Master Degree Project

UNIVERSITY OF SKÖVDE

Functional analysis of an *arsB* gene (gene-4251) presumably involved in accumulation of arsenics in *Lysinibacillus sphaericus*.

MB701A VT17: Master Degree Project in Molecular Biotechnology (30 credits)
Autumn term 2016

Vedant Subhash Borghate
a16vedbo@student.his.se

Supervisor: Prof. Abul Mandal
abul.mandal@his.se

Examiner: Patric Nilsson, PhD
patric.nilsson@his.se

ABSTRACT

Many regions of the world are facing the problem with arsenic toxicity. Arsenic contamination has become a considerable threat to the environment triggering various big health issues for every life in that contaminated environment. *Lysinibacillus sphaericus* (B1-CDA) is an arsenic tolerant strain of bacteria that has been reported and characterized before by the researchers of the University of Skövde, Sweden. The bacteria were found to contain many arsenic responsive genes such as *arsB*, *arsC*, and *arsR* which are responsible for arsenic tolerance in the bacterium. The main focus of the current study was to characterize one of the *arsB* genes (gene-4251) of *Lysinibacillus sphaericus* B1-CDA by *in silico* and *in vitro* analyses in order to determine the molecular function of this gene. The *in silico* studies conducted by using the Iterative Threading Assembly and Refinement (I-TASSER) server predicted the tertiary structure of the ArsB protein and suggested that this protein is an intrinsic component of the membrane which primarily helps in the binding of metal ions and liberation of metabolic energy. To validate this predictive results, several *in vitro* experiments were performed. For complementation studies, the *arsB* gene was cloned from *L. sphaericus* B1-CDA and transferred to an *arsB* knock-out mutant of *Escherichia coli JW3469-1*. Both, the transgenic and mutant strains were grown under the arsenic stress of 50 mM for 96 hrs followed by measuring their growth and arsenic tolerance after every 24 hrs. Statistical analysis confirmed that there was a significant difference in growth between the transgenic and the mutant *E. coli* strains. The ICP-MS (Inductive Coupled Plasma-Mass Spectroscopy) analysis revealed that after 24 hrs of culture, the arsenic content in the cell-free broth of transgenic strain was reduced from 50 mM to 9.10 mM (81.8%), whereas the reduction in arsenic content by the mutant strain was from 50 mM to 9.80 mM (80.2%). These results suggest that the *arsB* gene is partly involved in the accumulation of arsenic inside the cells and this feature could be used for a large scale removal of arsenic from the contaminated environment.

POPULAR SCIENTIFIC SUMMARY

Heavy metal poisoning is a severe threat to the human society as well as the environment. The poisoning is caused due to the accumulation of the toxic of heavy metals in soft tissues of the body. It is usually derived from industrial exposure, contaminated wastewater or air pollution, or ingestion of contaminated plant or animal-based food product. One of the naturally occurring and highly toxic chemicals is arsenic, also known as the king of poisons. Arsenic is widely distributed in soil and water and the extent of arsenic pollution is alarming. The maximum acceptable concentration of arsenic in water has been reduced from 50 parts per billion (ppb) to 10 ppb (EPA, 2001). Overexposure to inorganic arsenic severely impacts human health and may result in life-threatening complications such as many cancerous and heart diseases, melanosis like hyper- and hypo-pigmentation and cancers of skin, lung, and bladder, skin thickening, muscular spasms, loss of appetite as well as many neurological and vascular problems. Therefore, it is crucial to seek out ways to eliminate arsenic from the contaminated environment. Use of microorganisms for bioremediation of arsenic is one of the eco-friendly and cost-effective methods which makes it a successful approach for detoxifying the adulterated environment.

Several microbial strains have been reported to possess resistance to various heavy metals. *Lysinibacillus sphaericus* (B1-CDA) is one of the newly studied bacteria that can tolerate a very high concentration of arsenic and can grow well in the contaminated environment. This unique property is due to the presence of arsenic responsive genes that confer resistance mechanisms and tolerance against high levels of arsenic in the surrounding. These genes are usually located in a functional unit of the genomic DNA called as an operon and work together in a group. The *ars* operon consists of several structural genes- *arsB*, *arsC*, and *arsR* along with two additional components *arsA* and *arsD* and all collectively confer resistance to arsenic toxicity up to a certain level. Previous studies have acknowledged several genes of B1-CDA that are certainly responsible for accumulation and regulation of arsenic in the bacterial cells. One of the significant genes is an *arsB* gene which has been identified previously in various arsenic tolerant bacteria. The present study involves the functional analysis of *arsB* gene found in B1-CDA. The presumptive function of the gene was anticipated from *in silico* analyses and was established mainly on the basis of the tertiary structure of the ArsB protein. This was achieved by using the online server, I-TASSER. The gene function predicted by *in silico* studies was validated by performing *in vitro* experiments (studies on bacteria). These involved the isolation of an *arsB* gene from the whole genome of B1-CDA and cloning the gene into the 'mutant strain' of *Escherichia coli*- JW3470 lacking the *arsB* gene. This recombinant bacteria (cloned *E. coli* strain carrying the *arsB* gene) or 'transgenic strain' was used to carry out complementation studies, where the mutant strain served as a control. Both the transgenic and the mutant strain were exposed to different concentrations of arsenic. A significant difference could be seen, the transgenic bacteria (+*arsB*) were able to survive and grow in 50 mM arsenate unlike the mutant strain (-*arsB*) which could hardly survive in presence of arsenic.

Results from the study suggest that this gene from *Lysinibacillus sphaericus* (B1-CDA) accompanied by other arsenic responsive genes could be used to produce e. g., genetically modified plants that could serve as an eco-friendly and cost-effective means of bioremediation to detoxify the arsenic contaminated environment elsewhere.

ABBREVIATIONS

+*arsB*	*arsB* gene present
-*arsB*/Δ*arsB*	*arsB* gene deleted
As^{3+}	Arsenite
As^{5+}	Arsenate
ATPase	Adenosine triphosphatase
BLAST	Basic Local Alignment Search Tool
bp	Base pair
EPA	Environment Protection Agency
GO terms	Gene ontology terms
ICP-MS	Inductively Coupled Plasma – Mass Spectroscopy
I-TASSER	Iterative Threading ASSEmbly Refinement
LB media	Luria Bertani medium
LOMETS	Locally installed meta threading server
L. sphaericus	*Lysinibacillus sphaericus*
NCBI	National Center for Biotechnology Information
OD	Optical Density
PCR	Polymerase chain reaction
PDB	Protein Data Bank
ppb	parts per billion
PSI-BLAST	Position-specific iterated BLAST
RT-PCR	Reverse transcriptase polymerase chain reaction
Tm	Melting temperature
TM-Score	Template modeling score
X-Gal	5-Bromo-4-Chloro-3-Indolyl β-D-Galactopyranoside

TABLE OF CONTENTS

INTRODUCTION

Today, most of the severe environmental problems faced by many countries are due to heavy metal pollution. It has become a life-threatening challenge and a large number of people are affected each year by heavy metal pollution (Xiong et al., 2012; He et al., 2013). For most people, food, and water is the major source of exposure. Heavy metals have high atomic weight elements that occur naturally in the environment and also get widely distributed due to anthropogenic activities (Tchounwou et al., 2012). Arsenic is one such heavy metal known for its high toxicity and for the largest mass poisoning of a population in history and hence also referred as the king of poisons (Yosim et al., 2015). After years of evaluation, the United States environmental protection agency (EPA) dropped the maximum allowable level of arsenic in drinking water from 50 ppm to 10 ppm which could be considered safe (EPA, 2001).

Arsenic is significantly harmful in its inorganic form. It is found in all respects of the environment through release into the air by volcanoes and most of its dispersion is caused due to mining and industrial processes. The main source of exposure to arsenic in humans is food such as crops and sea-food which are contaminated with arsenic infested water (Chung et al., 2014). Acute or short-term inhalation of arsenic dust leads to nausea, diarrhea, abdominal pain and even central and peripheral nervous system disorders. Chronic or long-term exposure to inorganic arsenic results in irritation of the skin, skin lesions, hyperpigmentation, liver or kidney damage. Chronic exposure to elevated levels of arsenic is associated with life-threatening diseases like liver and lung cancer (Rossman et al., 2004; Watanabe et al., 2003; Yoshida et al., 2004).

Arsenic has a severe impact on human health and thus its elimination from the environment has become crucial for human welfare. There is a wide range of techniques and methods available to remove arsenic from surrounding but their impact is not significant and it has been scaled down to suit the rural environment. The most conventional methods include oxidation, precipitation, filtration, adsorption onto sorptive media, ion exchange, and membrane techniques (Ahmed, 2011). In recent years, the challenge of finding an adequate method to deal with the situation has drawn the attention of researchers towards the use of microbes as a means of bioremediation. Studies have revealed some arsenite-oxidizing and arsenate-reducing microbial community that is resistant to arsenic and could survive on arsenic-contaminated water or soil (Kao et al., 2013; Liao et al., 2011). Bioremediation integrates the use of microorganisms in degradation of the pollutants and contaminants and plays a vital role in cleaning the environment (EPA, 2001). Arsenic forms alloys with metals and covalent bonds with carbon, oxygen, hydrogen and other elements. Thus it has several structural forms and oxidation states but naturally, the most consistent and admissible forms of arsenic are inorganic and organic existing in the trivalent or pentavalent state as arsenite (As^{3+}) and arsenate (As^{5+}), respectively. Potentially, the arsenic trivalent form is considered more toxic than the pentavalent form. This is because of their reactivity with sulfur-containing compounds and generation of reactive oxygen species (ROS) (Hughes et al., 2011). Arsenic found in water is almost completely in the inorganic form and can be stable as both arsenite and arsenate. Groundwater arsenic contamination has become a serious threat to all mankind (Saxe et al., 2006). It has been accounted that at least 140 million people in 50 different countries including Argentina, Bangladesh, China, India, Mexico, and the United States of America have been drinking water having arsenic in it at levels higher than the WHO provisional guideline value of 10 µg/L (Chowdhury et al., 2000; Ravenscroft et al., 2009). The problem of Arsenic toxicity in groundwater in southern West Bengal (India) and southern and eastern parts of Bangladesh are at an alarming state that was first recognized about decades ago (Acharyya et al., 2000). Metal-

reducing microorganism can play a key role within the mobilization of arsenic and may alter it to completely different oxidation states in particular environmental conditions (Islam et al., 2004; Kuai et al., 2001).

Bioremediation of heavy metals using microorganisms has received a great deal of consideration during recent years mainly due to its advantages over conventional means. Researchers are interested in discovering and genetically modifying microorganisms for the removal of heavy metals (Clausen, 2000). Despite arsenic's toxicity, it has been demonstrated that some microorganisms are capable of both arsenite oxidation and arsenate reduction and can use arsenic compounds as electron donors, electron acceptors or possess a set of arsenic resistance genes that play a key role in detoxification mechanisms (Ahmann et al., 1994; Stolz et al., 2002). They are capable of resisting arsenic toxicity to some extent through the ars genetic system, ars operon, present in most bacterial genomes and many plasmids (Cervantes et al., 1994; Ji and Silver, 1992; Silver and Phung, 2005). The ars operon mainly encodes for five proteins, *arsA*, *arsB*, *arsC*, *arsD*, and *arsR*. Three of these cistrons, *arsA*, *arsB*, and *arsC* form a protein pump capable of extruding arsenite from the cells and the other two, *arsD* and *arsR* have a regulatory function (Dey and Rosen, 1995; Ramanathan et al., 1998; Silver et al., 1993). One of the highly studied genes in the row of ars genes is *arsB*. The ArsB protein in conjugation with *arsA* has been postulated to form a membrane complex which functions as an anion-translocating ATPase. The *ArsA* protein supports catalytic energy transducing of the anion pump, whereas the integral membrane protein, *arsB*, serves as both the anion channel and membrane binding site for the *arsA* protein. Studies show that the expression of *arsB* is essential to anchor the *arsA* protein to the inner membrane. The toxicity of the metalloids is related to uptake, whereas detoxification requires efflux. The *arsB* gene putatively encodes an arsenic efflux membrane protein (Meng et al., 2004; Tisa and Rosen, 1990). In a gram-negative bacteria called, *Campylobacter jejuni* it has been seen that inactivation of *arsB* resulted in a significant reduction in the minimal inhibitory concentration (MIC) of arsenite and arsenate, respectively. Additionally, overexpression of *arsB* in *C. jejuni* resulted in an increase in the MIC of arsenite to some extent (Shen et al., 2013).

Lysinibacillus sphaericus (B1-CDA) is an example of such heavy metal tolerant bacteria known for its arsenic resistance. This soil-borne bacterium was isolated from a cultivated land in the southwest region of Bangladesh. The bacteria were screened for the presence of any ars genes by previous researchers by employing bioinformatics tools. It came into knowledge that this bacteria harbors some arsenic marker genes such as acr3, *arsR*, *arsB* and *arsC* (Rahman et al., 2014; Rahman et al., 2015). The present study is targeted at determining the function of the *arsB* gene by *in silico* and *in vitro* experiments.

An online server, named, the Iterative implementation of the Threading ASSEmbly Refinement (I-TASSER) program was used to predict the three-dimensional structure of the protein (arsB) and the function associated with it. Following is a brief introduction of the simulations performed by the program during *in silico* analyses (Roy et al., 2010; Wu et al., 2007; Zhang, 2008) and the overall steps performed by I-TASSER in virtue of structure and function prediction for a given query sequence is shown in Figure 1.

i. Threading
   It is a knowledge-based tertiary structure prediction method. The program attempts to align the complete backbone plus side chains of a target sequence (here, *arsB*) with an unknown structure to just the backbone (no side chains) of a known structure through the available PDB structure library. The possible folds were searched by PSI-BLAST (position-specific iterated BLAST). From the significant homologous sequences, multiple alignments were

generated which were later used by PSI-PRED (PSI-blast based secondary structure prediction) for predicting the secondary structure of the protein. A sequence profile was created depending on the multiple alignments of homologous sequence and hence, the secondary structure was anticipated, followed by threading of *arsB* sequence through PDB structure database by LOMETS (Local Meta-Threading-Server). The structure and sequence-based scores were considered to rank the templates.

ii. Structure assembly

In this step, the continuous fragments were excised from the threading aligned regions which were used to rebuild full-length models while the ab initio modeling was used in threading unaligned regions. Then, the fragments assembly were constructed by replica-exchange Monte Carlo simulations. The refinement simulations generated conformations which were clustered by SPICKER to identify near-native protein folds.

iii. Model selection and refinement

The fragment assembly simulation was performed using the selected cluster centroids. The second round of simulations was performed by the structure alignment program (TM-align) where PDB structures and threading alignments by LOMETS anatomically nearest to the centroids were identified.

iv. Structure-based functional annotation

Based on the global topology, the structural trajectories of the query protein in the GO library were matched using TM-align. The mapping between the structural analogs in the EC (Enzyme Commission) number and binding site libraries was done on the basis of both global and local structural similarity. The functional annotations were presented on the basis of structural patterns preserved within the current model measured by sequence identity, structural alignment coverage, TM score, and REMO (reconstruct atomic model from reduced representation).

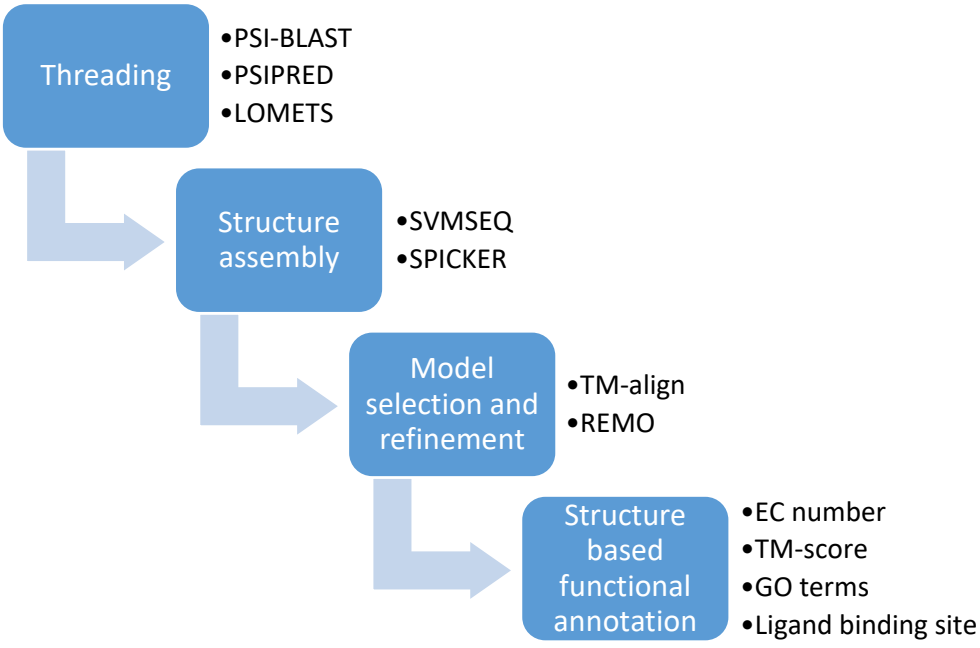


Figure 1. Schematic representation of the overall simulations performed by I-TASSER. Step 1, threading alignment employs PSI-BLAST, PSIPRED, and LOMETS. Step 2, SVMSEQ and SPICKER organizes structural assembly. Step 3, TM-score and REMO implement model selection and refinement. Step 4, the structure-based functional annotation is concluded based on Enzyme classification, TM-score, GO terms and ligand binding sites.

AIM

The primary aim of this study is to determine the molecular function of an arsenic responsive gene (*arsB* gene-4251) isolated from an arsenic tolerant soilborne bacterium *Lysinibacillus sphaericus* B1-CDA. Here, we investigated the involvement of the gene in uptake, accumulation, and sequestration of arsenic in the bacterial cells. The fundamental objectives of the study are:

- To clone a particular arsenic responsive gene *arsB* from *Lysinibacillus sphaericus* B1-CDA by PCR using database sequences as primers,
- To characterize the molecular function of the gene by using both *in vivo* and *in silico* methods, and
- To transfer the gene to a mutant strain lacking the *arsB* gene to carry out complementation studies for confirmation of gene function.

MATERIALS AND METHODS

3.1 *in silico* studies of the *arsB* gene:

The amino acid sequence (Appendix I) of *Lysinibacillus sphaericus* was retrieved from NCBI server (Accession: PRJNA296399). The ArsB protein comprises 431 amino acids (Appendix II). Prediction of the secondary and tertiary structures was done by submitting the *arsB* gene sequence to an automated online iterative threading server, I-TASSER. Results from *in silico* studies were validated by performing *in vitro* studies.

3.2 Chemicals and antibiotics:

All the bacterial strains used in the study were grown over LB media. Working solution of a selectable marker, Ampicillin (from Sigma-Aldrich), was prepared at a concentration of 100 µg/ml for transformation studies. The different concentrations of sodium arsenate used were prepared from the 500 mM of sodium arsenate (Sigma-Aldrich) solution provided by the University of Skövde, Sweden. All the dilutions were made using nuclease-free sterile water. Target specific primers (Appendix III) for the amplification of *arsB* gene were designed manually using a bioinformatic tool called SnapGene (SnapGene software, GSL Biotech) and they were custom made and ordered from Thermo Fischer Scientific. X-gal plates were made for screening purpose. All the chemicals were prepared and handled through sterile practice.

3.3 Bacterial strains and growth conditions:

The bacterial strains and plasmids used in the study are listed in Table 1. Sterilization of the growth media was achieved by autoclaving it at 121^0C for 15 min. Cultures were revived and grown in LB media at 37°C in a shaking incubator at 200 rpm. Same growth conditions were applicable for every type of bacteria used.

Table 1: Bacterial strains and plasmids used.

Organism	Description	Source
Lysinibacillus sphaericus B1-CDA	Isolated *arsB* Gene	University of Skövde
pGEM-T Easy Vector	Cloning Vector	Promega
E. coli JW3469-1	Δ(araD-araB)567, ΔlacZ4787(::rrnB-3), λ-, Δ*arsB*758::kan, rph-1, Δ(rhaD-rhaB)568, hsdR514	CGSC Yale university

3.4 Isolation of DNA and PCR:

Amplification of *arsB* gene was performed for cloning the gene into the vector. The genomic DNA was isolated from the overnight culture of B1-CDA by using MasterPure™ DNA Purification Kit (Epicentre) following the manufacturer's protocol. The DNA was dissolved and stored in TE buffer at -20°C.

Initially, a temperature gradient PCR was performed to optimize the annealing temperature using MasterAmp™ PCR Optimization Kits (Epicentre) and for this five different temperatures were tested in the presence of the target-specific primers. Then, for the amplification of *arsB* gene, a conventional

PCR was performed using the isolated DNA with the sequence-specific primers *arsB*-F (sense) and *arsB*-R (anti-sense) (Appendix III) using MasterAmp™ PCR Optimization Kits (Epicentre) and MasterAmp™ 2X PCR PreMixes (Epicentre) adhering to the provider's protocol (Appendix V). The cycling parameters were used as mentioned (Appendix VI). The PCR product obtained was purified by using QIAquick PCR purification kit (Qiagen) as per the kit's protocol and was verified by agarose gel electrophoresis.

3.5 Cloning and transformation:

The purified DNA was cloned into a pGEM-T Easy vector followed by transformation into the competent cells of *arsB* mutant *E. coli* JW3469-1. The ligation and transformation procedures were carried out in accordance with the pGEM-T Easy vector systems protocol (Promega). The purified *arsB* gene was ligated into the pGEM-T Easy vector (Appendix IV) by overnight ligation at 4^0C. The ligation mixture components were combined with an Insert: Vector ratio of 3:1 (Appendix VII). The mutant *E. coli* (-*arsB*) JW3469-1 strain (Coli Genetic Stock Centre, USA) were made competent using calcium chloride method as given by Tu et al., 2005. Aliquots were made and stored at -80^0C. The competent cells and the ligation mixture were mixed in a ratio suggested by the pGEM-T Easy vector systems protocol (Promega). The mixture was kept on ice for 20 mins. Then, without shaking, the tubes were suddenly subjected to heat shock at exactly 42°C for 45 seconds in a hot water-bath and immediately returned back to the ice for 2 mins. To condition and revive the cells from the heat shock, SOC medium (950 µl) at room temperature was added and incubated at 37°C in a shaking incubator at 200rpm for 90 mins. X-gal plates with ampicillin (100 µg/ml) were prepared for the blue-white screening of the recombinants. Transformation culture was then plated over X-gal agar plates (100 µg/ml Amp) according to the protocol (Promega). The same method was used for negative and positive control. The positive control carrying the control insert DNA was used to check the transformation potency whereas, on the opposite hand, the negative control offered the verification of background colonies emerging from vector alone. From the transformation plate, white colonies were selected which are suggestive of successful recombination.

DNA was isolated from the transformants using the same kit as mentioned earlier (Section 3.4) and PCR reaction was carried out with *arsB* primers. The cycling parameters were kept same as cited earlier (Section 3.4). Agarose gel electrophoresis was performed to confirm the presence of the *arsB* gene.

3.6 Analysis of gene expression by RT-PCR:

To evaluate the successful gene implantation and proper functioning of the *arsB* gene, expression analysis was carried out. The transgenic *E. coli* JW3469-1 (*arsB* +ve) and mutant *E. coli* JW3469-1 (*arsB* -ve) were grown overnight on LB plates in the presence of 50 mM sodium arsenate. Later, RNA was isolated using MasterPure™ RNA Purification Kit (Epicentre), followed by RT-PCR in presence of *arsB*-F and *arsB*-R primers (Appendix III) using MasterAmp™ High Fidelity RT-PCR Kit (Epicentre) in accordance with the company's protocol (Appendix VIII). The cycling parameters were used as mentioned (Appendix IX). The amplified PCR product was visualized by gel electrophoresis and hence the expression of the *arsB* gene in transgenic *E. coli* JW3469-1 was confirmed.

3.7 Gel electrophoresis:

The agarose gel electrophoresis was performed to visualize the output from PCR. 1X Tris-acetate EDTA (TAE) was prepared to use as a running buffer. 2% agarose gel was prepared using the 1X TAE. Electrophoresis was performed at 70 V to visualize the amplified products. GelgreenTm (Biotium) was used as gel dye at 1X concentration. For the loading of samples on the agarose gel, 6X gel loading dye

(NEB) was used at 1x concentration. As a reference to the molecular weight of obtained bands, 2log DNA ladder from NEB was used.

3.8 Arsenic tolerance:

Arsenic resistance assay for transgenic *E. coli* JW3469-1 strain bearing *arsB* gene and mutant *E. coli* JW3469-1 strain (lacking *arsB*) was performed. To find the optimum concentration of arsenic for growth studies and ICP analyses, the cells were cultured in LB medium having different concentrations (5 mM, 10 mM, 25 mM, 50 mM and 100 mM) of arsenic in it. The cells were grown at 30^0C. After overnight incubation, optical density at 600 nm (OD_{600}) was measured using a (WPA Bio wave CO8000) cell density spectrophotometer (Yan et al., 2017) and the difference in growth between the two groups was observed.

3.9 Growth curve and ICP-MS:

The transgenic *E. coli* JW3469-1 (*arsB* +ve) and mutant *E. coli* JW3469-1 (*arsB* -ve) were grown in LB medium with 100 µg/ml ampicillin and without ampicillin, respectively, in presence of 50 mM sodium arsenate as this concentration was optimum for both the strains to grow with a significant difference. For each group, three parallel sets of bacterial cultures were grown for 96 hr at 37°C with continuous shaking at 180 rpm. Each set contained four culture flasks. On interval of every 24 hrs, samples were removed and their optical cell density was measured at OD_{600} using a cell density spectrophotometer, as described previously by Yan et al., 2017.

After measuring absorbance at each time interval, the bacterial cultures were prepared for measurement of arsenic using Inductively coupled plasma mass spectrometry (ICP-MS) (Rahman et al., 2014). Cells were harvested by centrifugation at 10,000 g for 10 min from every culture flask. The quantity of arsenic present within the supernatant or cell-free media was evaluated by ICP-MS. For determination of arsenic in cells, the cell pellet was washed thrice with deionized water and then dried completely until a constant dry weight was achieved. Measurement of arsenic was carried out by Eurofins Environment Testing Sweden AB (Lidköping, Sweden) by using ICP-MS.

3.10 Statistical analysis:

Statistical analysis was carried out to compare the magnitude of the growth rate of mutant *E. coli* JW3469-1 and transgenic *E. coli* JW3469-1. SPSS statistics v.24 program was used to accomplish Mann-Whitney U test for analysis with a significance level set at 0.05.

RESULTS

The *in silico* analysis of the *arsB* gene by the I-TASSER server provided the secondary and tertiary structure and postulated its putative function as shown here:

4.1 Secondary structure prediction:

The query sequence was threaded through a representative PDB library and possible folds were searched by PSI-BLAST, one of the profile-profile alignment (PPA) method. Then, by applying different combinations of alignment algorithms, PSI-PRED summed up the raw probabilities based on helix, strand and coiling state. Selecting the state of the highest probability, the secondary structure of the ArsB protein was achieved, Figure 2.

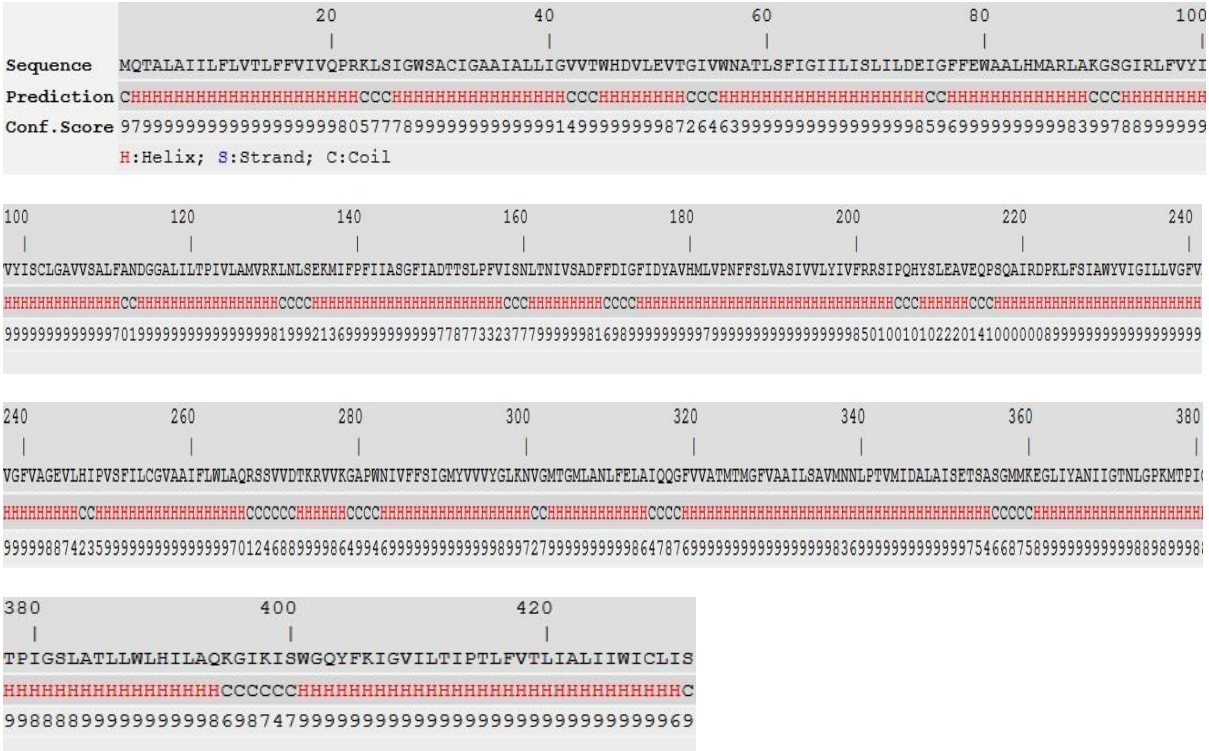
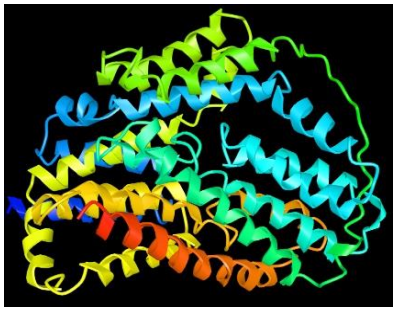


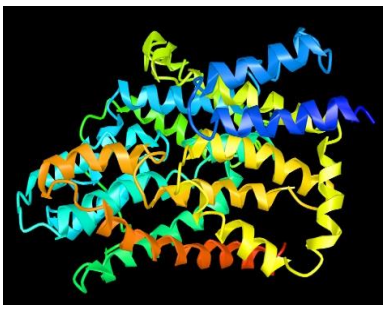
Figure 2. The secondary structure of the ArsB protein. C, H, and S denote coil, α-helix, and β-sheet, respectively. Conf. score refers to the confidence score of amino acid ranging from 1 to 9, where a higher score means a more confident prediction of secondary structure.

4.2 Tertiary structure prediction:

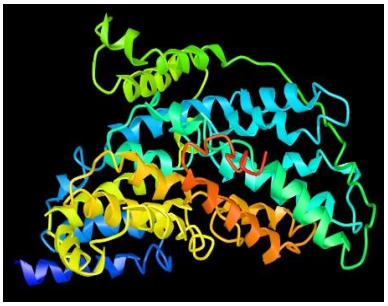
A series of simulations by I-TASSER produced a large number of possible alignments which were used to assemble structural conformations. The fragment assembly was performed using a modified replica-exchange Monte Carlo simulation technique. The conformations generated were clustered by SPICKER and cluster centroids were obtained by averaging the 3D coordinates. Top five cluster centroids were selected generating the full atomic models. They were ranked according to their C-score, which is a measure of quality estimation of predicted models. C-score is generally in the range of (-5 to 2). Higher the C-score, higher the quality of the anticipated model and vice-versa.



a) Model 1 (C-score: 0.37)



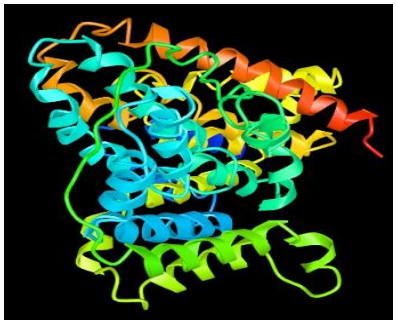
b) Model 2 (C-score: -0.38)



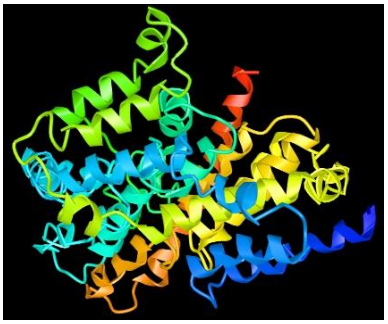
c) Model 3 (C-score: -2.69)

Estimated TM-score = 0.76±0.10

Estimated RMSD = 6.2±3.8Å



d) Model 4 (C-score: -2.43)



e) Model 5 (C-score: -2.52)

Figure 3 (a, b, c, d, and e): The figures represent the 3D structure of all the top five models of the ArsB protein generated by I-TASSER. The models are arranged according to their respective C-score.

Template modeling score (TM- score) and root-mean-square deviation value (RMSD value) were used to measure the structural similarity of the predicted conformations. TM-score ranges from (0 to 1). A TM-score<0.17 implies random similarity and a TM-score>0.5 implies the model of correct topology. Higher the score, better the similarity. Whereas, RMSD value signifies coverage of the structural similarity. The protein structure, shown in Figure 3(a) was the top-ranked conformation among all the native models. The accuracy of each model was determined based on their cluster density, C-score, TM-score, and RMSD value.

4.3 Structure-based functional annotation and prediction:

The top-ranked predicted model of the target protein was matched with the available known structural data in the PDB library. The TM-align program generated the top ten models based on nearest structural similarity resulting in the proteins possessing a similar function to the target protein. The predicted function of the target gene was classified into three divisions based on Enzyme Commission numbers, GO terms and ligand-binding sites. Based on global and local structural similarities the structural analogs of the predicted protein were matched against a large number of non-redundant entries with known Gene Ontology (GO) terms in the PDB library and then, the consensus was predicted based on the prevalence of GO terms. The molecular functions included: metal ion binding, interaction with heavy metal ion selectively and non-covalently and tetrapyrrole binding. The biological activity involved: generation of precursor metabolites, metabolic energy liberation through chemical pathways and covalent alteration or modification of macromolecule.

Found as an intrinsic component of the membrane and sustains as a part of a macromolecular complex.

4.4 Analysis of *arsB* in B1-CDA

DNA was isolated from B1-CDA and PCR was performed in presence of sequence-specific primers (*arsB*-F and *arsB*-R) to analyze the *arsB* gene and to optimize the annealing temperature using five different temperatures (51.2°C, 52.1°C, 53.5°C, 54.7°C, and 55.6°C). The amplified PCR product was visualized by agarose gel electrophoresis (Figure 4).

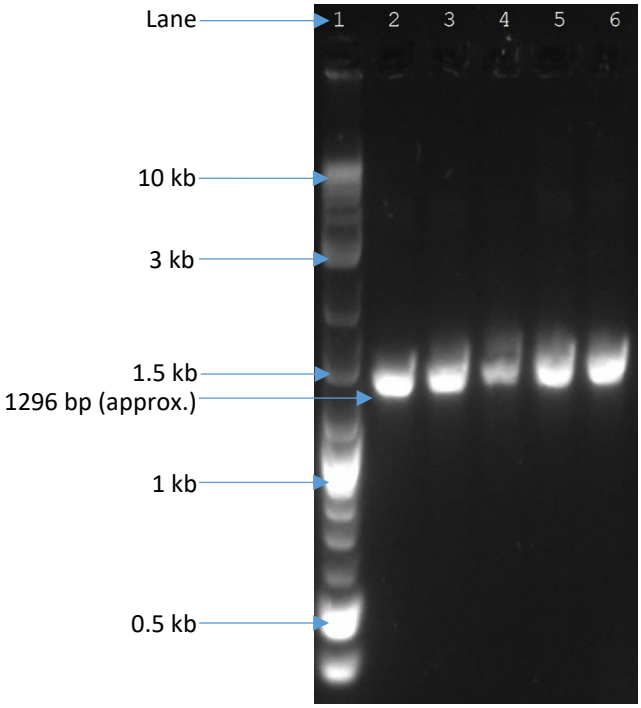


Figure 4. Image obtained from gel electrophoresis of PCR products (1296 bp long *arsB* gene) from temperature gradient PCR. Lane 1: 2log DNA ladder; Lane (2, 3, 4, 5, and 6): PCR product with different annealing temperatures as mentioned earlier, respectively.

All the different temperatures showed a band of good intensity. The lane 2 was considered best and 55.6°C was chosen as an optimum annealing temperature for further PCR programs.

4.5 Cloning and transformation of the *arsB* gene.

The DNA from B1-CDA was isolated again and PCR was performed using primers *arsB*-F and *arsB*-R. After purifying the PCR product, the *arsB* gene was cloned into a pGEM-T Easy vector. The cloned vector was then transformed into mutant *E. coli* JW3469-1 competent cells by the heat-shock method. The transformation was then plated over the X-gal plates for blue-white screening. White colonies were selected and DNA was isolated from them. Later, PCR analysis was performed using *arsB* specific (forward and reverse) primers for validation of successful transformation and PCR product was visualized by performing agarose gel electrophoresis (Figure 5).

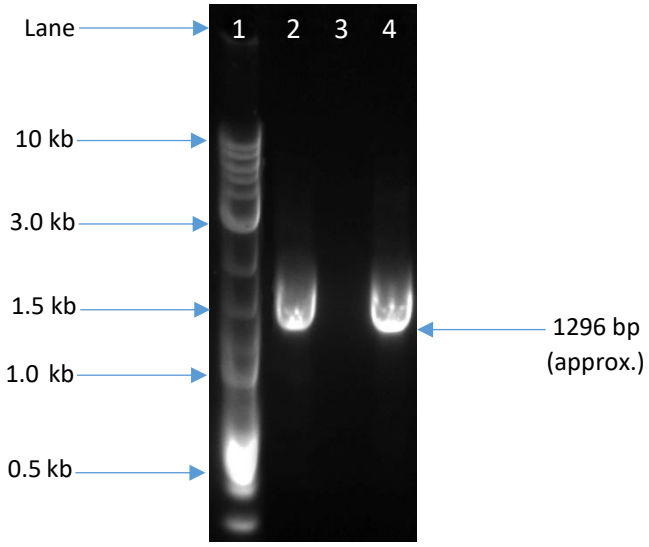


Figure 5. Transformed *arsB* gene into mutant *E. coli* strain. Lane 1: 2-log DNA ladder; Lane 2: Transformed *E. coli* strain carrying the *arsB* gene; Lane 3: mutant *E. coli* (negative control); Lane 4: B1-CDA (positive control).

In figure 5, a band was observed in lane 2, which corresponds to 1296 bp validating a successful transformation. The mutant *E. coli* strain lacking the *arsB* gene and B1-CDA strain was used as a negative and positive control, respectively.

4.6 Expression analysis of the functional *arsB* gene by RT-PCR:

The transgenic *E. coli* and mutant *E. coli* cells, both were grown in presence of arsenic to check the functionality of the implanted *arsB* gene. Both the strains were exposed to 50 mM sodium arsenate. Later, RNA was isolated from these strains and RT-PCR was performed to analyze the expression of the *arsB* gene. Gel electrophoresis was performed to validate the result (Figure 6).

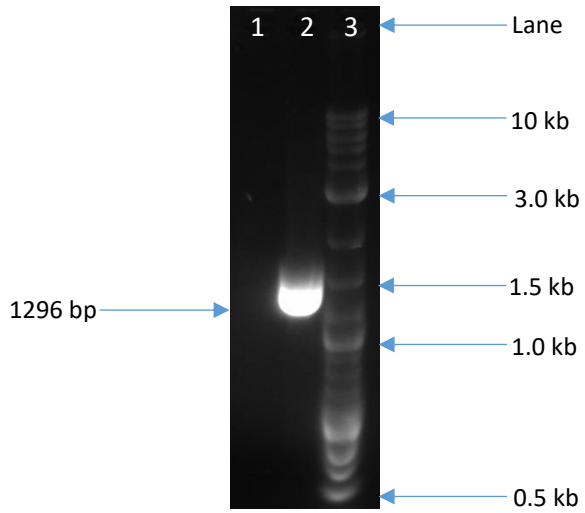


Figure 6. Expression of the *arsB* gene. In the gel, Lane 1 represents PCR sample from mutant *E. coli* (negative control), lane 2 represents *arsB* RNA (1296 bp) from transgenic *E. coli*, and lane 3 represents 2-log DNA ladder for reference.

4.7 Degree of arsenic tolerance:

The arsenic assay was performed in which the transgenic *E. coli* and the mutant *E. coli* strain was exposed to different concentrations of arsenic ranging from 5 mM to 100 mM. Measurements of optical density at OD_{600} were taken after 24 hrs. The results are shown in Table 2.

Table 2. The cell density of the transgenic *E. coli* JW3469-1 and mutant *E. coli* JW3469-1 at various concentrations of arsenic.

S. No.	Concentration of arsenic (mM)	Optical density at 600nm	
		Transgenic *E. coli* JW3469-1	Mutant *E. coli* JW3469-1
1	5	1.67	0.25
2	10	1.64	0.16
3	25	0.84	0.14
4	50	0.76	0.11
5	100	0.37	0.09

A significant difference was observed between the growth of the two strains. The growth of the transgenic strain was found to be much higher than that of the mutant strain suggesting a higher tolerance level in the transgenic *E. coli* strain.

4.8 Statistical analysis of bacterial growth rate:

To analyze the bacterial growth rate in the transgenic *E. coli* JW3469-1 and mutant *E. coli* J W3469-1, both the strains were exposed to a 50 mM concentration of sodium arsenate and OD was recorded by using a cell density spectrophotometer (CO 8000 Biowave Cell Density Meter) at varied time intervals ranging from 24 to 96 hrs. The Mann-Whitney U test was performed for statistical evaluation of the growth rate. Figure 7 shows the growth pattern obtained after a statistical evaluation.

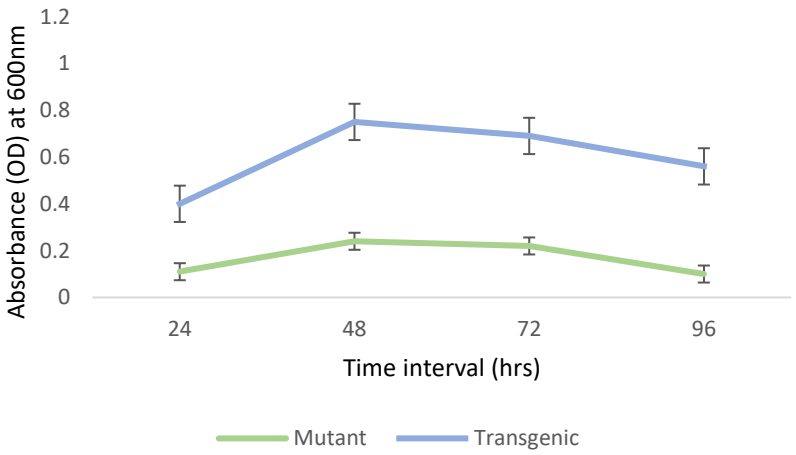


Figure 7. Growth pattern comparing mutant *E. coli* JW3469-1 and transgenic *E. coli* JW3469-1 in the presence of 50 mM sodium arsenate at varied time intervals. Error bars signify mean ± SE.

The median for the mutant cell growth (median =0.17, n=4) and the transgenic cell growth (median = 0.63, n=4) was significantly different according to the obtained p-value, 0.03. Since the p-value is less than the significant level of 0.05, the null hypothesis was rejected and a conclusion was made that due to the presence of active *arsB* gene the growth rate of the transgenic strain is higher than the mutant strain in which the *arsB* gene was omitted.

4.9 Analysis of arsenate accumulation by ICP-MS:

The transgenic *E. coli* JW3469-1 and the mutant *E. coli* JW3469-1 strains were grown in LB medium in presence of a 50 mM concentration of sodium arsenate for up to 96 hrs. After every 24 hrs, the cells were harvested from the broth and the concentration of arsenic in the cell-free medium of both the strains were analyzed by ICP-MS. The measure of residual arsenic content in the supernatant is given in Table 3.

Table 3. ICP-MS data showing the concentration of arsenic in cell-free broth. The assessment was done at various time intervals simultaneously for both the strains.

Time interval (hours)	**Arsenic concentration in cell-free broth (mM)**	
	Mutant *E. coli*	**Transgenic *E. coli***
24	9.90	9.10
48	10.46	9.88
72	10.53	10.20
96	11.18	11.00

The values in table 3 show a significant reduction in the arsenic concentration in the broth containing the transgenic *E. coli* JW3469-1 exhibiting the *arsB* gene compared to that of *arsB* mutant *E. coli* strain, suggesting the adsorption or accumulation of the dissolved arsenic by the activity of the *arsB* gene.

DISCUSSION

Arsenic poisoning is stated as a medical condition that develops due to the profound levels of arsenic in the body. The condition is caused by consumption of contaminated water, food, and exposure to industrial effluents. It is naturally present at toxic levels in groundwater of many countries (Choong et al., 2007; Thirunavukkarasu et al., 2003). The pentavalent state of arsenic is considered as most dangerous form. An arsenic affected body shows symptoms like a headache, vomiting, nausea, dizziness, abdominal pain and severe diarrhea. The chronic toxicity is associated with multisystem diseases and is also documented as carcinogenic affecting various organs in humans (He at al., 2013; Hughes, 2002; Ratnaike, 2003). Arsenic toxicity has become a major global health problem, currently estimated to affect ~150 million people around the world (Podgorski et al., 2017). This number is alarmingly high and indicate towards the need of an adequate and dynamic method to remove these heavy metals from the environment and efficient in detoxification of our surrounding. There are various methods available to treat contaminated soil and water such as bioremediation, filtration, ion exchange method, membrane filtration, and oxidation (Cheng et al., 1994; Mohan and Pittman, 2007). Among all of these, bioremediation is considered as the most eco-friendly and cost-effective method and hence in recent years, it has gained a lot of attention of the scientific community (Singh et al., 2008). Throughout the years, as an outcome of evolution, many microorganisms living in contaminated regions have developed mechanisms which make them resistant to heavy metal toxicity (Xiong et al., 2012).

Lysinibacillus sphaericus (B1-CDA) is one of the newly recognized bacteria which has shown to sustain up to 500 mM concentration of arsenic and have potential to serve as a tool for bioremediation (Rahman et al., 2014). Previous studies have documented various genes accountable for arsenic tolerance in B1-CDA (Rahman et al., 2015). The current study focuses on the *arsB* gene that is identified as an arsenic responsive gene previously by Rahman et al., 2015. In the current study, *in silico* study was done to determine the molecular function of the *arsB* gene. The obtained theoretical results were then validated by performing *in vitro* experiments. The *in silico* studies were carried out using I-TASSER software. I-TASSER is an online automated platform for full-length secondary and tertiary model prediction of protein structures and prediction of function based on a state-of-the-art query sequence-structure similarity prototype (Roy et al., 2010). The *In silico* results by I-TASSER proposed top five predicted models of the protein. Out of which, the first model (figure 3.a) was considered as the most appropriate one because it possessed largest cluster size and highest confidence score based on RMSD and Z-score co-relation. The functional similarities of the predicted structure and analogues were identified based on the highest occurrence of a GO-score (Roy et al., 2011). The predicted *arsB* gene molecular functions included metal ion binding, interaction with heavy metal ion selectively and non-covalently, and tetrapyrrole binding. Their biological activity involved generation of precursor metabolites, metabolic energy liberation through chemical pathways and covalent alteration or modification of macromolecule, and are primarily found as an intrinsic component of the membrane and sustains as a part of a macromolecular complex which is in resemblance to the earlier studies done on *E. coli* R773 (Dey and Rosen, 1995; Meng et al., 2004; Saltikov and Olson, 2002).

The anticipated function was verified through complementation studies during *in vitro* experiments. Firstly, the target-specific primers (Appendix III) were designed using a bioinformatic software called SnapGene (SnapGene software, GSL Biotech) (de Kruijf et al., 2017). The primers were designed especially to target *arsB* gene sequence. The genomic DNA was isolated from *Lysinibacillus sphaericus* (B1-CDA) and using the sequence-specific primers, the *arsB* gene was amplified in a PCR reaction. For complementation studies, a mutant strain of *E. coli* (*E. coli* JW3469-1) from Coli Genetic Stock Center

was evaluated for the absence of the *arsB* gene. Using the same primers (*arsB*-F and *arsB*-R) a PCR program was carried out where no amplification of *arsB* gene was found which confirmed the mutation. The *arsB* mutant *E. coli* was used as a negative control during the whole experiment. Following this, the *arsB* gene was cloned into pGEM-T Easy Vector and transformed into the mutant *E. coli* strain. The positive transformants were identified through blue-white screening method where the white colonies were an indication of positive transformation (Julin, 2014). Again a PCR was performed using isolated DNA from transformants in presence of *arsB* primers for further verification (Figure 5) of the insert of the *arsB* gene.

Gene expression studies were carried out on the *arsB* gene where the transformants were grown over 50mM concentration of arsenic. RNA was isolated from these bacteria and expression analysis was done using RT-PCR with *arsB* primers to generate cDNA. The results from the gel electrophoresis confirmed the presence of the *arsB* gene in its functional form under adverse condition (arsenic stress) (Figure 6).

To determine the degree of arsenic tolerance in the transgenic *E. coli* JW3469-1 due to the presence of an *arsB* gene, arsenic tolerance study was administered in the presence of sodium arsenate with serial increment in concentrations as mentioned in the previous study by Sunitha et al. (Sunitha et al., 2015). The analysis provided the minimum inhibitory concentration of arsenic for the transgenic and mutant *E. coli* JW3470-1 strains, which could be used as a benchmark for further growth studies (Table 2). Both the transgenic and mutant strains were grown for 24 hrs in the presence of various concentrations of arsenic (5 mM, 10 mM, 25 mM, 50 mM and, 100 mM). The absorbance values obtained at 600 nm revealed that the transgenic *E. coli* strain comparatively grew at a higher rate than the mutant *E. coli* strain with respect to all the concentrations of arsenic. The growth rate of the transgenic strain at 5 mM and 25 mM was six times higher than that of the mutant strain. The rate difference was found to be the highest at 10 mM and the lowest at 100 mM i.e. ten and four times higher, respectively. The growth rate being seven times higher was in between these ranges at a 50 mM concentration of arsenic. These results go in harmony with the study done by previous researchers, where it has been shown that the bacterial strains possessing the *arsB* gene were more arsenic tolerant than the bacteria lacking the *arsB* gene given to its role as an arsenite transporter (Cai et al., 2009; Chang et al., 2008). Hence, it would be appropriate to say that *arsB* plays an important and vital role in tolerating arsenic in arsenic-resistant bacteria.

After arsenic tolerance assay, a growth curve study was performed for the transgenic *E. coli* JW3469-1 and the mutated *E. coli* JW-3469-1 strains by exposing to a common arsenic concentration. The strains were cultured in a Luria Bertani (LB) broth and the concentration of arsenic was initially maintained at 50 mM for both the cultures as this concentration was considered to be optimum to study the difference in growth curves of both the strains. The absorbance was recorded at varying time intervals from 24 hrs to 96 hrs. The transgenic *E. coli* JW3469-1 strain showed an exponential growth till 48 hrs where its growth was almost doubled. With the increase in time, a gradual decrement was observed in the growth till the 72nd hour, probably due to increase in saturation of secondary metabolites causing inadequate aeration. As the time interval reached to 96 hrs, a heavy decrement was observed in the growth the reason being the exhaustion of growth nutrients suspended in the broth. A similar growth pattern was observed for the mutant *E. coli* strain. However, the transgenic strain showed a significantly higher growth rate than the mutant strain. Later, a statistical evaluation was done between these two groups using Mann-Whitney U test (Wang and Bushman, 2006). The results obtained from the statistical test validated the growth rate difference and a growth curve chart was made that indicates a significant difference in growth between the mutant *E. coli* JW3469-1 and the transgenic *E. coli* JW3469-1 strains.

The amount of arsenic accumulated in the cell-free broth of the mutant and the transgenic *E. coli* JW3469-1 strains were determined by performing ICP-MS analysis. The analysis was carried out at different time intervals i.e. from 24 hours to 96 hours and the initial concentration of arsenic was 50 mM. The ICP-MS analysis results (Table 3) proposed a minimal difference between the arsenic accumulation capacity of both the strains. In simple words, the concentration of arsenic in the supernatant of the transgenic *E. coli* was less than the concentration of arsenic in the supernatant of the mutant *E. coli* but, the difference was slight and not significant. The maximum difference in concentration of arsenic (in the cell-free broth) was observed at 24 hrs which was 0.8 mM where the amount of arsenic in supernatant of the mutant *E. coli* was 9.80 mM and that in the transgenic *E. coli* was 9.10 mM i.e. arsenic accumulation was slightly higher in the transgenic *E. coli* strain (due to the action of incorporated *arsB* gene). A simple representation of the results in Table 3 could be given as, in 24 hrs, the transgenic strain was able to decrease the arsenic concentration by 81.8% (50 mM to 9.10 mM) whereas that of the mutant strain was 80.2% (50 mM to 9.80 mM). However, it was observed that after 48 hrs the arsenic concentration in the supernatant was elevated for both the strains and it continued to rise slightly during the subsequent hours. This suggests that the arsenic was released outside the cells by efflux action as a response to the increased internal toxicity. This can be supported through the earlier studies done on B1-CDA and other species of bacteria by Banerjee et al. (2011) and Rahman et al. (2014) where similar results show that usually, the accumulation of arsenic in supernatant tends to increase after 48 hrs of incubation but again decreases after 72 hrs. Though the obtained results of arsenic accumulation studies are not rational as there was a slight but steady rise in the concentration of arsenic in the supernatant of both mutant *E. coli* JW3469-1 and transgenic *E. coli* JW3469-1. However, the current ICP-MS results are inconclusive as there were no replicates and hence any statistical evaluation could not be performed.

It can be concluded that the primary objectives of the conducted study were accomplished successfully. The *in silico* results suggest that the studied *arsB* gene is an intrinsic component of the membrane which primarily helps in metal-ion binding and serves a role in metabolic energy liberation and generation of precursor metabolites. The gene function predicted by *in silico* results were validated by performing *in vitro* experiments. Pursuing the goal, complementation studies were carried out successfully for which the *arsB* gene from *Lysinibacillus sphaericus* B1-CDA was transformed into a mutant *E. coli* strain that lacked the gene. The transgenic *E. coli* JW3469-1 was subjected to different growth studies and arsenic assays where the mutant *E. coli* JW3469-1 was used as a negative control. Unfortunately, the ICP-MS results were not rational and degree of arsenic accumulation in the bacteria could not be concluded due to the absence of sample replicates and hence arsenic accumulation assay could be repeated. However, the growth pattern showed that indeed there was a substantial difference in growth rate of the mutant *E. coli* JW3469-1 and the transgenic *E. coli* JW3469-1 which justifies the vital role of the *arsB* gene in arsenic tolerance.

ETHICAL ASPECTS AND IMPACT ON THE SOCIETY

The ethical concerns do not apply here as the study did not involve any human or animal model. The bacterial strains used here were non-hazardous in all manners. The complete study was done in accordance with the sterile lab practices. Use of antibiotics and arsenic solutions were carried out in a separate chamber. While handling arsenic, proper lab coats, gloves, safety glasses, and masks were worn. No powder form of arsenic was used as the pre-made stock solutions were provided by the University of Skövde. Before steam sterilization of arsenic-exposed bacteria, they were first disinfected using 1% virkon. The disposable materials that were used to handle arsenic were kept in a properly labeled biohazard waste box and it was kept separate from normal lab waste.

This study could help in better understanding of the *arsB* gene and its function. Bacterial strains with the *arsB*, coupled with other arsenic-responsive genes all together with a way of implementation could prove as a sustainable and economical way of bio-remediation which could make the environment free or less saturated with heavy metal contamination. Further taking genetic engineering into account, bioengineered crops resistant to arsenic can be produced which could detoxify the contaminated soil. After a certain period of time, these crops could be recycled for the production of bio-fuel.

FUTURE PERSPECTIVES

The primary goals of this research were attained with success. The current study showed that the *arsB* gene helps in metal-ion binding and functions as an arsenite efflux pump. The *in silico* results obtained from bioinformatic tools were validated by performing *in vitro* experiments. Complementation studies were accomplished and to achieve this, cloning and transformation of the *arsB* gene into a mutant *E. coli* JW3469-1 (deficit of *arsB*) were performed successfully. Through growth studies and arsenic assays, it was found that the transgenic *E. coli* strain was able to resist arsenic toxicity and grow significantly better than the mutant *E. coli*.

Further advancements can be done in this study my exploiting other genes of B1-CDA and exploring their biological behaviour and molecular functions. It would be more reliable and legitimate to analyze the expression levels of these genes through real-time studies. Unlike this experiment, the ICP-MS study could be performed with double or triple replicates of the sample to speculate the arsenic accumulation capacity of the bacteria even better. By using binary vector system, transgenic plants could be grown in large scale and this could help in reduction of arsenic contamination and thereby detoxifying our environment. This kind of bioremediation could prove as eco-friendly, economical and ergonomic in future.

ACKNOWLEDGEMENTS

First of all, thanks to my family for their eternal love and support. I owe them the highest degree of appreciation. Without their support, this journey would not have been possible.

I would like to express my heartfelt gratitude to my thesis advisor Dr. Abul Mandal, Professor of Molecular Biology at University of Skövde, Sweden. The door to Prof. Mandal was always open whenever I ran into a trouble or had a question about my research or writing. He consistently allowed this paper to be my own work but steered me in the right the direction whenever he thought I needed it.

I thank the faculty of Bioscience, University of Skövde, Sweden for nurturing my academic career and the management for facilitating this degree program and training.

Lastly, I am thankful to all of those who supported me in any respect during completion of my thesis.

Thank You All!

REFERENCES

Acharyya, S.K., Lahiri, S., Raymahashay, B.C. and Bhowmik, A., 2000. Arsenic toxicity of groundwater in parts of the Bengal basin in India and Bangladesh: the role of Quaternary stratigraphy and Holocene sea-level fluctuation. Environmental Geology, 39(10), pp.1127-1137.

Ahmann, D., Roberts, A.L., Krumholz, L.R. and Morel, F.M., 1994. Microbe grows by reducing arsenic. Nature, 371(6500), pp.750-750.

Ahmed, M.F., 2001, May. An overview of arsenic removal technologies in Bangladesh and India. In Proceedings of BUET-UNU international workshop on technologies for arsenic removal from drinking water, Dhaka (pp.5-7). Available at: http://www.rainforum.org/images/documents/rainforum_paper/an_overview_of_arsenic_removal_technologies_in_bangladesh_and_india.pdf [accessed 14 December 2017].

Cai, L., Liu, G., Rensing, C. and Wang, G., 2009. Genes involved in arsenic transformation and resistance associated with different levels of arsenic-contaminated soils. BMC microbiology, 9(1), p.4.

Cervantes, C., Ramirez, J. and Silver, S., 1994. Resistance to arsenic compounds in microorganisms. FEMS microbiology reviews, 15(4), pp.355-367.

Chang, J.S., Kim, Y.H. and Kim, K.W., 2008. The ars genotype characterization of arsenic-resistant bacteria from arsenic-contaminated gold–silver mines in the Republic of Korea. Applied microbiology and biotechnology, 80(1), pp.155-165.

Cheng, R.C., Liang, S., Wang, H.C. and Beuhler, M.D., 1994. Enhanced coagulation for arsenic removal. Journal of the American Water Works Association;(United States), 86(9).

Choong, T.S., Chuah, T.G., Robiah, Y., Koay, F.G. and Azni, I., 2007. Arsenic toxicity, health hazards and removal techniques from water: an overview. Desalination, 217(1-3), pp.139-166.

Chowdhury, U.K., Biswas, B.K., Chowdhury, T.R., Samanta, G., Mandal, B.K., Basu, G.C., Chanda, C.R., Lodh, D., Saha, K.C., Mukherjee, S.K. and Roy, S., 2000. Groundwater arsenic contamination in Bangladesh and West Bengal, India. Environmental health perspectives, 108(5), p.393.

Chung, J.Y., Yu, S.D. and Hong, Y.S., 2014. Environmental source of arsenic exposure. Journal of Preventive Medicine and Public Health, 47(5), p.253.

Clausen,C.A. (2000) Isolating metal-tolerant bacteria capable of removing copper, chromium, and arsenic from treated wood. Waste Manage. Res. 18, (p. 264–268).

de Kruijf, M., Govender, R., Yearsley, D., Coffey, A. and O'Mahony, J., 2017. A comparative study evaluating the efficacy of IS_MAP04 with IS900 and IS_MAP02 as a new diagnostic target for the detection of Mycobacterium avium subspecies paratuberculosis from bovine faeces. Veterinary Microbiology, 204, pp.104-109.

Dey, S. and Rosen, B.P., 1995. Dual mode of energy coupling by the oxyanion-translocating *ArsB* protein. Journal of Bacteriology, 177(2), pp.385-389.

EPA. Arsenic: Rule-Making History [website]. Washington, DC:Office of Water, U.S. Environmental Protection Agency (updated 6 March 2012). Available at: http://water.epa.gov/lawsregs/rulesregs/sdwa/arsenic/history.cfm [accessed 13 December 2017].

He, B., Yun, Z., Shi, J. and Jiang, G., 2013. Research progress of heavy metal pollution in China: sources, analytical methods, status, and toxicity. Chinese Science Bulletin, pp.1-7.X

Hughes, M.F., 2002. Arsenic toxicity and potential mechanisms of action. Toxicology letters, 133(1), pp.1-16.

Hughes, M.F., Beck, B.D., Chen, Y., Lewis, A.S. and Thomas, D.J., 2011. Arsenic exposure and toxicology: a historical perspective. Toxicological Sciences, 123(2), pp.305-332.

Islam, F.S., Gault, A.G., Boothman, C., Polya, D.A., Charnock, J.M., Chatterjee, D. and Lloyd, J.R., 2004. Role of metal-reducing bacteria in arsenic release from Bengal delta sediments. Nature, 430(6995), pp.68-71.

Ji, G. and Silver, S., 1992. Regulation and expression of the arsenic resistance operon from Staphylococcus aureus plasmid pI258. Journal of Bacteriology, 174(11), pp.3684-3694.

Julin, D., 2014. Blue-White Selection. In Molecular Life Sciences (pp. 1-2). Springer New York.

Kao, A.C., Chu, Y.J., Hsu, F.L. and Liao, V.H.C., 2013. Removal of arsenic from groundwater by using a native isolated arsenite-oxidizing bacterium. Journal of contaminant hydrology, 155, pp.1-8.

Kuai, L., Nair, A.A. and Polz, M.F., 2001. Rapid and simple method for the most-probable-number estimation of arsenic-reducing bacteria. Applied and environmental microbiology, 67(7), pp.3168-3173.

Liao, V.H.C., Chu, Y.J., Su, Y.C., Hsiao, S.Y., Wei, C.C., Liu, C.W., Liao, C.M., Shen, W.C. and Chang, F.J., 2011. Arsenite-oxidizing and arsenate-reducing bacteria associated with arsenic-rich groundwater in Taiwan. Journal of contaminant hydrology, 123(1), pp.20-29.

Meng, Y.L., Liu, Z. and Rosen, B.P., 2004. As (III) and Sb (III) uptake by GlpF and efflux by *ArsB* in Escherichia coli. Journal of Biological Chemistry, 279(18), pp.18334-18341.

Mohan, D. and Pittman, C.U., 2007. Arsenic removal from water/wastewater using adsorbents—a critical review. Journal of hazardous materials, 142(1), pp.1-53.

Podgorski, J.E., Eqani, S.A.M.A.S., Khanam, T., Ullah, R., Shen, H. and Berg, M., 2017. Extensive arsenic contamination in high-pH unconfined aquifers in the Indus Valley. Science advances, 3(8), p.e1700935.

Rahman, A., Nahar, N., Nawani, N.N., Jass, J., Desale, P., Kapadnis, B.P., Hossain, K., Saha, A.K., Ghosh, S., Olsson, B. and Mandal, A., 2014. Isolation and characterization of a Lysinibacillus strain B1-CDA showing potential for bioremediation of arsenics from contaminated water. Journal of Environmental Science and Health, Part A, 49(12), pp.1349-1360.

Rahman, A., Nahar, N., Nawani, N.N., Jass, J., Ghosh, S., Olsson, B. and Mandal, A., 2015. Comparative genome analysis of Lysinibacillus B1-CDA, a bacterium that accumulates arsenics. Genomics, 106(6), pp.384-392.

Ramanathan, S., Shi, W., Rosen, B.P. and Daunert, S., 1998. Bacteria-based chemiluminescence sensing system using β-galactosidase under the control of the ArsR regulatory protein of the ars operon. Analytica chimica acta, 369(3), pp.189-195.

Ratnaike, R.N., 2003. Acute and chronic arsenic toxicity. Postgraduate medical journal, 79(933), pp.391-396.

Ravenscroft, P., Brammer, H. and Richards, K., 2009. Arsenic in North America and Europe. Arsenic Pollution: A Global Synthesis, pp.387-454.

Rossman, T.G., Uddin, A.N. and Burns, F.J., 2004. Evidence that arsenite acts as a cocarcinogen in skin cancer. Toxicology and applied pharmacology, 198(3), pp.394-404.

Roy, A., Kucukural, A. and Zhang, Y., 2010. I-TASSER: a unified platform for automated protein structure and function prediction. Nature protocols, 5(4), pp.725-738.

Roy, A., Xu, D., Poisson, J. and Zhang, Y., 2011. A protocol for computer-based protein structure and function prediction. Journal of visualized experiments: JoVE, (57).

Saltikov, C.W. and Olson, B.H., 2002. Homology of Escherichia coli R773 *arsA*, *arsB*, and *arsC* genes in arsenic-resistant bacteria isolated from raw sewage and arsenic-enriched creek waters. Applied and environmental microbiology, 68(1), pp.280-288.

Saxe JK, Bowers TS, Reid KR. Arsenic. In: Morrison RD, Murphy BL, editors. Environmental Forensics: Contaminant Specific Guide. Burlington, MA: Academic Press; 2006. pp. 279–292.

Shen, Z., Han, J., Wang, Y., Sahin, O. and Zhang, Q., 2013. The contribution of *ArsB* to arsenic resistance in Campylobacter jejuni. PLoS One, 8(3), p.e58894.

Silver, S. and Phung, L.T., 2005. A bacterial view of the periodic table: genes and proteins for toxic inorganic ions. Journal of Industrial Microbiology and Biotechnology, 32(11-12), pp.587-605.

Silver, S., Ji, G., Bröer, S., Dey, S., Dou, D. and Rosen, B.P., 1993. Orphan enzyme or patriarch of a new tribe: the arsenic resistance ATPase of bacterial plasmids. Molecular microbiology, 8(4), pp.637-642.

Singh, S., Kang, S.H., Mulchandani, A. and Chen, W., 2008. Bioremediation: environmental clean-up through pathway engineering. Current opinion in biotechnology, 19(5), pp.437-444.

Stolz, J., Basu, P. and Oremland, R., 2002. Microbial transformation of elements: the case of arsenic and selenium. International Microbiology, 5(4), pp.201-207.

Sunitha, M.S.L., Prashanth, S. and Kishor, P.K., 2015. Characterization of Arsenic-Resistant Bacteria and their ars Genotype for Metal Bioremediation. International Journal of Scientific & Engineering Research 6 (2), pp.304-309.

Tchounwou, P.B., Yedjou, C.G., Patlolla, A.K. and Sutton, D.J., 2012. Heavy metal toxicity and the environment. In Molecular, clinical and environmental toxicology (pp. 133-164). Springer Basel.

Thirunavukkarasu, O.S., Viraraghavan, T. and Subramanian, K.S., 2003. Arsenic removal from drinking water using granular ferric hydroxide. Water Sa, 29(2), pp.161-170.

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. Journal of Biological Chemistry, 265(1), pp.190-194.

United States Environmental Protection Agency. Citizen's Guide to Bioremediation (U.S. Environmental Protection Agency, Washington, DC, 2001). Available at: https://www.epa.gov/remedytech/citizens-guide-bioremediation. [accessed 15 December 2017].

Wang G.P. and Bushman F.D. (2006) A statistical method for comparing viral growth curves. J Virol Methods. 135 (p. 118–123).

Watanabe, C., Inaoka, T., Matsui, T., Ishigaki, K., Murayama, N. and Ohtsuka, R., 2003. Effects of arsenic on younger generations. Journal of Environmental Science and Health, Part A, 38(1), pp.129-139.

Wu, S., Skolnick, J. and Zhang, Y., 2007. Ab initio modeling of small proteins by iterative TASSER simulations. BMC biology, 5(1), p.17.

Xiong, J., He, Z., Van Nostrand, J.D., Luo, G., Tu, S., Zhou, J. and Wang, G., 2012. Assessing the microbial community and functional genes in a vertical soil profile with long-term arsenic contamination. PloS one, 7(11), p.e50507.

Yan, Y., Xue, X.M., Guo, Y.Q., Zhu, Y.G. and Ye, J., 2017. Co-expression of cyanobacterial genes for arsenic methylation and demethylation in Escherichia coli offers insights into arsenic resistance. Frontiers in microbiology, 8.

Yoshida, T., Yamauchi, H. and Sun, G.F., 2004. Chronic health effects in people exposed to arsenic via the drinking water: dose–response relationships in review. Toxicology and applied pharmacology, 198(3), pp.243-252.

Yosim, A., Bailey, K. and Fry, R.C., 2015. Arsenic, the" King of Poisons," in Food and Water. American Scientist, 103(1), p.34.

Zhang, Y., 2008. I-TASSER server for protein 3D structure prediction. BMC bioinformatics, 9(1), p.40.

APPENDICES

I. FASTA sequence of the *ArsB* (1296 nucleotides long).

ATGCAGACTGCTTTAGCCATTATTTTATTTTTAGTCACATTATTTTTTGTCATAGTGCAA
CCACGAAAACTATCAATTGGCTGGTCAGCCTGTATAGGAGCAGCTATTGCGCTATTAATA
GGTGTAGTAACCTGGCATGATGTGTTAGAGGTAACGGGGATTGTTTGGAATGCAACGCTA
TCATTTATTGGTATTATTTTAATTTCACTCATCTTAGATGAAATTGGCTTTTTCGAATGG
GCTGCTCTACATATGGCACGACTTGCAAAGGGAAGTGGAATTCGCTTATTTGTTTATATT
AGTTGCCTTGGAGCTGTTGTATCGGCTTTGTTTGCGAACGATGGGGGTGCTTTAATTCTA
ACGCCGATCGTTTTAGCGATGGTGAGAAAACTGAACTTGAGTGAAAAAATGATCTTTCCA
TTTATTATTGCAAGTGGATTTATAGCAGATACAACCTCGTTACCGTTCGTTATTAGCAAT
TTAACGAATATCGTATCCGCTGACTTTTTCGATATAGGCTTTATAGATTATGCAGTTCAT
ATGCTTGTACCGAATTTCTTTTCGTTAGTGGCAAGTATAGTAGTACTTTACATTGTTTTT
CGCCGTTCCATACCACAACATTATTCTCTTGAGGCCGTTGAGCAACCAAGTCAGGCAATC
CGAGATCCCAAATTATTTTCGATCGCTTGGTATGTAATTGGGATTCTATTAGTGGGCTTC
GTAGCTGGTGAGGTTTTACACATCCCTGTATCTTTTATTTTATGTGGGGTCGCTGCAATA
TTCTTATGGCTAGCACAGCGTAGCTCTGTGGTAGATACAAAAAGAGTTGTAAAAGGAGCG
CCATGGAATATTGTGTTCTTCTCAATTGGCATGTATGTGGTTGTCTATGGGTTAAAAAAT
GTAGGGATGACAGGAATGTTAGCAAATTTATTTGAACTTGCTATACAACAGGGCTTCGTA
GTAGCAACGATGACAATGGGATTTGTCGCTGCCATTTTGTCTGCTGTCATGAACAATTTA
CCTACTGTTATGATAGATGCTTTGGCGATTTCAGAAACCAGCGCATCGGGCATGATGAAA
GAGGGACTGATTTATGCAAATATAATTGGTACGAATTTAGGTCCAAAAATGACACCGATT
GGTTCACTGGCAACATTGCTATGGCTTCATATATTAGCCCAGAAAGGCATTAAAATTAGC
TGGGGACAATATTTTAAAATCGGTGTTATTCTTACCATCCCGACACTGTTCGTCACTTTA
ATTGCATTAATTATTTGGATATGCTTAATATCATAG

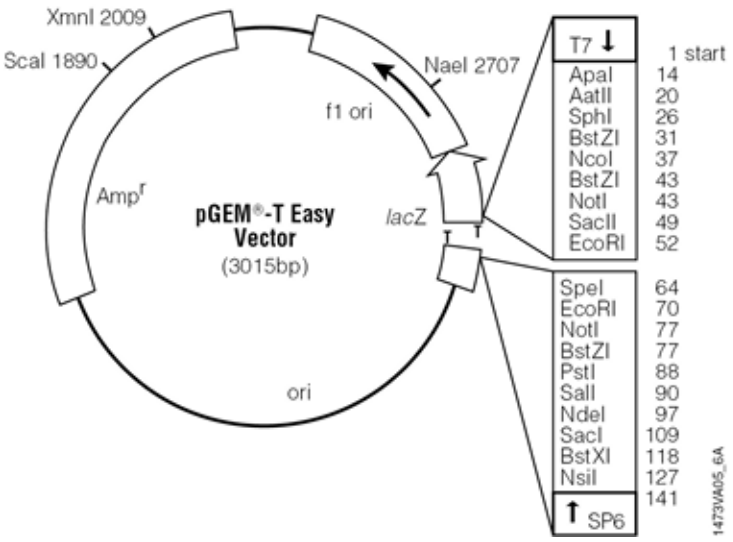
II. Protein sequence of the *ArsB*.

MQTALAIILFLVTLFFVIVQPRKLSIGWSACIGAAIALLIGVVTWHDVLEVTGIVWNATL
SFIGIILISLILDEIGFFEWAALHMARLAKGSGIRLFVYISCLGAVVSALFANDGGALIL
TPIVLAMVRKLNLSEKMIFPFIIASGFIADTTSLPFVISNLTNIVSADFFDIGFIDYAVH
MLVPNFFSLVASIVVLYIVFRRSIPQHYSLEAVEQPSQAIRDPKLFSIAWYVIGILLVGF
VAGEVLHIPVSFILCGVAAIFLWLAQRSSVVDTKRVVKGAPWNIVFFSIGMYVVVYGLKN
VGMTGMLANLFELAIQQGFVVATMTMGFVAAILSAVMNNLPTVMIDALAISETSASGMMK
EGLIYANIIGTNLGPKMTPIGSLATLLWLHILAQKGIKISWGQYFKIGVILTIPTLFVTL
IALIIWICLIS

III. *ArsB* gene-specific primers and its properties.

Primers	Sequence (5' to 3')	Length (bp)	Tm (°C)	GC (%)
ArsB-F	ATGCAGACTGCTTTAGCCATTAT	23	59.3	39
ArsB-R	CTATGATATTAAGCATATCCAAATAAT	27	57.1	22

IV. pGEM®-T Easy Vector map.



V. Preparation of reaction mixture for PCR.

Reaction components	Volume per reaction
Deionized water	8 µl
10µM Forward primer	1.25 µl
10µM Reverse primer	1.25 µl
DNA template	1 µl
MasterAmp TAQurate DNA Polymerase Mix (1 unit)	1 µl
MasterAmp 2X PCR PreMix	12.5 µl
Total volume	25 µl

The table represents reaction components and their respective volumes taken for one reaction of 25 µl total volume. The forward primer *arsB*-F and the reverse primer *arsB*-R were used for the reaction. The reaction mixture was set up on the ice and also the reaction components in the PCR vials were mixed thoroughly before placing it in the PCR.

VI. PCR program parameters.

Step	Temperature	Time	Cycles
Initial denaturation	95°C	5 minutes	1
Denaturation	95°C	1 minute	35
Annealing	55.6°C	45 seconds	
Extension	72°C	1 minute	
Final extension	72°C	10 minutes	1
Hold	4°C	∞	1

VII. Ligation reaction mixture for cloning of the *arsB* gene into pGEM-T Easy Vector.

Reaction component	Standard reaction	Positive control	Negative control
2X Rapid ligation buffer	5 µl	5 µl	5 µl
pGEM-T Easy Vector	1 µl	1 µl	1 µl
Purified DNA (50ng)	2 µl	-	-
Control insert DNA	-	2 µl	-
T4 DNA ligase	1 µl	1 µl	1 µl
Nuclease-free water	1 µl	1 µl	3 µl
Total volume	10 µl	10 µl	10 µl

Standard reaction represents cloning of the *arsB* gene into pGEM-T Easy Vector. Positive control represents cloning of control insert DNA to check transformation efficiency. The negative control is to verify the background control resulting from vector alone.

VIII. RT-PCR reaction components and their volumes for a single reaction of 50 µl.

Reaction components	Volume per reaction
Deionized water	20 µl
12.5 µM Forward primer	1 µl
12.5 µM Reverse primer	1 µl
RNA template	1 µl
MasterAmp TAQurate DNA Polymerase Mix (1 unit)	1 µl
MMLV RT-Plus (40 units)	1 µl
MasterAmp 2X RT-PCR PreMix	25 µl
Total volume	50 µl

The table represents the RT-PCR reaction components and their volumes performed for gene expression studies. The *arsB*-F and *arsB*-R primers were used as forward and reverse primers, respectively. All the reaction components were mixed thoroughly and kept on ice before the beginning of the reaction.

IX. RT-PCR program parameters.

Step	Temperature	Time	Cycles
Rverse transcription	37°C	30 minutes	1
Initial denaturation	95°C	5 minutes	1
Denaturation	95°C	1 minute	35
Annealing	55.6°C	45 seconds	
Extension	72°C	1 minute	
Final extension	72°C	10 minutes	1
Hold	4°C	∞	1