In *silico* analysis of bacterial arsenic islands reveals remarkable synteny and functional relatedness between arsenate and phosphate

**Hang Li**††, **Mingshun Li**††, **Yinyan Huang**†, **Christopher Rensing**‡* and **Gejiao Wang**‡*

1 State Key Laboratory of Agricultural Microbiology, College of Life Sciences and Technology, Huazhong Agricultural University, Wuhan, P. R. of China
2 Department of Plant and Environmental Sciences, University of Copenhagen, Frederiksberg, Denmark

**Keywords**: arsenic islands, arsenite oxidase, AioBA, phosphorus, synteny

In order to construct a more universal model for understanding the genetic requirements for bacterial AsIII oxidation, an *in silico* examination of the available sequences in the GenBank was assessed and revealed 21 conserved 5–71 kb arsenic islands within phylogenetically diverse bacterial genomes. The arsenic islands included the AsIII oxidase structural genes *aioBA*, *ars* operons (e.g., *arsRCB*) which code for arsenic resistance, and *pho*, *pst*, and *phn* genes known to be part of the classical phosphate stress response and that encode functions associated with regulating and acquiring organic and inorganic phosphorus. The regulatory genes *aioXSR* were also an island component, but only in *Proteobacteria* and orientated differently depending on whether they were in *α-Proteobacteria* or *β/-γ-Proteobacteria*. Curiously though, while these regulatory genes have been shown to be essential to AsIII oxidation in the *Proteobacteria*, they are absent in most other organisms examined, inferring different regulatory mechanism(s) yet to be discovered. Phylogenetic analysis of the *aio*, *ars*, *pst*, and *phn* genes revealed evidence of both vertical inheritance and horizontal gene transfer (HGT). It is therefore likely the arsenic islands did not evolve as a whole unit but formed independently by acquisition of functionally related genes and operons in respective strains. Considering gene synteny and structural analogies between arsenate and phosphate, we presumed that these genes function together in helping these microbes to be able to use even low concentrations of phosphorus needed for vital functions under high concentrations of arsenic, and defined these sequences as the arsenic islands.

**INTRODUCTION**

Microbial arsenite (AsIII) oxidation converts the more toxic AsV to the less toxic AsV which is known to be catalyzed by a molybdenum-containing enzyme. The AsIII oxidase enzyme is encoded by *aioBA* (previously referred to as *aooAB*) (Silver and Phung, 2005). Putative *AioA* were found to be specific for AsIII-oxidizing bacteria, therefore the usefulness of the *aioBA* as a functional marker indicating the ability to oxidize AsIII of a strain was proposed (Inskeep et al., 2007; Quéméneur et al., 2008; Hamamura et al., 2009). However, *aioA* is not a suitable marker for microbial diversity studies because its phylogeny does not always strictly correlate with that of the 16S rRNA genes, due to horizontal gene transfer (HGT) (Heinrich-Salmeron et al., 2011). In addition, in some AsIII-oxidizing strains, an *aioA* sequence was not detected either by PCR nor genome sequencing (Richy et al., 2009). It is now known that a new type of AsIII oxidase gene *arxA*, which was only distantly related to *aioA*, could be identified in a number of strains (Zargar et al., 2010, 2012).

Currently, the genetics underlying AsIII oxidation and its regulation are perhaps best understood in *Agrobacterium tumefaciens* 5A. The *aioBA* genes are part of the *aioX-aioS-aioR-aioB-aioA-cytc2-chlE* operon that has been shown to be regulated by a two-component regulatory pair comprised of the sensor kinase *AioS* and its cognate response regulator *AioR* in conjunction with the periplasmic *AsIII*-binding protein *AioX* (Kashyap et al., 2006; Koechler et al., 2010; Liu et al., 2012). In addition, the σ^54^ factor (RpoN) has been shown by two different groups to play a role in *aioBA* regulation (Koechler et al., 2010; Kang et al., 2012). A −24/−12 box for RpoN binding has been detected upstream of *aioB* and shown to be important for *aioBA* expression by 5′ RACE (rapid-amplification of cDNA ends) (Sardiwal et al., 2010) and precision deletion experiments (Koechler et al., 2010). Furthermore, *AioR* also contains a conserved domain for response regulators that could regulate σ^54^-type promoters (Sardiwal et al., 2010). RpoN is viewed to form a close complex with RNA polymerase, which requires energy provided by regulators for transcriptional initiation. The σ^54^-dependent regulators such as *AioR* may bind to upstream activation sequences (UAS) of σ^54^-type promoters for energy conservation (Shingler, 2011).

Regarding sequences in the vicinity of the *aio* operon, Silver and Phung (2005) first proposed the concept of an “arsenic island” based on a 71-kb DNA region of the *Alcaligenes fae- calis* genome which contains over 20 functionally related genes.
such as those encoding AsIII oxidase AioBA, ArsAB for AsIII efflux, and a variety of oxyanion ABC transporters. Muller et al. (2007) reported the gene sequences in the vicinity of aioBA in *Harmininomus arsenicoxydans* ULPAs1 and several other strains. Later, Arsène-Ploetze et al. (2010) proposed that aioBA was located in a genomic island which may have been acquired by HGT in *Thiomonas* sp. 3A. However, since only a few aio operons were known until recently, the definition and distribution of such arsenic islands was unclear and speculative. Due to the development and usage of high-throughput sequencing, more aio operons could be identified in microbial genomes (Hao et al., 2012; Huang et al., 2012; Li et al., 2012; Lin et al., 2012; Phung et al., 2012). From visual inspection, gene patterns associated with the aioBA genes became apparent and warranted a more detailed characterization.

One such pattern is the frequent physical association of genes involved with As and phosphorus (P) metabolism (Moreno-Sanchez et al., 2012). As and P are both members of Group V of the periodic table, and they are involved with As and phosphorus (P) metabolism (Moreno-Sanchez et al., 2012). From visual inspection, gene patterns associated with the aioBA genes became apparent and warranted a more detailed characterization.

The primary aim of this study was to characterize the physical arrangement and functional relatedness of aio, pho, pst, and ars genes among the available arsenic islands. In addition, we also examined the phylogenetic relationships of these genes so as to assess mode of inheritance and used this information to begin visualizing a broader picture of how AsIII oxidation is regulated in different organisms and speculate the functional relationships of some genes that appear to have been repeatedly co-inherited in nature.

**RESULTS**

**DETECTION OF GENES IN THE VICINITY OF aio OPERON REVEALED PUTATIVE ARSENIC ISLANDS**

A total of 57 full-length aioBA operons encoding arsenite oxidase were detected in 55 strains using a BLAST search in GenBank. Among them, genes in the vicinity of the aioBA operons showed significant synteny among 21 genome sequences (ranging 5–71 kb). These genes that are all responsible for AsIII oxidation (21 aio operons, aioBA, or aioXSRBACD), arsenic resistance (23 ars operons, e.g., arsR, arsC, arsB, and acr3), phosphate transport (10 pst1 operons, e.g., pstS, pstC, pstA, and pstB) and phosphonate transport (6 phn1 operons, e.g., phnC, phnD, phnE, and phnF) were frequently detected (Figure 1). Considering the major function for AsIII’s oxidation, we refer to these 21 sequences as arsenic oxidase gene islands (Figure 1). In some closely related bacterial strains, gene arrangements of the islands showed excellent synteny as similar gene and operon arrangements were found in *A. tumefaciens* 5A, *Agrobacterium* sp. GW4 and *Sinorhizobium* sp. M14. In addition, the same arrangement was also shared by *Acidiphilium multivorum* AUF301 and *Acidiphilium* sp. PM. However, other distantly related bacteria containing an arsenic island did not show a similar arrangement (Figure 1). Later on, *Agrobacterium albertimagni* strain AOL15 (Trimble et al., 2012a) and *Achromobacter pichiaudii* strain HLE (Trimble et al., 2012b), were sequenced, but showed similar arsenic gene islands to *A. arsenitoxydans* SY8 (Li et al., 2012) and *A. tumefaciens* 5A, respectively (Hao et al., 2012) (data not shown).

By scanning the genomes, we found that aio operons were only present as a single copy and often located within the arsenic islands. To better interpret the results of this analysis, it is important to point out that in addition to the pst or phn operons on the arsenic islands (here referred to as pst1 and phn1), almost all strains possessed another pst or phn operon located distantly (here referred to as pst2 or phn2). The phylogenies of the representative amino acid sequences (AioA, PstS, and PhnC) for aio, pst and phn operons were compared to their 16S rDNAs in order to determine whether possible HGT had taken place. Since ars operons have frequently been shown to associate with HGT events (Tuffin et al., 2005; Čai et al., 2009a), only the representative ars genes (acr3 or arsB) on the arsenic islands were analyzed in this study (see following sections for detailed results), although there are orthologs in many other phyla.

**THE ARSENIC ISLANDS ARE LOCALIZED ON CHROMOSOMES OR ON PLASMIDS**

Of the 21 arsenic islands analyzed in this study, 3 have been shown to be localized on a plasmid (GenBank GU990088, CP000321 and AP012037) and 8 on a chromosome (GenBank CP000781, CP000312, AP012037, NC_009138, FP475956, NC_010087, FP929003, and CP001097) (Figure 1). To determine where all of the 21 arsenic islands are located, we performed a bioinformatics analysis to predict localization on a plasmid using the cBar program (Zhou and Xu, 2010). According to our analysis, nine arsenic islands were predicted to be located on a plasmid and 12 on a chromosome (Figure 1). The previously determined localization of 3 arsenic islands on a plasmid and eight on a chromosome were all correctly predicted, demonstrating good reliability of plasmid prediction using cBar. The plasmid-borne arsenic islands were prevalent in α-Proteobacteria (7/11). Notably, strains *Agrobacterium* sp. GW4, *A. tumefaciens* 5A, and *Sinorhizobium* sp. M14 shared similar “arsenic island” arrangements, all predicted to be located...
FIGURE 1 | Gene arrangements of the 21 arsenic islands. Arrows with different colors represent the following genes: blue for aiobA, black for aiokSR, pink for aioCD or nitR (encoding cytochrome c and molybdenum biosynthesis protein or nitroreductase, respectively), red for pst operon, yellow for phn operon, orange for ars operon in which green for arsB, purple for acet, light blue for mobile element. ■ and □ represent the reported and predicted plasmid-originated sequences, respectively. ● and ○ represent the reported and predicted chromosome-originated sequences, respectively. GenBank accession numbers are as follows: Acidovorax sp. NO1 (AGTS01000000), Herminiimonas arsenicoxydans ULPAs1 (CU207211), Alcaligenes faecalis NCIB 8687 (AY297781), Achromobacter arsenitoxydans SY8 (AGUF01000000), Sinorhizobium sp. M14 (GU990088), Agrobacterium sp. GW4 (UQ423942), Roseomonas cervicalis ATCC 49957 (NZ_AVD/L0000677), Xanthobacter autotrophicus Py2 (NC_0009720), Rhodobacter sp. SV2 (NZ_ACY01000001), Starkeya novella DSM 506 (NC_014217), Nitrobacter hamburgensis X14 (NC_007960), Halomonas sp. HAL1 (EU651834), Pseudomonas sp. TS44 (EU311944), Candidatus nitrospira defluvi (NC_014355), Chlorobium limicola DSM 245 (CP001097), Burkholderia multivorans ATCC17616 (NC_010087), Acidiphilium multivorum AIU301 (NC_015186 and NC_015187).

on a plasmid. Again, these predictions were in agreement with the known localization of the arsenic gene island on plasmid pSinA (GenBank GU990088) of strain Sinorhizobium sp. M14 (Figure 1). Furthermore, the three arsenic islands from Acidiphilium multivorum AIU301 and Acidiphilium sp. PM had a similar arrangement of their genes and were all localized on a plasmid (Figure 1), two of the arsenic islands were predicted by cBar whereas one occurs on pACMV2 (AP012037). It appears likely that these strains may have acquired their respective arsenic islands by HGT.

WIDE SPREAD DISTRIBUTION AND GENOMIC STABILITY OF AioBA

The phylogenetic tree of AioA was generally in accordance with the 16S rDNA phylogeny which yielded the two clades of Proteobacteria, Chlorobi, Deinococcus-Thermus, Chloroflexi, and Archaee (Figure 2), and the AioBA phylogeny was similar to that of AioA (data not shown). Most of the sequences encoding AioA were found in Proteobacteria, and mainly be divided into two groups. Group I is made up of sequences of α-Proteobacteria and Group II is comprised of β- and γ-Proteobacteria. This distribution was consistent with a previous analysis in which partial AioA sequences obtained by degenerate primers was distributed along a similar pattern (Heinrich-Salmeron et al., 2011). However, eight AioA-like proteins from marine α- or γ-Proteobacteria clustered together into a separate branch and exhibited a unique arrangement. The genes encoding this subfamily of AioA were all located downstream of two genes encoding the cytochrome c peroxidase MauG and were arranged in the gene order of mauG-mauG-aioBA. This phylogenetically distinct clade of these AioA-like proteins may have evolved in their marine environment due to unique conditions. Compared to the 16S rDNA phylogenetic tree, there were three conflicts with the AioA phylogenetic tree, which suggests the occurrence of HGT. For example, the two AioAs in Acidiphilium multivorum AIU301 (one located from the chromosome and one from a plasmid) and the AioA in Acidiphilium sp. PM fell into a clade together with Chlorobi or Deinococcus-Thermus strains appears to have transferred the aiaA into A. multivorum AIU301 and Acidiphilium sp. PM since they all have been isolated from a similar acidic environment (San Martin-Uriz et al., 2011). In addition, a HGT might also be more likely since the two AioAs in A. multivorum AIU301 and Acidiphilium sp. PM were also predicted by cBar to be located on plasmids (Figure 1). Chloroflexus aggregans DSM 9485,
Chloroflexus sp. J-10-fl and Y-400-fl are closely related based on 16S rDNA analysis, while the AioA of C. aggregans DSM 9485 clustered with the AioAs from Deinococcus-Thermus (Figure 2). It appears that A. arsenitoxydans SY8 had transferred aioA into Ralstonia sp. 22 by HGT (Lieutaud et al., 2010), and here again we found that the aioA of A. arsenitoxydans SY8 is located on plasmid.

**ANALYSIS OF THE PUTATIVE REGULATORS FOR aioBA OPERONS**

**SUGGESTED DIFFERENT MECHANISMS OF aio REGULATION**

The genes encoding regulators AioXSR located upstream of aioBA were only identified in 12 strains of Proteobacteria among the 21 analyzed strains encoding aioBA as part of their arsenic island. All of these 12 strains belonged to either α-Proteobacteria or β-Proteobacteria (Figure 1). The transcriptional orientation of aioXSR genes differed between α- and β-Proteobacteria, which is in agreement with the AioA phylogeny (Figure 2). The aioXSR genes from α-Proteobacteria displayed the same transcriptional orientation as aioBA, while those from β-Proteobacteria displayed the opposite orientation (Figure 1). The other nine sequences without aioXSR genes were distributed in different taxonomic groups such as Proteobacteria, Chlorobi and Nitrospira. In some of these identified species such as Pseudomonas sp. TS44 and Halomonas sp. HAL1, AsIII oxidation could be verified (Cai et al., 2009b; Lin et al., 2012). The mode and mechanism regulating expression of aioBA in these strains is unknown but might involve distantly located regulators. It is interesting that there are nitRs (encoding nitroreductases) after the aioC instead of aioD in Acidovorax sp. NO1 (AGTS01000000), Herminiimonas arsenitoxydans ULPAs1 (Figure 1). Recently, a disruption of the nitR in strain NO1 resulted in the delay of AsIII oxidation indicating that the nitR may participate to the electron transfer in the strain (data not shown).

**PHYLOGENETICAL ANALYSES OF ARSENITE EFFLUX PROTEINS ArsB OR ACR3 ENCODED IN ars OPERONS OF THE ARSENIC ISLANDS.**

A total of 23 ars operons were detected in the arsenic islands and their inheritance models were analyzed by phylogenetic analyses.
of ArsB or ACR3, and comparing their phylogeny with those obtained using 16S rDNA. Among the 23 ars operons, eight ArsB and 13 ACR3 were detected. Some ars operons without the arsB or arc3 genes such as those from N. hamburgensis X14 and A. arsenitoxydans SY8 could not be analyzed in this context.

Phylogenetic analysis suggested most ArsB arsenite efflux proteins were congruent with 16S rDNAs (Supplementary materials, Figure S1). However, a notable exception included the ArsBs from Acidovorax sp. NO1, Thiomonas sp. 3A and A. faecalis NCIB8687, which clustered together. All of these ArsB proteins were encoded as part of a transposon, again indicating HGT events by transposon insertion accounted for acquisition of these ars operons in the respective arsenic islands.

The ACR3s separated into two clades in previous studies (Achour et al., 2007) and we also found that ACR3s on the arsenic islands could be divided into ACR3 (1) and ACR3 (2). In the respective ACR3 clades, their phylogenies were both in accordance with 16S rDNA phylogeny, therefore, suggesting genomic stability (Figure S2).

**PHYLGENETIC ANALYSES OF THE PHOSPHORUS RELATED pst AND phn OPERONS**

The pst1 or phn1 genes are localized within arsenic islands, while pst2 and phn2 genes are localized distantly on the respective chromosomes. Phylogenetic analysis indicated that all of the Pst2 branched in accordance with the 16S rDNAs (Figure S3). Therefore, pst2 operons appear to follow vertical inheritance. However, Pst1 did not strictly branch as the phylogenetic tree based on the 16S rDNA sequences (Figure 3). The Pst1 of Alcaligenes faecalis NCIB 8687 (β-Proteobacteria) clustered together with the Pst1 of α-Proteobacteria strains Agrobacterium tumefaciens 5A, Agrobacterium sp. GW4, Sinorhizobium sp. M14 and Xanthobacter autotrophicus Py2. The Pst1 of A. arsenitoxydans SY8 and H. arsenicoxydans ULPAs1 (β-Proteobacteria) were more related to those from γ-Proteobacteria. These results suggest that HGT may have occurred in transmission of the pst1 operon.

The phylogenies of Phn2 were in accordance with those calculated for the 16S rDNAs (Figure S4). However, Phn1 showed some conflicts (Figure 4). A Phn1 (A. faecalis NCIB 8687) from β-Proteobacteria clustered with the α-Proteobacteria. The phn2 loci were usually arranged as phnCDEE and located in the vicinity of other phosphonate utilizing genes, such as phnFGHIJKLMNOP (Jochimsen et al., 2011). The phn1 locus A. faecalis NCIB 8687 was arranged as phnDCEE, which had no other functional related genes in vicinity. Thus, the phn1 and phn2 may be functional different operons.

**DISCUSSION**

This study provides a comprehensive analysis of most of the available full-length AioBA sequences. Large scale scanning of the sequences in the vicinity of aioBA operons revealed the frequent occurrence of genes related to arsenic and phosphorous metabolism, such as the regulatory aioXSR operon and pst, phn, and ars operons (Silver and Phung, 2005). Considering gene synteny and structural analogies between arsenate and phosphate, we presumed that these genes function together in helping these
Li et al. Syntheny of bacterial arsenic islands

**FIGURE 4 | Phylogenetic trees of PhnC1 and 16S rDNA sequences.** Bold and * symbol represent proteins from the strains of the arsenic islands while the others are not. Putative horizontal gene transfer events (connected lines) have been compared based on the inconsistency of the amino acid sequence tree (on the left) and the 16S rDNA tree (on the right).

microbes to be able to use even low concentrations of phosphorus needed for vital functions under high concentrations of arsenic, and defined these sequences as the arsenic islands. The aioBA operons function to convert AsIII to the less toxic AsV but frequently also use this as a chemolithotrophic energy source. In contrast, ars operons are responsible for arsenic efflux after arsenate reduction and have a purely protective role (Silver and Phung, 2005). We found that some strains contain pst1 or phn1 operons encoding putative phosphate and phosphonate uptake transport systems in the vicinity of the aio operons in addition to the distantly located pst2 or phn2 operons, which raises the question about the functional role of pst1 and phn1 operons. Previous results indicated that arsenate can increase the $V_{\text{max}}$ of Pst2 for phosphate uptake (Moreno-Sanchez et al., 2012). AsIII may induce phosphate starvation as a competitive inhibitor of phosphate uptake, and cells may need to express more of these transporters or possibly more specific transporters for phosphate uptake. Similarly, we conjecture that the AsV generated by the AioBA may lead to phosphate starvation and the pst1 and phn1 may encode additional more specific uptake systems for P assimilation. This proposition is in accordance with the transcriptional profile of *H. arsenicoxydans* ULPAs1, in which pst1 operon was induced under conditions of As exposure (Cleiss-Arnold et al., 2010). Recently, it was reported that a pst1-like protein discriminated P from AsV 500–850-fold in phosphate-limited condition (Elias et al., 2012). In this study, we analyzed the localization of AsIII oxidation genes and found that aioBA of the α-Proteobacteria was prevalently localized on plasmids. As many arsenic islands were localized on plasmids, we predict that plasmids played a role in the widespread distribution of aioBA. Most of the aioBA sequences

<table>
<thead>
<tr>
<th>Strain</th>
<th>aioBA</th>
<th>aioXSR</th>
<th>pst1</th>
<th>phn1</th>
<th>arsB</th>
<th>acr3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agrobacterium tumefaciens 5A</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Agrobacterium sp. GV4</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Sinorhizobium sp. M14</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Nitrobacter hamburgensis X14</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Roseomonas cervicalis ATCC 49957</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Acidiphilium multivorans AIU301 chromosome</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Acidiphilium multivorans AIU301 pACMV2</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Acidiphilium sp. PM</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Herrnmimonas arsenicoxydans ULPAs1</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Achromobacter arsenitoxydans SY8</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Alcaligenes faecalis NCIB 8687</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
</tbody>
</table>

+ Represents the genes present in the arsenic island; △ Represents putative horizontal gene transfer event suggested by phylogenetic analysis or the presence of transposon elements; ■ Represents the plasmid origin of the genes.
analyzed here could be retrieved from Proteinobacteria and these sequences could be assigned to two groups, α-Proteobacteria and β-/γ-Proteobacteria, consistent with a previous analysis (Hamamura et al., 2009). The AioAs generally showed similar phylogeny as their 16S rDNA sequences (Figure 2) indicating an ancient origin of the enzyme (Cai et al., 2009b; Zhou and Xu, 2010). However, several strains showed putative HGT events with AioAs (Figure 2; Table 1) suggesting HGT also play a role during inheritance process (Arsène-Ploetze et al., 2010; Heinrich-Salmeron et al., 2011).

Unlike aioBAs, which are widely distributed among Proteinobacteria, Chlorobi, Deinococcus-Thermus, Chloroflexi, and even Archaea, the three component regulator genes genes aioXSR were only found in Proteinobacteria and displayed opposite transcriptional orientation between α- and β-Proteobacteria. It was possible that the aioXSR genes emerged in Proteinobacteria after the introduction of aioBA. The regulation of these aioBA operons with no aioXSR genes is not clear, but they may be controlled by distantly located regulators, or quorum sensing, as proposed by Kashyap et al. (Kashyap et al., 2006). Thus, the regulatory genes aioXSR may have evolved independently from aioBA. In a few strains including A. tumefaciens 5A, AioSR regulation of aioBA was RpoN-dependent, and the -24/-12 region for RpoN (σ54 factor for RNA polymerase) binding was also detected (Kang et al., 2012). The arsenite oxidase regulator AioR belonged to the NtrC family indicating that aioBA may be under the regulation of RpoN-dependent σ54-type promoter. However, the molecular details of AioR interacting with the promoter, and of the RpoN-RNA polymerase complex initiating transcription are still not known. Here we identified two tandem repeats of palindrome-like sequences which are located 100–200 nt upstream of the aioB start codon (Figure 5). The palindrome-like sequences are probably the upstream activating sequences (UAS) of σ54-type promoters which function in binding of AioR. The two palindromes and the -24/-12 regions were detected in all of the 12 aioBA operons that contained the aioXSR three-component system, but absent in other aioBA operons without aioXSR. Thus, we have to propose that the aioBAs without the upstream sequences of aioXSR are regulated differently.

The ars, pst, and phn operons were frequently detected on the arsenic gene islands but did not display a similar arrangement in various strains. Some plasticity was found even in the taxonomically closely related strains A. tumefaciens 5A, Agrobacterium sp.
GW4 and Sinorhizobium sp. M14. These strains shared the same arrangement in aio, pst, and phn operons, but not in ars operons. The large scale synteny of aio, pst and phn operons in these three strains may be due to vertical inheritance, while ars operons were integrated independently into the arsenic islands.

The ars operons encoded either ArsB or ACR3 as the AsIII efflux pump but did not display the same arrangement of the remaining genes such as arsC or arsH. This does not indicate a common origin of the different ars operons on the respective arsenic islands. The phylogenetic relatedness of ArsB or ACR3 seems to be in accordance with the corresponding tree predicted by 16S rDNA comparison. This suggests that ArsB or ACR3 were both mostly vertically inherited from the gene pool of the respective taxonomic clade. Vertical inheritance and HGT may have contributed to the origin of pst1 and phn1 operons (Table 1). It is therefore likely the arsenic islands did not evolve as a whole unit but formed independently by acquisition of functionally related genes and operons in respective strains. The elucidation of the phylogeny and distribution of aio genes might provide further insight into the evolution of the aioBA operon, and lead to better understanding of the arsenic island.

METHODS

DATA SOURCES

The amino acids sequence of AioA from Agrobacterium tumefaciens 5A was used as the initial query for a BLASTP search at the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov). Partial AioA sequences obtained from degenerate primers were ignored, as there was usually no flanking sequence information for them. We selected the full-length AioA sequences with the following threshold: sequence identity > 30%, coverage > 80%, starting with methionine and harboring the conserved domain TIGR02693 specific for arsenite oxidase. The selected BLASTP hits were used as query sequences for additional BLASTP searches, until no more full-length AioA hits were found. The corresponding nucleotide sequences where aioA was located, as well as the gene annotation information were downloaded in GenBank format for further analysis.

DETECTION OF GENE SYNTENY IN THE ARSENIC ISLANDS

The GenBank formatted sequences containing 57 aioA genes were loaded in the CLC sequence viewer program (http://www.clcbio.com). And the downstream and upstream sequences were scanned over 100 kb. Twenty-one sequences were found in vicinity of aioBA which were called arsenic islands, the others are single aioBA. The genes in the arsenic islands were exported as image files with the same genes represented by the same colors to detect synteny.

PHYLGENETIC ANALYSIS OF NUCLEOTIDE OR AMINO ACID SEQUENCES

All of the gene sequences were searched in the GenBank using the aioA sequence and a neighbor-joining (NJ) phylogenetic tree was constructed using ClustalX analysis (Thompson et al., 1997) and MEGA 4.0 software (Tamura et al., 2007). The parameters are as follows: phylogeny test and options (Bootstrap, 1000 replicates), Gaps/Missing Data (Pairwise Deletion), Substitution Model (Poisson correction for amino acids, Kimura 2P for nucleotides). Later on, other phylogenetic comparisons were made using the some methods for 16S rDNA, ArsB, Acr3, pstS, phnC, and the sequences were extracted from the corresponding genomes or other related genomes when necessary.

PREDICTING OF THE CHROMOSOME AND PLASMID LOCATION OF aioA GENES

The information on chromosome or plasmid location for all the 21 arsenic gene islands, if existing, was identified from the strain notes in GenBank. However, many of the 21 arsenic islands had no information on chromosome or plasmid location because they were from draft genomes. We predicted the chromosome and plasmid location of all the 21 “arsenic islands” by the cBar program (Zhou and Xu, 2010). The cBar program was developed for classifying metagenomes into chromosomal and plasmid sequences based on their different nucleotide pentamer frequencies.

DETECTION OF CONSERVED SEQUENCE MOTIFS

The upstream 300 bp sequences of all the 57 aioBA genes were selected. The conserved motifs were detected by The MEME Suite motif-based sequence analysis tools (http://meme.sdsc.edu/meme/intro.html). The sequence logo which graphically represents the sequence conservation was also automatically generated by MEME on-line program.

AUTHORS’ CONTRIBUTIONS

Hang Li carried out data collection, participated in the bioinformatic analyses and wrote the draft of the manuscript. Mingshun Li participated in bioinformatic analyses and helped to draft the manuscript. Yinyan Huang participated in sequence alignment study. Christopher Rensing participated in the design of the study and drafted the manuscript. Gejiao Wang coordinated the study, participated in its design and wrote the draft of the manuscript. All authors read and approved the final manuscript. We thank Dr. Timothy McDermott for discussion of the study design, editing, and comments on the manuscript.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China 31010103903 and 31170106.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://www.frontiersin.org/journal/10.3389/fmicb.2013.00347/abstract

Supplementary Figure S1 | Phylogenetical trees of ArsB and 16S rDNA sequences. Bold and *symbol represent proteins from the strains of the arsenic islands while the others are not. Phylogenetic relationship have been compared based on the amino acid sequence tree (on the left) and the 16S rDNA tree (on the right).

Supplementary Figure S2 | Phylogenetical trees of ACR3 and 16S rDNA sequences. Bold and *symbol represent proteins from the strains of the arsenic islands while the others are not. Phylogenetic relationship have been compared based on the amino acid sequence tree (on the left) and a 16S rDNA tree (on the right).
Supplementary Figure S3 | Phylogenetic trees based on PstS2 and 16S rRNA sequences. Bold and * symbol represent proteins from the strains of the arsenic islands while the others are not. Phylogenetic relationship have been compared based on the inconsistency of the amino acid sequence tree (on the left) and the 16S rDNA tree (on the right).

Supplementary Figure S4 | Phylogenetic trees of PhnC2 and 16S rDNA sequences. Bold and * symbol represent proteins from the strains of the arsenic islands while the others are not. Phylogenetic relationship have been compared based on the amino acid sequence tree (on the left) and the 16S rDNA tree (on the right).

REFERENCES


Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 29 July 2013; accepted: 31 October 2013; published online: 20 November 2013.

This article was submitted to Microbiological Chemistry, a section of the journal Frontiers in Microbiology.
Copyright © 2013 Li, Li, Huang, Rensing and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.