTCGCGTCTGCCGTGAAAGTCCGGGGCTCAACTCCGGATCTGCGGTGGGTACGG

                                   661       671       681       691       701       711
O.grignonense                      GAAGTCTTGAGTATGGTAGAGGTGAGTGGAATTCCGAGTGTAGAGGTGAAATTCGTAGAT
Ochrobactrum sp LMG20564           GAAGTCTTGAGTATGGTAGAGGTGAGTGGAATTCCGAGTGTAGAGGTGAAATTCGTAGAT
Ochrobactrum sp                    GAAGTCTTGAGTATGGTAGAGGTGAGTGGAATTCCGAGTGTAGAGGTGAAATTCGTAGAT
CDB2                               GAAGTCTTGAGTATGGTAGAGGTGAGTGGAATTCCGAGTGTAGAGGTGAAATTCGTAGAT
B.melitens                         GAAGTCTTGAGTATGGTAGAGGTGAGTGGAATTCCGAGTGTAGAGGTGAAATTCGTAGAT
B.canis                            GAAGTCTTGAGTATGGTAGAGGTGAGTGGAATTCCGAGTGTAGAGGTGAAATTCGTAGAT
O.anthropi                         GAAGTCTTGAGTATGGTAGAGGTGAGTGGAATTCCGAGTGTAGAGGTGAAATTCGTAGAT
```

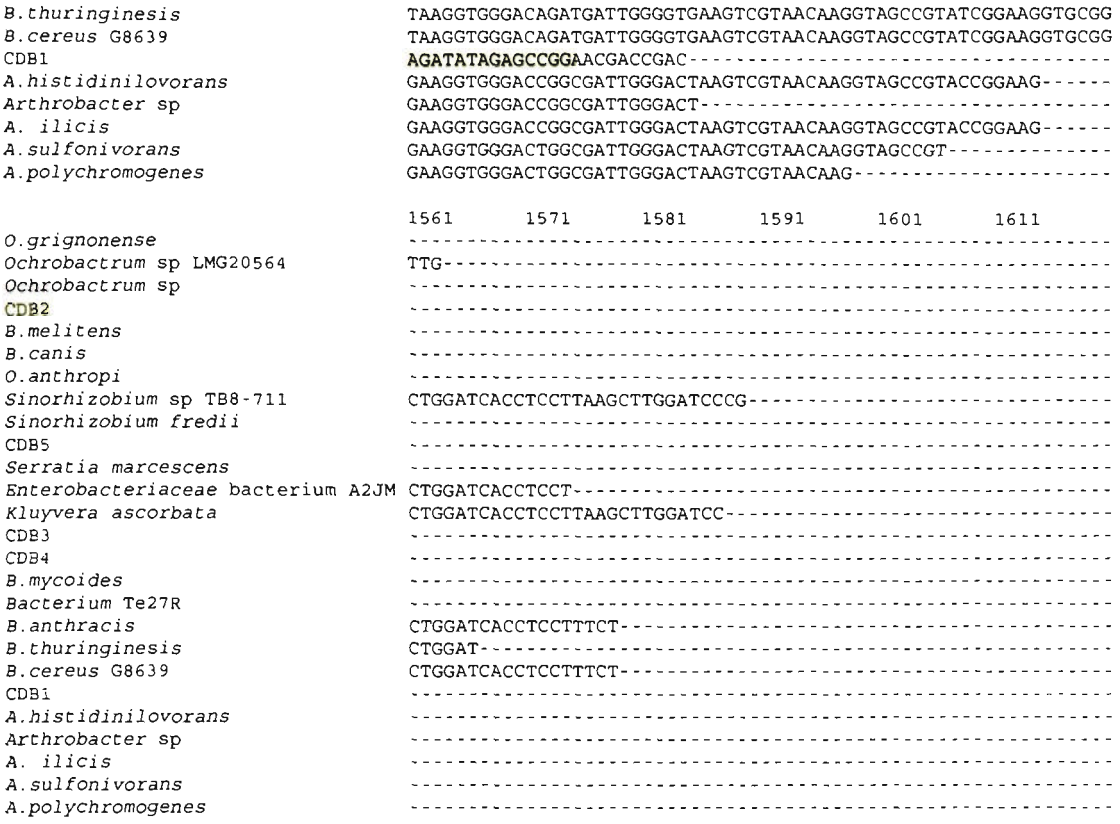
```
Sinorhizobium sp TB8-711          GGTGTCTAGAGTATGGAAGAGGTGAGTGGAATTCCGAGTGTAGAGGTGAAATTCGTAGAT
Sinorhizobium fredii              GGTGTCTAGAGTCCGGAAGAGGTGAGTGGAATTCCGAGTGTAGAGGTGAAATTCGTAGAT
CDB5                              GCAAGCTAGAGTCTCGTAGAGGGGGGTAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAG
Serratia marcescens               GCAAGCTAGAGTCTCGTAGAGGGGGGTAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAG
Enterobacteriaceae bacterium A2JM GCAAGCTAGAGTCTTGTAGAGGGGGGTAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAG
Kluyvera ascorbata                GCAGGCTAGAGTCTTGTAGAGGGGGGTAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAG
CDB3                              GGAGACTTGAGTGCAGAAGAGGAAAGTGGAATTCCATGTGTAGCGGTGAAATGCGTAGAG
CDB4                              GGAGACTTGAGTGCAGAAGAGGAAAGTGGAATTCCATGTGTAGCGGTGAAATGCGTAGAG
B.mycoides                        GGAGACTTGAGTGCAGAAGAGGAAAGTGGAATTCCATGTGTAGCGGTGAAATGCGTAGAG
Bacterium Te27R                   GGAGACTTGAGTGCAGAAGAGGAAAGTGGAATTCCATGTGTAGCGGTGAAATGCGTAGAG
B.anthracis                       GGAGACTTGAGTGCAGAAGAGGAAAGTGGAATTCCATGTGTAGCGGTGAAATGCGTAGAG
B.thuringinesis                   GGAGACTTGAGTGCAGAAGAGGAAAGTGGAATTCCATGTGTAGCGGTGAAATGCGTAGAG
B.cereus G8639                    GGAGACTTGAGTGCAGAAGAGGAAAGTGGAATTCCATGTGTAGCGGTGAAATGCGTAGAG
CDB1                              GCAGACTAGAGTGCAGTAGGGGAGACTGGAATTCCTGGTGTAGCGGTGAAATGCGCAGAT
A.histidinilovorans               GCAGACTAGAGTGCAGTAGGGGAGACTGGAATTCCTGGTGTAGCGGTGAAATGCGCAGAT
Arthrobacter sp                   GCAGACTAGAGTGCAGTAGGGGAGACTGGAATTCCTGGTGTAGCGGTGAAATGCGCAGAT
A. ilicis                         GCAGACTAGAGTGCAGCAGGGGAGACTGGAATTCCTGGTGTAGCGGCGAAATGCGCAGAT
A.sulfonivorans                   GCAGACTAGAGTGATGTAGGGGAGACTGGAATTCCTGGTGTAGCGGTGAAATGCGCAGAT
A.polychromogenes                 GCAGACTAGAGTGATGTAGGGGAGACTGGAATTCCTGGTGTAGCGGTGAAATGCGCAGAT

                                  721       731       741       751       761       771
O.grignonense                     ATTCGGAGGAACACCAGTGGCGAAGGCGGCTCACTGGACCATTACTGACGCTGAGGTGCG
Ochrobactrum sp LMG20564          ATTCGGAGGAACACCAGTGGCGAAGGCGGCTCACTGGACCATTACTGACGCTGAGGTGCG
Ochrobactrum sp                   ATTCGGAGGAACACCAGTGGCGAAGGCGGCTCACTGGACCATTACTGACGCTGAGGTGCG
CDB2                              ATTCGGAGGAACACCAGTGGCGAAGGCGGCTCACTGGACCATTACTGACGCTGAGGTGCG
B.melitens                        ATTCGGAGGAACACCAGTGGCGAAGGCGGCTCACTGGACCATTACTGACGCTGAGGTGCG
B.canis                           ATTCGGAGGAACACCAGTGGCGAAGGCGGCTCACTGGACCATTACTGACGCTGAGGTGCG
O.anthropi                        ATTCGGAGGAACACCAGTGGCGAAGGCGGCTCACTGGACCATTACTGACGCTGAGGTGCG
Sinorhizobium sp TB8-711          ATTCGGAGGAACACCAGTGGCGAAGGCGGCTCACTGGTCCATTACTGACGCTGAGGTGCG
Sinorhizobium fredii              ATTCGGAGGAACACCAGTGGCGAAGGCGGCTCACTGGTCCGGTACTGACGCTGAGGTGCG
CDB5                              ATCTGGAGGAATACCGGTGGCGAAGGCGGCCCCCTGGACGAAGACTGACGCTCAGGTGCG
Serratia marcescens               ATCTGGAGGAATACCGGTGGCGAAGGCGGCCCCCTGGACGAAGACTGACGCTCAGGTGCG
Enterobacteriaceae bacterium A2JM ATCTGGAGGAATACCGGTGGCGAAGGCGGCCCCCTGGACAAAGACTGACGCTCAGGTGCG
Kluyvera ascorbata                ATCTGGAGGAATACCGGTGGCGAAGGCGGCCCCCTGGACAAAGACTGACGCTCAGGTGCG
CDB3                              ATATGGAGGAACACCAGTGGCGAAGGCGACTTTCTGGTCTGTAACTGACACTGAGGCGCG
CDB4                              ATATGGAGGAACACCAGTGGCGAAGGCGACTTTCTGGTCTGTAACTGACACTGAGGCGCG
B.mycoides                        ATATGGAGGAACACCAGTGGCGAAGGCGACTTTCTGGTCTGTAACTGACACTGAGGCGCG
Bacterium Te27R                   ATATGGAGGAACACCAGTGGCGAAGGCGACTTTCTGGTCTGTAACTGACACTGAGGCGCG
B.anthracis                       ATATGGAGGAACACCAGTGGCGAAGGCGACTTTCTGGTCTGTAACTGACACTGAGGCGCG
B.thuringinesis                   ATATGGAGGAACACCAGTGGCGAAGGCGACTTTCTGGTCTGTAACTGACACTGAGGCGCG
B.cereus G8639                    ATATGGAGGAACACCAGTGGCGAAGGCGACTTTCTGGTCTGTAACTGACACTGAGGCGCG
CDB1                              ATCAGGAGGAACACCGATGGCGAAGGCAGGTCTCTGGGCTGTAACTGACGCTGAGGAGCG
A.histidinilovorans               ATCAGGAGGAACACCCATGGCGAAGGCAGGTCTCTGGGCTGTAACTGACGCTGAGGAGCG
Arthrobacter sp                   ATCAGGAGGAACACCGATGGCGAAGGCAGGTCTCTGGGCTGTAACTGACGCTGAGGAGCG
A. ilicis                         ATCAGGAGGAACACCCATGGCGAAGGCAGGTCTCTGGGCTGTAACTGACGCTGAGGAGCG
A.sulfonivorans                   ATCAGGAGGAACACCGATGGCGAAGGCAGGTCTCTGGGCATTAACTGACGCTGAGGAGCG
A.polychromogenes                 ATCAGGAGGAACACCGATGGCGAAGGCAGGTCTCTGGGCATTAACTGACGCTGAGGAGCG

                                  781       791       801       811       821       831
O.grignonense                     AAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGAATG
Ochrobactrum sp LMG20564          AAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGAATG
Ochrobactrum sp                   AAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGAATG
CDB2                              AAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGAATG
B.melitens                        AAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGAATG
B.canis                           AAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGAATG
O.anthropi                        AAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGAATG
Sinorhizobium sp TB8-711          AAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGAATG
Sinorhizobium fredii              AAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGAATG
CDB5                              AAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCTGTAAACGATGTCGA
Serratia marcescens               AAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCTGTAAACGATGTCGA
Enterobacteriaceae bacterium A2JM AAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCTGTAAACGATGTCGA
Kluyvera ascorbata                AAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGTCTA
CDB3                              AAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGAGTG
CDB4                              AAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGAGTG
B.mycoides                        AAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGAGTG
Bacterium Te27R                   AAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGAGTG
B.anthracis                       AAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGAGTG
B.thuringinesis                   AAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGAGTG
B.cereus G8639                    AAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGAGTG
CDB1                              AAAGCATGGGGAGCGAACAGGATTAGATACCCTGGTAGTCCATGCCGTAAACGTTGGGCA
A.histidinilovorans               AAAGCATGGGGAGCGAACAGGATTAGATACCCTGGTAGTCCATGCCGTAAACGTTGGGCA
Arthrobacter sp                   AAAGCATGGGGAGCGAACAGGATTAGATACCCTGGTAGTCCATGCCGTAAACGTTGGGCA
A. ilicis                         AAAGCTAGGGGAGCGAACAGGATTAGATACCCTGGTAGTCCATGCCGTAAACGTTGGGCA
A.sulfonivorans                   AAAGCATGGGGAGCGAACAGGATTAGATACCCTGGTAGTCCATGCCGTAAACGTTGGGCA
A.polychromogenes                 AAAGCATGGGGAGCGAACAGGATTAGATACCCTGGTAGTCCATGCCGTAAACGTTGGGCA

                                  841       851       861       871       881       891
O.grignonense                     TTAGCCG-TCGGGGAGTTT-ACTCTT-CGGTGGCGCAGCTAA-CGCATTAAACATTCCGC
Ochrobactrum sp LMG20564          TTAGCCG-TCGGGGAGTTT-ACTCTT-CGGTGGCGCAGCTAA-CGCATTAAACATTCCGC
Ochrobactrum sp                   TTAGCCG-TCGGGGTGTTT-ACACTT-CGGTGGCGCANNTAA-CGCATTAAACATTCCGC
CDB2                              TTAGCCG-TCGGGGTGTTT-ACACTTACGGTGGCGCAGCTAATCGCATTAAACATTCCGC
```


B.melitens TTAGCCG-TCGGGGTGTTT-ACACTT-CGGTGGCGCAGCTAA-CGCATTAAACATTCCGC
B.canis TTAGCCG-TCGGGGTGTTT-ACACTT-CGGTGGCGCAGCTAA-CGCATTAAACATTCCGC
O.anthropi TTAGCCG-TTGGGGAGTTT-ACTCTT-CGGTGGCGCAGCTAA-CGCATTAAACATTCCGC
Sinorhizobium sp TB8-711 TTAGCCG-TCGGGCAGTTT-ACTGTT-CGGTGGCGCAGCTAA-CGCATTAAACATTCCGC
Sinorhizobium fredii TTAGCCG-TCGGGCAGTTT-ACTGTT-CGGTGGCGCAGCTAA-CGCATTAAACATTCCGC
CDB5 TTTGGAGGTTGTGCCCTTG-AGGCGT-GGCTTCCGGAGCTAA-CGCGTTAAATCGACCGC
Serratia marcescens TTTGGAGGTTGTGCCCTTG-AGGCGT-GGCTTCCGGAGCTAA-CGCGTTAAATCGACCGC
Enterobacteriaceae bacterium A2JM TTTGGAGGTTGTGCCCTTG-AGGCGT-GGCTTCCGGAGCTAA-CGCGTTAAATCGACCGC
Kluyvera ascorbata TTTGGAGGTTGTTCCCTTG-AGGAGT-GGCTTCCGGAGCTAA-CGCGTTAAATAGACCGC
CDB3 CTAAGTGTTAGAGGGTTTCCGCCCTTTA-GTGCTGAAGTTAA-CGCATTAAGCACTCCGC
CDB4 CTAAGTGTTAGAGGGTTTCCGCCCTTTA-GTGCTGAAGTTAA-CGCATTAAGCACTCCGC
B.mycoides CTAAGTGTTAGAGGGTTTCCGCCCTTTA-GTGCTGAAGTTAA-CGCATTAAGCACTCCGC
Bacterium Te27R CTAAGTGTTAGAGGGTTTCCGCCCTTTA-GTGCTGAAGTTAA-CGCATTAAGCACTCCGC
B.anthracis CTAAGTGTTAGAGGGTTTCCGCCCTTTA-GTGCTGAAGTTAA-CGCATTAAGCACTCCGC
B.thuringinesis CTAAGTGTTAGAGGGTTTCCGCCCTTTA-GTGCTGAAGTTAA-CGCATTAAGCACTCCGC
B.cereus G8639 CTAAGTGTTAGAGGGTTTCCGCCCTTTA-GTGCTGAAGTTAA-CGCATTAAGCACTCCGC
CDB1 CTAGGTGTGGGGGACATTCCACGTTTTCCGCGCCGTAGCTAA-CGCATTAAGTGCCCCGC
A.histidinilovorans CTAGGTGTGGGGGACATTCCACGTTTTCCGCGCCGTAGCTAA-CGCATTAAGTGCCCCGC
Arthrobacter sp CTAGGTGTGGGGGACATTCCACGTTTTCCGCGCCGTAGCTAA-CGCATTAAGTGCCCCGC
A. ilicis CTAGGTGTGGGGGACATTCCACGTTTTCCGCGCCGTAGCTAA-CGCATTAAGTGCCCCGC
A.sulfonivorans CTAGGTGTGGGGGACATTCCACGTTTTCCGCGCCGTAGCTAA-CGCATTAAGTGCCCCGC
A.polychromogenes CTAGGTGTGGGGGACATTCCACGTTTTCCGCGCCGTAGCTAA-CGCATTAAGTGCCCCGC
901 911 921 931 941 951
O.grignonense CTGGGGAGTACGGTCGCAAGATTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGG
Ochrobactrum sp LMG20564 CTGGGGAGTACGGTCGCAAGATTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGG
Ochrobactrum sp CTGGGGAGTACGGTCGCAAGATTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGG
CDB2 CTGGGGAGTACGGTCGCAAGATTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGG
B.melitens CTGGGGAGTACGGTCGCAAGATTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGG
B.canis CTGGGGAGTACGGTCGCAAGATTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGG
O.anthropi CTGGGGAGTACGGTCGCAAGATTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGG
Sinorhizobium sp TB8-711 CTGGGGAGTACGGTCGCAAGATTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGG
Sinorhizobium fredii CTGGGGAGTACGGTCGCAAGATTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGG
CDB5 CTGGGGAGT-CGGCCGCAAGGTTAAA-CTCAAATGAATTGACGGGGGCCCGCACAAGCGG
Serratia marcescens CTGGGGAGTACGGCCGCAAGGTTAAAACTCAAATGAATTGACGGGGGCCCGCACAAGCGG
Enterobacteriaceae bacterium A2JM CTGGGGAGTACGGCCGCAAGGTTAAAACTCAAATGAATTGACGGGGGCCCGCACAAGCGG
Kluyvera ascorbata CTGGGGAGTACGGCCGCAAGGTTAAAACTCAAATGAATTGACGGGGGCCCGCACAAGCGG
CDB3 CTGGGGAGTACGGCCGCAAGGCTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGG
CDB4 CTGGGGAGTACGGCCGCAAGGCTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGG
B.mycoides CTGGGGAGTACGGCCGCAAGGCTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGG
Bacterium Te27R CTGGGGAGTACGGCCGCAAGGCTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGG
B.anthracis CTGGGGAGTACGGCCGCAAGGCTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGG
B.thuringinesis CTGGGGAGTACGGCCGCAAGGCTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGG
B.cereus G8639 CTGGGGAGTACGGCCGCAAGGCTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGG
CDB1 CTGGGGAGTACGGCCGCAAGGCTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGG
A.histidinilovorans CTGGGGAGTACGGCCGCAAGGCTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGG
Arthrobacter sp CTGGGGAGTACGGCCGCAAGGCTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGG
A. ilicis CTGGGGAGTACGGCCGCAAGGCTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGG
A.sulfonivorans CTGGGGAGTACGGCCGCAAGGCTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGG
A.polychromogenes CTGGGGAGTACGGCCGCAAGGCTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGG
961 971 981 991 1001 1011
O.grignonense TGGAGCATGTGGTTTAATTCGAAGCAACGCGCAGAACCTTACCAGCCCTTGAGATACCGG
Ochrobactrum sp LMG20564 TGGAGCATGTGGTTTAATTCGAAGCAACGCGCAGAACCTTACCAGCCCTTGACATACCGG
Ochrobactrum sp TGGAGCATGTGGTTTAATTCGAAGCAACGCGCAGAACCTTACCAGCCCTTGACATACCGG
CDB2 TGGAGCATGTGGTTTAATTCGAAGCAACGCGCAGAACCTTACCAGCCCTTGACATACCGG
B.melitens TGGAGCATGTGGTTTAATTCGAAGCAACGCGCAGAACCTTACCAGCCCTTGACATCCCGG
B.canis TGGAGCATGTGGTTTAATTCGAAGCAACGCGCAGAACCTTACCAGCCCTTGACATCCCGG
O.anthropi TGGAGCATGTGGTTTAATTCGAAGCAACGCGCAGAACCTTACCAGCCCTTGACATACCGG
Sinorhizobium sp TB8-711 TGGAGCATGTGGTTTAATTCGAAGCAACGCGCAGAACCTTACCAGCCCTTGACATCCCGA
Sinorhizobium fredii TGGAGCATGTGGTTTAATTCGAAGCAACGCGCAGAACCTTACCAGCCCTTGACATCCCGA
CDB5 TGGAGCATGTGGTTTAATTCGATGCAACGCGAAGAACCTTACCTACTCTTGACATCCAGA
Serratia marcescens TGGAGCATGTGGTTTAATTCGATGCAACGCGAAGAACCTTACCTACTCTTGACATCCAGA
Enterobacteriaceae bacterium A2JM TGGAGCATGTGGTTTAATTCGATGCAACGCGAAGAACCTTACCTACTCTTGACATCCAGA
Kluyvera ascorbata TGGAGCATGTGGTTTAATTCGATGCAACGCGAAGAACCTTACCTACTCTTGACATCCAGA
CDB3 TGGAGCATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCAGGTCTTGACATCCTCT
CDB4 TGGAGCATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCAGGTCTTGACATCCTCT
B.mycoides TGGAGCATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCAGGTCTTGACATCCTCT
Bacterium Te27R TGGAGCATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCAGGTCTTGACATCCTCT
B.anthracis TGGAGCATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCAGGTCTTGACATCCTCT
B.thuringinesis TGGAGCATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCAGGTCTTGACATCCTCT
B.cereus G8639 TGGAGCATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCAGGTCTTGACATCCTCT
CDB1 CGGAGCATGCGGATTAATTCGATGCAACGCGAAGAACCTTACCAAGGCTTGACATGAACC
A.histidinilovorans CGGAGCATGCGGATTAATTCGATGCAACGCGAAGAACCTTACCAAGGCTTGACATGAACC
Arthrobacter sp CGGAGCATGCGGATTAATTCGATGCAACGCGAAGAACCTTACCAAGGCTTGACATGAACC
A. ilicis CGGAGCATGCGGATTAATTCGATGCAACGCGAAGAACCTTACCAAGGCTTGACATGAACC
A.sulfonivorans CGGAGCATGCGGATTAATTCGATGCAACGCGAAGAACCTTACCAAGGCTTGACATGGACT
A.polychromogenes CGGAGCATGCGGATTAATTCGATGCAACGCGAAGAACCTTACCAAGGCTTGACATGAACC

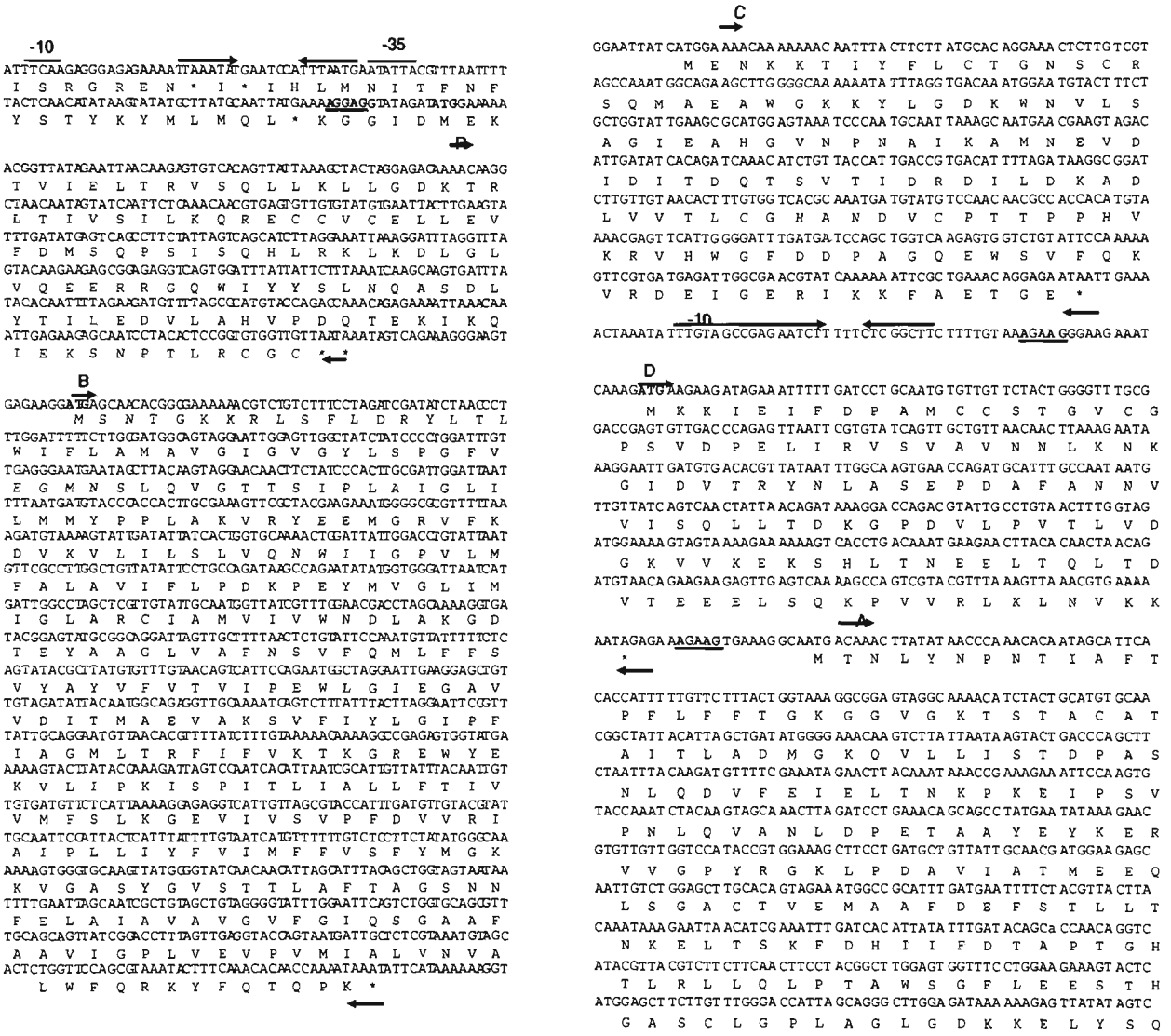
```
                                    1021      1031      1041      1051      1061      1071
O.grignonense                       TCGCGGACACAG-AGATG-TGTCTTTCAGTTCGGCTGGAC-CGGATACAGGTGCTGCATG
Ochrobactrum sp LMG20564            TCGCGGACACAG-AGATG-TGTCTTTCAGTTCGGCTGGAC-CGGATACAGGTGCTGCATG
Ochrobactrum sp                     TCGCGGACACAG-AGATG-TGTCTTTCAGTTCGGCTGGAC-CGGATACAGGTGCTGCATG
CDB2                                TCGCGGACACAG-AGATG-TGTCTTTCAGTTCGGCTGGAC-CGGATACAGGTGCTGCATG
B.melitens                          TCGCGGTTAGTGGAGACACTATCCTTCAGTTAGGCTGGAC-CGGAGACAGGTGCTGCATG
B.canis                             TCGCGGTTAGTGGAGACACTATCCTTCAGTTAGGCTGGAC-CGGAGACAGGTGCTGCATG
O.anthropi                          TCGCGGACACAG-AGATG-TGTCTTTCAGTTCGGCTGGAC-CGGATACAGGTGCTGCATG
Sinorhizobium sp TB8-711            TCGCGGATTACAGAGATGTGATCCTTCAGTTCGGCTGGAT-CGGAGACAGGTGCTGCATG
Sinorhizobium fredii                TCGCGGATACGAGAGATCGTATCCTTCAGTTCGGCTGGAT-CGGAGACAGGTGCTGCATG
CDB5                                GAACTTAG-CAG-AGATGCTTTGGTGCC-TTCGG--GAACTCTGAGACAGGTGCTGCATG
Serratia marcescens                 GAACTTAG-CAG-AGATGCTTTGGTGCC-TTCGG--GAACTCTGAGACAGGTGCTGCATG
Enterobacteriaceae bacterium A2JM   GAACTTAG-CAG-AGATGCTTTGGTGCC-TTCGG--GAACTCTGAGACAGGTGCTGCATG
Kluyvera ascorbata                  GAACTTAG-CAG-AGATGCTTTGGTGCC-TTCGG--GAACTCTGAGACAGGTGCTGCATG
CDB3                                GAAAAC-C-CTAGAGATAGG-GCTTCTCCTTCGG--GAGCAGAGTGACAGGTGGTGCATG
CDB4                                GAAAAC-C-CTAGAGATAGG-GCTTCTCCTTCGG--GAGCAGAGTGACAGGTGGTGCATG
B.mycoides                          GAAAAC-T-CTAGAGATAGA-GCTTCTCCTTCGG--GAGCAGAGTGACAGGTGGTGCATG
Bacterium Te27R                     GACAAC-C-CTAGAGATAGG-GCTTCTCCTTCGG--GAGCAGAGTGACAGGTGGTGCATG
B.anthracis                         GACAAC-C-CTAGAGATAGG-GCTTCTCCTTCGG--GAGCAGAGTGACAGGTGGTGCATG
B.thuringinesis                     GAAAAC-C-CTAGAGATAGG-GCTTCTCCTTCGG--GAGCAGAGTGACAGGTGGTGCATG
B.cereus G8639                      GAAAAC-C-CTAGAGATAGG-GCTTCTCCTTCGG--GAGCAGAGTGACAGGTGGTGCATG
CDB1                                GGAAAGAC-CTGGAAACAGGTGCCCC-GCTTGCG--G--TCGGTTTACAGGTGGTGCATG
A.histidinilovorans                 GGAAAGAC-CTGGAAACAGGTGCCCC-GCTTGCG--G--TCGGTTTACAGGTGGTGCATG
Arthrobacter sp                     GGTAATAC-CTGGAAACAGGTGCCCC-GCTTGCG--G--TCGGTTTACAGGTGGTGCATG
A. ilicis                           GGTAATAC-CTGGAAACAGGTGCCCC-GCTTGTG--G--TCGGTTTACAGGTGGTGCATG
A.sulfonivorans                     GGAAATAC-CTGGAAACAGGTGCCCCCGCTTGCG--G--CCGGTTCACAGGTGGTGCATG
A.polychromogenes                   GGTAATAC-CTGGAGACAGGTGCCCC-GCTTGCG--G--TCGGTTTACAGGTGGTGCATG

                                    1081      1091      1101      1111      1121      1131
O.grignonense                       GCTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTCG
Ochrobactrum sp LMG20564            GCTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTCG
Ochrobactrum sp                     GCTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTCG
CDB2                                GCTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTCG
B.melitens                          GCTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTCG
B.canis                             GCTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTCG
O.anthropi                          GCTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTCG
Sinorhizobium sp TB8-711            GCTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTCG
Sinorhizobium fredii                GCTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTCG
CDB5                                GCTGTCGTCAGCTCGTGTTGTGAAATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTA
Serratia marcescens                 GCTGTCGTCAGCTCGTGTTGTGAAATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTA
Enterobacteriaceae bacterium A2JM   GCTGTCGTCAGCTCGTGTTGTGAAATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTA
Kluyvera ascorbata                  GCTGTCGTCAGCTCGTGTTGTGAAATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTA
CDB3                                GTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTG
CDB4                                GTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTG
B.mycoides                          GTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTG
Bacterium Te27R                     GTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTG
B.anthracis                         GTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTG
B.thuringinesis                     GTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTG
B.cereus G8639                      GTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTG
CDB1                                GTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTCG
A.histidinilovorans                 GTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTCG
Arthrobacter sp                     GTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTCG
A. ilicis                           GTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTCG
A.sulfonivorans                     GTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTCG
A.polychromogenes                   GTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTCG

                                    1141      1151      1161      1171      1181      1191
O.grignonense                       CCTTTAGTTGCCATCA--TTTAGTTGGGCACTCTAAAGGGACTGCCAGTGATAAGCTG-G
Ochrobactrum sp LMG20564            CCTTTAGTTGCCATCA--TTTAGTTGGGCACTCTAAAGGGACTGCCAGTGATAAGCTG-G
Ochrobactrum sp                     CCTTTAGTTGCCATCA--TTTAGTTGGGCACTCTAAAGGGACTGCCAGTGATAAGCTG-G
CDB2                                CCTTTAGTTGCCATCA--TTTAGTTGGGCACTCTAAAGGGACTGCCAGTGATAAGCTG-G
B.melitens                          CCCTTAGTTGCCAGCA--TTCAGTTGGGCACTCTAAGGGGACTGCCGGTGATAAGCCGAG
B.canis                             CCCTTAGTTGCCAGCA--TTCAGTTGGGCACTCTAAGGGGACTGCCGGTGATAAGCCGAG
O.anthropi                          CCCTTAGTTGCCAGCA--TTTAGTTGGGCACTCTAAGGGGACTGCCGGTGATAAGCCGAG
Sinorhizobium sp TB8-711            CCCTTAGTTGCCAGCA--TTTAGTTGGGCACTCTAAGGGGACTGCCGGTGATAAGCCGAG
Sinorhizobium fredii                CCCTTAGTTGCCAGCA--TTTAGTTGGGCACTCTAAGGGGACTGCCGGTGATAAGCCGAG
CDB5                                TCCTTTGTTGCCAGCGG-TTCGGCCGGGAACTCAAAGGAGACTGCCAGTGATAAACTG-G
Serratia marcescens                 TCCTTTGTTGCCAGCGG-TTCGGCCGGGAACTCAAAGGAGACTGCCAGTGATAAACTG-G
Enterobacteriaceae bacterium A2JM   TCCTTTGTTGCCAGCGG-TTCGGCCGGGAACTCAAAGGAGACTGCCAGTGATAAACTG-G
Kluyvera ascorbata                  TCCTTTGTTGCCAGCGG-TTCGGCCGGGAACTCAAAGGAGACTGCCAGTGATAAACTG-G
CDB3                                ATCTTAGTTGCCATCA--TTAAGTTGGGCACTCTAAGGTGACTGCCGGTGACAAACCG-G
CDB4                                ATCTTAGTTGCCATCA--TTAAGTTGGGCACTCTAAGGTGACTGCCGGTGACAAACCG-G
B.mycoides                          ATCTTAGTTGCCATCA--TTAAGTTGGGCACTCTAAGGTGACTGCCGGTGACAAACCG-G
Bacterium Te27R                     ATCTTAGTTGCCATCA--TTAAGTTGGGCACTCTAAGGTGACTGCCGGTGACAAACCG-G
B.anthracis                         ATCTTAGTTGCCATCA--TTWAGTTGGGCACTCTAAGGTGACTGCCGGTGACAAACCG-G
B.thuringinesis                     ATCTTAGTTGCCATCA--TTAAGTTGGGCACTCTAAGGTGACTGCCGGTGACAAACCG-G
B.cereus G8639                      ATCTTAGTTGCCATCA--TTAAGTTGGGCACTCTAAGGTGACTGCCGGTGACAAACCG-G
CDB1                                TTCTATGTTGCCAGCG-GTTCGGCCGGGGACTCATAGGAGACTGCCGGGGTCAACTCG-G
A.histidinilovorans                 TTCTATGTTGCCAGCG-GTTCGGCCGGGGACTCATAGGAGACTGCCGGGGTCAACTCG-G
Arthrobacter sp                     TTCTATGTTGCCAGCG-GTTCGGCCGGGGACTCATAGGAGACTGCCGGGGTCAACTCG-G
A. ilicis                           TTCTATGTTGCCAGCG-GTTCGGCCGGGGACTCATAGGAGACTGCCGGGGTCAACTCG-G
```

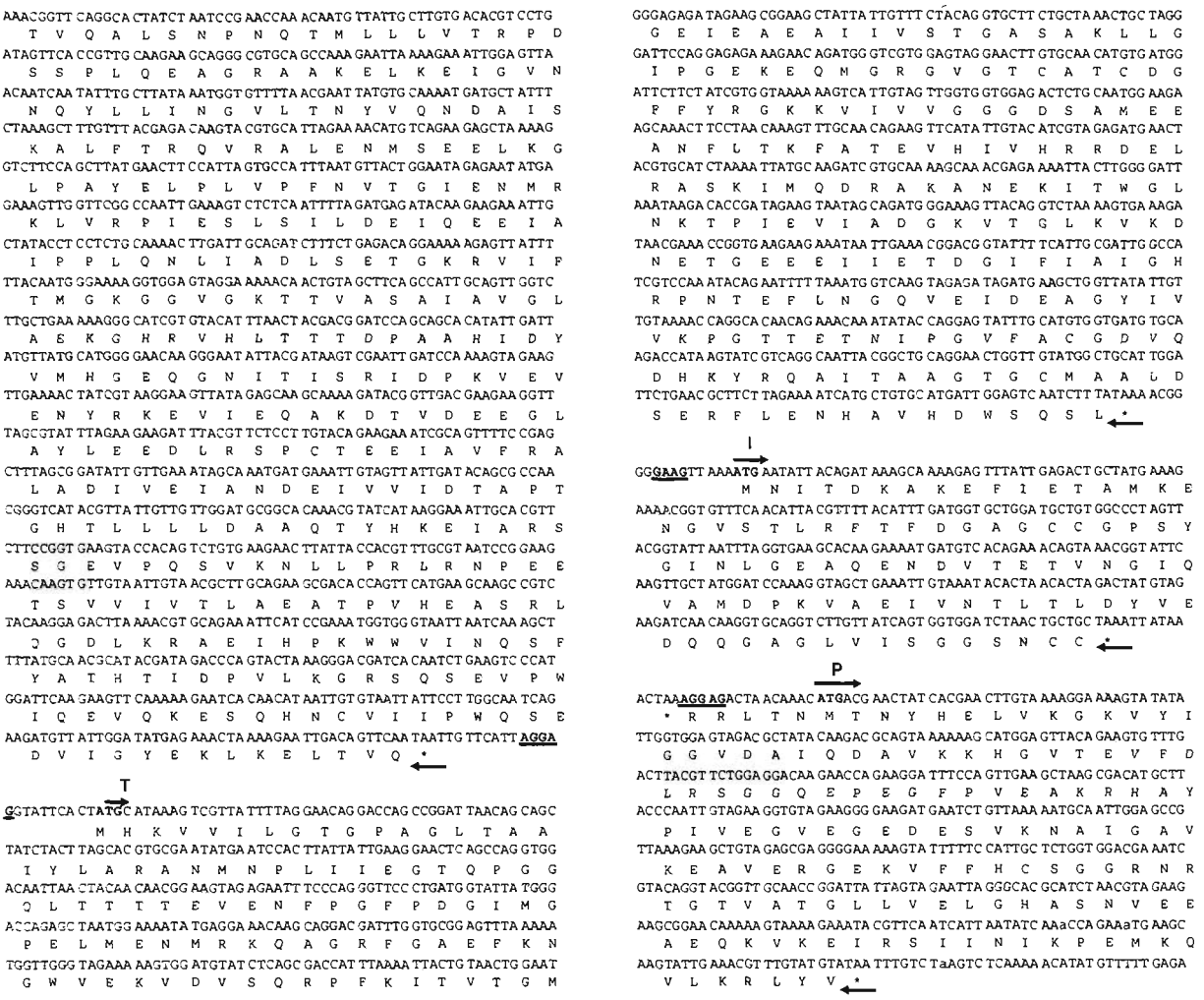
```
A.sulfonivorans                    TTCTATGTTGCCAGCGCGTGATGGCGGGGACTCATAGGAGACTGCCGGGGTCAACTCG-G
A.polychromogenes                  TTCTATGTTGCCAGCACGTGATGGTGGGGACTCATAGGAGACTGCCGGGGTCAACTCG-G

                                   1201      1211      1221      1231      1241      1251
O.grignonense                      AGGAAGGTGGGGATGACGTCAAGTCC-TCATGGCCCTTACGGGCTGGGCTACACACGTGC
Ochrobactrum sp LMG20564           AGGAAGGTGGGGATGACGTCAAGTCC-TCATGGCCCTTACGGGCTGGGCTACACACGTGC
Ochrobactrum sp                    AGGAAGGTGGGGATGACGTCAAGTCC-TCATTGCCNTTACGGGCTGGGCTACACACGTGC
CDB2                               AGGAAGGTGGGGATGACGTCAAGTCCCTCATGGCCCTTACGGGCTGGGCTACACACGTGC
B.melitens                         AGGAAGGTGGGGATGACGTCAAGTCC-TCATGGCCCTTACGGGCTGGGCTACACACGTGC
B.canis                            AGGAAGGTGGGGATGACGTCAAGTCC-TCATGGCCCTTACGGGCTGGGCTACACACGTGC
O.anthropi                         AGGAAGGTGGGGATGACGTCAAGTCC-TCATGGCCCTTACGGGCTGGGCTACACACGTGC
Sinorhizobium sp TB8-711           AGGAAGGTGGGGATGACGTCAAGTCC-TCATGGCCCTTACGGGCTGGGCTACACACGTGC
Sinorhizobium fredii               AGGAAGGTGGGGATGACGTCAAGTCC-TCATGGCCCTTACGGGCTGGGCTACACACGTGC
CDB5                               AGGAAGGTGGGGATGACGTCAAGTCA-TCATGGCCCTTACGAGTAGGGCTACACACGTGC
Serratia marcescens                AGGAAGGTGGGGATGACGTCAAGTCA-TCATGGCCCTTACGAGTAGGGCTACACACGTGC
Enterobacteriaceae bacterium A2JM  AGGAAGGTGGGGATGACGTCAAGTCA-TCATGGCCCTTACGAGTAGGGCTACACACGTGC
Kluyvera ascorbata                 AGGAAGGTGGGGATGACGTCAAGTCA-TCATGGCCCTTACGAGTAGGGCTACACACGTGC
CDB3                               AGGAAGGTGGGGATGACGTCAAATCAT-CATGCCCCTTATGACCTGGGCTACACACGTGC
CDB4                               AGGAAGGTGGGGATGACGTCAAATCAT-CATGCCCCTTATGACCTGGGCTACACACGTGC
B.mycoides                         AGGAAGGTGGGGATGACGTCAAATCAT-CATGCCCCTTATGACCTGGGCTACACACGTGC
Bacterium Te27R                    AGGAAGGTGGGGATGACGTCAAATCAT-CATGCCCCTTATGACCTGGGCTACACACGTGC
B.anthracis                        AGGAAGGTGGGGATGACGTCAAATCAT-CATGCCCCTTATGACCTGGGCTACACACGTGC
B.thuringinesis                    AGGAAGGTGGGGATGACGTCAAATCAT-CATGCCCCTTATGACCTGGGCTACACACGTGC
B.cereus G8639                     AGGAAGGTGGGGATGACGTCAAATCAT-CATGCCCCTTATGACCTGGGCTACACACGTGC
CDB1                               AGGAAGGTGGGGACGACGTCAAATCATACATGCCCCTTATGTCTTGGGCTTCACGCATGC
A.histidinilovorans                AGGAAGGTGGGGACGACGTCAAATCAT-CATGCCCCTTATGTCTTGGGCTTCACGCATGC
Arthrobacter sp                    AGGAAGGTGGGGACGACGTCAAATCAT-CATGCCCCTTATGTCTTGGGCTTCACGCATGC
A. ilicis                          AGGAAGGTGGGGACGACGTCAAATCAT-CATGCCCCTTATGTCTTGGGCTTCACGCATGC
A.sulfonivorans                    AGGAAGGTGGGGACGACGTCAAATCAT-CATGCCCCTTATGTCTTGGGCTTCACGCATGC
A.polychromogenes                  AGGAAGGTGGGGACGACGTCAAATCAT-CATGCCCCTTATGTCTTGGGCTTCACGCATGC

                                   1261      1271      1281      1291      1301      1311
O.grignonense                      TACAATGGTGGTGACAGTGGGCAGCAAGCACGCGAGTGTGAGCTAATCTCCAAAAGC-CA
Ochrobactrum sp LMG20564           TACAATGGTGGTGACAGTGGGCAGCAAGCACGCGAGTGTGAGCTAATCTCCAAAAGC-CA
Ochrobactrum sp                    TACAATGGTGGTGACAGTGGGCAGCAAGCACGCGAGTGTGAGCTAATCTCCAAAAGC-CA
CDB2                               TACAATGGTGGTGACAGTGGGCAGCAAGCACGCGAGTGTGAGCTAATCTCCAAAAGC-CA
B.melitens                         TACAATGGTGGTGACAGTGGGCAGCGAGCACGCGAGTGTGAGCTAATCTCCAAAAGC-CA
B.canis                            TACAATGGTGGTGACAGTGGGCAGCGAGCACGCGAGTGTGAGCTAATCTCCAAAAGC-CA
O.anthropi                         TACAATGGTGGTGACAGTGGGCAGCGAGCACGCGAGTGTGAGCTAATCTCCAAAAGC-CA
Sinorhizobium sp TB8-711           TACAATGGTGGTGACAGTGGGCAGCGAGACCGCGAGGTCGAGCTAATCTCCAAAAGC-CA
Sinorhizobium fredii               TACAATGGTGGTGACAGTGGGCAGCGAGACCGCGAGGTCGAGCTAATCTCCAAAAGC-CA
CDB5                               TACAATGGCGTATACAAAGAGAAGCGACCTCGCGAGAGCAAGCGGACCTCATAAAGTACG
Serratia marcescens                TACAATGGCGTATACAAAGAGAAGCGACCTCGCGAGAGCAAGCGGACCTCATAAAGTACG
Enterobacteriaceae bacterium A2JM  TACAATGGCGTATACAAAGAGAAGCGACCTCGCGAGAGCAAGCGGACCTCATAAAGTACG
Kluyvera ascorbata                 TACAATGGCATATACAAAGAGAAGCGACCTCGCGAGAGCAAGCGGACCTCATAAAGTATG
CDB3                               TACAATGGACGGTACAAAGAGCTGCAAGACCGCGAGGTGGAGCTAATCTCATAAAACCGT
CDB4                               TACAATGGACGGTACAAAGAGCTGCAAGACCGCGAGGTGGAGCTAATCTCATAAAACCGT
B.mycoides                         TACAATGGACGGTACAAAGAGCTGCAAGACCGCGAGGTGGAGCTAATCTCATAAAACCGT
Bacterium Te27R                    TACAATGGACGGTACAAAGAGCTGCAAGACCGCGAGGTGGAGCTAATCTCATAAAACCGT
B.anthracis                        TACAATGGACGGTACAAAGAGCTGCAAGACCGCGAGGTGGAGCTAATCTCATAAAACCGT
B.thuringinesis                    TACAATGGACGGTACAAAGAGCTGCAAGACCGCGAGGTGGAGCTAATCTCATAAAACCGT
B.cereus G8639                     TACAATGGACGGTACAAAGAGCTGCAAGACCGCGAGGTGGAGCTAATCTCATAAAACCGT
CDB1                               TACAATGGCCGGTACAAAGGGTTGCGATACTGTGAGGTGGAGCTAATCCCAAAAAGCCGG
A.histidinilovorans                TACAATGGCCGGTACAAAGGGTTGCGATACTGTGAGGTGGAGCTAATCCCAAAAAGCCGG
Arthrobacter sp                    TACAATGGCCGGTACAAAGGGTTGCGATACTGTGAGGTGGAGCTAATCCCAAAAAGCCGG
A. ilicis                          TACAATGGCCGGTACAAAGGGTTGCGATACTGTGAGGTGGAGCTAATCCCAAAAAGCCGG
A.sulfonivorans                    TACAATGGCCGGTACAAAGGGTTGCGATACTGTGAGGTGGAGCTAATCCCAAAAAGCCGG
A.polychromogenes                  TACAATGGCCGGTACAAAGGGTTGCGATACTGTGAGGTGGAGCTAATCCCAAAAAGCCGG

                                   1321      1331      1341      1351      1361      1371
O.grignonense                      TCTCAGTTCGGATTGCACTCTGCAACTCGAGTGCATGAAGTTGGAATCGCTAGTAATCGC
Ochrobactrum sp LMG20564           TCTCAGTTCGGATTGCACTCTGCAACTCGAGTGCATGAAGTTGGAATCGCTAGTAATCGC
Ochrobactrum sp                    TCTCAGTTCGGATTGCACTCTGCAACTCGAGTGCATGAAGTTGGAATCGCTAGTAATCGC
CDB2                               TCTCAGTTCGGATTGCACTCTGCAACTCGAGTGCATGAAGTTGGAATCGCTAGTAATCGC
B.melitens                         TCTCAGTTCGGATTGCACTCTGCAACTCGAGTGCATGAAGTTGGAATCGCTAGTAATCGC
B.canis                            TCTCAGTTCGGATTGCACTCTGCAACTCGAGTGCATGAAGTTGGAATCGCTAGTAATCGC
O.anthropi                         TCTCAGTTCGGATTGCACTCTGCAACTCGAGTGCATGAAGTTGGAATCGCTAGTAATCGC
Sinorhizobium sp TB8-711           TCTCAGTTCGGATTGCACTCTGCAACTCGAGTGCATGAAGTTGGAATCGCTAGTAATCGC
Sinorhizobium fredii               TCTCAGTTCGGATTGCACTCTGCAACTCGAGTGCATGAAGTTGGAATCGCTAGTAATCGC
CDB5                               TCGTAGTCCGGATTGGAGTCTGCAACTCGACTCCATGAAGTCGGAATCGCTAGTAATCGT
Serratia marcescens                TCGTAGTCCGGATTGGAGTCTGCAACTCGACTCCATGAAGTCGGAATCGCTAGTAATCGT
Enterobacteriaceae bacterium A2JM  TCGTAGTCCGGATTGGAGTCTGCAACTCGACTCCATGAAGTCGGAATCGCTAGTAATCGT
Kluyvera ascorbata                 TCGTAGTCCGGATTGGAGTCTGCAACTCGACTCCATGAAGTCGGAATCGCTAGTAATCGT
CDB3                               TCTCAGTTCGGATTGTAGGCTGCAACTCGCCTACATGAAGCTGGAATCGCTAGTAATCGC
CDB4                               TCTCAGTTCGGATTGTAGGCTGCAACTCGCCTACATGAAGCTGGAATCGCTAGTAATCGC
B.mycoides                         TCTCAGTTCGGATTGTAGGCTGCAACTCGCCTACATGAAGCTGGAATCGCTAGTAATCGC
Bacterium Te27R                    TCTCAGTTCGGATTGTAGGCTGCAACTCGCCTACATGAAGCTGGAATCGCTAGTAATCGC
B.anthracis                        TCTCAGTTCGGATTGTAGGCTGCAACTCGCCTACATGAAGCTGGAATCGCTAGTAATCGC
B.thuringinesis                    TCTCAGTTCGGATTGTAGGCTGCAACTCGCCTACATGAAGCTGGAATCGCTAGTAATCGC
B.cereus G8639                     TCTCAGTTCGGATTGTAGGCTGCAACTCGCCTACATGAAGCTGGAATCGCTAGTAATCGC
CDB1                               TCTCAGTTCGGATTGGGGTCTGCAACTCGACCCCATGAAGTCGGAGTCGCTAGTAATCGC
```

```
A.histidinilovorans                 TCTCAGTTCGGATTGGGGTCTGCAACTCGACCCCATGAAGTCGGAGTCGCTAGTAATCGC
Arthrobacter sp                     TCTCAGTTCGGATTGGGGTCTGCAACTCGACCCCATGAAGTCGGAGTCGCTAGTAATCGC
A. ilicis                           TCTCAGTTCGGATTGGGGTCTGCAACTCGACCCCATGAAGTCGGAGTCGCTAGTAATCGC
A.sulfonivorans                     TCTCAGTTCGGATTGGGGTCTGCAACTCGACCCCATGAAGTCGGAGTCGCTAGTAATCGC
A.polychromogenes                   TCTCAGTTCGGATTGGGGTCTGCAACTCGACCCCATGAAGTCGGAGTCGCTAGTAATCGC

                                    1381      1391      1401      1411      1421      1431
O.grignonense                       GGATCAGCAT-GCCGCGGTGAATACGTT-CCCGGGCCTTGTACACACCGCCCGTCACACC
Ochrobactrum sp LMG20564            GGATCAGCAT-GCCGCGGTGAATACGTT-CCCGGGCCTTGTACACACCGCCCGTCACACC
Ochrobactrum sp                     GGATCAGCAT-GCCGCGGTGAATACGTT-CCCGGGCCTTGTACACACCGCCCGTCACACC
CDB2                                GGATCAGCAT-GCCGCGGTGAATACGTT-CCCGGGCCTTGTACACACCGCCCGTCACACC
B.melitens                          GGATCAGCAT-GCCGCGGTGAATACGTT-CCCGGGCCTTGTACACACCGCCCGTCACACC
B.canis                             GGATCAGCAT-GCCGCGGTGAATACGTT-CCCGGGCCTTGTACACACCGCCCGTCACACC
O.anthropi                          GGATCAGCAT-GCCGCGGTGAATACGTT-CCCGGGCCTTGTACACACCGCCCGTCACACC
Sinorhizobium sp TB8-711            AGATCAGCAT-GCTGCGGTGAATACGTT-CCCGGGCCTTGTACACACCGCCCGTCACACC
Sinorhizobium fredii                AGATCAGCAT-GCTGCGGTGAATACGTT-CCCGGGCCTTGTACACACCGCCCGTCACACC
CDB5                                AGATCAGAAT-GCTACGGTGAATACGTT-CCCGGGCCTTGTACACACCGCCCGTCACACC
Serratia marcescens                 AGATCAGAAT-GCTACGGTGAATACGTT-CCCGGGCCTTGTACACACCGCCCGTCACACC
Enterobacteriaceae bacterium A2JM   AGATCAGAAT-GCTACGGTGAATACGTT-CCCGGGCCTTGTACACACCGCCCGTCACACC
Kluyvera ascorbata                  AGATCAGAAT-GCTACGGTGAATACGTT-CCCGGGCCTTGTACACACCGCCCGTCACACC
CDB3                                GGATCAGCAT-GCCGCGGTGAATACGTT-CCCGGGCCTTGTACACACCGCCCGTCACACC
CDB4                                GGATCAGCAT-GCCGCGGTGAATACGTT-CCCGGGCCTTGTACACACCGCCCGTCACACC
B.mycoides                          GGATCAGCAT-GCCGCGGTGAATACGTT-CCCGGGCCTTGTACACACCGCCCGTCACACC
Bacterium Te27R                     GGATCAGCAT-GCCGCGGTGAATACGTT-CCCGGGCCTTGTACACACCGCCCGTCACACC
B.anthracis                         GGATCAGCAT-GCCGCGGTGAATACGTT-CCCGGGCCTTGTACACACCGCCCGTCACACC
B.thuringinesis                     GGATCAGCAT-GCCGCGGTGAATACGTT-CCCGGGCCTTGTACACACCGCCCGTCACACC
B.cereus G8639                      GGATCAGCAT-GCCGCGGTGAATACGTT-CCCGGGCCTTGTACACACCGCCCGTCACACC
CDB1                                AGATCAGCAACGCTGCGGTGAATACGTTACCCGGGCCTTGTACACACCGCCCGTCAAGTC
A.histidinilovorans                 AGATCAGCAACGCTGCGGTGAATACGTT-CCCGGGCCTTGTACACACCGCCCGTCAAGTC
Arthrobacter sp                     AGATCAGCAACGCTGCGGTGAATACGTT-CCCGGGCCTTGTACACACCGCCCGTCAAGTC
A. ilicis                           AGATCAGCAACGCTGCGGTGAATACGTT-CCCGGGCCTTGTACACACCGCCCGTCAAGTC
A.sulfonivorans                     AGATCAGCAACGCTGCGGTGAATACGTT-CCCGGGCCTTGTACACACCGCCCGTCAAGTC
A.polychromogenes                   AGATCAGCAACGCTGCGGTGAATACGTT-CCCGGGCCTTGTACACACCGCCCGTCAAGTC

                                    1441      1451      1461      1471      1481      1491
O.grignonense                       ATGGGAGTTGGTTCTGCCCGAAGGC---ACTGTGCTAACC-GTA----------------
Ochrobactrum sp LMG20564            ATGGGAGTTGGTTCTGCCCGAAGGC---ACTGTGCTAACC--GTAAGGAG-GCAGGTGAC
Ochrobactrum sp                     ATGGGAGTTGGTTCTGCCCGAAGGC---ACTGTGCTAACC--GTAAGGAG-GCAGGTGAC
CDB2                                ATGGGAGTTGGTTCTGCCCGAAGGCTTGAGAGTTTGATTC--CT--GGCT-CCGGATGAG
B.melitens                          ATGGGAGTTGGTTTTACCCGAAGGC---GCTGTGCTAACC--GCAAGGAG-GCAGGCGAC
B.canis                             ATGGGAGTTGGTTTTACCCGAAGGC---GCTGTGCTAACC--GCAAGGAG-GCAGGCGAC
O.anthropi                          ATGGGAGTTGGTTTTACCCGAAGGC---GCTGTGCTAACC--GCAAGGAG-GCAGGCGAC
Sinorhizobium sp TB8-711            ATGGGAGTTGGTTCTACCCGAAGGT---AGTGCGCTAACC--GCAAGTAG-TCAGCTAAC
Sinorhizobium fredii                ATGGGAGTTGGTTCTACCCGAAGGT---AGTGCGCTAACC--GCAAGGAG-GCAGCTAAC
CDB5                                ATGGGAGTGGGTTGCAAAAGAAGTA---GGTAGCTTAACC--TTCGGGAG-GGCGTTTAG
Serratia marcescens                 ATGGGAGTGGGTTGCAAAAGAAGTA---GGTAGCTTAACC--TTCGGGAG-GGCGCTACC
Enterobacteriaceae bacterium A2JM   ATGGGAGTGGGTTGCAAAAGAAGTA---GGTAGCTTAACC--TTCGGGAG-GGCGCTTAC
Kluyvera ascorbata                  ATGGGAGTGGGTTGCAAAAGAAGTA---GGTAGCTTAACC--TTCGGGAG-GGCGCTTAC
CDB3                                ACGAGAGTTTGTAACACCCGAAGTC---GGTGGGGTAACCT-TTATGGAG-CCAGCCGCC
CDB4                                ACGAGAGTTTGTAACACCCGAAGTC---GGTGGGGTACC----TATGGAG-CCAGCC---
B.mycoides                          ACGAGAGTTTGTAACACCCGAAGTC---GGTGGGGTAACCT-TTATGGAG-CCAGCCGCC
Bacterium Te27R                     ACGAGAGTTTGTAACACCCGAAGTC---GGTGGGGTAACCT-TTTTGGAG-CCAGCCCCC
B.anthracis                         ACGAGAGTTTGTAACACCCGAAGTC---GGTGGGGTAACCT-TTTTGGAG-CCAGCCGCC
B.thuringinesis                     ACGAGAGTTTGTAACACCCGAAGTC---GGTGGGGTAACCT-TTTTGGAG-CCAGCCGCC
B.cereus G8639                      ACGAGAGTTTGTAACACCCGAAGTC---GGTGGGGTAACCT-TTATGGAG-CCAGCCGCC
CDB1                                ACGAAAGTTGGTAACACCCGAAGCC---GGTAGCCTAACCC-TTGTGGGG-GGAGC--TT
A.histidinilovorans                 ACGAAAGTTGGTAACACCCGAAGCC---GGTGGCCTAACCC-TTGTGGGG-GGAGCCGTC
Arthrobacter sp                     ACGAAAGTTGGTAACACCCGAAGCC---GGTGGCCTAACCC-TTGTGGGG-GGAGCCGTC
A. ilicis                           ACGAAAGTTGGTAACACCCGAAGCC---GGTGGCCTAACCC-TTGTGGGG-GGAGCCGTC
A.sulfonivorans                     ACGAAAGTTGGTAACACCCGAAGCC---GGTGGCCTAACCCCTTATGGGGAGGAGCTGTC
A.polychromogenes                   ACGAAAGTTGGTAACACCCGAAGCC---GGTGGCCTAACCCCTTGTGGGAGGGAGCTGTC

                                    1501      1511      1521      1531      1541      1551
O.grignonense                       ------------------------------------------------------------
Ochrobactrum sp LMG20564            CACGGTAGGGTCAGCGACTGGGGTGAAGTCGTAACAAGGTAGCCGTAGGGGAACCTGCGG
Ochrobactrum sp                     CACGGTAGGGTCAGCGACTGGGGTGAAGTCGTAACAAGGTAGCCGTAGGGGAAC-TGC--
CDB2                                ------------------------------------------------------------
B.melitens                          CACGGTAGGGTCAGCGACTGGGGTG-----------------------------------
B.canis                             CACGGTAGGKTCAGCGACTGGGGTGAAGTCGC-----------------------------
O.anthropi                          CACGGTAGGGTCAGCGACTGGGGTGAAG---------------------------------
Sinorhizobium sp TB8-711            CACGGTAGGGTCAGCGACTGGGGTGAAGTCGTAACAAGGTAGCCGTAGGGGAACCTGCGG
Sinorhizobium fredii                CACGGTAGGGTCAGCGACTGGGGTGAAGTCGTAACAAGGTAGCCGTAGGGGAACC-----
CDB5                                A--TTTATTATACTGCTCCGCTGA-------------------------------------
Serratia marcescens                 A--CTT-------------------------------------------------------
Enterobacteriaceae bacterium A2JM   CACTTTGTGATTCATGACTGGGGTGAAGTCGTAACAAGGTAACCGTAGGGGAACCTGCGG
Kluyvera ascorbata                  CACTTTGTGATTCATGACTGGGGTGAAGTCGTAACAAGGTAACCGTAGGGGAACCTGCGG
CDB3                                TYYGTTG--ACAGATGATTGGGG--------------------------------------
CDB4                                ------------------------------------------------------------
B.mycoides                          TAAGGTGGGACAGATGATTGGGGTGAAGTCGTAACAAGGTATCCGT---------------
Bacterium Te27R                     TAAGGTGGGACAGATGATTGGGGTGAAGTCGTAACAAGGGCCA-----------------
B.anthracis                         TAAGGTGGGACAGATGATTGGGGTGAAGTCGTAACAAGGTAGCCGTATCGGAAGGTGCGG
```


B.thuringinesis TAAGGTGGGACAGATGATTGGGGTGAAGTCGTAACAAGGTAGCCGTATCGGAAGGTGCGG
B.cereus G8639 TAAGGTGGGACAGATGATTGGGGTGAAGTCGTAACAAGGTAGCCGTATCGGAAGGTGCGG
CDB1 AGATATAGAGCCGGAACGACCGAC
A.histidinilovorans GAAGGTGGGACCGGCGATTGGGACTAAGTCGTAACAAGGTAGCCGTACCGGAAG
Arthrobacter sp GAAGGTGGGACCGGCGATTGGGACT
A. ilicis GAAGGTGGGACCGGCGATTGGGACTAAGTCGTAACAAGGTAGCCGTACCGGAAG
A.sulfonivorans GAAGGTGGGACTGGCGATTGGGACTAAGTCGTAACAAGGTAGCCGT
A.polychromogenes GAAGGTGGGACTGGCGATTGGGACTAAGTCGTAACAAG
1561 1571 1581 1591 1601 1611
O.grignonense
Ochrobactrum sp LMG20564 TTG
Ochrobactrum sp
CDB2
B.melitens
B.canis
O.anthropi
Sinorhizobium sp TB8-711 CTGGATCACCTCCTTAAGCTTGGATCCCG
Sinorhizobium fredii
CDB5
Serratia marcescens
Enterobacteriaceae bacterium A2JM CTGGATCACCTCCT
Kluyvera ascorbata CTGGATCACCTCCTTAAGCTTGGATCC
CDB3
CDB4
B.mycoides
Bacterium Te27R
B.anthracis CTGGATCACCTCCTTTCT
B.thuringinesis CTGGAT
B.cereus G8639 CTGGATCACCTCCTTTCT
CDB1
A.histidinilovorans
Arthrobacter sp
A. ilicis
A.sulfonivorans
A.polychromogenes

APPENDIX 4: Nucleotide and protein sequences of the *ars* gene cluster 1 *arsRYCDATIP* from CDB3. Start and stop codons are indicated. Shine-Dalgarno (SD) sequences are printed in bold, highlighted and underlined. The putative promoter sequences for RNA polymerase are printed in bold, underlined and identified as -10 and -35.





APPENDIX 5: Multiple sequence alignment of the arsenite membrane transporter proteins belonging to *E. coli* type ArsB and 3rd sub-group.

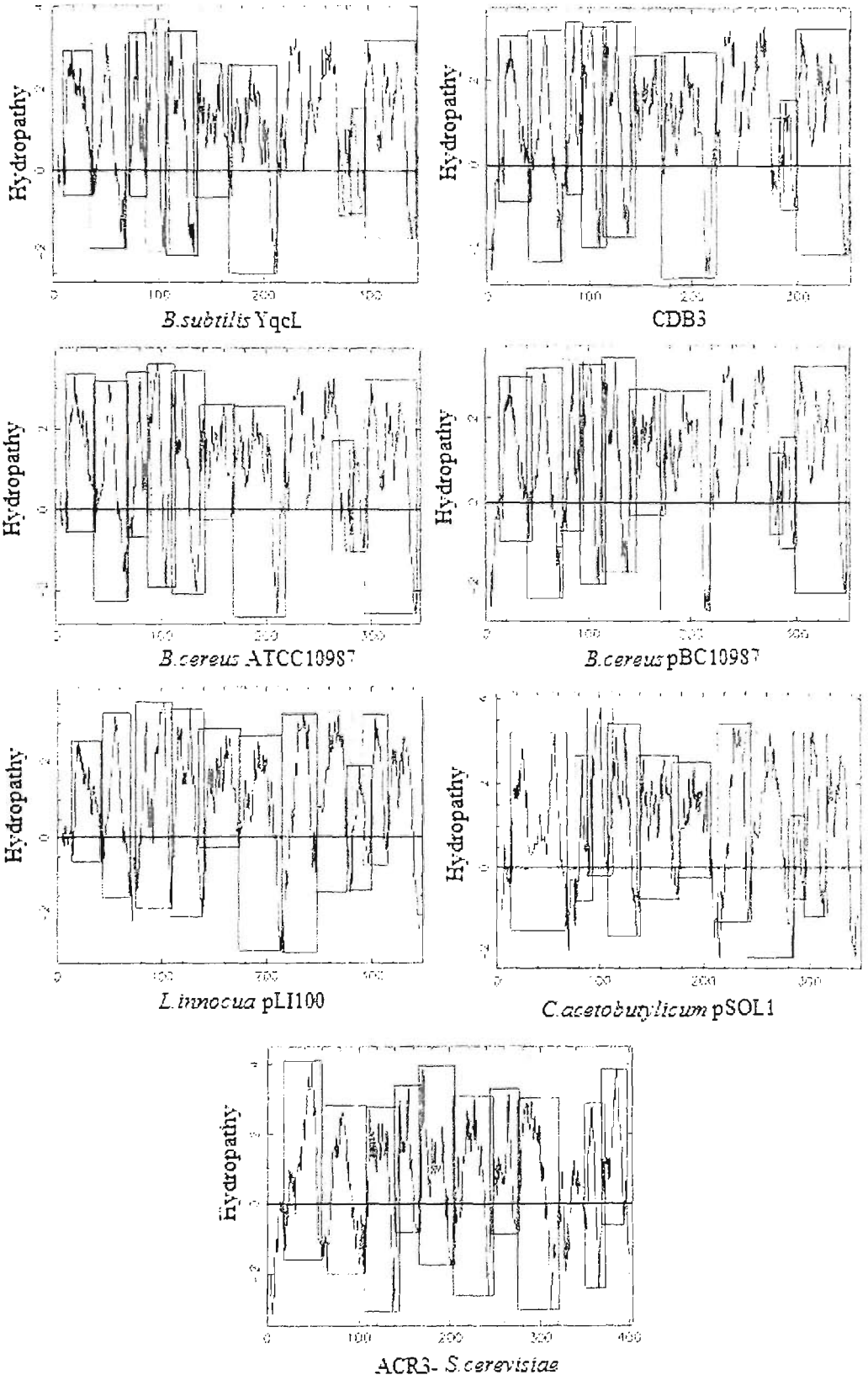
Appendix 5.1: Multiple sequence alignments of ArsB- like proteins from different strains compared with related species acquired from Genbank through blastx. The 16 related species were chosen according to their pairwise alignment score and identity. Alignments were made using the ClustalW program (Thompson et. al., 1994). Protein sequences are indicated by species name. Amino acid residues highlighted in yellow refers to identical conserved residues shared. Asterisk "*" refers to strains harboring Na^+/H^+ antiporter and related AsIII Permeases.

```
                                1         11        21        31        41        51
Y. pseudotuberculosis           ----------------------------------------------MLLAGAIFILTI
Y. enterocolitica pVYe227       ----------------------------------------------MLLAGAIFILTI
S. marsecens R478               ----------------------------------------------MLLAGAIFLFTL
E.  coli R773                   ----------------------------------------------MLLAGAIFILTI
S. typhi R64                    ----------------------------------------------MLLAGAIFILTI
E.coli INc R46                  ----------------------------------------------MLLAGAIFVLTI
Y. pestis CO92                  ----------------------------------------------MLLAGSIFVLTL
R. palustris CGA009             ----------------------------------------------MLIALAIFVATI
B. halodurans                   ---------------------------------------------MSS--VIVASLIFLFTL
B. japonicum USDA110            ---------------MEPCRRRGRQRRRHRLSQPNEPTKPQGLNVPSDAIWAWSIIVAAT
Bradyrhirhizobium sp BTAi1*     ------------------------------------MSLHDP---WLAYAIIIPAT
R. palustris CGA009*            ----------------------------------MPSSMLSHDP---VLTWSIVALAT
B. vietnamiensis G4             ---------------------------------------MNSI---FLSWGIAAVAT
B. fungorum LB400*              MRARCARRAASRRAARTAVSRPTCRSHRDEARVDALQPPGMPMNSA---PFAWSIAALAT
R. eutropha JMP134              ------------------------------------MSSVSA---VLIWAIAALTT
R. metallidurans CH34           ------------------------------------MPAYAT---PLIWSVAALST

                                61        71        81        91        101       111
Y. pseudotuberculosis           VLVIWQPKGLGIGWSAMLGAGLALLTGVVHVSDIPVVWQIVWNATATFIAIIISLLLDE
Y. enterocolitica pVYe227       VLVIWQPKGLGIGWSAMLGAGLALLTGVVHVSDIPVVWQIVWNATATFIAIIISLLLDE
S. marsecens R478               VLVIWQPRGLGIGWSASLGAILALLTGVVHLGDIPVVWQIVWNATATFIAVIISLLLDE
E. coli R773                    VLVIWQPKGLGIGWSATLGAVLALASGVIHIADIPVVWNIVWNATATFIAVIISLLLDE
S. typhi R64                    VLVIWQPKGLGIGWSATLGAVLALASGVIHIADIPVVWNIVWNATATFIAVIISLLLDE
E. coli INc R46                 VLVIWQPKGLGIGWSATLGAVLALISGVVHIGDIPVVWNIVWNATATFIAVIISLLLDE
Y. pestis CO92                  ILAIWQPRGLSIGWSASICAALALVSGVIHVGDIPVVWNIVWNATATFIAVIISLLLDE
R. palustris CGA009             ALVIWQPKGLGIGWSALGGAAIALLVGVVSLSDVPVVWGIVWNATATFVAIIIVSLLLDE
B. halodurans                   VLVIWQPRGLSIGWSATIGAVLALLAGVVDFADVQAVTGIVWNATLTFVAIIISLILDE
B. japonicum USDA110            AGVIIRPFRLPEAIWAVIGAGALVLLGLLPWQDALLGIGKGVDVYLFLIGMMLIAELARL
Bradyrhirhizobium sp BTAi1*     AGVIIRPFRLPEAVWAVAGAILLVALGLITTTDALAGIRKGLDVYLFLIGMMLIAELARR
R. palustris CGA009*           AGVIIRPMRWPEAVWAVAGAVALVALGLLSWQDALAGVRNGIDVYLFLIGMMLIAELAQR
B. vietnamiensis G4             AGVITRPFKWPEAVWAVAGAVLLVSLGLLPVHLAIEAIGKGTDVYLFLFGMMLLSEVARR
B. fungorum LB400*              AGVIVRPFGWPEAIWAVAGAVLLVALGLLPADAALAAVAKGGDVYPFLTGMMLLSEVARR
R. eutropha JMP134              AGVLLRPFRLSEAFWAVAGAVLLCVGGLLAVPDALAAVMRGLDVYLFLAGMMVSSELARK
R. metallidurans CH34           AGVLFRPFRLPEPFWAMAGALVLCVAGLLPWRDALQAVARGNDVYLFLAGMILISELARK

                                121       131       141       151       161       171
Y. pseudotuberculosis           SGFFEWAALHVARWGNGHGRRLFTYIILLGAAVAALFANDGAALILTPIVIAMLLALGFG
Y. enterocolitica pVYe227       SGFFEWAALHVARWGNGHGRRLFTYIILLGAAVAALFANDGAALILTPIVIAMLLALGFG
S. marsecens R478               SGFFEWAALHVARWGNGRGRLLFTWIVLLGAMVAALFANDGAALILTPIVIAMLLALGFS
E. coli R773                    SGFFEWAALHVSRWGNGRGRLLFTYIVLLGAAVAALFANDGAALILTPIVIAMLLALGFS
S. typhi R64                    SGFFEWAALHVSRWGNGRGRLLFTYIVLLGAAVAALFANDGAALILTPIVIAMLLALGFS
E. coli INc R46                 SGFFEWAALHVSRWGNGRGRLLFTYIVLLGAAVAALFANDGAALILTPIVIAMLLALGFS
Y. pestis CO92                  SGFFEWAALHVSRWGNGRGRLLFTWIVLLGAAVAALFANDGAALILTPIVIAMLLALGFS
R. palustris CGA009             AGFFEWAALHVARWGGGHGRRLFVLIVLLGAAVSALFANDGAALILTPIVIAMLRALGYK
B. halodurans                   IGFFEWSALHMARVANGSGIKMFLYITLLGAIVAALFANDGAALILTPIVLAMVRALKFK
B. japonicum USDA110            EGLFDYLAALAVEYADGSPQRLFLLIYLVGTLVTVLLSNDATAIVLTPAVYAATRAAGAK
Bradyrhirhizobium sp BTAi1*     EGLFDYLAAYAVEHARGSPQLLFLLVYVVGTLVTVLLSNDATAIVLTPAVYAATRAAGAS
R. palustris CGA009*           EGLFDYLAAYAVEHARGSPQRLFLLVYAVGVLVTVLLSNDATAIVLTPAVYAATRAAGAA
B. vietnamiensis G4             EGLFDWVAVLAVNHAQGSPRKLFLLVYLVGVVITAFLSNDATAVVLTPAVFAAARKAKTD
B. fungorum LB400*              EGLFDWVAAHVVNGARGSPRRLFALVYLVGTVTTVFLSNDATAVVLTPAVFAAARRAKTD
R. eutropha JMP134              TGLFDHVAALAVRRAGGSARALFLLVYGFGTLVTAFMSNDATAVVLTPAVLAATRAAKVQ
R. metallidurans CH34           TGLFDHVAALAVRAARGSARKLFALVYGFGIAVTAFMSNDATAVVLTPAVIAATRAARVK
```

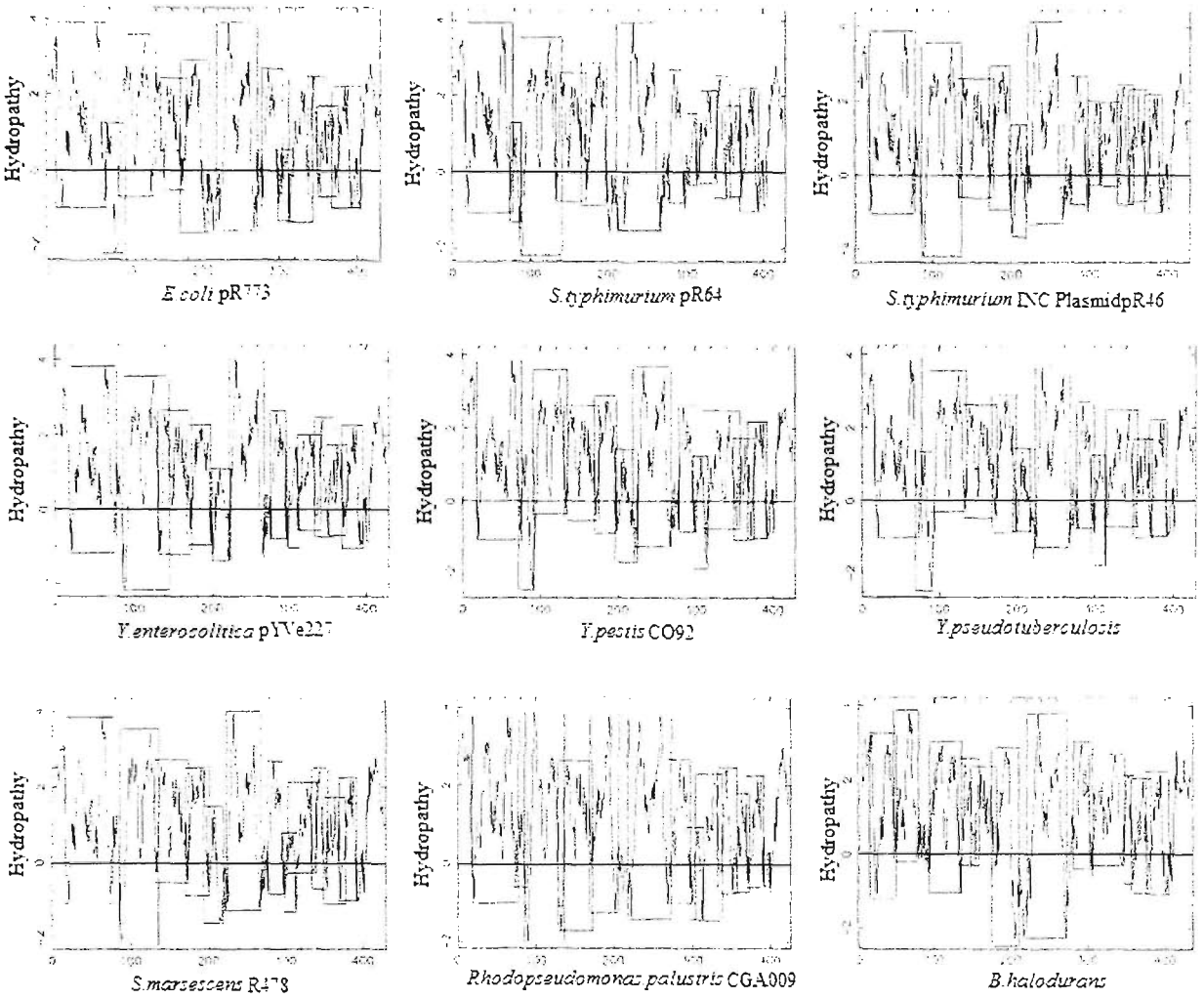
```
                                181       191       201       211       221       231
Y. pseudotuberculosis           RGATLAFVMAAGFIADTASLPLIVSNLVNIVSADFFNLGFTEYAAVMVPVNLVAVIATLA
Y. enterocolitica pVYe227       RGATLAFVMAAGFIADTASLPLIVSNLVNIVSADFFNLGFTEYAAVMVPVNLVAVIATLA
S. marsecens R478               RGATLAFIMAAGFIADTASLPLIVSNLVNIVSADFFKLGFSEYAAVMVPVNLAAIAATLV
E. coli R773                    KSTTLAFVMAAGFISDTASLPLIVSNLVNIVSADFFKLGFTEYASVMVPVDIAAIIATLV
S. typhi R64                    KSTTLAFVMAAGFIADTASLPLIVSNLVNIVSADFFKLGFTEYASVMVPVDIAAIVATLV
E. coli INc R46                 KGTTLAFVMAAGFIADTASLPLIVSNLVNIVSADFFGLGFTEYASVMVPVDIAAIIATLV
Y. pestis CO92                  KRTTLAFVMAAGFIADTASLPLIVSNLVNIVSADFFRLGFTEYASVMLPVDIAAIAATLG
R. palustris CGA009             DKATLAFVMAAGFIADTASLPLIVSNLVNIVSADFFRIGFADYASVMVPVDLASIAATLV
B. halodurans                   EAMVLPFVMASGFIADTTSLPLIVSNLVNIVSADFFNIGFVEYATRMVVPNLFALAASIV
B. japonicum USDA110            ---PLPYLFVCAFIANAASFVLPISNPANLVVFGARMPHLTEWLRLFALPSAASILLTYV
Bradyrhirhizobium sp BTAi1*     ---PLPYLYVCAFIANAASFVLPISNPANLVVFGARMPHLTEWLRQFTLPSITAIALTYA
R. palustris CGA009*            ---PLPYLFVCAFIANAASFVLPISNPANLVVFGEHMPHLGGWLQQFTLPSLAAILATYL
B. vietnamiensis G4             ---PLPLLFVCAFIANAASFVLPISNPANIVLYGNHTPALGAWLLRFTLPSLLSIIATYV
B. fungorum LB400*              ---PLPLLYVCAFVANAASFVLPISNPANLVLYGAHMPALGTWLARFALPSIASIACTYA
R. eutropha JMP134              N--PLPHLFACAFIANAASFVLPISNPANLVIFGTRMPPLAGWLAHFALPSLVAVVTTLA
R. metallidurans CH34           H--PLPYLYACAFIANAASFLLPISNPANLVLFGDRMPPLTSWLARFTLPSVVAIAMTFI

                                241       251       261       271       281       291
Y. pseudotuberculosis           MLHLFFRKDIPAVYDLSLLKAPQAAIRDMNTFKTGWLVLALLLIGFFALEPLGVPVSLVA
Y. enterocolitica pVYe227       MLHLFFRKDIPAVYDLSLLKAPQAAIRDMNTFKTGWLVLALLLIGFFALEPLGVPVSLVA
S. marsecens R478               MLHLFFRKDIPAVYDVSLLKEPKDAIRDVNTFKTGWLVLVLLLVGFFGLEPLGVPVSLVA
E. coli R773                    MLHLFFRKDIPPTYELARLKEPAKAIKDPATFRTGWVVLLLLLVGFFVLEPMGIPVSAIA
S. typhi R64                    MLHLFFRKDIPPTYELARLKEPAKAIKDPATFRTGWVVLLLLLVGFFVLEPMGIPVSAIA
E. coli INc R46                 MLHLFFRKDIPPTYDLALLKAPAKAIKDLATFRTGWIVLILLLVGFFVLEPLGIPVSAIA
Y. pestis CO92                  MLHLFFRRDIPATYDVSLLKMPASAIKDPATFRAGWIVLVLLLVGFFVLEPLGIPVSAIA
R. palustris CGA009             VLLLFFRRDIPASYDVAQLRAPADAIRDPATFKAGWVVLALLLAGFFLLEPIGVPVSAVA
B. halodurans                   VLYLYFRKDMPKTYDLSQLKKPHEAIKDKAMFKLSWVILAVLLIGYFVSEFVHIPVSFVA
B. japonicum USDA110            VLRLTQHRALKEETIARSVPHPKLSRGGKLTAVGIIAIGIVLVTASALDRQLGLPTFICG
Bradyrhirhizobium sp BTAi1*     ALWLTQRRALHAETLRTDVDKPRLSREGRLAAMGILAIAAVLLTASALDIQLGLPTFVCG
R. palustris CGA009*            VLRLALRRSLAGQTIETRVETPHLSRGGKLAAVGIGAIAIVLLTASALNVALGLPTFVCG
B. vietnamiensis G4             MLRWTQREALAG-TCEANLEALELSSSGRVALAGIAVTAAVLLTVSAFDIALGLPTAILG
B. fungorum LB400*              MLRATQRHALRG-RCACDLPVPSLSAGARIALAGIAATAAALMIVSLLDRPLGLPTALAG
R. eutropha JMP134              ALYWTQREALAE-RIASDVPVPPLSLQAWLTAVGIVLTGVALLAASLRGSDLGWPTCAAG
R. metallidurans CH34           VLYWTQRDALAE-PIENDVPTPPLTLQAWLTTLGIMLTGAALLTASLHGQDLGWPTFIGG

                                301       311       321       331       341       351
Y. pseudotuberculosis           AVGALILLAVAKKG----HAINTGKVLRGAPWQIVIFSLGMYLVVYGLRNAGLTHYISAL
Y. enterocolitica pVYe227       AVGALILLAVAKKG----HAINTGKVLRGAPWQIVIFSLGMYLVVYGLRNAGLTHYISAL
S. marsecens R478               AAGALLLFAVAKKG----HAINTGKVLRGAPWQIVIFSLGMYLVVYGLRNAGLTHYLSSL
E. coli R773                    AVGAAVLFAVAKKG----HGINTGKVLRGAPWQIVIFSLGMYLVIYGLRNAGLTDYLSDV
S. typhi R64                    AVGAAILFAVAKKG----HGINTGKVLRGAPWQIVIFSLGMYLVVYGLRNAGLTDYLSDV
E. coli INc R46                 AVGAVILFAVANRG----HAINTGKVLRGAPWQIVIFSLGMYLVVYGLRNAGLTEYLSGV
Y. pestis CO92                  AVGTVVLFAVAKKG----HAINTGKVLRGAPWQIVVFSLGMYLVVYGLRNAGLTEYLSGV
R. palustris CGA009             AAGAVLLLIIAARG----HVIQTRKVLAGAPWQVVIFSLGMYLVVYGLRNAGLTDHLAGL
B. halodurans                   MPIAFVFYLIFRRASGKNKALNTTGIIKGAPWAIVVFSIGMYVVVYGLRNVGLTNILADI
B. japonicum USDA110            VATSAIVLLISQQS----PLPVLRGVS----WSVLPLVGGLFVLVEALVKTGVIGQLSAL
Bradyrhirhizobium sp BTAi1*     VVTAAAILLAGRQS----PWPVVKGVS----WGVLPLVAGLFVMVEGLSRAGAIGQLSAW
R. palustris CGA009*            LVTTAVVLAIARQS----PVPVLKGVS----WSVVPLVAGLFVMVEALSRTGVIGALGGL
B. vietnamiensis G4             ALTALIVLILERKS----PLPMIREIS----WSVLPLVAALFVLVEMLDHTGVIAALAQL
B. fungorum LB400*              ALTAACVLARDRAA----CGPTLRAIS----WSVLPLVAGLFVLVEALERTGAVDALAAL
R. eutropha JMP134              AATLLIVCATKRDL----LAPALREIS----WGVLPMVAGLFVLVAALEHTGVIHWLADI
R. metallidurans CH34           LLTLAVVCATQPRL----LVPALKEVS----WGVLPLVAGLFVLVAGLAQTGLTAQLAHW

                                361       371       381       391       401       411
Y. pseudotuberculosis           LNRLGEQGLWASTLGTGFLTAFLS---SIMNNLPTVLVGALSIEGSTATGLIKEAMIYAN
Y. enterocolitica pVYe227       LNRLGEQGLWASTLGTGFLTAFLS---SIMNNLPTVLVGALSIEGSTATGLIKEAMIYAN
S. marsecens R478               LNQLAEQGLWAATLGTGFLTAFLS---SVMNNMPTVLVGALSIDGSTATGVIKEAMIYAN
E. coli R773                    LNELADKGLWAATLGTGFLTALLS---SIMNNMPTVLIGALSIDGSTATGVIKEAMIYAN
S. typhi R64                    LNVLADKGLWAATLGTGFLTALLS---SIMNNMPTVLIGALSIDGSTATGVIKEAMIYAN
E. coli INc R46                 LNVLADKGLWAATFGTGFLTAFLS---SIMNNMPTVLVGALSIDGSTATGVIKEAMIFAN
Y. pestis CO92                  LNMLADKGLLAATFGTGFLTAFLS---SVMNNMPTVLVGALSIDGSTASGVIKEAMIYAN
R. palustris CGA009             LDHTAQGGVWGAAFGTGTIAALLS---SVMNNMPTVLVGALSIDATHATGAIKEAMIYAN
B. halodurans                   IELAAAQGLFAGTMAMGFIAAILS---SLMNNMPTVMINALAIADTNTMGIMREALIYAN
B. japonicum USDA110            LHQAVAQSIPKAAWSVGIATAFAT---NVANNLPVGLVAGSVAASDHLP----APVVSAI
Bradyrhirhizobium sp BTAi1*     LHQGVTASPAGTAWAAGLITAVAD---NIANNLPVGLVAGSVAASDHLP----HEVVRTI
R. palustris CGA009*            LREAVADAPRLGSWAAGIVTAIAC---NIGNNLPVGLVAGSVAADNHLP----RETIAAL
B. vietnamiensis G4             AQRAAQHNELAAAGWAGSIIAIGS---NLMNNLPAGLIASSTVLQAHSP----ERVIDAL
B. fungorum LB400*              LRTLTAGGTAHAAAASGVAIALAS---NLVNNLPAGLVAGAIVGVAHSP----QPVVDAL
R. eutropha JMP134              IQAASRQDSLARLWGIGAAVALIG---NVANNLPAGLIASSALAAGHAS----NAVTGAV
R. metallidurans CH34           VRMLSGLQGPEAVLGAGVAGVLVGITSNIVNNLPAGLFAASALAAGHAS----DTVTAAV
```

```
                                421       431       441       451       461       471
Y. pseudotuberculosis           VIGSDLGPKITLIGSLATLLWLHVLSQKNMTITWGYYFRVGIIMTLPVLLVTLSALALRL
Y. enterocolitica pVYe227       VIGSDLGPKITPIGSLATLLWLHVLSQKNMTITWGYYFRVGIIMTLPVLLVTLSALALRL
S. marsecens R478               VIGSDLGPKITPIGSLATLLWLHVLSQKNIKITWGYYFRVGIVMTIPVLFVTLAALALRL
E. coli R773                    VIGCDLGPKITPIGSLATLLWLHVLSQKNMTITWGYYFRTGIVMTLPVLFVTLAALALRL
S. typhi R64                    VIGCDLGPKITPIGSLATLLWLHVLSQKNMTITWGYYFRTGIVMTLPVLFVTLAALALRL
E. coli INc R46                 VIGCDLGPKITPIGSLATLLWLHVLSQKNMTITWGYYFRTGIVMTLPVLFVTLAALALRL
Y. pestis CO92                  VIGCDLGPKITPIGSLATLLWLHVLSQKNMTISWGYYFRTGIIMTLPVLFVTLAALALRL
R. palustris CGA009             VIGCDLGPKITPIGSLATLLWLHVLGQKGIRIGWGYYFKVGATLTIPVLLVTLAALAIRL
B. halodurans                   VIGSDLGPKITPIGSLATLLWLHVLSKKGMKITWGYYFKVGIILTVPTLFITLVGLYLWL
B. japonicum USDA110            LIGVDLGPNLSVTGSLATILWLVALRREKIEVGAWPFLKLGLLVTPPALVAALAAAIR--
Bradyrhirhizobium sp BTAi1*     LIGVDLGPNISVTGSLATILWLTALRREKIQVSAWQFLKIGLVATPPALIAALAAAMW--
R. palustris CGA009*            LIGVDLGPNFSVTGSLATILWLVVLRRNQIEVTAWRFLKLGLVVTPPALIAALAAAVW--
B. vietnamiensis G4             LIGVDLGPNLSITGSLATILWLSAIRREGEDVSFMKFLKVGSVVMLPALVLALGARILTG
B. fungorum LB400*              LIGVDLGPNLSVTGSLATILWLAAIRRDGLQVSAGKFLAIGACVMLPALVAALAVRLV--
R. eutropha JMP134              LVGIDLGPNLSATGSLATLLWLTALRRDGHMVTAGQFLRLGAVVMPTAMVPTLAVLALAR
R. metallidurans CH34           LIGVDLGPNLSITGSLATLLWLTALRREGHMVGAGTFLKTGALVMPLALLPALAVLR---

                                481
Y. pseudotuberculosis           SVSM
Y. enterocolitica pVYe227       SVSM
S. marsecens R478               SFTL
E. coli R773                    SVTL
S. typhi R64                    SVTL
E. coli INc R46                 SFTL
Y. pestis CO92                  SVTL
R. palustris CGA009             SL--
B. halodurans                   MLIS
B. japonicum USDA110            ----
Bradyrhirhizobium sp BTAi1*     ----
R. palustris CGA009*            ----
B. vietnamiensis G4             ----
B. fungorum LB400*              ----
R. eutropha JMP134              ----
R. metallidurans CH34           ----
```

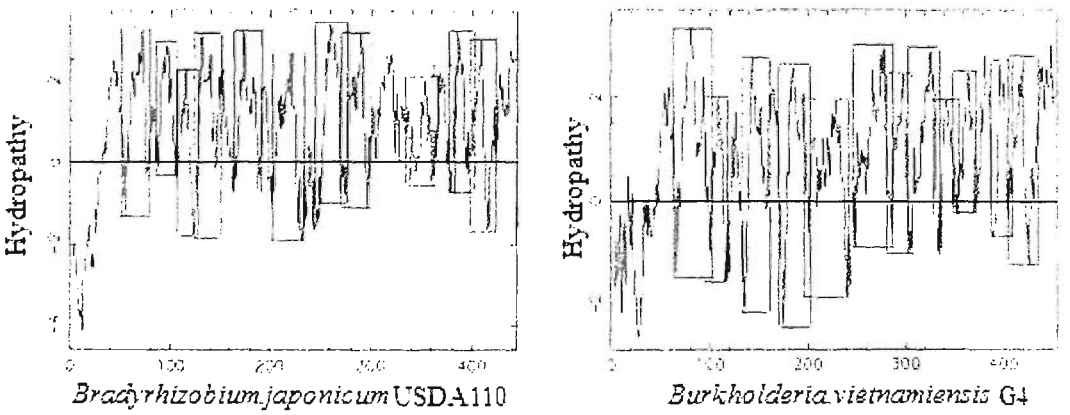
Appendix 5.2: Kyte Doolittle hydropathy plots generated by PepWindow showing the three groups of arsenite membrane transporter proteins. (a): represents the 10 TMD ArsB protein plots; (b): represents the 12 TMD ArsB protein plots and (c): represents the 14 TMD ArsB protein plots

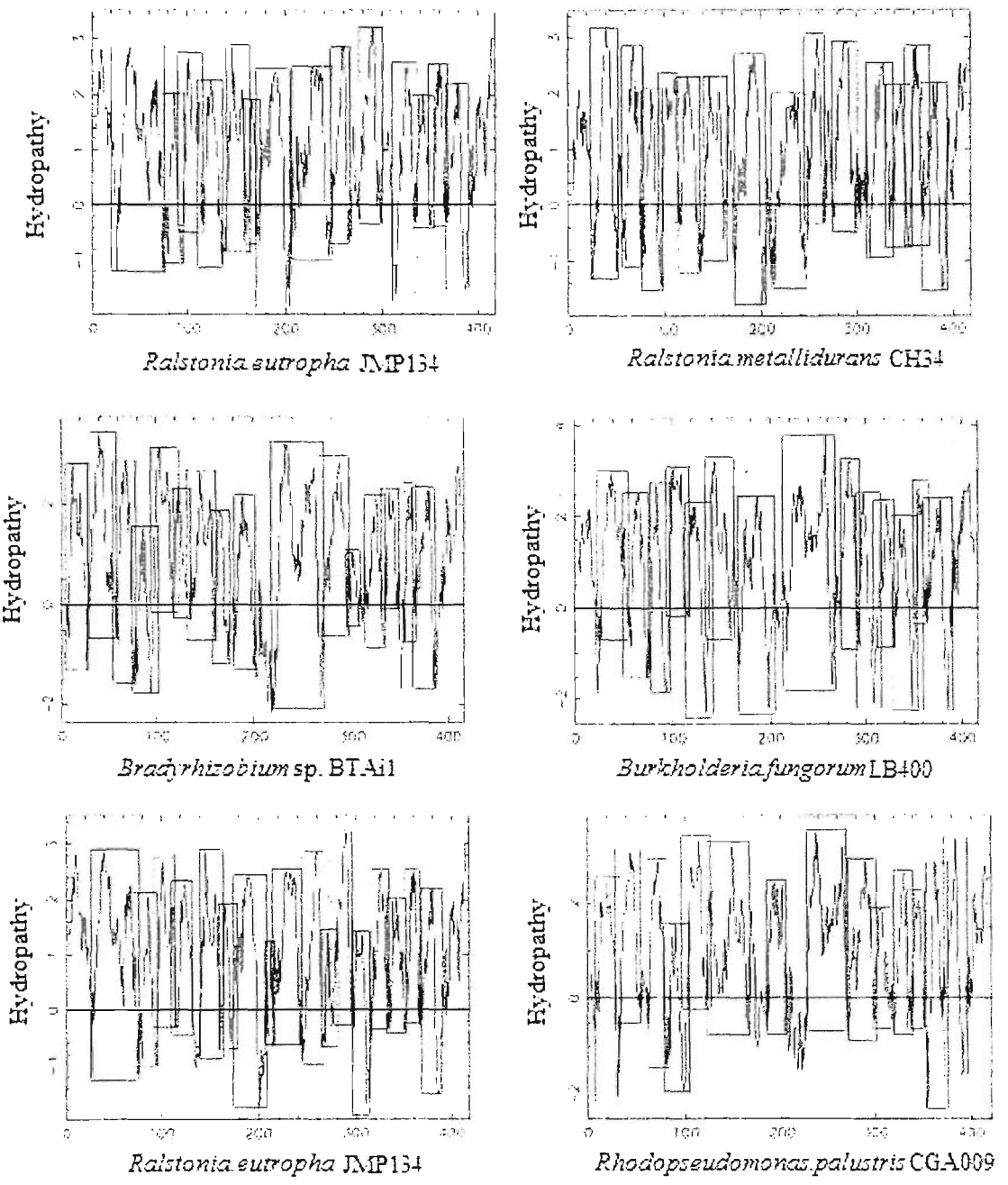


(a)


Hydropathy
E.coli pR
S.typhimurium pR64
S.typhimurium INC PlasmidpR46
Y.enterocolitica pYVe227
Y.pestis CO92
Y.pseudotuberculosis
S.marsescens R
Rhodopseudomonas palustris CGA009
B.halodurans

(b)

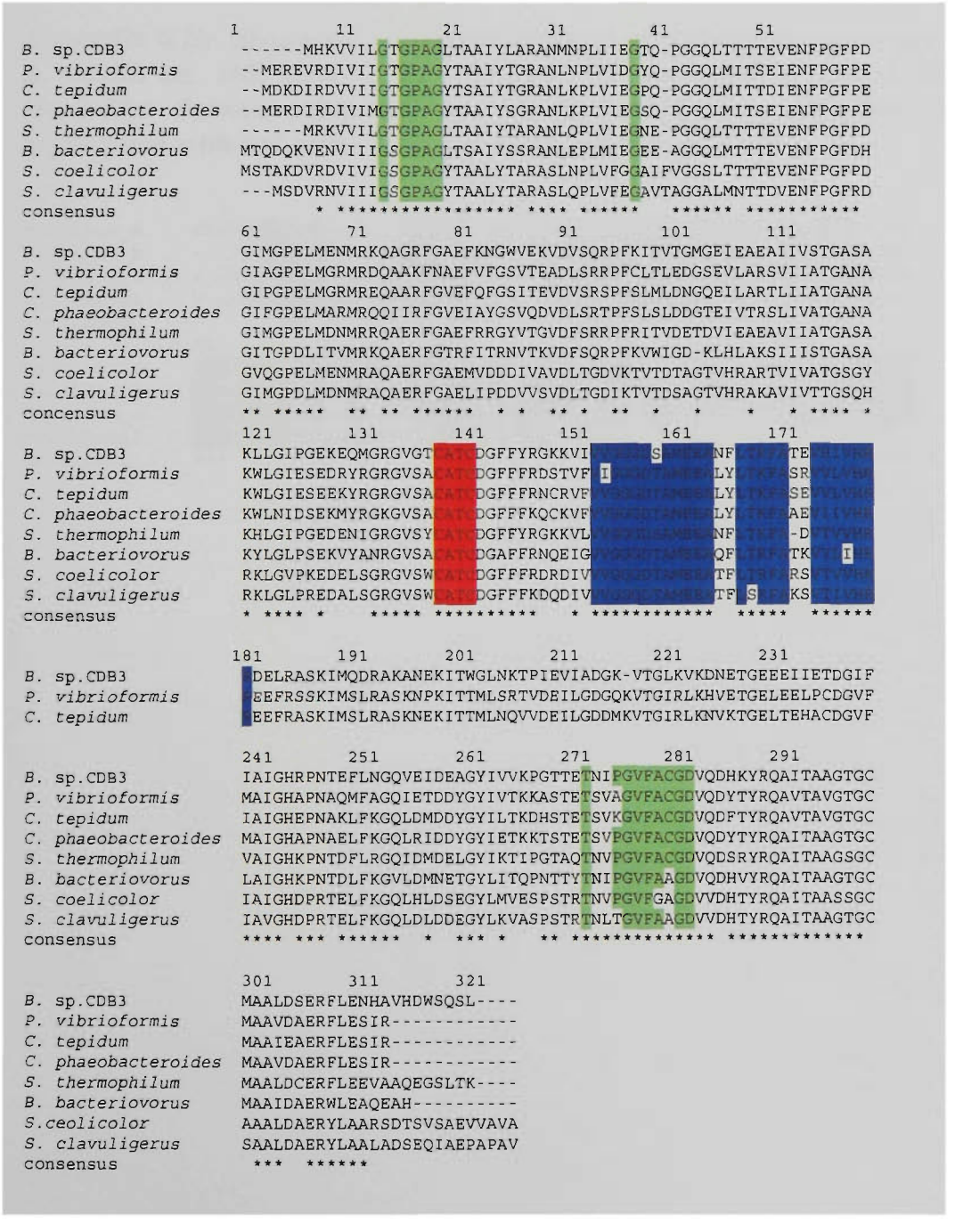

Hydropathy
Bradyrhizobium japonicum USDA110
Burkholderia vietnamiensis G4


Hydropathy
Ralstonia eutropha JMP134
Ralstonia metallidurans CH34
Bradyrhizobium sp. BTAi1
Burkholderia fungorum LB400
Ralstonia eutropha JMP134
Rhodopseudomonas palustris CGA009

(c)

APPENDIX 6: Multiple sequence alignments of CDB3 ArsT, ArsI and ArsP with those of related bacteria.

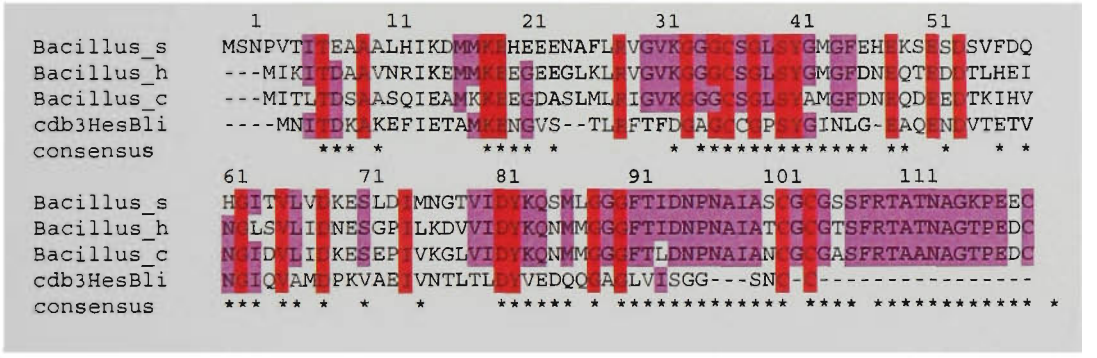
Appendix 6.1: *Multiple alignment of the amino acid sequences of* putative TrxB/ *Bacillus* sp. CDB3 with some other bacterial TrxB proteins (Xi, 2006). TrxB/*Prosthecochloris vibrioformis* DSM 265, trxB/*Chlorobium tepidum* TLS, trxB/*Chlorobium phaeobacteroides* DSM 266, trxB/*Symbiobacterium thermophilum* IAM 14863, trxB/*Bdellovibrio bacteriovorus* HD100, trxB/*Streptomyces ceolicolor* A3(2) and trxB/*Streptomyces clavuligerus* respectively. The trxB have 4 conserved domains, 2 FAD binding sites, one reactive site and one NADPH binding site which are marked as green, red and blue respectively. Consensus is labeled as asterisk.



Appendix 6.2a: Structural multiple alignment of the amino acid sequence of HesB-like domain protein (*arsI*)/*B.* sp. CDB3 (ORF7), HesB-like domain protein/*B.* sp. MB24 Tn ars1, HesB-like domain protein/*B. cereus* ATCC 10987 pBC10987. Consensus is labeled as asterisk.

```
                    1          11         21         31         41         51
B. sp. MB24          MNITDKAKEFIETAMKENGVTTLRFTFDGAGCCGPSYGINLGEAQENDVTETVNGIEVAM
B. sp. CDB3          MNITDKAKEFIETAMKENGVSTLRFTFDGAGCCGPSYGINLGEAQENDVTETVNGIQVAM
B. cereus pBc10987    MNITDKAKEFIENAMKENGVSTLRFTFEGAGCCGPSYGINLGEAEESDITETVNGVQVAM
consensus           ************ ******* ****** ***************** * * ******  ***
                    61         71         81         91
B. sp. MB24          DPKVVEIVNTLTLDYVEDQQGAGLVISGGSNCC
B. sp. CDB3          DPKVAEIVNTLTLDYVEDQQGAGLVISGGSNCC
B. cereus pBc10987    DPKVAEIVNTLTLDYVEDQQGAGLVISGGSNCC
consensus           **** ****************************
```

Appendix 6.2b: *Structural multiple alignment of the amino acid sequence of IscA/ B. subtilis* str. 168, *IscA/ B. halodurans* C-125 and *IscA/ B. clausii* KSM-K16 *and HesB-like domain protein/B. sp. CDB3 (arsI). The conserved cysteines are highlighted with red color.* Consensus is labeled as asterisk.



Appendix 6.3: Structural multiple alignment of the amino acid sequence of orf8 (*arsP*)/*B.* sp. CDB3, orf12/*B.* sp. MB24 Tn ars1, orf2/*B. cereus* ATCC 10987 pBC10987. The signature motif of dual specificity phosphatases is shadowed which are HCXXGXXRT. The conserved amino acids are marked with different colors. Consensus is labeled as asterisk.

```
MB24_hypot      MTNYHELVKGKVYIGGVDAIQEAVKKHGVTEVFDLRAGGEEPEGFPVETTRHAYPIVEGV
CDB3orf8hy      MTNYHELVKGKVYIGGVDAIQDAVKKHGVTEVFDLRSGGQEPEGFPVEAKRHAYPIVEGV
pBC10987_B      MTNYHELVKGKVYIGGVDAIQDAVNKHGITEVFDLRAGGEEPEGFPSGTKRHAYPIVEGV
consensus       ****************** **** ****** ** *****   ********

                61         71         81         91         101        111
MB24_hypot      EGQDESVKTAIAAVKEAVEQDKKVFFHCSGGRNRTGTVATGLLVELGHASNVEEAEQKAK
CDB3orf8hy      EGEDESVKNAIGAVKEAVERGEKVFFHCSGGRNRTGTVATGLLVELGHASNVEEAEQKVK
pBC10987_B      EGQDESVRNAIAAVKEAVEQDKKVFFHCSGGRNRTGTVATGLLVELGHASNVEEAEQKAK
consensus       *********** *******  ************************************ *

                121        131        141
MB24_hypot      EIRSSINIKPEMKQVLKRLYV
CDB3orf8hy      EIRSIINIKPEMKQVLKRLYV
pBC10987_B      EIRSSISIKPEMKQVLNRLYV
consensus       *** ********** ***
```

APPENDIX 7: Flow chart showing the steps followed in assessing the β-galactosidase activity of *lacZ* fusion constructs.

Single colony was taken from LB plate (with appropriate antibiotic) and inoculated into LB medium

↓

Incubated with shaking overnight at 37°C

↓

Overnight cultures were diluted 100 times with fresh LB medium with appropriate antibiotic and incubated until the cultures reached logarithmic growth phase (OD_{600} = 0.2-0.7)

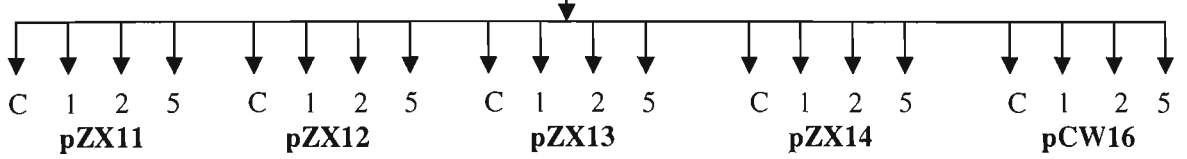
↓

Cultures were distributed into tubes, and divided into several groups:

↓

Control and Induced tubes

↓



↓

One group was control with no arsenite, while the other groups were induced by different final concentrations of arsenite

After incubation for additional 30mins, the cultures were immersed in ice for 20 minutes to prevent further growth

The cell density (OD_{600nm}) was measured with a spectrophotometer

0.1mL of culture was mixed in a tube with 0.9mL of Z buffer, 60μL of chloroform and 30μL of 0.1% SDS solution followed by vortexing for 10 seconds

Place the tubes in a 28°C water bath for 5 min, the reaction was started by adding 0.2mL of ONPG (4mg/mL) to each tube and shake the tubes for a few seconds

The tubes were incubated at 28°C water bath until sufficient yellow color was observed

After the reaction was stopped by adding 0.5mL of a 1M Na_2CO_3 solution and total assay time was recorded for each sample, the tubes were centrifuged at 4000g for 5 minutes

Finally, the absorbance of each assay tube at both 420nm and 550nm was measured by using UV-spectrophotometer

↓

The "Miller Units" of β-galactosidase activity was calculated using the following equation:

$$\text{Miller Units} = 1000 \times \frac{A_{420} - 1.75(OD_{550})}{t \times v \times OD_{600}}$$