Arsenic resistance strategy in *Pantoea* sp. IMH: Organization, function and evolution of *ars* genes

Liyong Wang¹,², Xuliang Zhuang¹,², Guoqiang Zhuang¹,² & Chuanyong Jing¹,²

*Pantoea* sp. IMH is the only bacterium found in genus *Pantoea* with a high As resistance capacity, but its molecular mechanism is unknown. Herein, the organization, function, and evolution of *ars* genes in IMH are studied starting with analysis of the whole genome. Two *ars* systems - *ars1* (*arsR1B1C1H1*) and *ars2* (*arsR2B2C2H2*) - with low sequence homology and two *arsC*-like genes, were found in the IMH genome. Both *ars1* and *ars2* are involved in the As resistance, where *ars1* is the major contributor at 15 °C and *ars2* at 30 °C. The difference in the behavior of these two *ars* systems is attributed to the disparate activities of their *arsR* promoters at different temperatures. Sequence analysis based on concatenated *ArsRBC* indicates that *ars1* and *ars2* clusters may be acquired from *Franconibacter helveticus* LMG23732 and *Serratia marcescens* (plasmid R478), respectively, by horizontal gene transfer (HGT). Nevertheless, two *arsC*-like genes, probably arising from the duplication of *arsC*, do not contribute to the As resistance. Our results indicate that *Pantoea* sp. IMH acquired two different As resistance genetic systems by HGT, allowing the colonization of changing ecosystems, and highlighting the flexible adaptation of microorganisms to resist As.

Arsenic (As) is a toxic element present in many environmental biotopes. Inorganic As exists primarily in two valence states: As(III) and As(V). To resist the disruptive effects of As, microbes have evolved a variety of mechanisms, including As(III) oxidation through the activity of As(III) oxidase and As methylation by methyltransferase. Microorganisms can also utilize As in metabolism either as a terminal electron acceptor in dissimilatory As(V) respiration or as an electron donor in chemoautotrophic As(III) oxidation. Nevertheless, the most universal and well-characterized As resistance mechanism is induced by the *ars* system.

The content and organization of the *ars* system vary greatly between strains. Most of the core genes in *ars* operons contain *arsR*, *arsB* and *arsC*, and other genes are also reported, such as *arsA*, *arsD*, *arsT*, *arsX*, *arsH*, and *arsN*. A common organization of the *ars* cluster is *arsRBC*, whereas duplicate *ars* operons can also be found in a single strain, such as *Pseudomonas putida* KT2440⁴. Interestingly, in this case more than one *ars* cluster with different structures is observed in the same strain. In all, the content and organization of *ars* operons exhibit great diversity and complexity, and subsequently contribute to the As resistance capability of strains as summarized in Table S1.

The complexity of the *ars* system in diverse bacteria raises the question of its origin and evolution. Different evolution theories have been advanced for the evolutionarily old proteins, efflux pump protein (*ArsB*) and As(V) reductase (*ArsC*). For example, *ArsB* and *ArsC* may have evolved convergently, as evidenced by sequence analyses⁵. In contrast, *arsC* genes are reported to have a common origin and may have been transferred to other domains by HGT in early times, followed by subsequent divergence to the current phylogeny⁶. Follow-up studies suggest that the HGT events of *ars* genes may be common in nature⁷.

*Pantoea* is a genus of Gram-negative, facultative anaerobic bacteria. This genus belongs to gamma *Proteobacteria*, family *Enterobacteriaceae*, and was recently separated from the genus *Enterobacter*. Currently, the genus contains twenty-six species (http://www.bacterio.net/pantoea.html). Members of this genus are found in various environmental matrices⁸. In 2013, strain *Pantoea* sp. IMH was isolated for the first time from As-polluted groundwater and reported to resist high concentrations of As, up to 150 mM As(V) and 20 mM As(III)¹⁰. However, the hyper As-resistance strategy employed by *Pantoea* sp. IMH remains unclear.

¹State Key Laboratory of Environmental Chemistry and Ecotoxicology, Research Center for Eco—Environmental Sciences, Chinese Academy of Sciences, P.O. Box 2871, Beijing 100085, China. ²University of Chinese Academy of Sciences, Beijing 100049, China. Correspondence and requests for materials should be addressed to C.J. (email: cyjing@rcees.ac.cn)
Herein, we present the first study of the molecular mechanism of As resistance in strain Pantoea sp. IMH. Two different ars systems - ars1 (arsR1B1C1H1) and ars2 (arsR2B2C2H2) - were identified as being responsible for As resistance, contributing in different ways under changing temperature. In addition, we determined that the ars genes in IMH were probably acquired by HGT. The insights gained in this study improve our understanding of the flexible adaptation of microorganisms to resist As.

**Results**

**As resistance systems in Pantoea sp. IMH.** Strain Pantoea sp. IMH was able to resist up to 150 mM As(V) and 20 mM As(III), whereas E. coli W3110 with an arsRBC operon did not survive at concentrations above 50 mM As(V) and 5 mM As(III) (Fig. S1). To explore the molecular basis for its hyper-resistance to As, we determined the genome sequence of IMH and identified eight ars genes, including two arsR encoding a self-repressed transcriptional regulator, two arsB encoding a membrane-bound transporter that extrudes As(III) out of the cell, two arsC encoding a cytoplasmic As(V) reductase, and two arsH encoding an NADPH-dependent FMN reductase with an unknown biological function. These ars genes were organized as an ars1 cluster (arsR1B1C1H1) and ars2 cluster (arsR2B2C2H2) scattered on the chromosome (Fig. 1a). The genes in each ars cluster were separated by a short sequence of only a few nucleotides, suggesting they were organized in the same operon. To justify this hypothesis, we performed RT-PCR experiments using primers across intergenic regions (Table S3). The results indicate that the genes arsRBC within the arsRBCH cluster were organized as a co-transcribed operon, whereas arsH was in another operon (Fig. 1b).

The degree of DNA sequence identity between homologous genes underscores the appreciable differences between these two ars clusters. Specifically, as shown in Fig. S2, arsR1 and arsR2 shared 50% sequence identity, arsB1 and arsB2 shared 75%, arsC1 and arsC2 shared 60%, and arsH1 and arsH2 shared 70%. Moreover, two arsC-like genes with just 25% homology (arsC1-like and arsC2-like) were found in the genome. It is an exceptional circumstance that IMH contains two ars systems and two As resistance molecular bases, considering that most bacteria have just one such cluster11. Therefore, we were motivated to investigate the functional contributions of each ars resistance system and molecular base to the As resistance in Pantoea sp. IMH.

**Contribution of two ars systems and two arsC-like genes to As resistance.** We first examined the transcription levels of ars genes in each ars operon and arsC-like gene by performing reverse transcription quantitative PCR (RT-Q-PCR), using 16S rRNA as an internal control. As shown in Fig. 2, all genes of ars1 and ars2 clusters were completely transcribed. Notably, ars2 genes exhibited about 2–4 fold higher transcription levels than ars1 genes.

Figure 1. Analysis of co-transcript unit in the ars clusters of Pantoea sp. IMH by RT-PCR. (a) Map position of ars genes and the primers for RT-PCR analysis. Primers used and amplified products (numbered) are given below the schematic representation of the genes. (b) Result of RT-PCR reactions with RNA from IMH grown in 1 mM As(V) condition. The numbering on the top of the gels corresponds to the product numbers drawn schematically in the outline given above. M, DNA mark; (+), positive control in which genomic DNA was used as template in the RT-PCR; RT, standard RT-PCR reaction; (−), negative control in which no reverse transcriptase was added to the RT reaction.
To further identify the role of each ars cluster, we generated strains Δars1 and Δars2 lacking ars1 and ars2 clusters, respectively, based on the genome of IMH as described in the Methods section. The growth of Δars1, Δars2, and the wild type strain IMH was then monitored in the LB medium with 50 mM As(V) and 5 mM As(III).

As shown in Fig. 3c, the growth of both Δars1 and Δars2 was substantially suppressed compared to the wild type strain IMH. Specifically, the deletion of ars1 resulted in more suppression of As resistance than that of ars2, implying that ars2 contributes to a greater extent to the overall As resistance. Moreover, we constructed the functional complementary plasmids pLGM1-ars1 and pLGM1-ars2 (see Methods for details) and then introduced them into the Δars1 and Δars2 strains, respectively. As shown in Fig. S4, the As resistance capabilities of complementary strains Δars1/pLGM1-ars1 and Δars2/pLGM1-ars2 were appreciably improved compared with the deleted strains (Δars1 and Δars2), which confirms the origin of As resistance in the corresponding operons.

Heterologous expression experiments were carried out to study the functional contributions of the two ars clusters to the As resistance. Recombinant plasmids for expression of ars1 and ars2 clusters were constructed according to the procedure described in the Methods. These plasmids were separately introduced in E. coli AW3110 (lacking any As resistance system), yielding the recombinant E. coli AW3110-ars1 and E. coli AW3110-ars2 strains. The growth of these strains together with the wild type strain E. coli W3110 (containing...
one ars operon) was then examined in LB medium containing 5 mM As(V) and 1 mM As(III). The results show that the heterologous host with the ars2 system acquired a higher resistance to As than that with the ars1 system (Fig. 3a).

In sum, the two ars clusters of strain IMH together contribute to its As resistance, in which the ars2 cluster is the major contributor. This observation raises a follow-up question: why did Pantoea sp. IMH evolve two ars systems to resist As? When bacteria species survive under changing environmental circumstances, some proteins may not function under all physicochemical conditions. Bacteria may meet this challenge by having two or more copies of genes to realize the same function under different conditions13. An example includes strain Pseudomonas putida KT2440, possessing two copies of equivalent ars operons to expand its functional scope14. Similarly, we propose that possessing two ars systems is an evolved strategy for IMH to survive in different ecological niches. To validate our speculation, the function of these two ars systems in different ecological conditions was further investigated.

Functioning of two ars systems under different environmental conditions. Major environmental factors, including concentrations of As(V) and As(III), pH, and temperature, may affect As resistance. We therefore tested the performance of each ars system in IMH at different pH values (pH 5, 7, and 9), As(V) concentrations (1 and 10 mM), As(III) concentrations (1 and 10 mM), and temperatures (15 and 30 °C). We first examined the transcriptional levels of arsC and arsC using RT-Q-PCR as proxies for the expression of the ars systems. Figure S5 shows that the transcription level of arsC2 was higher than that of arsC1 for different pH values and concentrations of As(V) and As(III). In contrast, an opposite result (i.e., arsC1 > arsC2) was obtained at a lower temperature (15 °C). This observation suggests that pH, As concentration, and speciation have no influence on the major contribution of ars2 to As resistance, but a low environmental temperature enables ars1 to be the predominant contributor.

Furthermore, the growth of E. coli AW3110-ars1, E. coli AW3110-ars2, Δars1, and Δars2 in LB medium with As at 15 °C was tested, where E. coli W3110 and IMH were used as controls. Figure 3b,d shows that E. coli-ars1 and Δars2 grew better than E. coli-ars2 and Δars1, respectively, at 15 °C, contrary to the results at 30 °C (Fig. 3a,c). In agreement with the RT-Q-PCR results, the above observations confirm that arsI contributed more than ars2 at the lower temperature (15 °C).

Why is this phenotype endowed with two ars systems dominant at different temperatures? We hypothesized that the activity of the arsR promoter should regulate its expression at different temperatures. To test this hypothesis, we first determined the transcription start site (TSS) using 5'-RACE, and predicted the −35 and −10 sequences, as well as ribosomal binding site (RBS) sequence, of each of the promoter regions of arsR using SoftBerry software. As shown in Fig. 4a,b, the TSS was located 21 bp upstream of the translational start site of arsR1 in ars1, and 136 bp in ars2. The distance between the −10 region and start codon was 28 bp in the arsRI promoter and
164 bp in the **arsR2** promoter. The pronounced difference in the organization of these two **arsR** promoters may contribute to their different functions.

To examine the activity of the **arsR** promoter at different temperatures, we assembled equivalent reporter gene fusions between the predicted promoter regions of each **ars** cluster and a lacZ reporter gene without its promoter, producing plasmids pPR9TT-Pars1 and pPR9TT-Pars2 according to the procedure described in the Methods. The two constructed plasmids were transferred into *E. coli* AW3110, and their β-galactosidase levels were measured at 15 and 30°C in LB media with and without As(III) (1 mM), which is the effective trigger for the **ars** operon. As shown in Fig. 4c, at the higher temperature (30°C), the β-galactosidase activity of pPR9TT-Pars2 was higher than that of pPR9TT-Pars1 with and without As(III) induction. In contrast, the activity of Pars2 was noticeably inhibited, while Pars1 showed comparatively high activity with and without As(III), even higher than that of Pars1 at the low temperature (15°C). These results explain the different behaviors of **ars1** and **ars2** expression regarding their temperature dependence. The different performance of the two **ars** promoters was one of the important factors influencing the function of the two **ars** clusters in response to different temperatures. On this basis, we propose that the evolutionary reason for maintaining two **ars** systems in IMH is that their combination facilitates the survival of this strain over an extended range of temperatures in arsenic-polluted niches. This result raises a further question: what is the evolutionary origin of these two **ars** systems and two **arsC**-like genes?

### Evolution of **ars** clusters and **arsC**-like genes in *Pantoea* sp. IMH genome.

Substantial differences were observed between the homologous genes, **ars1** and **ars2**, in their sequence identities and As resistance capabilities. We proposed that the differential origins of the two **ars** clusters may have derived from HGT. To validate our speculation, we first compared the genome of IMH to those of other bacterial strains including *Ps. IMH* (Fig. S7). A summary of the features for each genome is shown in Table S4. Notably, the IS elements are responsible for transferring genetic information between different cells15, and their abundance in the genome of IMH is that their combination and two **arsC**-like genes?

Figure 5. Neighbor-joining tree based on ArsRBCH protein sequences of IMH and representative microorganisms. The numbers at the nodes indicate levels of bootstrap support (%) based on analysis of 100 assembled datasets. Only values at or above 50% are given, bar 0.1 substitutions per amino acid position.
As(V) and 1 mM As(III) in 96-microwell plates at either 15 °C or 30 °C, and OD600 was evaluated at 12 h. For monitoring the growth of IMH in its early evolution, the genomes of Pantoea sp. IMH have been recently sequenced and annotated. More than one As resistance system in a strain can elevate its As resistance, which explains why IMH can survive in such high As concentrations.

Our genome sequencing showed that there are two arsC-like genes in the genome of IMH. Transcription levels of the two arsC-like genes were not detected when the strain encountered As (Fig. 2). This result suggests that arsC-like genes did not contribute to the As resistance. It is rare for arsC-like genes to show no As resistance capability. We further investigated the reasons for this phenomenon. It was reported that Cys-12, Arg-60, Arg-94, and Arg-107 were four conserved residues of the ArsC protein in the process of As resistance. Cys-12 was identified as a catalytic residue and was activated by nearby residues Arg-60, Arg-94, and Arg-107. Alignment analysis showed that Cys-12 and Arg-94 residues were conserved, but residues Arg-107 and Arg-60 in two ArsC-like proteins were not conserved respectively (Fig. S3). These changes in the amino acid sequence further verified that the two ArsC-like proteins did not contribute to the As resistance. Interestingly, using phylogenetic analysis, we found that arsC-like sequences fell into distinct groups when compared to arsC genes. This suggests that multiple arsC-like genes may have resulted from arsC duplication and had already evolved with deviation. To clarify their relations with As resistance, the structures of ArsC-like proteins are worthy of further study.

It is thought that variants of a core arsRBC operon are common in the genomes of various bacteria, and it is rare that more than one ars operon appears in the same genome. Bacterial species usually adapt to changing environments by evolving two or more copies of genes, each one performing the same function under differential conditions. Thus, the composition and gene distribution of a genome usually reflect the capacity for adaptation to different ecological niches. In the IMH strain, we indeed found two ars systems with different patterns of expression and efficiency at different temperatures. This is a strategy for strain IMH to expand the scope of the encoded function to a wider range of physicochemical settings. Interestingly, our result was consistent with the report on Pseudomonas putida KT2440. We speculate that temperature is the most important environmental factor in the evolutionary history of ars clusters. Of course, if more strains with two or more copies of ars clusters are found, further mechanistic research should be carried out to support this hypothesis.

HGT is an important adaptation strategy to efficiently obtain ‘alien’ DNA. To readily adapt to diverse and stringent growing conditions, IMH obtained two ars clusters by HGT. To identify the transfer of genetic information between genomes, we applied three commonly used methods including identification of IS elements and deviant G+C content, and phylogenetic analysis. The existence of IS elements in the flanking region of ars clusters (Fig. S7) together with the greatly different G+C content in the IMH genome (54.74%) and ars2 (56.28%) and ars2 (51.86%) clusters, suggests that the two ars clusters (arsR1B1C1H1 and arsR2B2C2H2) may have been acquired by HGT. Phylogenetic analysis further revealed that the ars1 cluster may have been acquired via HGT from a source related to Franciscubacter helveticus LMG23732 in its early evolution, and the ars2 cluster from Serratia marcescens (plasmid R478). Our result is in agreement with previous reports where microorganisms can obtain the same functional genes from different sources.

Methods

Genome sequencing, genome annotation and analysis. The genome of strain IMH was sequenced using the IlluminaHiSeq 2000 sequencing platform at the Beijing Genomics Institute (BGI) (Shenzhen, China). Genes were predicted from the assembled result using Glimmer 3.02. The rRNA and tRNA genes were identified with RNAmmer and tRNAscan-SE, respectively. Genome annotation was accomplished by analyzing protein sequences. The resulting translations were aligned with databases, including KEGG 59, GO 1.419 and Swiss-Prot 201206. The draft genome has been deposited in GenBank and the accession number used is JFGT01000000.

Strains, plasmids and culture conditions. The stains and plasmids used in this work are summarized in Table S2. E. coli and Pantoea strains were grown in LB medium (per liter contains: 10 g tryptone, 5 g yeast and 10 g NaCl) or LB plates (LB medium with w/v 1.5% agar) at either 15 °C or 30 °C as indicated in each case. When appropriate, antibiotics were added in the following concentrations: 100 μg/mL ampicillin, 100 μg/mL kanamycin, and 100 μg/mL streptomycin. For testing of minimal inhibitory concentrations (MICs), strains were incubated in LB medium with a series of concentrations of As(V) and As(III) as shown in Fig. S1. For monitoring the growth of E. coli AW3110-arsI, E. coli AW3110-ars2 and E. coli AW3110, strains were cultured in LB medium with 5 mM As(V) and 1 mM As(III) in 96-microwell plates at either 15 °C or 30 °C, and OD600 was evaluated at 12 h. For monitoring the growth of ΔarsI, Δars2 and IMH, strains were cultured in LB medium with 50 mM As(V) and 5 mM As(III) in 96-microwell plates at either 15 °C or 30 °C, and OD600 was evaluated at 12 h. When detecting the As resistance under differential pH conditions, the pH of the LB medium was adjusted to pH 5.0, 7.0 and 9.0.

Construction of recombinant plasmids for expression in E. coli. In order to construct the plasmids used in the heterologous expression experiments, genomic DNA of Pantoea sp. IMH was used as a template for
cloning the two \( \text{ars} \) clusters. A 3.9 kb \( \text{BamHI-XbaI} \) DNA fragment containing the complete \( \text{ars1} \) cluster (promoter region, 360 bp upstream of the start codon ATG of \( \text{arsR} \), the contiguous four genes \( \text{arsRIBC1HI} \) and 310 bp upstream of the start codon ATG of \( \text{arsH1} \)) was PCR-amplified with primers \( \text{ArsR1-F} \) and \( \text{ArsR1-R} \) (Table S3). A 3.6 kb \( \text{BamHI-XbaI} \) DNA fragment containing the complete \( \text{ars2} \) cluster (a 301 bp region downstream of the stop codon TAA of \( \text{arsR2} \) and the contiguous ten genes \( \text{arsR2R2C2H2} \) and 361 bp downstream of the stop codon TAA of \( \text{arsH2} \)) was PCR-amplified with primers \( \text{Ars2-F} \) and \( \text{Ars2-R} \) (Table S3). The PCR products were ligated to the \( \text{BamHI-XbaI} \) site of plasmid \( \text{pUC18} \), yielding plasmids \( \text{pUC18-ars1} \) and \( \text{pUC18-ars2} \). Then the plasmids were transferred to \( \text{E. coli} \) \( \text{AW3110} \), yielding the recombinant \( \text{E. coli} \) \( \text{AW3110-ars1} \) and \( \text{E. coli} \) \( \text{AW3110-ars2} \) strains, respectively.

**Construction of \( \Delta \text{ars1} \) and \( \Delta \text{ars2} \).** To obtain the deleted mutants of \( \text{ars1} \) and \( \text{ars2} \) clusters in \( \text{Pantoea sp.} \) IMH, the suicide vector \( \text{pARS10} \) was constructed by inserting the Invitrogen Gateway \( \text{attR-Cmr}^\text{K} \) cassette into the backbone of \( \text{SmaI-SmaI} \) digested plasmid \( \text{pKNG101} \), where \( \text{E. coli DH5} \alpha \) (\( \text{pir} \)) was used as the host of \( \text{pARS10} \). The \( \Delta \text{ars1} \) and \( \Delta \text{ars2} \) mutated strains were created using a modified Gateway method described by \( \text{Cho}^{27} \). To delete the \( \text{ars1} \) cluster, the flanking regions of the \( \text{ars1} \) cluster were amplified by PCR using primers as summarized in Table S3. A kanamycin resistance cassette derived from plasmid \( \text{pKD4} \) was inserted between the flanking regions of the \( \text{ars1} \) cluster using a PCR overlap technique with the primers in Table S3. The resulting PCR products containing the Km-resistance cassette flanked by \( \text{ars1} \) cluster were cloned into the Gateway Entry vectors \( \text{pDONR221} \). The construct was transferred into the suicide vector \( \text{pARS10} \), obtaining plasmid \( \text{pARS1-1} \). The plasmid \( \text{pARS1-1} \) was transferred into \( \text{E. coli} \) \( \text{S17-1} \) and conjugally introduced into \( \text{Pantoea sp.} \) IMH. An allelic replacement event was selected based on double resistance. PCR with primers listed in Table S3 was used for the verification of the allelic replacement. Generation of the \( \Delta \text{ars2} \) strain followed the same method.

**Construction of plasmids for complementation studies.** In order to verify the As resistance function of the two \( \text{ars} \) clusters, plasmids \( \text{pLGM1-ars1} \) and \( \text{pLGM1-ars2} \) were constructed for complementation studies. As described in the previous section titled “Construction of recombinant plasmids for expression in \( \text{E. coli} \)”, \( \text{ars1} \) and \( \text{ars2} \) were amplified and then ligated to the \( \text{BamHI-EcoRI} \) site of plasmid \( \text{pLGM1} \), yielding \( \text{pLGM1-ars1} \) and \( \text{pLGM1-ars2} \).

**RT-PCR analysis.** In order to determine the operons in \( \text{ars} \) clusters, an RT-PCR experiment with primers designed to span across intergenic regions (Table S3) was carried out. A culture of \( \text{Pantoea sp.} \) IMH was grown in LB medium with 1 mM As(V). After 8h, the IMH strains were harvested by centrifugation at 4 °C, and the total RNA was isolated using the PrimeScript" RT reagent Kit with gDNA Eraser (Takara Bio) according to the manufacturer’s instructions. The possibility of contamination of genomic DNA was eliminated by digestion with RNase-free DNase I (Takara Bio). The integrity and size distribution of the RNA were verified by agarose gel electrophoresis, and the concentration was determined spectrophotometrically. Synthesis of cDNA was carried out using RT Prime Mix according to the manufacturer’s specification (Takara Bio). 1.0 \( \mu \)g of cDNA was used for the template of RT-PCR.

**RT-Q-PCR analysis.** In order to understand the differences in each gene's transcription level in \( \text{ars1} \) and \( \text{ars2} \) clusters under differential environmental factors, RT-Q-PCR analysis was used. \( \text{Pantoea sp.} \) IMH was grown in LB medium with different As(V) or As(III) concentrations (1 mM and 10 mM), in LB medium with 1 mM As(V) at different temperatures (15 °C and 30 °C), and in LB medium with different pH (5, 7, and 9). Then the cDNA was obtained as described in the RT-PCR analysis. Specific cDNA was employed to quantify the transcriptional signals of the \( \text{ars} \) genes and \( \text{arsC-like} \) genes, where 16 S rRNA gene was used as an internal reference. Primers used are listed in Table S3. RT-Q-PCR reactions were performed with three replicates using the ABI applied Biosystems vii A7.

**Transcription start site identification.** To determine the transcription start site (TSS) of the two \( \text{ars} \) operons, the 5′-RACE method was employed using the SMARTer™ RACE cDNA Amplification Kit (Clontech). Gene-specific primers are listed in Table S3. The PCR product was cloned into the pMD18-T Vector and then sequenced.

**\( \text{P}_{\text{ars}}-\text{lacZ} \) transcriptional fusions and \( \beta \)-galactosidase assays.** To explain the reason for the distinctly different performance of \( \text{ars1} \) and \( \text{ars2} \) clusters at 15°C and 30°C, the promoter activities of the two \( \text{ars} \) clusters were determined. The promoter of the \( \text{ars1} \) cluster - a 110 bp DNA fragment (\( \text{P}_{\text{ars1}} \)) (from -107 to +3 relative to the \( \text{arsR1} \) transcription start codon) - and the promoter of the \( \text{ars2} \) cluster - a 241 bp DNA fragment (\( \text{P}_{\text{ars2}} \)) (from -238 to +3 relative to the \( \text{arsR2} \) transcription start codon) - were amplified from the total DNA of \( \text{Pantoea sp.} \) IMH using primers listed in Table S3. The amplified fragments were then cloned into the promoter vector \( \text{pPR9TT} \), generating transcriptional fusions between the inserted promoter regions and a promoterless, complete \( \text{lacZ} \) gene, \( \text{pPR9TT-P}_{\text{ars1}} \) and \( \text{pPR9TT-P}_{\text{ars2}} \). The plasmids were then transformed into \( \text{E. coli DH5}\alpha \), yielding \( \text{E. coli DH5}\alpha /\text{P}_{\text{ars1}}::\text{lacZ} \) and \( \text{E. coli DH5}\alpha /\text{P}_{\text{ars2}}::\text{lacZ} \), respectively. For \( \beta \)-galactosidase activity assays, strains \( \text{E. coli DH5}\alpha /\text{P}_{\text{ars1}}::\text{lacZ} \) and \( \text{E. coli DH5}\alpha /\text{P}_{\text{ars2}}::\text{lacZ} \) were grown in LB medium with 1 mM As(III) for 12h at 15°C and 30°C with shaking. \( \beta \)-galactosidase activity was measured according to the method described by \( \text{Miller}^{28} \). Briefly, a 100 \( \mu \)L sample was mixed with 900 \( \mu \)L of Z buffer and shaken for 20 sec. Then, 200 \( \mu \)L o-nitrophenyl\( \beta \)-D-galactopyranoside (ONPG) (4 mg/mL) was added and incubated for 20 min at 30°C. To stop the above reaction, 500 \( \mu \)L of 1 M Na2CO3 solution was used. Finally, the OD\( _{420} \) and OD\( _{550} \) values were measured after the mixture was centrifuged. \( \beta \)-galactosidase (1 unit) = [1000 × (OD\( _{420} \) - 1.7 × OD\( _{550} \))]/[Time (min) × Vol (mL) × OD\( _{400} \)].
Comparative genomics. All orthologous pairs between tests of Pantoae genomes were identified by Pan Genome Analysis Pipeline29. The common dataset of shared genes among test strains was defined as their core genome. The total set of genes with test genomes was defined as the pan-genome. The unshared genes in each strain were defined as unique genes. The genomes used in this study are listed in Table S4.

Phylogenetic analysis. In order to analyze the evolution of the two ars clusters of strain IMH, phylogenetic trees were constructed using the neighbor-joining method. Evolutionary distances were calculated according to Kimura’s two-parameter model. Bootstrap analysis was performed on the basis of 1000 replications. The software package MEGA version 5.0 was used.

References
1. Oremland, R. S. & Stolz, J. F. The ecology of arsenic. Science 300, 939–944 (2003).
2. Páez-Espino, D., Tamames, J., De, L. V. & Cánovas, D. Microbial responses to environmental arsenic. Biometals 22, 117–130 (2009).
3. Silver, S. & Phung, L. T. Genes and enzymes involved in bacterial oxidation and reduction of inorganic arsenic. Appl. Environ. Microbiol. 71, 599–608 (2005).
4. Cuebas, M., Villafane, A., McBride, M., Yee, N. & Bini, E. Arsenate reduction and expression of multiple chromosomal ars operons in Geobacillus kaustophilus A1. Microbiology 157, 2004–2011 (2011).
5. Mukhopadhyay, R., Rosen, B. P., Phung, L. T. & Silver, S. Microbial arsenic: from geocycles to genes and enzymes. FEMS Microbiol. Rev. 26, 311–325 (2002).
6. Jackson, C. R. & Dugas, S. L. Phylogenetic analysis of bacterial and archaean arsC gene sequences suggests an ancient, common origin for arsenate reductase. BMC Evol. Biol. 3, 18 (2003).
7. Fernandez, M., Udaondo, Z., Niqui, J. L., Duque, E. & Ramos, J. L. Synergic role of the two ars operons in arsenic tolerance in Pseudomonas putida KT2440. Environ. Microbiol. Rep. 6, 483–489 (2014).
8. Zhang, L. & Birch, R. G. Arsenate and expression of multiple chromosomal ars operons. J. Appl. Microbiol. 97, 1071–1081 (2004).
9. Wu, Q., Du, J., Zhaob, G. & Jing, C. Bacillus sp. SXB and Pantoae sp. IMH, aerobic As(V)–reducing bacteria isolated from arsenic-contaminated soil. J. Appl. Microbiol. 114, 713–721 (2013).
10. Shen, W. et al. Microbial responses to environmental arsenic. Biometals 22, 117–130 (2009).
11. Martin, P. et al. Insights into the structure, solvation, and mechanism of ArsC arsenate reductase, a novel arsenic detoxification enzyme. Structure 9, 1071–1081 (2001).
12. Sanchez-Perez, G., Mira, A., Nyiro, G., Pasic, L. & Rodriguez-Valera, F. Adapting to environmental changes using specialized paralogs. Trends Genet. 24, 154–158 (2008).
13. Páez-Espino, D., Durante-Rodriguez, G. & Lorenzo, V. Functional coexistence of twin arsenic resistance systems in Pseudomonas putida KT2440. Environ. Microbiol. 17, 229–238 (2015).
14. Touchon, M. & Rocha, E. P. Causes of insertion sequences abundance in prokaryotic genomes. Mol. Biol. Evol. 24, 969–981 (2007).
15. Francino, M. P. An adaptive radiation model for the origin of new gene functions. Nat. Genet. 37, 573–578 (2005).
16. Tian, Y. et al. Nitrogen fixation island and rhizosphere competence traits in the genome of root-associated Pseudomonas stutzeri A1501. Proc. Natl. Acad. Sci. USA 105, 7564–7569 (2008).
17. Aréiz-Ploetz, F. et al. Structure, function, and evolution of the Thiomonas spp. genome. PLoS Genet. 6, e1000859 (2010).
18. Potapova, O. & Doolittle, W. F. Lateral gene transfer. Curr. Biol. 21, R242–R246 (2011).
19. Delcher, A. L., Bratke, K. A., Powers, E. C. & Salzberg, S. L. Identifying bacterial genes and endosymbiont DNA with Glimmer. Bioinformatics 23, 673–679 (2007).
20. Zhao, Y. & Rosen, B. P. His-8 lowers the pKa of the essential Cys-12 residue of the ArsC arsenate reductase of plasmid R773. J. Biol. Chem. 271, 33256–33260 (1996).
21. Zhaxybayeva, O. & Doolittle, W. F. Lateral gene transfer. Curr. Biol. 21, R242–R246 (2011).
22. Kunc, P. & Stolz, J. F. Microbial arsenic: from geocycles to genes and enzymes. Biometals 22, 117–130 (2009).
23. Kanehisa, M. et al. KEGG: genome databases for complete genomes. Nucleic Acids Res. 35, D354–D357 (2007).
24. Ashburner, M., et al. Gene Ontology: tool for the unification of biology. Nat. Genet. 35, 25–29 (2000).
25. Wu, Q., Du, J., Zhaob, G. & Jing, C. Bacillus sp. SXB and Pantoae sp. IMH, aerobic As(V)–reducing bacteria isolated from arsenic-contaminated soil. J. Appl. Microbiol. 114, 713–721 (2013).
26. Millar, I. H. Experiments in molecular genetics. (Cold Spring Harbor Laboratory, 1972).
27. Zhan, Y. et al. PGAP: Pan-genomes analysis pipeline. Bioinformatics 28, 416–418 (2011).

Acknowledgements
We acknowledge the financial support of the National Basic Research Program of China (2015CB932003), the Strategic Priority Research Program of the Chinese Academy of Sciences (XDB14020302), and the National Natural Science Foundation of China (41373123, 41425016, 41503094 and 21321004). We thank Yongguan Zhu for the strain E. coli AW3110, and Sanfeng Chen for pPr9TT vector and strain E. coli S17-1.

Author Contributions
L.W. and C.J. conceived and designed the study. L.W. performed the laboratory work and data analysis, with assistance from X.Z. and G.Z., L.W. and C.J. drafted the tables and figures, and prepared the main manuscript.

Additional Information
Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Wang, L. et al. Arsenic resistance strategy in Pantoae sp. IMH: Organization, function and evolution of ars genes. Sci. Rep. 6, 39195; doi: 10.1038/srep39195 (2016).

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
