

# Discerning the role of a functional arsenic-resistance cassette in the evolution and adaptation of a rice pathogen

Amandeep Kaur, Rekha Rana, Tanu Saroha and Prabhu B. Patil\*

## Abstract

Arsenic is highly toxic element to all forms of life and is a major environmental contaminant. Understanding acquisition, detoxification and adaptation mechanisms in bacteria that are associated with the host in arsenic-rich conditions can provide novel insights into the evolutionary dynamics of host–microbe–environment interactions. In the present study, we have investigated an arsenic-resistance mechanism acquired during the evolution of a particular lineage in the population of *Xanthomonas oryzae* pv. *oryzae*, which is a serious plant pathogen infecting rice. Our study revealed the horizontal acquisition of a novel chromosomal 12 kb *ars* cassette in *X. oryzae* pv. *oryzae* IXO1088 that confers high resistance to arsenate/arsenite. The *ars* cassette comprises several genes that constitute an operon induced in the presence of arsenate/arsenite. Transfer of the cloned *ars* cassette to *X. oryzae* pv. *oryzae* BXO512, which lacks the cassette, confers an arsenic-resistance phenotype. Furthermore, the transcriptional response of *X. oryzae* pv. *oryzae* IXO1088 under arsenate/arsenite exposure was analysed using RNA sequencing. Arsenic detoxification and efflux, oxidative stress, iron acquisition/storage, and damage repair are the main cellular responses to arsenic exposure. Our investigation has provided insights into the existence of a novel detoxification and adaptation mechanism within the *X. oryzae* pv. *oryzae* population to deal with high-arsenic conditions outside the rice plant.

## DATA SUMMARY

The genome sequence of *Xanthomonas oryzae* pv. *oryzae* strain BXO512 has been submitted to the National Center for Biotechnology Information (NCBI) GenBank with accession number CP065228. The transcriptome data of *X. oryzae* pv. *oryzae* strain IXO1088 has been submitted to the NCBI Gene Expression Omnibus (GEO) database under accession number GSE163417.

Supplementary information can be found with the online version of this article and via the following Figshare links.

- (1) Data file (.xlsx) Tables S1–S3 ([https://figshare.com/articles/dataset/Discerning\\_role\\_of\\_a\\_functional\\_arsenic\\_resistance\\_cassette\\_in\\_evolution\\_and\\_adaptation\\_of\\_a\\_rice\\_pathogen/14501055](https://figshare.com/articles/dataset/Discerning_role_of_a_functional_arsenic_resistance_cassette_in_evolution_and_adaptation_of_a_rice_pathogen/14501055)).
- (2) Data file (.pdf) Figs S1–S4 ([https://figshare.com/articles/figure/Discerning\\_role\\_of\\_a\\_functional\\_arsenic\\_resistance\\_cassette\\_in\\_evolution\\_and\\_adaptation\\_of\\_a\\_rice\\_pathogen/14501121](https://figshare.com/articles/figure/Discerning_role_of_a_functional_arsenic_resistance_cassette_in_evolution_and_adaptation_of_a_rice_pathogen/14501121)).

[resistance\\_cassette\\_in\\_evolution\\_and\\_adaptation\\_of\\_a\\_rice\\_pathogen/14501121](https://figshare.com/articles/figure/Discerning_role_of_a_functional_arsenic_resistance_cassette_in_evolution_and_adaptation_of_a_rice_pathogen/14501121)).

## INTRODUCTION

Arsenic is a highly toxic metalloid that occurs in the natural environment either through natural processes (volcanic eruption, mineral activities) or by anthropogenic activities (use of pesticides, fertilizers, industrial activities and agricultural practices) [1, 2]. Arsenic exists in different forms in the environment (organic and inorganic), but arsenate [As(V)] and arsenite [As(III)] are two predominant inorganic forms that are toxic to cells. As(III) is more toxic than As(V) and mostly predominates under anoxic and reduced environments, whereas arsenate predominates in soil and oxygenated surface water [3]. Accordingly, bacteria have evolved different mechanisms for resistance and detoxification,

Received 16 March 2021; Accepted 11 May 2021; Published 13 July 2021

**Author affiliations:** <sup>1</sup>Bacterial Genomics and Evolution Laboratory, CSIR-Institute of Microbial Technology, Chandigarh, India.

**\*Correspondence:** Prabhu B. Patil, [pbpatil@imtech.res.in](mailto:pbpatil@imtech.res.in)

**Keywords:** *ars* cassette; arsenic; heavy-metal resistance; stress response; transcriptome; *Xanthomonas oryzae* pv. *oryzae*.

**Abbreviations:** COG, cluster of orthologous groups; CSIR, Council of Scientific and Industrial Research; DEG, differentially expressed gene; GEI, genomic island; Grx, glutaredoxin; Hsp, heat shock protein; NCBI, National Center for Biotechnology Information; qPCR, quantitative real-time PCR; qRT-PCR, quantitative reverse-transcriptase PCR; ROS, reactive oxygen species; Trx, thioredoxin.

The GenBank/EMBL/DDBJ accession number for the complete genome sequence of *X. oryzae* pv. *oryzae* BXO512 is CP065228. The transcriptome data for *X. oryzae* pv. *oryzae* strain IXO1088 have been deposited at the NCBI Gene Expression Omnibus under accession number GSE163417.

**Data statement:** All supporting data, code and protocols have been provided within the article or through supplementary data files. Four supplementary figures and three supplementary tables are available with the online version of this article.

000608 © 2021 The Authors



This is an open-access article distributed under the terms of the Creative Commons Attribution NonCommercial License.

including As reduction, As oxidation, extrusion of As(III) out of cells mediated by efflux transporters, As methylation, etc. [4–7]. These resistance mechanisms in the bacterial population are encoded by the *ars* operons that include *arsRBC*, *arsRABC* and *arsRDABC* [6, 8]. In some cases, *ars* genes also exist in a single copy. These operons are either chromosomally encoded or plasmid encoded, the latter often acquired through horizontal gene transfer events [8]. The major genes in the operon include the *arsR* gene that encodes a transcriptional regulator, *arsC* that encodes arsenate reductase which transforms arsenate to arsenite, *arsB* that encodes a membrane-bound arsenite carrier protein ArsB which causes efflux of arsenite out of the cell, and *arsA* that encodes an ATPase subunit which forms a complex with ArsB to form an arsenite efflux pump [4, 6]. The complex operons also include *arsD*, which encodes a metallo-chaperone to deliver arsenite to ArsA. Other than these, recently characterized *ars* genes include *arsH*, *arsM*, *arsP* and *arsI*, which broaden resistance towards organic arsenicals [9, 10]. The *arsH* gene encodes an organoarsenical oxidase enzyme that confers resistance to methyl As(III) [11]. The *arsM* gene encoding ArsM (arsenite S-adenosylmethionine methyltransferase) methylates arsenite forming trimethyl arsenite [12].

Arsenic contamination from agricultural sources is a growing concern. In countries like India, Bangladesh, Argentina and China, As-contaminated groundwater used for irrigation adds to the accumulation of As in soil and water, and uptake of As by crop plants [13–16]. Accumulation of As in rice has become a great disaster, as rice is a staple food for half of the world population [17–20]. Unlike in other crops, the anaerobic conditions prevailing in the paddy field make the environment suitable for As accumulation [21, 22]. Usually, plant species have successfully adapted different mechanisms to cope with the toxic metal's effects, such as detoxification, elimination and heavy-metal accumulation. Studies have elucidated the mechanisms of As uptake and its translocation to different parts of rice plants, including roots, shoots, grains and seeds [20, 23]. Besides affecting the plants' systems, arsenic present in the environment can also positively or negatively affect the bacterial population that grows in association with plants [24]. There is a response to environmental stress specifically posed by toxin pressure and selection for bacteria to evolve under high selection pressure or acquire resistance against toxic substances [25]. Some of the bacteria that co-evolve with plants or exposure to heavy metals have acquired resistance mechanisms to cope with metal toxicity [26, 27]. These resistance mechanisms help bacteria to not only prevent metal toxicity, but also provide cross-protection against plant defence responses. Studies have shown that arsenite exposure can induce oxidative stress in bacteria, leading to the expression of stress-inducible genes [28]. Induction of such responses can provide a survival advantage to the bacteria during plant–pathogen interactions. Various antioxidant enzymes that are induced on exposure to arsenite can provide cross-protection against reactive oxygen species (ROS), which are generated by plants as an initial defence response against invading pathogens [29, 30].

### Impact Statement

Arsenic accumulation in rice is a serious and unique agronomic issue. Arsenic-contaminated groundwater used for irrigation purposes is adding to the accumulation of arsenic in rice. Submerged conditions in the paddy fields further induce the prevalence of toxic inorganic arsenic species in the environment. Our genomics- and transcriptomics-based study reveals how a rice pathogen is coping with the lethal concentrations of arsenic by acquiring a novel resistance cassette during diversification into lineages. Presence and selection of arsenic operons in a lineage suggest that a rice pathogen is exposed to high-arsenic conditions, probably outside the rice plant. Our study highlights the importance of genome-based studies on the role of environmental conditions in the evolution of host-adapted pathogens.

In the present study, we demonstrate the arsenic-resistance mechanism in *Xanthomonas oryzae* pv. *oryzae*, which causes bacterial blight disease in rice plants and is a serious threat to rice cultivation. As *X. oryzae* pv. *oryzae* is a pathogen of rice, it is likely that the *X. oryzae* pv. *oryzae* population is exposed to an arsenic environment (through either paddy field water or soil) and is under selection for adaptation to a higher arsenic concentration, which is particularly found in the submerged rice cultivation conditions. Inspection of the genomic sequence of *X. oryzae* pv. *oryzae* strain IXO1088 revealed the presence of a 12 kb *ars* cassette acquired through horizontal gene transfer events and expanded in a particular lineage of the *X. oryzae* pv. *oryzae* population. The cassette confers high arsenic resistance to *X. oryzae* pv. *oryzae* strain IXO1088 carrying the *ars* cassette, whereas a *X. oryzae* pv. *oryzae* strain BXO512 lacking this cassette was sensitive to arsenic ions. Cloning and transfer of the *ars* cassette rescued the sensitive strain BXO512 from a lethal concentration of arsenic. The transcriptional analysis demonstrated that the expression of *ars* operon is induced by arsenate and arsenite, and is regulated by *arsR*. Furthermore, to investigate the genome-scale transcriptional response of *X. oryzae* pv. *oryzae* to arsenic stress, we performed RNA sequencing under both As(V) and As(III) exposure. Our results provided the set of differentially expressed genes (DEGs) under both conditions, highlighting the mechanisms employed by *X. oryzae* pv. *oryzae* in response to arsenic treatment. Overall, our study shows how arsenic resistance mediated by horizontal gene transfer of a novel arsenic-resistance cassette has shaped the population structure and lineage of a rice pathogen.

## METHODS

### Bacterial growth and arsenic-sensitivity assay

*X. oryzae* pv. *oryzae* strains IXO1088 and BXO512 were used in the present study to represent arsenic-resistant and -sensitive strains, respectively. A list of strains, vectors and primers

used in the study is given in Table S3. The complete genome sequence of IXO1088 is available from the National Center for Biotechnology Information (NCBI) GenBank under accession number CP040687, and the complete genome of BXO512 was sequenced using Nanopore sequencing technology, as described in our previous study [31]. The BXO512 genome was submitted to the NCBI GenBank with accession number CP065228. For the sensitivity assay, strains were grown in nutrient broth (NB) media at 28 °C at 180 r.p.m. Cultures were diluted to an optical density at 600 nm of 1.0. One per cent of the culture was added to fresh 10 ml NB aliquots containing increasing concentrations of sodium arsenate ( $\text{Na}_2\text{HAsO}_4 \cdot 7\text{H}_2\text{O}$ ) or sodium arsenite ( $\text{NaAsO}_2$ ) (purchased from Himedia) added at the zero time point. Then, cultures were incubated for 48 h, and after that growth was measured at  $\text{OD}_{600}$ . Cultures without arsenate/arsenite treatment were used as controls. For the growth curve assay, cultures were treated with given concentrations of arsenate and arsenite, and incubated for 78 h. Samples were taken at regular intervals (6 h), and growth was measured as  $\text{OD}_{600}$  values.

### Sequence characterization and distribution of a novel *ars* cassette in the *X. oryzae* pv. *oryzae* population

The genomic islands (GEIs) were identified with GIPSy (Genomic Island Prediction Software) (v1.1.3) [32]. *X. oryzae* pv. *oryzae* BXO512 was used as a reference genome. Putative promoter sequences were identified using BPROM software (<http://www.softberry.com/berry.phtml?topic=bprom&group=programs&subgroup=gfindb>). Protein BLAST (basic local alignment search tool) was used to find out the homologues of the ArsR protein. Homologue proteins obtained from different species were then used for multiple sequence alignment using ClustalW (<https://www.genome.jp/tools-bin/clustalw>). Phylogenetic analysis was performed using MEGA6 software by the neighbour-joining method [33]. A core-genome tree of all *X. oryzae* pv. *oryzae* genomes available at the NCBI was reconstructed using PhyML [34]. Briefly, a core-genome alignment was obtained using Roary v3.11.2 [35] and converted to Phylip format using SeaView v4.4.2-1 [36]. Then, PhyML was used to obtain a Newick tree file, which was visualized using iTOL [37]. The presence of the *ars* cassette in other *X. oryzae* pv. *oryzae* was scanned by taking the cassette sequence as a query and performing nucleotide BLAST against all the *X. oryzae* pv. *oryzae* genomes available in the NCBI database.

### RNA sequencing and data analysis

Sublethal concentrations of arsenate (0.5 and 1 mM) and arsenite (0.6 mM) were selected by growing cultures in the different concentrations and measuring growth at regular time intervals. Strains were grown in NB media at 28 °C at 180 r.p.m. to an  $\text{OD}_{600}$  of 1.0. One per cent of the aliquot was added to fresh media and grown up to  $\text{OD}_{600}$  0.4; then, cultures were treated with different concentrations of arsenate (0.5 and 1 mM) or arsenite (0.6 mM) for 30 min. Culture without treatment was used as a control. The cells

were harvested by centrifugation at 6000 g for 5 min for RNA isolation. Total RNA was isolated using a Direct-zol RNA miniprep kit (Zymo Research), and samples were shipped to Agrigenome Labs (India). Briefly, the purity of isolated RNA was determined using a NanoDrop spectrophotometer (ThermoFisher Scientific) and quantified using the Qubit4 system (Invitrogen). RNA integrity was checked using an Agilent 2200 Tape Station system (Agilent Technologies). The rRNA was removed using si-TOOLS pan-prokaryote riboPOOL probes, and further library preparation was done using a TruSeq stranded RNA library prep kit. Prepared libraries were sequenced using the Illumina HiSeq X Ten platform, and 150 bp paired-end reads were generated. Clean reads that passed the quality filter were mapped to the reference genome (*X. oryzae* pv. *oryzae* IXO1088, accession no. CP040687) using EDGE pro (version 1.3.1). Differential-gene-expression analysis of treated samples with respect to the control was performed using the DESeq program [38]. DEGs with adjusted *P* values < 0.05 and  $\log_2$  fold change values  $\geq 1$  or  $\leq -1$  were used for further analysis. Common and unique DEGs were visualized by Venn diagram and volcano plots using the R software package. Heat maps were generated using  $\log_2$  fold change values of treated versus control samples using GENE-E software. DEGs with  $\log_2$  fold change values  $\geq 1$  or  $\leq -1$  were classified into different COG (cluster of orthologous groups) classes using the eggNOG tool [39].

### Validation of RNA-Seq data by quantitative real-time PCR (qPCR)

To check the induction of the *ars* cassette and to validate RNA-Seq data, qPCR experiments were performed. The primers used are listed in Table S3. The qPCR assay was performed with a SuperScript III Platinum SYBR Green one-step qRT-PCR kit (ThermoFisher Scientific). Three technical replicates were included for each sample, and reactions were set up according to the manufacturer's instructions. The amplification conditions were: cDNA synthesis at 50 °C for 45 min; initial denaturation at 95 °C for 5 min; 40 cycles of denaturation at 95 °C for 15 s, followed by annealing at 60 °C for 30 s and extension at 40 °C for 30 s. *ftsZ* was used as an endogenous control. Each gene's relative expression in the treated and untreated sample was expressed as a fold change calculated using the  $2^{-\Delta\Delta\text{Ct}}$  method. Three biological replicates were included for each condition and used for statistical analysis. Student's paired *t*-test was performed to calculate *P* values. *P* < 0.05 was considered statistically significant.

### Cloning of the *ars* cassette

*X. oryzae* pv. *oryzae* strain IXO1088 genomic DNA was isolated from an overnight grown culture using a Zymo kit. The complete *ars* gene cassette was amplified using primers listed in Table S3. PCR conditions were as follows: initial denaturation 95 °C for 30 s; 35 cycles of denaturation 95 °C for 10 s, annealing 60 °C for 30 s, extension 72 °C for 3 min; and final extension 72 °C for 10 min. The amplified fragment was gel extracted using the QIAquick gel extraction kit, given kinase treatment, purified and ligated into the pUFR034 vector. The

vector was linearized using inverse M13 primers. The ligation mix was used to transform *Escherichia coli* Top10 cells. The recombinant plasmid was verified using M13 forward and reverse universal primers and insert-specific primers. For transformation of *E. coli* cells, ligation mix was added to 100 µl competent cells and incubated on ice for 15 min. Then, the competent cells were subjected to heat shock at 42 °C for 60 s. Cells were immediately shifted to ice for 2 min. Then, 1 ml LB media was added, and cells were incubated at 37 °C at 200 r.p.m. for 1 h. After incubation, cells were plated on LB agar plates containing kanamycin (50 µg ml<sup>-1</sup>). Plates were incubated at 37 °C for 12 h for the growth of recombinants.

### Preparation of electrocompetent cells and transformation of *X. oryzae* pv. *oryzae*

For the preparation of electrocompetent cells, *X. oryzae* pv. *oryzae* strain BXO512 was inoculated in 6 ml NB media and cultured until late exponential phase (OD=0.8–1.0) at 28 °C with 180 r.p.m. shaking. The culture was then equally divided into four centrifuge tubes, and cells were harvested by centrifugation at 16000 g for 2 min at room temperature. The pellets in each tube were then washed with 1 ml of 300 mM sucrose. The washing step was repeated three times at the same speed. Finally, the cells in the four tubes were resuspended with a total of 100 µl of 300 mM sucrose. A 100 µl aliquot of the competent cells was used for one electroporation experiment.

For electro-transformation, 100 µl electrocompetent cells was mixed with 10 µl (500 ng–1 µg) plasmid DNA and transferred to 2 mm ice-cold electroporation cuvettes [Gene Pulser cuvettes (BioRad)] and tapped gently to remove air bubbles. Cells without plasmid were used as a negative control. Furthermore, electrocompetent cells were subjected to electroporation at 2.5 kV, 25 µF and 200 Ω using the Gene Pulser Xcell electroporation system (BioRad). After the electric shock, 1 ml NB broth was added immediately to the cuvette, mixed properly and transferred to 5 ml micro-centrifuge tubes. The electroporated cells were incubated at 28 °C at 180 r.p.m. for 6 h. After recovery, cells were plated on nutrient agar plates containing kanamycin antibiotic (15 µg ml<sup>-1</sup>) and incubated at 28 °C until colonies appeared (usually 72 h).

## RESULTS

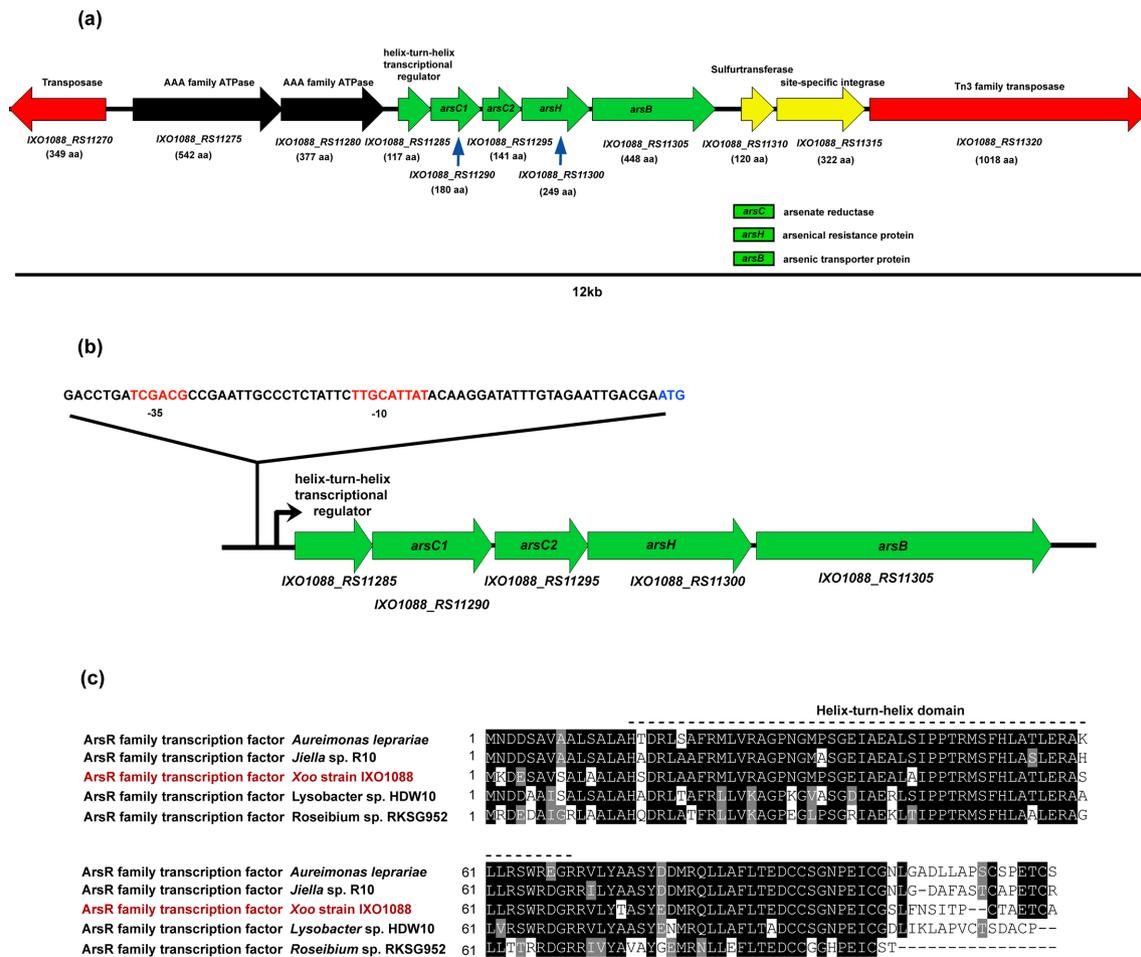
### *X. oryzae* pv. *oryzae* IXO1088 harbours a novel arsenic-resistance cassette

In the present study, we identified a unique 12214 bp *ars* cassette in the *X. oryzae* pv. *oryzae* strain IXO1088 genome related to arsenic resistance (Fig. 1a). Genes in the cassette include a transcriptional regulator, two *arsC* genes, which encode arsenate reductase, the *arsH* gene that encodes arsenical-resistance protein, an arsenic transporter, two AAA family ATPase, a sulfurtransferase and a site-specific integrase, and the cassette is flanked by transposases on both sides. Investigation of upstream and downstream regions of this cassette revealed that it is part of a GEI. The identification of GEIs in the IXO1088 genome was performed with GIPSY based on the criteria: deviation in nucleotide composition

(anomalous G+C and codon usage), presence of transposase and virulence or flanking tRNA genes (Table S1). One GEI (IXO1088\_RS11210–IXO1088\_RS11325) was predicted as harbouring both pathogenicity- and metabolic-related factors and comprises our identified *ars* cassette. The features of GEIs are listed in Table S1. The presence of this island probably suggests the acquisition of this *ars* operon through horizontal gene transfer events. Then, we investigated the cassette's G+C content, and the presence of genes with atypical G+C content [IXO1088\_RS11270 (66%), IXO1088\_RS11275 (59.6%), IXO1088\_RS11310 (60.2%), IXO1088\_RS11315 (66.1%) and IXO1088\_RS11320 (60.9%)] that differ from the mean G+C content of the IXO1088 genome 63.7% (±2.5%) further suggested the acquisition of this cassette from other organisms (Table S1). A more detailed analysis of the cassette revealed a putative promoter region recognized by a σ70 transcription factor in the 5' upstream of the helix-turn-helix transcriptional regulator (Fig. 1b). The other *ars* genes in the cassette, *arsC*, *arsH* and the gene encoding the arsenic transporter, do not contain promoter regions as predicted by BPROM tool. This gene organization suggested that the *ars* cassette constitutes an operon, whose expression is regulated by a helix-turn-helix transcriptional regulator. The transcriptional regulator present in the cassette showed homology to the ArsR/Smt family of transcriptional regulators. Sequence alignment of the regulator with ArsR family proteins from other bacteria indicates conservation of the helix-turn-helix domain (Fig. 1c). This suggested that the transcriptional regulator in the cassette belongs to ArsR family regulators and its regulation is by arsenic ions.

### Phylogenetic relatedness of arsenate reductase and arsenic transporter encoding genes

To identify the related neighbours of arsenate reductase encoding genes (IXO1088\_11290 and IXO1088\_11295) and the arsenite transporter encoding gene (IXO1088\_11305), we performed phylogenetic analysis of these proteins with the ArsC and ArsB proteins from various arsenic-resistant micro-organisms, respectively. The phylogenetic tree of ArsC protein sequences revealed that the two ArsC proteins encoded in the cassette fall into two groups, one ArsC1 protein (IXO1088\_11290) falls in the ArsC arsenate reductases of the thioredoxin (Trx) family, whereas the other ArsC2 protein (IXO1088\_11295) grouped with the arsenate reductases of the glutaredoxin (Grx) family (Fig. 2a). Usually, the primitive *arsRBC* operon comprises an ArsB arsenite efflux pump that extrudes arsenite from the cells. Later, it was found that *Bacillus subtilis* contained a novel arsenite transporter that showed homology to the Acr3 protein from yeast *Saccharomyces cerevisiae*, which confers arsenic resistance. The *acr3* gene was subsequently found in many micro-organisms such as *Rhodospseudomonas palustris*, *Campylobacter jejuni*, *Herminiimonas arsenicoxydans*, *Microbacterium* sp., etc. [8]. The phylogenetic analysis of the arsenic transporter (IXO1088\_11305) present in the IXO1088 cassette revealed that it is closely related to ArsB family transporters and all Acr3 transporters fall in the other clade (Fig. 2b).



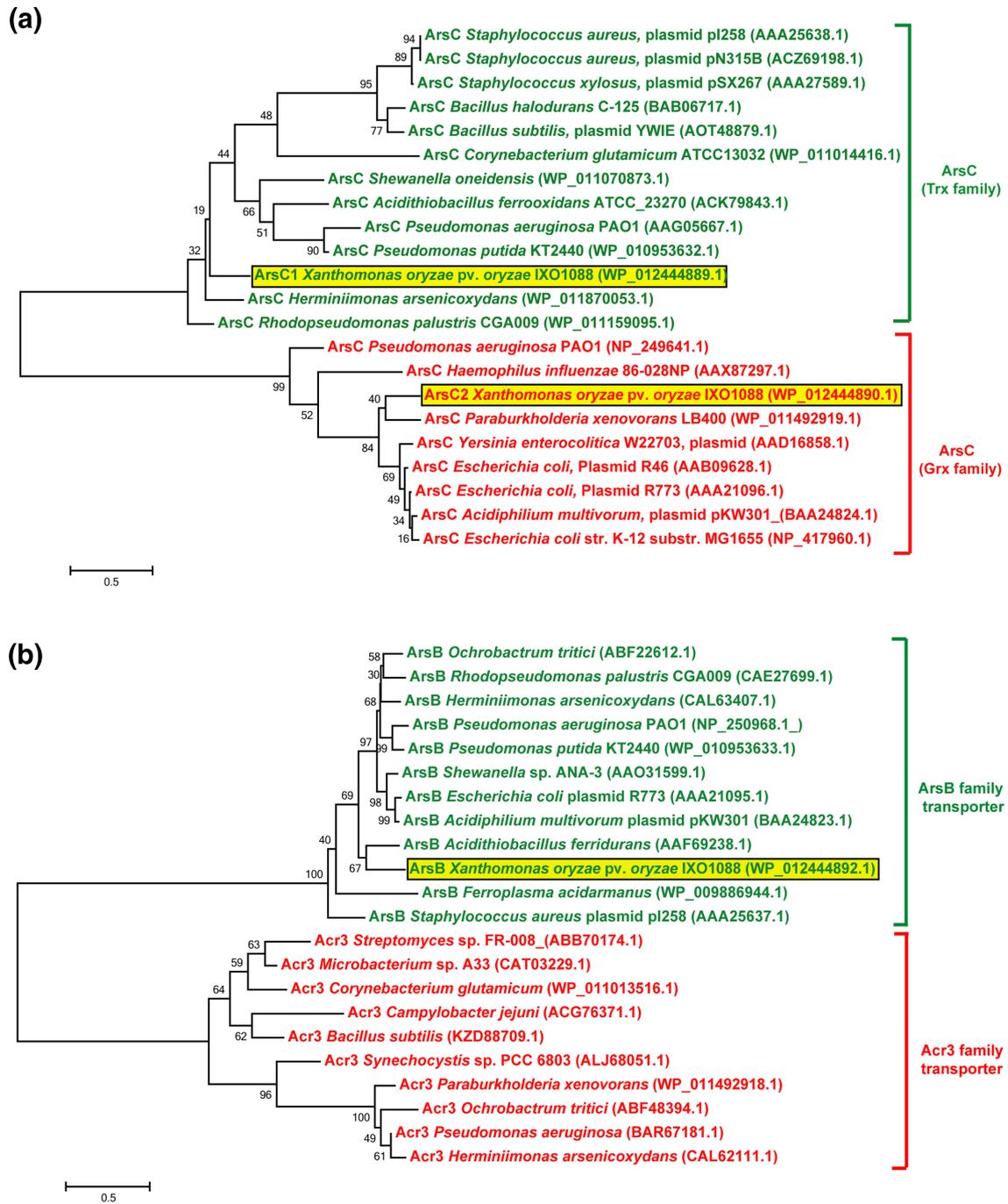
**Fig. 1.** Genetic organization of the *ars* operon present in *X. oryzae* pv. *oryzae* strain IXO1088. (a) The arrows represent genes present in the cassette and the direction of the transcription. Below the arrows, the locus tags are given for each gene along with the size (number of amino acid residues). (b) The putative promoter sequence upstream of the *arsR* gene with the predicted conserved regions -35 and -10 highlighted in red, and the initiation codon in blue. (c) Multiple sequence alignment of the ArsR family regulator of IXO1088 with the ArsR family transcription factor homologues from other bacterial species. The dashed line represents the helix-turn-helix domain. The black background indicates identical amino acid residues. *Xoo*, *X. oryzae* pv. *oryzae*.

## Distribution of the *ars* cassette in the *X. oryzae* pv. *oryzae* population

Core-genome phylogenetic analysis of 413 *X. oryzae* pv. *oryzae* genomes available in the NCBI database from different geographical regions was performed, and then we looked for the presence of the *ars* cassette in all the genomes. Phylogenetic analysis revealed six distinct lineages of the *X. oryzae* pv. *oryzae* population. Interestingly, we observed that the cassette was restricted to a particular lineage comprising 55 isolates (Fig. 3). The lineage predominantly constitutes strains from Asia, including 22 strains from India, 21 from China, 9 from Philippines and 1 from Nepal. One strain each from Colombia and Bolivia was also present. Furthermore, we observed that in India, the strains carrying the *ars* cluster were isolated from areas where arsenic contamination in groundwater or irrigation water is a problem (Fig. S1).

## Regulation and functional characterization of the arsenic-resistance gene cassette present in *X. oryzae* pv. *oryzae* IXO1088

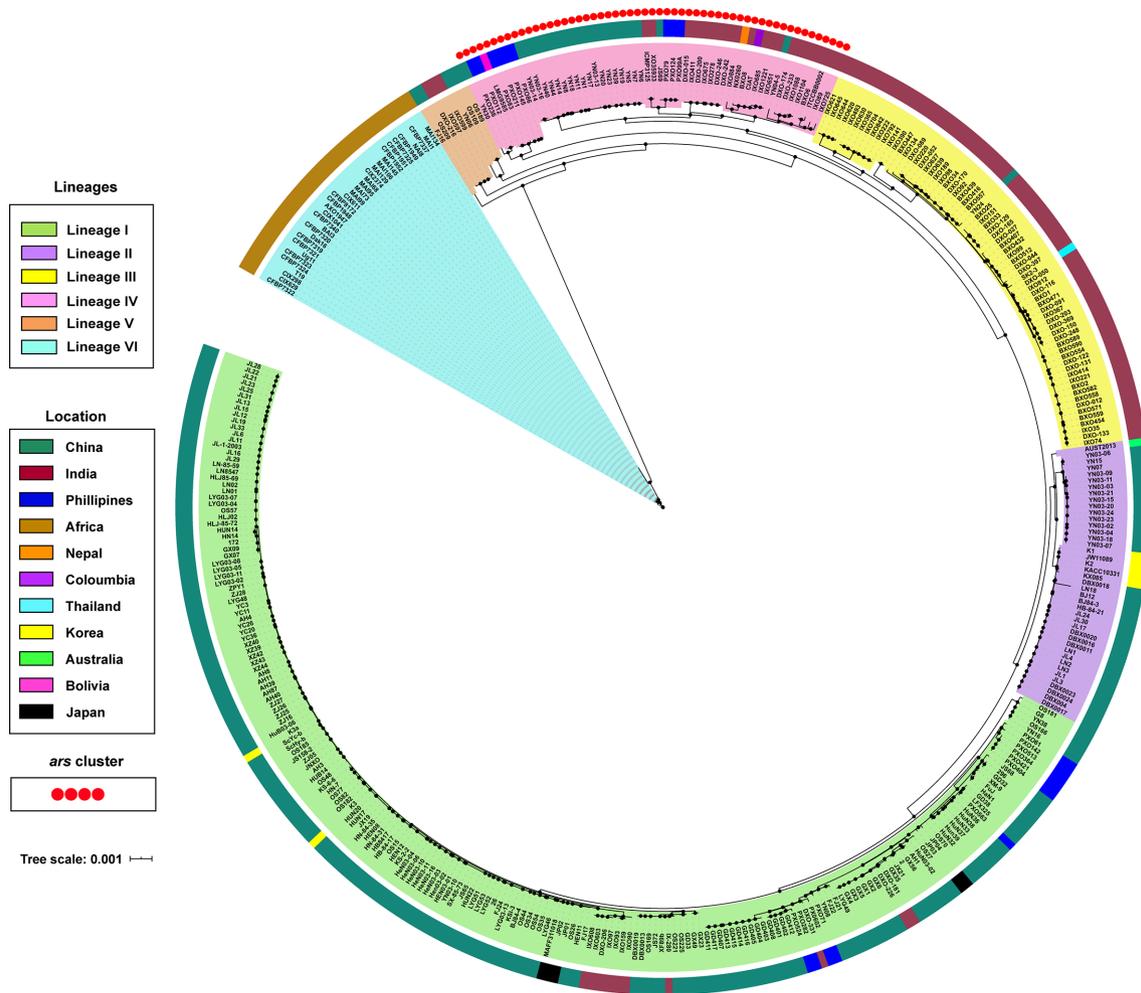
Arsenic sensitivity assay showed that *X. oryzae* pv. *oryzae* IXO1088 could tolerate significantly higher concentrations of arsenate and arsenite compared to another *X. oryzae* pv. *oryzae*, BXO512, that lacks the *ars* cassette. Significant growth inhibition of the IXO1088 strain with arsenic ions was only observed above 12 mM arsenate and 0.6 mM arsenite (Fig. 4a, b). Compared to the IXO1088 strain, the BXO512 strain could not grow even in the presence of 0.1 mM arsenite (Fig. 4b). However, the BXO512 strain tolerated a slight concentration of arsenate, but significant inhibition was observed even at the 0.5 mM concentration (Fig. 4a). To check whether the arsenic-resistance phenotype is directly linked with the *ars* cassette, we cloned the complete *ars* gene cluster in the pUFR034 cosmid



**Fig. 2.** Phylogenetic analysis of arsenate reductase (ArsC) and arsenic transporter proteins encoded in the *ars* cassette of *X. oryzae* pv. *oryzae* strain IXO1088. Phylogenetic trees of (a) ArsC protein and (b) the arsenic transporter protein from IXO1088 with other bacterial species reconstructed using MEGA6 software. The *X. oryzae* pv. *oryzae* IXO1088 ArsC and ArsB proteins are highlighted in yellow boxes, and accession numbers for all sequences are given in parentheses. Bootstrap values are shown at each branch point. The scale bar indicates number of substitutions per site.

vector, transferred recombinant plasmid (pUFR034+*ars*) to the BXO512 strain and performed an arsenic-sensitivity assay. As expected, the BXO512 strain with this cassette was resistant to both arsenate and arsenite as compared to the strain with only empty vector (pUFR034) (Fig. 4c, d). The

plasmid presence not only increased arsenate resistance, but also conferred an increased level of resistance against arsenite. The data indicated that the *X. oryzae* pv. *oryzae* IXO1088 *ars* cassette is functional when transferred to the other *X. oryzae* pv. *oryzae* strain.



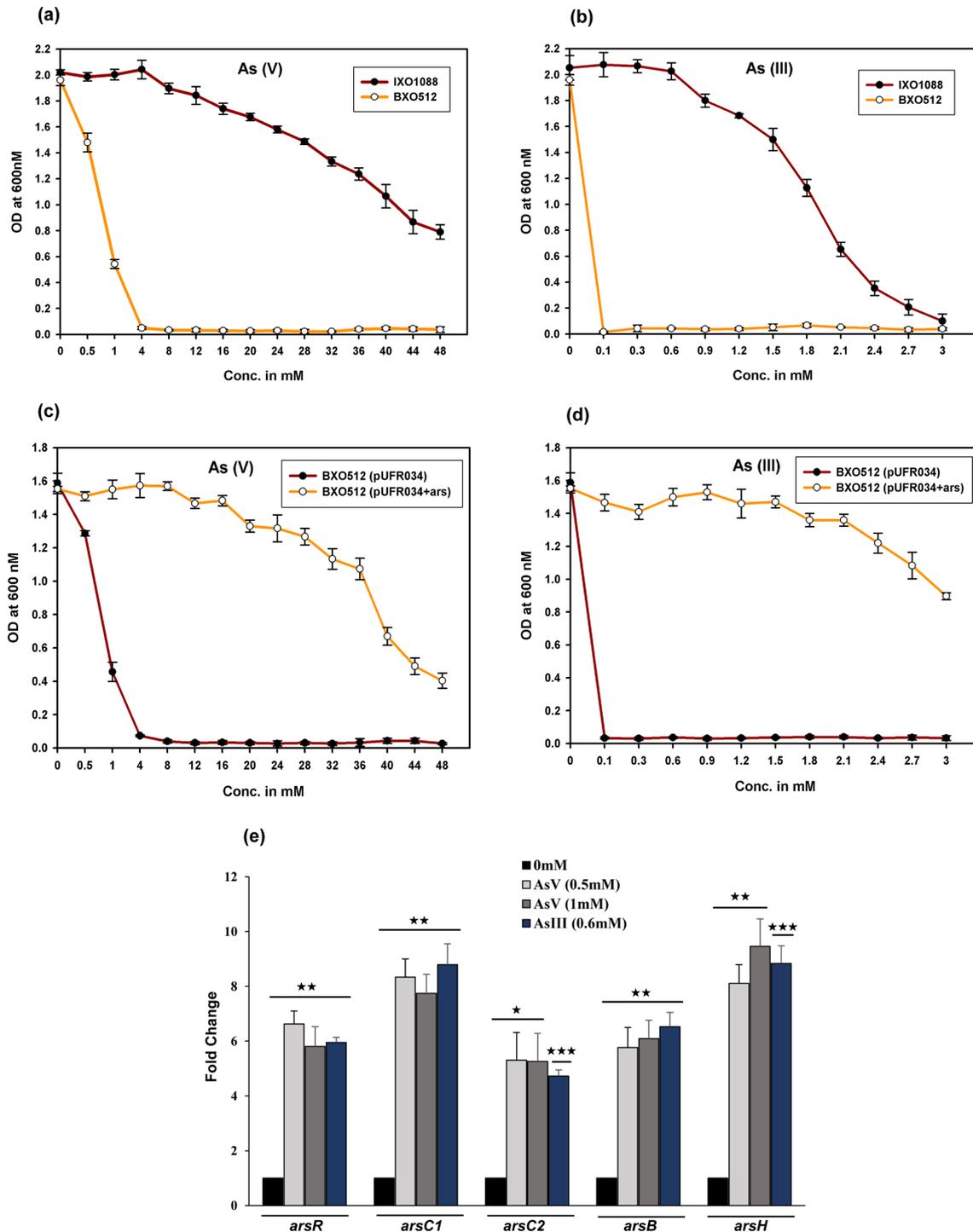
**Fig. 3.** Distribution of the *ars* cassette among the *X. oryzae* pv. *oryzae* population. The core-genome tree was reconstructed using PhyML and visualized using iTOL software. Background colours of nodes represent different lineages. The innermost circle represents different countries from where strains were isolated, and the outermost circle (red circles) refers to strains in which the *ars* cassette is present. The scale bar indicates number of substitutions per site.

To determine whether the arsenic ions regulate the expression of the *ars* cassette, quantitative reverse-transcriptase PCR (qRT-PCR) was used to compare the transcription of the *arsR*, *arsC1*, *arsC2*, *arsH* and *arsB* genes in *X. oryzae* pv. *oryzae* IXO1088 grown in the presence and absence of subinhibitory concentrations of arsenate (0.5 and 1 mM) and arsenite (0.6 mM) (as described in Methods). The subinhibitory concentration was decided by performing a growth curve assay by growing cells in the presence of different concentrations of arsenate and arsenite (Fig. S2). The transcription of the *ars* operon was regulated in the presence of both arsenate and arsenite. The expression of *arsR*, *arsC1*, *arsC2*, *arsH* and *arsB* genes was induced sixfold to eightfold in arsenate- and arsenite-treated samples compared to the control cells without any treatment (Fig. 4e). The data suggested that the presence of arsenic ions induces

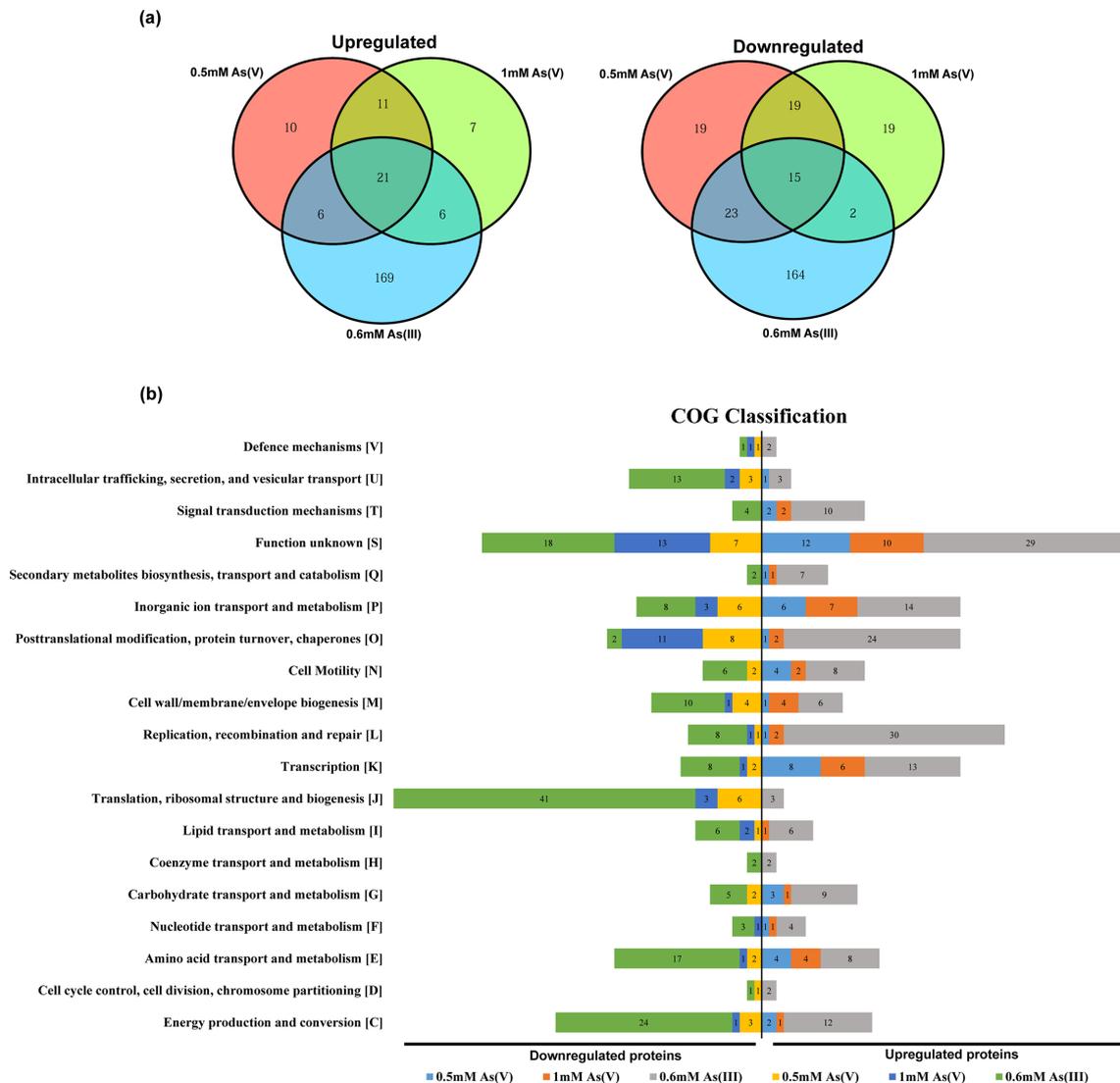
the *ars* operon. The expression level for the other two genes encoding the AAA family ATPase remained unchanged (data not shown).

### Transcriptomics of *X. oryzae* pv. *oryzae* IXO1088 under arsenic stress

To study the defence mechanism employed by *X. oryzae* pv. *oryzae* in response to arsenic treatment, we carried out RNA sequencing at three concentrations, two for arsenate (0.5 and 1 mM) and one (0.6 mM) for arsenite, as described in Methods. On average, 30 million reads were obtained for each sample, and more than 98% of the reads were mapped to the reference genome. We further analysed the DEGs. Genes with an adjusted *P* value < 0.05 were considered significant, genes with  $\log_2$  fold change  $\geq 1$  were considered upregulated, and those with  $\log_2$  fold change  $\leq -1$  were considered



**Fig. 4.** Arsenic sensitivity assay of *X. oryzae* pv. *oryzae* strain IXO1088 (contains the *ars* cassette) and strain BXO512 (does not contain the *ars* cassette). Cultures were grown in NB with different concentrations of (a) sodium arsenate and (b) sodium arsenite. The panels underneath show the susceptibility of strain BXO512 containing the IXO1088 *ars* cassette against (c) sodium arsenate and (d) sodium arsenite: wild-type BXO512 (pUFR034) and BXO512 (pUFR034+*ars* cassette). Data values shown are means with standard deviation from three independent experiments. (e) The *ars* operon is induced in the presence of sodium arsenate and sodium arsenite. Transcription of *arsR*, *arsC1*, *arsC2*, *arsB* and *arsH* genes was determined using qRT-PCR in cells grown in the presence of arsenate (0.5 and 1 mM) and arsenite (0.6 mM) during the exponential phase. Relative expression was calculated using the  $2^{-\Delta\Delta Ct}$  method. *ftsZ* was used as an endogenous control. Error bars represent the standard deviation from three independent biological replicates (each performed in three technical replicates) (\* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ ).



**Fig. 5.** (a) Venn diagrams showing DEGs upregulated or downregulated under arsenate As(V) and arsenite As(III) treatment. (b) COG classification of DEGs upregulated or downregulated under arsenate As(V) and arsenite As(III) treatment. The numbers within the bars represent the number of proteins present in that particular category.

downregulated (Fig. S3). The complete list of DEGs under all three conditions is shown in Table S2. Further analysis of DEGs revealed that a total of 21 upregulated and 15 downregulated genes were shared by all three conditions. As As(III) is highly toxic compared to As(V), the number of DEGs in the As(III)-treated samples was more, with 202 upregulated and 204 downregulated genes. Whereas, in the 0.5 mM As(V)-treated samples, 48 upregulated and 76 downregulated genes were found, and in the sample treated with 1 mM As(V), 45 upregulated and 55 downregulated genes were found (Fig. 5a). This shows that 0.5 and 1 mM As(V) concentrations could be less stressful conditions for the cells than As(III).

The DEGs were further classified into different COGs. A total of 552 genes were classified into 19 COG categories (Fig. 5b). The COG analysis revealed that the maximum number of

DEGs fall in the function unknown category (89; 16.1%); followed by other categories including translation, ribosomal structure and biogenesis (53; 9.6%); posttranslational modification, protein turnover, chaperones (48; 8.7%); inorganic ion transport and metabolism (44; 8.0%); energy production and conversion (43; 7.8%); replication, recombination and repair (43; 7.8%); transcription (38; 6.9%); amino acid transport and metabolism (36; 6.5%); cell wall/membrane/envelope biogenesis (26; 4.7%); intracellular trafficking, secretion and vesicular transport (22; 4.0%); cell motility (22; 4.0%); carbohydrate transport and metabolism (20; 3.6%); and signal transduction mechanisms (18; 3.3%). The other categories with the lowest numbers include nucleotide transport and metabolism, lipid transport and metabolism, secondary metabolites biosynthesis, transport and catabolism, defence

mechanisms, coenzyme transport and metabolism, cell cycle, which contributes (50; 9.0%) in total.

### Gene expression pattern common to both arsenate As(V) and arsenite As(III)

Significantly, upregulated genes induced by all three conditions include the *ars* operon encoding a helix-turn-helix transcriptional regulator (*RS\_11285*), arsenate reductase (*RS\_11290*), arsenate reductase (glutaredoxin) (*RS\_11295*), arsenical-resistance protein (*RS\_11300*) and arsenic transporter (*RS\_11305*) (Fig. 6a). The upregulation of the operon in the RNA-Seq experiment was consistent with our induction results. Additionally, other genes that were upregulated encode DNA binding protein (*RS\_14295*), Abr family transcriptional regulator (*RS\_14300*), MarR transcriptional regulator (*RS\_01475*), (2Fe-2S)-binding protein (*RS\_02860*), TonB-dependent siderophore receptor (*RS\_06840*), hemin transporter HemP (*RS\_19500*) and BON-domain-containing protein (*RS\_20250*) (Fig. 6a). In contrast, significantly downregulated genes encode bacterioferritin (*RS\_02865* and *RS\_09635*), potassium ATPase subunit F (*RS\_19820*), potassium ATPase subunit A (*RS\_19815*), RND-efflux transporter subunits (*RS\_15130* and *RS\_15135*) and methionine synthase (*RS\_10005*) (Fig. 6a).

### Gene expression pattern specific to arsenite As(III) exposure

More genes were differentially expressed in response to As(III) treatment, and their expression levels were different compared to after As(V) treatment (Fig. 6b, c). The genes regulated in response to As(III) treatment were grouped into different categories as below.

#### Oxidative stress proteins

Several genes encoding stress-related proteins and which are known to protect against oxidative damage were upregulated in response to As(III) treatment. These genes encode superoxide dismutase (*RS\_13380*), flavin-dependent oxidoreductase (*RS\_14885*), serine endopeptidase (*RS\_00295*), organic hydroperoxide resistance protein (*RS\_01470*), peptide methionine sulfoxide reductase MsrA and MsrB (*RS\_19100* and *RS\_03230*), alkyl hydroperoxide reductase subunit F (*RS\_19060*), oxidative stress transcriptional regulator (*RS\_19065*), nucleoside diphosphate kinase regulator (*RS\_19070*) and catalase (*RS\_22515*) (Fig. 6b, c).

#### Heat shock proteins (Hsps)

Hsps function as chaperones, and are considered as general markers of cell damage and stress conditions. Hsps facilitate proper folding of newly or already translated proteins and prevent aggregation of inappropriately folded proteins. We observed the upregulation of major Hsps under arsenic stress conditions. Many genes including those encoding heat-inducible transcriptional repressor HrcA (*RS\_09765*), chaperonin GroEL (*RS\_02265*), Hsp20/alpha\_crystallin\_family\_protein (*RS\_04330*), ATP-dependent Clp endopeptidase proteolytic subunit ClpP (*RS\_05105*), endopeptidase La

(*RS\_05115*), ATP-dependent chaperone ClpB (*RS\_06715*), Hsp33 family molecular chaperone HslO (*RS\_08700*), molecular chaperone DnaK (*RS\_09775*), molecular chaperone DnaJ (*RS\_09780*), molecular chaperone HtpG (*RS\_10250*), protease HtpX (*RS\_13455*), and ATP-dependent protease ATPase subunit HslU and HslV (*RS\_20455* and *RS\_20460*) were found to be upregulated (Fig. 6b). In addition, we observed downregulation of protein translocase subunits YajC, SecD, F, G and SecE involved in intracellular protein transport (*RS\_10440*, *RS\_10445*, *RS\_10450*, *RS\_16485* and *RS\_18785*) (Fig. 6b).

#### Iron acquisition, storage or production of siderophores

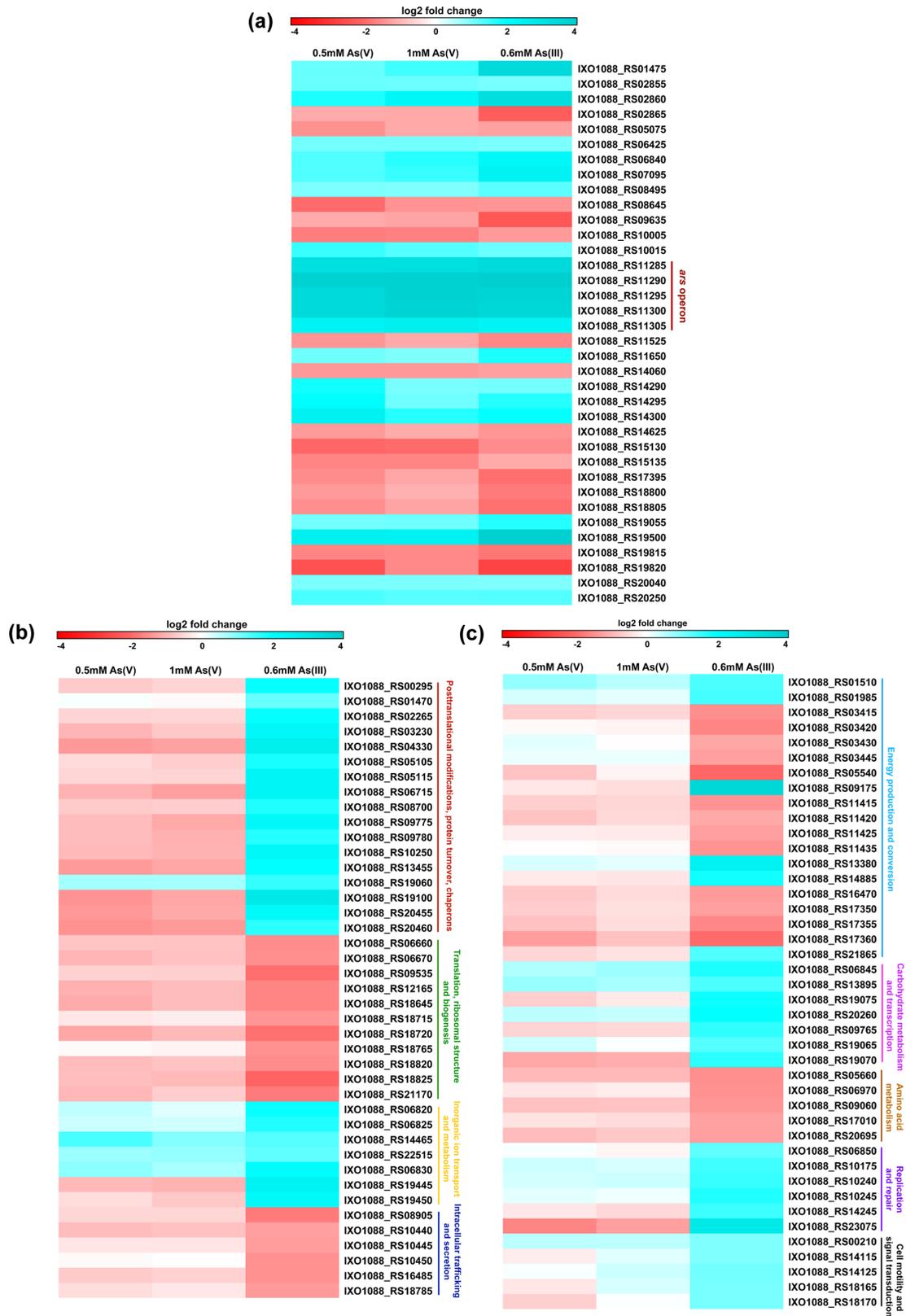
RNA-Seq results also revealed significant upregulation of genes related to Fe-S cassette biogenesis and iron recruitment to alleviate the arsenic stress. Genes related to iron storage were differentially expressed under both As(V)- and As(III)-treated conditions. Fe-S binding protein (*RS\_02860*), TonB-dependent siderophore receptor (*RS\_06840*), hemin uptake protein HemP (*RS\_19500*) and MarR family transcriptional regulator (*RS\_01475*) were significantly upregulated in all conditions. In contrast, iron transporter (*RS\_06820*), MFS (major facilitator superfamily) transporter (*RS\_06825*), IucA/IucC family siderophore biosynthesis protein (*RS\_06830*), and iron-sulfur cassette insertion protein ErpA (*RS\_21865*) were significantly upregulated in As(III)-treated samples (Fig. 6c, b).

#### Recombination and energy metabolism

Interestingly, genes involved in repair and recombination encoding various transposases such as IS3 family transposase (*RS\_06850*), IS 701 family transposase (*RS\_10175* and *RS\_10240*), IS630 family transposase (*RS\_10245* and *RS\_14245*) and bifunctional DNA-formamidopyrimidine glycosylase (*RS\_23075*) were found to be upregulated (Fig. 6c). Also, genes involved in the electron transfer pathway in oxidative phosphorylation including NADH:flavin oxidoreductase (*RS\_01985*), alkene reductase (*RS\_09175*), LLM class flavin dependent oxidoreductase (*RS\_14885*), ATP-binding cassette protein (*RS\_19445*) and FtsX-like permease family protein (*RS\_19450*) were found to be upregulated. In contrast, FOF1 ATP synthase subunits (*RS\_03415*, *RS\_03420*, *RS\_03430* and *RS\_03445*), and cytochrome ubiquinol oxidase subunits (*RS\_17350*, *RS\_17355* and *RS\_17360*) were found to be downregulated in response to arsenite (Fig. 6c). Other than this, genes involved in translation machinery encoding 50S ribosomal proteins (L36, L19, L10, L25) (*RS\_12165*, *RS\_06670*, *RS\_18765* and *RS\_18825*) and 30S ribosomal proteins (S8, S10, S21) (*RS\_18645*, *RS\_18720* and *RS\_21170*) were also found to be downregulated (Fig. 6b).

#### qRT-PCR validation

To confirm the RNA-Seq data reliability, the expression level of eight genes, which included four upregulated genes (*RS\_01475*, *RS\_02860*, *RS\_06840* and *RS\_19500*) and four



**Fig. 6.** Heat map visualization of upregulated and downregulated genes. (a) Common DEGs (upregulated and downregulated) in response to both arsenate and arsenite. (b, c) Heat map visualization of DEGs in response to arsenite. The scale represents the log<sub>2</sub> fold change value.

downregulated genes (*RS\_02865*, *RS\_19815*, *RS\_19820* and *RS\_15130*) was examined using qRT-PCR, as shown in Fig. S4. For all the genes, the same expression trend was detected for both qRT-PCR and RNA-Seq analyses. Upregulation of genes that constitute the *ars* operon in the qRT-PCR analysis (Fig. 4e) further confirmed the reproducibility and reliability of the RNA-Seq data.

## DISCUSSION

Heavy-metal contamination in the environment is a serious threat to human health. Metals such as arsenic, cadmium, chromium, lead and mercury are commonly occurring metals, and rank among the priority metals of public-health risk. Among these, arsenic contamination in drinking water and food crops, especially in rice, is a growing concern, particularly in countries like India, Bangladesh and China [18, 40, 41]. Unlike in any other crops, the water embedded conditions in the paddy fields make the environment suitable for the accumulation of inorganic forms of arsenic, specially arsenite As(III) that is highly toxic to cells [42, 43]. Studies have shown the transport and accumulation of arsenic to different parts of rice plants, including roots, shoots, straw and grains [20, 23, 44]. Rice is an efficient accumulator of arsenic compared to other crops, due to arsenite's mobilization and sequestration of arsenite through iron oxides under anaerobic conditions [45]. Rice is a staple food and human exposure to arsenic through rice intake is causing serious health issues. In the present study, we have elucidated an acquired arsenic-resistance mechanism of *X. oryzae* pv. *oryzae*, which is a serious and highly specific rice pathogen. The presence of metals in the environment can influence the survival of pathogens, as well as plants. Our study highlights the important and specific role of environmental conditions, such as those found in the submerged water conditions in paddy fields, can lead to high arsenic stress to both the host and the microbes associated with it. In this context, acquired resistance against high arsenic stress by acquiring novel genomic cassettes in an emerging lineage indicates ongoing selection pressure on this host-specific pathogen. The absence of the cassette in other lineages of *X. oryzae* pv. *oryzae* and at the same time in a lineage that is predominant to the Asian continent reiterates the unique selection pressure on the *X. oryzae* pv. *oryzae* population.

We observed the presence of a GEI comprising an *ars* cassette in the *X. oryzae* pv. *oryzae* IXO1088 genome, encoding genes involved in arsenic metabolism. The cassette includes an *ars* operon regulated by an ArsR family regulator, and is induced in the presence of both arsenate and arsenite. The cassette provides high-level resistance to *X. oryzae* pv. *oryzae* IXO1088 against both arsenate and arsenite. Phylogenetic analysis was carried out to check the relatedness of the arsenate reductase (ArsC) or arsenite transporter (ArsB) proteins present in the *ars* cassette with the already reported or characterized ArsC or ArsB proteins [8]. The ArsC proteins have been shown to use either Trx or Grx as an electron source. The *ars* cassette we found encodes two copies of arsenate reductase, ArsC1

and ArsC2. From the phylogenetic analysis, we observed that ArsC1 protein displayed a phylogenetic relationship to ArsC reductase from the Trx family and ArsC2 protein to ArsC reductase from the Grx family. The presence of two copies of arsenate-reductase-encoding genes in the operon seems to display differential capabilities to use different metabolic pathways for arsenic detoxification, further enhancing the efficiency of this process. Similarly, phylogenetic analysis of the arsenite transporter present in the cassette revealed that it belongs to ArsB family transporters. Furthermore, core-genome-based phylogenetic analysis showed grouping of *X. oryzae* pv. *oryzae* strains from different geographical locations into six distinct lineages. The *ars* cassette is present only in a subset of strains or a particular lineage in the *X. oryzae* pv. *oryzae* population. This may be because the GEI comprising the *ars* cassette was probably acquired only once by the ancestor of lineage IV and has persisted by vertical transmission. The fact that any extant members of this lineage have not lost this operon indicates the strong selective advantage it provides to the host. The cassette is not present in any other *Xanthomonas*, including those that infect xylem, that infects other plants, suggesting the importance of the operon outside the host environment. Rice cultivation is water intensive and rice is grown in flooded conditions. Hence, the acquisition of this cassette might be an adaptation to survive in paddy water fields. To better understand the adaptive response of the *X. oryzae* pv. *oryzae* population to arsenic exposure and the toxicity coping mechanisms, we performed a transcriptome-sequencing study. RNA-Seq is an efficient high-throughput sequencing method to study bacterial transcriptomes, and our transcriptome-based study of *X. oryzae* pv. *oryzae* under arsenic stress highlights major pathways differentially expressed in strain IXO1088 in response to arsenate and arsenite.

Our RNA-Seq analysis revealed that the gene expression patterns in *X. oryzae* pv. *oryzae* IXO1088 were different in both arsenate- and arsenite-treated samples. Commonly upregulated genes under arsenate and arsenite treatment include genes of the *ars* operon that encode proteins involved in arsenic detoxification. Other than this, MarR and Abr family transcriptional regulators, proteins involved in iron acquisition, transport and storage such as siderophores and bacterioferritin, were also upregulated, consistent with previous studies [46]. In arsenic conditions, more iron is recruited to alleviate stress. During non-flooded conditions, arsenic is incorporated in Fe(hydr)oxide in soil, whereas in the flooded period it is released from soil to water due to reductive dissolution of Fe(hydr)oxides [43]. Production of siderophores is the common strategy for iron acquisition under Fe-limiting conditions. Some metals also possess affinity towards siderophores. The complex formation of siderophores with metals reduces complexation with iron; thus, decreasing the concentration of soluble iron. Iron deficiency subsequently leads to the production of siderophores [47].

A relatively large number of genes were expressed in response to arsenite compared to arsenate treatment. Many DEGs specific to arsenite encoding proteins involved in the oxidative response, protein damage, ion transport, recombination and

energy generation, cell motility and signal transduction were found to be upregulated. Arsenic exposure induces ROS and nitric oxide (NO) production inside the cell and, to reduce the oxidative damage, bacteria induce various antioxidant enzymes [28, 48]. We observed upregulation of superoxide dismutase, catalase, organic hydroperoxide protein, alkyl hydroperoxide reductase subunit F (AphF) and oxidative stress transcriptional regulator that participate in the control of elevated ROS levels. Another adaptive response to arsenite treatment in *X. oryzae* pv. *oryzae* IXO1088 includes the upregulation of various Hsps. Hsps help with survival under stress conditions by facilitating the proper folding of newly or already translated proteins and preventing the accumulation of aggregated proteins. Upregulation of chaperones has also been observed in response to other metals [49, 50]. Chaperones can contribute by converting metal-induced misfolded proteins to active native proteins; thus, enhancing cellular tolerance to heavy metals. Various transporters, including ATP-binding cassette proteins, MFS (major facilitator superfamily) and ferrous ion transport protein, were also upregulated. Additionally, chemotaxis is one of the most important features of bacteria that regulates their movement towards or away from specific substances, either attractants or repellents. The transcriptome data also showed upregulation of four DEGs, CheW, CheA, MotD and methyl-accepting chemotaxis protein, suggesting a role in the chemotaxis response towards arsenite. Upregulation of genes involved in carbohydrate metabolism pathways was also in agreement with previous studies [51, 52].

The majority of the genes encoding protein translational machinery (ribosomal proteins) and proteins involved in amino acid metabolism were downregulated in response to arsenite. This finding is consistent with previous studies showing a decline in the rate of protein synthesis in response to arsenite As(III) exposure [52–54]. In addition, genes involved in oxidative phosphorylation pathways such as ATP synthase subunits A, C,  $\delta$  and  $\beta$ , cytochrome ubiquinol oxidase subunits, and genes involved in intracellular protein transport, including *secD*, *secF*, *secG* and *secE*, were also downregulated in response to arsenite treatment. Multi-drug RND-efflux transporter subunits and potassium ATPase subunit F and A were downregulated in both the conditions. The downregulation of these genes indicates a switch from aerobic to anaerobic respiration to promote energy conservation following arsenic exposure.

In conclusion, we have identified acquisition of a novel cassette in a major lineage of the *X. oryzae* pv. *oryzae* population that confers arsenic resistance. The arsenic-rich environment has possibly led to the acquisition of such a cassette in the *X. oryzae* pv. *oryzae* population, thereby providing selective/adaptive advantage to the strains from this lineage. Using RNA-Seq technology, we have identified the adaptive and detoxification pathways adapted by *X. oryzae* pv. *oryzae* to evade arsenic toxicity. The study highlights the evolutionary pathway of pathogens to rapidly adapt to unique environmental conditions during the course of evolution and diversification into different lineages. Altogether, the study could be

useful for further studies to elucidate molecular mechanisms under arsenic stress conditions and shows the need to tackle arsenic contamination in rice fields.

#### Funding information

This work was supported by CSIR projects entitled 'High throughput and integrative genomic approaches to understand adaptation of probiotic and pathogenic bacterium' (OLP-148) and 'Mega-genomic insights into coevolution of rice and its microbiome (MICRA)' (MLP0020). A.K is supported by a DST-INSPIRE fellowship. R.R and T.S. are supported by CSIR fellowship. We acknowledge the funding through CSIR-FBR project "Mega-genomic insights into co-evolution of rice and its Microbiome".

#### Author contributions

A. K. performed all the experiments and analysed the results with the help of R. R. R. performed the qRT-PCR experiment to validate RNA-Seq data. T. S. helped in analysing RNA-Seq data. A. K. drafted the manuscript with input from R. R., T. S. and P. B. P. P. B. P. and A. K. conceived and participated in designing the study. P. B. P. applied for the funding. All authors have read and approved the manuscript.

#### Conflicts of interest

The authors declare that there are no conflicts of interest.

#### References

- Zhu YG, Yoshinaga M, Zhao FJ, Rosen BP. Earth abides arsenic biotransformations. *Annu Rev Earth Planet Sci* 2014;42:443–467.
- Naidu R, Bhattacharya P. Arsenic in the environment—risks and management strategies. *Environ Geochem Health* 2009;31:1–8.
- Hughes MF. Arsenic toxicity and potential mechanisms of action. *Toxicol Lett* 2002;133:1–16.
- Yan G, Chen X, Du S, Deng Z, Wang L, et al. Genetic mechanisms of arsenic detoxification and metabolism in bacteria. *Curr Genet* 2019;65:329–338.
- Slyemi D, Bonnefoy V. How prokaryotes deal with arsenic. *Environ Microbiol Rep* 2012;4:571–586.
- Yang HC, Fu HL, Lin YF, Rosen BP. Pathways of arsenic uptake and efflux. *Curr Top Membr* 2012;69:325–358.
- Oremland RS, Stolz JF. The ecology of arsenic. *Science* 2003;300:939–944.
- Fekih IB, Zhang C, Li YP, Zhao Y, Alwathnani HA, et al. Distribution of arsenic resistance genes in prokaryotes. *Front Microbiol* 2018;9:2473.
- Li J, Pawitwar SS, Rosen BP. The organoarsenical biocycle and the primordial antibiotic methylarsenite. *Metallomics* 2016;8:1047–1055.
- Yang HC, Rosen BP. New mechanisms of bacterial arsenic resistance. *Biomed J* 2016;39:5–13.
- Chen J, Bhattacharjee H, Rosen BP. ArsH is an organoarsenical oxidase that confers resistance to trivalent forms of the herbicide monosodium methylarsenate and the poultry growth promoter roxarsone. *Mol Microbiol* 2015;96:1042–1052.
- Qin J, Rosen BP, Zhang Y, Wang G, Franke S, et al. Arsenic detoxification and evolution of trimethylarsine gas by a microbial arsenite S-adenosylmethionine methyltransferase. *Proc Natl Acad Sci USA* 2006;103:2075–2080.
- Rahaman S, Sinha A, Pati R, Mukhopadhyay D. Arsenic contamination: a potential hazard to the affected areas of West Bengal, India. *Environ Geochem Health* 2013;35:119–132.
- Ahmad SA, Khan MH, Haque M. Arsenic contamination in groundwater in Bangladesh: implications and challenges for healthcare policy. *Risk Manag Healthc Policy* 2018;11:251–261.
- Meharg AA, Rahman MM. Arsenic contamination of Bangladesh paddy field soils: implications for rice contribution to arsenic consumption. *Environ Sci Technol* 2003;37:229–234.
- Alam M, Sattar M. Assessment of arsenic contamination in soils and waters in some areas of Bangladesh. *Water Sci Technol* 2000;42:185–192.

17. Zhu YG, Williams PN, Meharg AA. Exposure to inorganic arsenic from rice: a global health issue. *Environ Pollut* 2008;154:169–171.
18. Mondal D, Polya DA. Rice is a major exposure route for arsenic in Chakdaha block, Nadia district, West Bengal, India: a probabilistic risk assessment. *Appl Geochem* 2008;23:2987–2998.
19. Meharg AA, Williams PN, Adomako E, Lawgali YY, Deacon C, et al. Geographical variation in total and inorganic arsenic content of polished (white) rice. *Environ Sci Technol* 2009;43:1612–1617.
20. Ma JF, Yamaji N, Mitani N, Xu X-Y, Su Y-H, et al. Transporters of arsenite in rice and their role in arsenic accumulation in rice grain. *Proc Natl Acad Sci USA* 2008;105:9931–9935.
21. Meharg AA, Zhao FJ. *Arsenic and Rice*. Dordrecht: Springer; 2012.
22. Xu X, McGrath S, Meharg A, Zhao F. Growing rice aerobically markedly decreases arsenic accumulation. *Environ Sci Technol* 2008;42:5574–5579.
23. Wu Z, Ren H, McGrath SP, Wu P, Zhao FJ. Investigating the contribution of the phosphate transport pathway to arsenic accumulation in rice. *Plant Physiol* 2011;157:498–508.
24. Poschenrieder C, Tolrà R, Barceló J. Can metals defend plants against biotic stress. *Trends Plant Sci* 2006;11:288–295.
25. Prabhakaran P, Ashraf MA, Aqma WS. Microbial stress response to heavy metals in the environment. *RSC Adv* 2016;6:109862–109877.
26. Chen Y, Chao Y, Li Y, Lin Q, Bai J, et al. Survival strategies of the plant-associated bacterium *Enterobacter* sp. strain EG16 under cadmium stress. *Appl Environ Microbiol* 2016;82:1734–1744.
27. Lu M, Jiao S, Gao E, Song X, Li Z, et al. Transcriptome response to heavy metals in *Sinorhizobium meliloti* CCNWSX0020 reveals new metal resistance determinants that also promote bioremediation by *Medicago lupulina* in metal-contaminated soil. *Appl Environ Microbiol* 2017;83:e01244–17.
28. Andres J, Bertin PN. The microbial genomics of arsenic. *FEMS Microbiol Rev* 2016;40:299–322.
29. Hrimpeng K, Prapagdee B, Banjerdki P, Vattanaviboon P, Dubbs JM, et al. Challenging *Xanthomonas campestris* with low levels of arsenic mediates cross-protection against oxidant killing. *FEMS Microbiol Lett* 2006;262:121–127.
30. Sukchawalit R, Prapagdee B, Charoenlap N, Vattanaviboon P, Mongkolsuk S. Protection of *Xanthomonas* against arsenic toxicity involves the peroxide-sensing transcription regulator OxyR. *Res Microbiol* 2005;156:30–34.
31. Kaur A, Bansal K, Kumar S, Sonti RV, Patil PB. Complete genome dynamics of a dominant-lineage strain of *Xanthomonas oryzae* Pv. *Oryzae* harbouring a novel plasmid encoding a type iv secretion system. *Access Microbiol* 2019;1:e000063.
32. Soares SC, Geyik H, Ramos RT, de Sá PH, Barbosa EG, et al. GIPSY: genomic island prediction software. *J Biotechnol* 2016;232:2–11.
33. Tamura K, Stecher G, Peterson D, Filipiński A, Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* 2013;30:2725–2729.
34. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, et al. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 2010;59:307–321.
35. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, et al. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 2015;31:3691–3693.
36. Gouy M, Guindon S, Gascuel O. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol* 2010;27:221–224.
37. Letunic I, Bork P. Interactive Tree of Life (iTOL) v3: An online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res* 2016;44:W242–W245.
38. Anders S, Huber W. *Differential Expression of RNA-Seq Data at the Gene Level – the DESeq Package*. Heidelberg: European Molecular Biology Laboratory; 2012.
39. Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, et al. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Mol Biol Evol* 2017;34:2115–2122.
40. Mondal D, Rahman MM, Suman S, Sharma P, Siddique AB, et al. Arsenic exposure from food exceeds that from drinking water in endemic area of Bihar, India. *Sci Total Environ* 2020;754:142082.
41. Nickson R, McArthur J, Ravenscroft P, Burgess W, Ahmed K. Mechanism of arsenic release to groundwater, Bangladesh and West Bengal. *Appl Geochem* 2000;15:403–413.
42. Mitra A, Chatterjee S, Moogouei R, Gupta DK. Arsenic accumulation in rice and probable mitigation approaches: a review. *Agronomy* 2017;7:67.
43. Takahashi Y, Minamikawa R, Hattori KH, Kurishima K, Kihou N, et al. Arsenic behavior in paddy fields during the cycle of flooded and non-flooded periods. *Environ Sci Technol* 2004;38:1038–1044.
44. Suriyagoda LD, Dittert K, Lambers H. Mechanism of arsenic uptake, translocation and plant resistance to accumulate arsenic in rice grains. *Agriculture, Ecosystems & Environment* 2018;253:23–37.
45. Yamaguchi N, Nakamura T, Dong D, Takahashi Y, Amachi S, et al. Arsenic release from flooded paddy soils is influenced by speciation, Eh, pH, and iron dissolution. *Chemosphere* 2011;83:925–932.
46. Zhang Y, Chen S, Hao X, J-Q S, Xue X, et al. Transcriptomic analysis reveals adaptive responses of an *Enterobacteriaceae* strain LSJC7 to arsenic exposure. *Front Microbiol* 2016;7:636.
47. Schalk IJ, Hannauer M, Braud A. New roles for bacterial siderophores in metal transport and tolerance. *Environ Microbiol* 2011;13:2844–2854.
48. Maity JP, Nath B, Kar S, Chen C-Y, Banerjee S, et al. Arsenic-induced health crisis in peri-urban Moyna and Ardebok villages, West Bengal, India: an exposure assessment study. *Environ Geochem Health* 2012;34:563–574.
49. Tamás MJ, Fauvet B, Christen P, Goloubinoff P. Misfolding and aggregation of nascent proteins: a novel mode of toxic cadmium action in vivo. *Curr Genet* 2018;64:177–181.
50. Elran R, Raam M, Kraus R, Brekhman V, Sher N, et al. Early and late response of *Nematostella vectensis* transcriptome to heavy metals. *Mol Ecol* 2014;23:4722–4736.
51. Cleiss-Arnold J, Koechler S, Proux C, Fardeau ML, Dillies M-A, et al. Temporal transcriptomic response during arsenic stress in *Hermiimonas arsenicoxydans*. *BMC genomics* 2010;11:709.
52. Shah S, Damare S. Cellular response of *Brevibacterium casei* #NIO5BA88 to arsenic and chromium – a proteomic approach. *Braz J Microbiol* 2020;51:1885–1895.
53. Wang A, Crowley DE. Global gene expression responses to cadmium toxicity in *Escherichia coli*. *J Bacteriol* 2005;187:3259–3266.
54. Henne KL, Turse JE, Nicora CD, Lipton MS, Tollaksen SL, et al. Global proteomic analysis of the chromate response in *Arthro-bacter* sp. strain FB24. *J Proteome Res* 2009;8:1704–1716.