Full Length Research Paper

Insights into the characterization of *ars* genes in the biomining bacterium *Acidithiobacillus caldus* SM-1

Shuijing Yu, Yangwu Deng, Yonghong Zhao, Changli Liang, Yanhua guo and Xianping Luo*

School of Resources and Environment, Jiangxi University of Science and Technology, Ganzhou, Jiangxi, 341000, People's Republic of China.

Accepted 3 September, 2012

*Corresponding author. E-mail: lxp9491@163.com.Tel/Fax: +86-797-8312-331.

*Acidithiobacillus caldus* is an extremely acidophilic, moderately thermophilic, chemoautotrophic bacterium, which has been used to treat gold-bearing arsenopyrite ores. The arsenic resistance system (*ars* operon) was responsible for arsenic resistance of *A. caldus*. To investigate the characterization of *ars* genes, we analyzed 12 *ars* genes in *A. caldus* SM-1 using the bioinformatics database and softwares. Their amino acid composition and physical and chemical characteristics were predicted. Secondary structure simulations revealed that the dominant patterns were predicted to be alpha helix and random coil among the 12 *ars* proteins. Three-dimensional structure analysis showed that there totally existed three major types of three-dimensional structure of the 12 *ars* proteins in *A. caldus* SM-1. Subcellular localization of these proteins indicated that these *ars* proteins were mainly located in the bacterial cytoplasm, while Atc_0977, Atc_1809 and Atc_m110 were especially localized in bacterial inner membrane. Furthermore, DNA-binding residues in ArsR proteins and binding sites of ArsC-As were predicted. Phylogeny analysis revealed that *A. caldus*, *A. ferrooxidans*, *A. ferrivorans* and *A. thiooxidans* were well-supported group based on ArsB and ArsC sequences data. This bioinformatics analysis of *ars* genes could help in probing to the arsenic resistance of *A. caldus* SM-1.

**Key words:** *Acidithiobacillus caldus*, *ars* gene, acidophile, arsenic resistance, bioinformatics analysis, biomining bacterium.

**INTRODUCTION**

*Acidithiobacillus caldus* is a moderately thermophilic, acidophilic, sulphur-oxidizing, Gram-negative bacterium (Hallberg and Lindstrom, 1994). It lives in extremely acidic environments (pH 1 to 3) typically associated with the bioleaching and natural acid drainage systems. *A. caldus* could increase the arsenopyrite-leaching efficiency in arsenopyrite leaching in combination with *Sulfobacillus thermosulfidooxidans* (Dopson and Lindstrom, 1999). Continuous-flow tanks, which were used for the bio-oxidation of arsenopyrite concentrates and operated at 40°C, were dominated by a mixture of the sulphur-oxidizing bacterium *A. caldus* and the iron-oxidizing bacterium *Leptospirillum ferripilum* (Rawlings, 1999). *A. caldus* KU was found to be resistant to the arsenical ions arsenate, arsenite, and antimony via an inducible, chromosomally encoded resistance mechanism, and induced *A. caldus* KU could transport arsenate and arsenite out of the cell against a concentration gradient (Dopson et al., 2001). The well-characterized microbial arsenic detoxification pathway involves the arsenic resistance system (*ars*) operon which codes for a regulatory protein (ArsR), an arsenate permease (ArsB), and an arsenate reductase (ArsC). The *arsR* gene coded for arsenite (As (III))-responsive repressor of transcription, the *arsB* gene coded for an arsenite specific transmembrane pump,
and the arsC gene coded for an arsenate reductase that converted arsenate to arsenite (Cervantes et al., 1994; Ji and Silver, 1992; Saltikov and Newman, 2003). The response of L. ferriphilum to arsenic stress was analyzed and three arsenic response proteins were ars member proteins (Li et al., 2010). The arsenic resistance gene cluster of Microbacterium sp. A33 contained an unusual arsRC2 fusion gene, ACR3, and arsC1 in an operon. ArsRC2 negatively regulated the expression of the pentacistronic operon. ArsC1 and Arsc3 were related to thioredoxin-dependent arsenate reductases (Achour-Rokbani et al., 2010). The chromosomally located arsenic resistance operon from A. caldus has been described previously, and it consists of three genes: arsR, arsB, and arsC (Kotze et al., 2006). The arsenic operon of transposon origin, TnAtcArs, that carries a set of arsenic-resistance genes was isolated from a strain of A. caldus (Tuffin et al., 2005). The chromosomal A. caldus ars genes were cloned and found to consist of arsR and arsC genes transcribed in one direction, and arsB in the opposite direction. The TnAtcArs was expressed at a higher level, and was less tightly regulated in Escherichia coli than were the A. caldus ars genes of chromosomal origin (Kotze et al., 2006).

The ars operon provides arsenic resistance to a variety of microorganisms and can be chromosomal or plasmid-borne. However, no studies regards to physical or chemical properties and structural analysis of ars genes in A. caldus whole genome were mentioned. Recent description of the whole genome sequences of A. caldus SM-1 isolated from a pilot bioleaching reactor offers the opportunity to conduct detailed investigation on the characterization of ars genes (You et al., 2011). This study was done to analyze the characterization of ars genes in A. caldus SM-1 using the bioinformatics database and softwares.

MATERIALS AND METHODS

Sources of sequence data

The sequences (NC_015850, NC_015851, NC_015852 and NC_015853) of A. caldus SM-1 were downloaded from National Center for Biotechnology Information (NCBI) genome database, including .faa, .fna and .ptt files. The ars genes were searched in .ptt file according to annotation information, and then verified by online BLAST searching at NCBI website.

Bioinformatics analysis of ars genes

Physical and chemical properties of ars proteins were calculated with the help of ProtParam online (http://www.expasy.ch/tools/protparam.html) (Wilkins et al., 1999), related indexes including theoretical pI, molecular weight, formula, aliphatic index and instability index. Prediction of protein secondary structure was carried out via program of HNN: Secondary Structure Prediction Method protocols (Combet et al., 2000) (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=NPSA/npsa_hnn.html) and transmembrane domains were analyzed by TMHMM Server v.2.0 (http://www.cbs.dtu.dk/services/TMHMM/). SignalP v. 4.0 (http://www.cbs.dtu.dk/services/SignalP/) and PSORT (http://psort.hgc.jp/form.html) were used, respectively for signal peptide prediction and subcellular localization (Petersen et al., 2011; Nakai and Horton, 1999). The prediction of three-dimensional structure of ars proteins was carried out with the help of SWISS-MODEL online server (Benkert et al., 2011) (http://swissmodel.expasy.org/workspace/index.php?func=modelling_simple1).

Prediction of DNA-binding residues in ArsR proteins

Sequence-based prediction of DNA-binding residues in ArsR proteins of A. caldus SM-1 was carried out with the help of the DNABR web-server (Ma et al., 2012) (http://www.cbi.seu.edu.cn/DNABR/).

Prediction of binding sites of ArsC-As

ArsC sequence was used to PSI-BLAST against the PDB database, so as to retrieve PDB-ID with relevant homology for a specific template. The prediction of molecular modeling of ArsC was carried out with the help of SWISS-MODEL online server (Benkert et al., 2011). The prediction of protein–ligand binding sites of ArsC was carried out via the Q-SiteFinder web-server (Laurie and Jackson, 2005) (http://www.modelling.leeds.ac.uk/qsitefinder/).

Phylogenetic tree construction of ars genes

Phylogenetic tree was constructed by using MEGA 4.0 software based on the sequences of ArsB, ArsC and ArsR (Tamura et al., 2007). The query sequences of ArsB and ArsC in A. caldus were BLASTed against the NCBI protein database, respectively. Highly similar sequences from NCBI output were selected for constructing phylogenetic tree.

RESULTS

Annotation information of ars genes

Using information retrieval to ars genes, 12 ars genes of A. caldus SM-1 were obtained as shown in Table 1. Among them, Atc_0975 and Atc_0977 genes were coded for ArsC and ArsB proteins, respectively while the others were coded for ArsR protein. Only Atc_m066 and Atc_m110 were plasmid-borne from the plasmid (NC_015851) of A. caldus SM-1. Furthermore, according to BLASTN output, the nucleotide sequence of Atc_0558 shared little or no sequence similarity to any nucleotide sequence in the nr/nt database at NCBI and was considered as specific marker gene of A. caldus SM-1, which had potential to be applied in molecular identification of A. caldus SM-1.

Physical and chemical analysis of ars genes

As shown in Table 1, the number of amino acids coded by the 12 ars genes varied from 60 to 435, in which
Table 1. Annotation information and physical and chemical properties of *ars* genes in *Acidithiobacillus caldus* SM-1.

<table>
<thead>
<tr>
<th>Protein ID</th>
<th>PID</th>
<th>Product</th>
<th>aa</th>
<th>Theoretical pI</th>
<th>Molecular weight (kD)</th>
<th>Aliphatic index</th>
<th>Instability index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atc_0550</td>
<td>340781293</td>
<td>ArsR family transcriptional regulator</td>
<td>119</td>
<td>7.74</td>
<td>13270.3</td>
<td>103.28</td>
<td>35.14</td>
</tr>
<tr>
<td>Atc_0558</td>
<td>340781301</td>
<td>ArsR family transcriptional regulator</td>
<td>129</td>
<td>8.42</td>
<td>14640.9</td>
<td>99.15</td>
<td>35.93</td>
</tr>
<tr>
<td>Atc_0951</td>
<td>340781693</td>
<td>ArsR family transcriptional regulator</td>
<td>97</td>
<td>10.48</td>
<td>10854.8</td>
<td>117.63</td>
<td>28.91</td>
</tr>
<tr>
<td>Atc_0975</td>
<td>340781717</td>
<td>ArsC</td>
<td>163</td>
<td>5.49</td>
<td>17525.8</td>
<td>89.39</td>
<td>51.78</td>
</tr>
<tr>
<td>Atc_0976</td>
<td>340781718</td>
<td>ArsR family transcriptional regulator</td>
<td>121</td>
<td>6.58</td>
<td>13417.2</td>
<td>93.55</td>
<td>43.50</td>
</tr>
<tr>
<td>Atc_0977</td>
<td>340781719</td>
<td>arsenic efflux pump protein</td>
<td>435</td>
<td>9.78</td>
<td>46770.1</td>
<td>137.47</td>
<td>37.35</td>
</tr>
<tr>
<td>Atc_1751</td>
<td>340782493</td>
<td>ArsR family transcriptional regulator</td>
<td>102</td>
<td>7.83</td>
<td>11493.3</td>
<td>102.25</td>
<td>54.55</td>
</tr>
<tr>
<td>Atc_1809</td>
<td>340782551</td>
<td>ArsR family transcriptional regulator</td>
<td>260</td>
<td>7.76</td>
<td>28931.1</td>
<td>90.19</td>
<td>45.07</td>
</tr>
<tr>
<td>Atc_1895</td>
<td>340782637</td>
<td>ArsR family transcriptional regulator</td>
<td>129</td>
<td>8.42</td>
<td>14640.9</td>
<td>99.15</td>
<td>35.93</td>
</tr>
<tr>
<td>Atc_1903</td>
<td>340782645</td>
<td>ArsR family transcriptional regulator</td>
<td>119</td>
<td>7.74</td>
<td>13270.3</td>
<td>103.28</td>
<td>35.14</td>
</tr>
<tr>
<td>Atc_m066</td>
<td>340783691</td>
<td>ArsR family transcriptional regulator</td>
<td>60</td>
<td>6.82</td>
<td>6848.8</td>
<td>102.50</td>
<td>40.56</td>
</tr>
<tr>
<td>Atc_m110</td>
<td>340783735</td>
<td>putative ArsR family transcriptional regulator</td>
<td>254</td>
<td>8.39</td>
<td>28238.3</td>
<td>89.65</td>
<td>46.96</td>
</tr>
</tbody>
</table>

Atc_0977 had the most amino acids, 435, and the highest molecular weight, while Atc_m066 the least. Aliphatic index analysis illustrated that Atc_0550, Atc_0951, Atc_0977, Atc_1751 and Atc_1903 had high values above 100, which were indicated as hydrophilic proteins. And the others were hydrophobic proteins. Moreover, instability index assay implicated that these proteins with high values above 40 were instable proteins, including Atc_0975, Atc_0976, Atc_1751, Atc_1809, Atc_m066 and Atc_m110, and the others were stable proteins. Most ArsR proteins were instable proteins, and metabolic instability of these proteins may involve into an inducible, chromosomally encoded arsenic resistance mechanism of *A. caldus* SM-1. It is worth mentioning that theoretical isoelectric point (pI) of the majority of *ars* proteins was > 7, which seemed to be inconsistent with extremely acidophilic characterization of *A. caldus* SM-1.

**Secondary structure simulations of *ars* genes**

As is illustrated in Table 2, peptide prediction analysis showed that all of the 12 *ars* genes were non secreted proteins. And only Atc_0977 possessed 11 transmembrane regions by transmembrane prediction. In addition, secondary structure simulations showed that all of the *ars* genes possessed three patterns, that is, most alpha helix, less random coil, and the least extended strand.

**Tertiary structure prediction of *ars* genes**

The tertiary structure of *ars* genes were predicted via
SWISS-MODEL online server. As shown in Figure 1, there totally existed 3 major types of three-dimensional structure, in which Atc_0976, Atc_1809 and Atc_m110 made up one class that exhibited similar structure. Atc_0550, Atc_1751 and Atc_1903 were classified into another class. Atc_0558 and Atc_1895 also exhibited similar structure. In addition, tertiary structure prediction of Atc_0977 found no suitable templates, so its tertiary structure could not be given. These findings implied that ars genes may play different roles in inducible arsenic-resistance mechanism in *A. caldus*, which needs further studies.

**Subcellular localization of ars proteins**

Presumably, as shown in Table 2, these ars protein members were mainly localized in bacterial cytoplasm of *A. caldus* SM-1. Seemingly, Atc_0550 and Atc_1903, Atc_0558 and Atc_1895 had similar localization levels, respectively. Especially, Atc_0977, Atc_1809 and Atc_m110 were localized in bacterial inner membrane. These results implied that ars proteins had extensive and delicate localizations in *A. caldus* SM-1 cell.

**Prediction of DNA-binding residues in ArsR proteins**

DNA-binding residues in 10 ArsR proteins of *A. caldus* SM-1 were predicted using the DNABR web-server. As shown in Table 3, DNA-binding residues of six ArsR proteins (Atc_0550, Atc_0558, Atc_1809, Atc_1895, Atc_1903, and Atc_m110) contained two cysteine residues, respectively probably provided with the ability to interact with arsenite.

**Prediction of binding sites of ArsC-As**

Crystal structure (PDB ID:1JL3-A) of ArsC in *Bacillus subtilis* was retrieved with relevant homology to ArsC protein of *A. caldus* SM-1 via PSI-BLAST searching. Using PDB 1JL3-A as a specific template, PDB file of ArsC protein was obtained with the help of SWISS-MODEL online server. With the searching of the Q-SiteFinder web-server, the structure of binding sites of ArsC model was obtained as shown in Figure 2. There were ten binding active sites in ArsC model, and among of them, second binding site was composed of six residues Leu 52, Arg 55, Glu 56, Tyr 125, Arg 126 and Arg 129 (Figure 2). For ArsC arsenate reductase, three of the basic residues, Arg 60, Arg 94, and Arg 107, are particularly significant because they interact directly with the arsenate and arsenite intermediates (Martin et al., 2001). It implied that the second binding site was binding site of ArsC-As in *A. caldus* SM-1 because the other seven binding sites cannot be provided with Arg residue and the two binding sites (first and tenth binding sites) were provided with only one Arg residue (data not shown).

**Phylogeny analysis of ars genes**

Phylogenetic tree was constructed by using MEGA 4.0 software between the sequences of ArsRs in *A. caldus*. Atc_0550, Atc_1809 and Atc_0558 were strictly clustered with Atc_1903, Atc_m110 and Atc_1895, correspondingly (Figure 3). *A. caldus*, *A. ferrooxidans* ATCC 53993, and *A. ferrivorans* SS3 were well-supported group using ArsB sequence data (Figure 4). The three sampled species of *A. thiooxidans* ATCC 19377, *A. ferrooxidans* ATCC 53993, *Acidithiobacillus* sp. GGI-221 formed a supported

### Table 2. Higher structure simulations of ars genes and subcellular localization of ars proteins in Acidithiobacillus caldus SM-1.

<table>
<thead>
<tr>
<th>Protein ID</th>
<th>Transmembrane region</th>
<th>Secondary structure</th>
<th>Bacterial cytoplasm</th>
<th>Bacterial inner membrane</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Alpha helix</td>
<td>Extended strand</td>
<td>Random coil</td>
</tr>
<tr>
<td>Atc_0550</td>
<td>0</td>
<td>56</td>
<td>15</td>
<td>48</td>
</tr>
<tr>
<td>Atc_0558</td>
<td>0</td>
<td>86</td>
<td>5</td>
<td>38</td>
</tr>
<tr>
<td>Atc_0951</td>
<td>0</td>
<td>60</td>
<td>14</td>
<td>23</td>
</tr>
<tr>
<td>Atc_0975</td>
<td>0</td>
<td>50</td>
<td>15</td>
<td>93</td>
</tr>
<tr>
<td>Atc_0976</td>
<td>0</td>
<td>62</td>
<td>7</td>
<td>52</td>
</tr>
<tr>
<td>Atc_0977</td>
<td>11</td>
<td>256</td>
<td>56</td>
<td>123</td>
</tr>
<tr>
<td>Atc_1751</td>
<td>0</td>
<td>57</td>
<td>13</td>
<td>32</td>
</tr>
<tr>
<td>Atc_1809</td>
<td>0</td>
<td>130</td>
<td>26</td>
<td>104</td>
</tr>
<tr>
<td>Atc_1895</td>
<td>0</td>
<td>86</td>
<td>5</td>
<td>38</td>
</tr>
<tr>
<td>Atc_1903</td>
<td>0</td>
<td>56</td>
<td>15</td>
<td>48</td>
</tr>
<tr>
<td>Atc_m066</td>
<td>0</td>
<td>26</td>
<td>13</td>
<td>21</td>
</tr>
<tr>
<td>Atc_m110</td>
<td>0</td>
<td>128</td>
<td>22</td>
<td>104</td>
</tr>
</tbody>
</table>
group with *A. caldus* based on the analysis of ArsC sequence data (Figure 5). These results suggest that the *Acidithiobacillus* genus seems to have the same evolutionary origin in arsenic resistance system.

**DISCUSSION**

Microorganisms have evolved a variety of mechanisms for coping with arsenic toxicity, including minimizing the amount of arsenic that enters the cell, example through increased specificity of phosphate uptake (Cervantes et al., 1994), arsenite oxidation through the activity of arsenite oxidase (Cervantes et al., 1994; Muller et al., 2003). Some microorganisms utilize arsenic in metabolism, either as a terminal electron acceptor in dissimilatory arsenate respiration (Dianne, 1998; Stolz and Oremland, 1999; Huber et al., 2000) or as an electron donor in chemoautotrophic arsenite oxidation (Santini et al., 2000). In this paper, *ars* genes of *A. caldus* SM-1 coded for ArsB, ArsC and ArsR, implying an arches ABC transporters. Therefore, we had interests in grapevine genome database, with the intent to give bioinformatical support to further grapevine biological researches.

The structure characteristics analysis of *ars* proteins was of considerable importance for the investigation into biological mechanism of arsenic resistance of *A. caldus*. The ArsC arsenate reductase from *E. coli* plasmid R773

<table>
<thead>
<tr>
<th>Protein ID</th>
<th>amino acid sequences of ArsRs*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atc_0550</td>
<td>VRYALFSEVFALAHPKRLEIIHLYLEGKNAELISKANLCSKWSLKIARKGLHVHECGTFCYHRRTSPKVLATCEIVRDLIDQMTTQQLHELATVVPLR</td>
</tr>
<tr>
<td>Atc_0558</td>
<td>TDPRLCNCFEPGKVAAARMRLEQDSEIKAVRIFDVLGNRTRLRILALASEELCVDIALHLISISAASHQLRLCHDRWLMRMDGKMVYRTDPQPKVQLSMTNAFLEARIA</td>
</tr>
<tr>
<td>Atc_0951</td>
<td>LSDREIPRVAIAKIIYKIICLLSKGEMSMQNLKVINTHSNASQHAMLQDSGVLARKVASTVYRYRDRDTRNLAVS</td>
</tr>
<tr>
<td>Atc_0976</td>
<td>SDRIVARLEASPVRALFIWVEHEPEGLFSQIAERIGOLHNVSFHLKVLQADELVHVREGFQYRAOMPVPNVRSLAYLTENCECCRGTSCTIPDESESSAAATAS</td>
</tr>
<tr>
<td>Atc_1751</td>
<td>TREDIEOASRLKAMHPRLKVLVGLSGDEMSVQDIVAAVGTOSNISQHAILREKDLRARKDANKVYRVDPRTRLRISMSEVFCC</td>
</tr>
<tr>
<td>Atc_1809</td>
<td>GMCRNAAELVYFEGISGFDERSVELHDAKPAEMHEAALLGVPARAMVWALIAGDALPASELAYRAIOTPQASHHLAMVAGGLAVERCLRYRILRALFPQVEAVLENMISSPDPDLDRADSAIDPLSHVRSFYCMLAQLAVMLDQGGLTDRGWIALQDRAVAYSEVGERFAALGLDLPNLRQRHFLARRCDSWERRPHIGGALGALA MCKRFQGEDGWLRKSAQQRQWVTSVQGQFKA</td>
</tr>
<tr>
<td>Atc_1895</td>
<td>TDPRLCNCFEPGKVAAARMRLEQDSEIKAVRIFDVLGNRTRLRILALASEELCVDIALHLISISAASHQLRLCHDRWLMRMDGKMVYRTDPQPKVQLSMTNAFLEARIA</td>
</tr>
<tr>
<td>Atc_1903</td>
<td>VRYALFSEVFALAHPKRLEIIHLYLEGKNAELISKANLCSKWSLKIARKGLHVHECGTFCYHRRTSPKVLATCEIVRDLIDQMTTQQLHELATVVPLR</td>
</tr>
<tr>
<td>Atc_m066</td>
<td>ELSIAGASHQLRLHMRMDGKMVYRLSPDALROVLEQEGRFIE</td>
</tr>
<tr>
<td>Atc_m110</td>
<td>ELVYFEGISGFDERSVELHDAKPAEMHEAALLGVPARAMVWALIAGDALPASELAYRAIOTPQASHHLAMVAGGLAVERCLRYRILRALFPQVEAVLENMISSPDPDLDRADSAIDPLSHVRSFYCMLAQLAVMLDQGGLTDRGWIALQDRAVAYSEVGERFAALGLDLPNLRQRHFLARRCDSWERRPHIGGALGALA MCKRFQGEDGWLRKSAQQRQWVTSVQGQFKA</td>
</tr>
</tbody>
</table>

*Predicted DNA-binding residues in protein sequences were labeled using the red and underscore.*
has a catalytic cysteine, Cys 12, in the active site, surrounded by an arginine triad composed of Arg 60, Arg 94, and Arg 107 (Martin et al., 2001). This native structure utilized the chemistry of the Cys in concert with at least three arginines to trap the arsenate in three binding pockets. Compared with our predicted result of binding sites of ArsC-As, there was a certain similarity between amino acid residues, and they both provided with three Arg residues. Transcription of the ars operon was negatively regulated by the ArsR proteins and induced by the arsenite. Cysteines in the ArsR protein comprised part of a metal binding motif found in members of the ArsR family of metalloregulatory proteins (Shi et al., 1994). In this study, DNA-binding residues of six ArsR proteins contained two cysteine residues, respectively. However, it was reported that an atypical ArsR regulator from A. ferrooxidans, which was able to respond to arsenic, did not contain the conserved metal-binding motif (Butcher and Rawlings, 2002). It implied that ArsR regulators may have a different method of binding the inducer.

A. caldus is one of the four members (A. caldus, A. thiooxidans, A. ferrooxidans and A. ferrivorans) of the genus Acidithiobacillus characterized to date whose shared metabolic and functional capabilities allow them to survive in extremely acidic environments (Valdes et al., 2011; Liljeqvist et al., 2011; Valdes et al., 2008; You et al., 2011). Genome analysis revealed a closer functional relatedness of A. caldus to A. thiooxidans than to A. ferrooxidans and A. ferrivorans (Valdes et al., 2011). In this paper, phylogeny analysis showed that these four members of the genus Acidithiobacillus seems to share the same evolutionary origin in arsenic resistance system. However, the iron-oxidizing bacterium of the genus Leptospirillum formed a weak supported group with A. caldus based on phylogeny analysis of ArsB and ArsC sequences data in this study. These results provide new opportunities for experimental research and contribute to a better understanding of arsenic resistance system of A. caldus SM-1.
Figure 2. The structure of ten binding sites of ArsC model in *A. caldus* SM-1. Second binding site was composed of six residues Leu 52, Arg 55, Glu 56, Tyr 125, Arg 126 and Arg 129, and implied as the binding site of ArsC-As in *A. caldus* SM-1.

Figure 3. Phylogenetic tree based on the sequences of ArsR proteins in *Acidithiobacillus caldus* SM-1.
To study the ecological relationships of biomining bacteria and the population dynamics during the bioleaching processes, specific methods for their identification and enumeration are required. Conventional plate count methods and biochemical identification methods described previously could not circumvent the problems linked to the long wait for the colony to develop and/or the inability of some bacteria to grow on solid media (Johnson, 1991; Ahmad, 1993). In recent years, various nucleic acid-based molecular methods, such as Polymerase chain reaction (PCR) method (Feng et al., 2012; Escobar et al., 2008; Kamimura et al., 2001; DeWulfDurand et al., 1997) and fluorescent in situ hybridization (FISH) (Mahmoud et al., 2005), have been developed for the rapid detection and identification of Acidithiobacillus strains because of simplicity in operation, stable detection results, and savings in time. A high level of marker specificity is crucial for various nucleic acid-based molecular methods. In this study, the nucleotide sequence of Atc_0558 was identified as

Figure 4. Phylogenetic tree based on the sequences of ArsB protein.
specific marker gene of *A. caldus* SM-1, which had a potential to be developed for new molecular methods for rapid identification of *A. caldus* SM-1. Further researches on this strain-specific marker will be carried out subsequently.

The distribution of genes on chromosome was one of decisive factors of gene functions. In this study, we found that 10 *ars* genes were intently localized in the chromosomes of *A. caldus* SM-1, implying an inducible, chromosomally encoded arsenic-resistance mechanism. This ancestral arrangement suggested that *A. caldus* has evolved a variety of mechanisms for coping with arsenic toxicity. Moreover, different dimensional structure determined functional discriminations. *In silico* three-dimensional structure analysis revealed that three major types of three-dimensional structures, interestingly, broadened horizons for further physiological or functional studies of *ars* genes in *A. caldus* SM-1.

In conclusion, we present systematic bioinformation of *in silico* 12 *ars* genes that might be involved in arsenic resistance system in *A. caldus* SM-1. The secondary structure analysis and physical and chemical properties comparative information of *ars* genes could be a valuable resource for further molecular functional studies and electrophysiological researches of *ars* genes. Our bioinformatics analysis of *ars* genes should help in probing to the arsenic resistance of *A. caldus* SM-1.

**ACKNOWLEDGEMENTS**

This work was jointly supported by the grant No. 2012BAC11B07 and 51064007 from the National Key Technology R&D Program of China, and the grant No. 20121BBG7004, 2011bdh80032 and 20114BAB213019 from the Natural Science Foundation of Jiangxi Province, China.

**REFERENCES**


