Effect of arsenite and growth in biofilm conditions on the evolution of Thiomonas sp. CB2

Kelle C Freel, Stéphanie Fouteau, David Roche, Julien Farasin, Aline Huber, Sandrine Koechler, Martina Peres, Olfa Chiboub, Stephane Cruveiller, Hugo Varet, et al.

MICROBIAL GENOMICS

RESEARCH ARTICLE

Freel *et al.*, *Microbial Genomics*

DOI 10.1099/mgen.0.000447

MICROBIOLOGY SOCIETY

OPEN DATA OPEN ACCESS

Effect of arsenite and growth in biofilm conditions on the evolution of *Thiomonas* sp. CB2

Kelle C. Freel[1]†, Stephanie Fouteau[2], David Roche[2], Julien Farasin[1], Aline Huber[1], Sandrine Koechler[1]‡, Martina Peres[1], Olfa Chiboub[1], Hugo Varet[3,4], Caroline Proux[3], Julien Deschamps[5], Romain Briandet[5], Rachel Torchet[2], Stephane Cruveiller[2], Didier Lièvremont[1], Jean-Yves Coppée[3], Valérie Barbe[2] and Florence Arsène-Ploetze[1,*,‡]

Abstract

Thiomonas bacteria are ubiquitous at acid mine drainage sites and play key roles in the remediation of water at these locations by oxidizing arsenite to arsenate, favouring the sorption of arsenic by iron oxides and their coprecipitation. Understanding the adaptive capacities of these bacteria is crucial to revealing how they persist and remain active in such extreme conditions. Interestingly, it was previously observed that after exposure to arsenite, when grown in a biofilm, some strains of *Thiomonas* bacteria develop variants that are more resistant to arsenic. Here, we identified the mechanisms involved in the emergence of such variants in biofilms. We found that the percentage of variants generated increased in the presence of high concentrations of arsenite (5.33 mM), especially in the detached cells after growth under biofilm-forming conditions. Analysis of gene expression in the parent strain CB2 revealed that genes involved in DNA repair were upregulated in the conditions where variants were observed. Finally, we assessed the phenotypes and genomes of the subsequent variants generated to evaluate the number of mutations compared to the parent strain. We determined that multiple point mutations accumulated after exposure to arsenite when cells were grown under biofilm conditions. Some of these mutations were found in what is referred to as ICE19, a genomic island (GI) carrying arsenic-resistance genes, also harbouring characteristics of an integrative and conjugative element (ICE). The mutations likely favoured the excision and duplication of this GI. This research aids in understanding how *Thiomonas* bacteria adapt to highly toxic environments, and, more generally, provides a window to bacterial genome evolution in extreme environments.

DATA SUMMARY

This study utilizes sequences previously generated from other studies. The accession number for the sequence data of the whole population genome is ERR3040228 (project number PRJEB29999). Accession numbers for the genome sequences of Bio17B3, Sup16B3, OC7 and the new sequence of CB2 are LR131947-LR131949, LR131950-LR131952, LR131953-LR131955 and LR131956-LR131958, respectively (project number PRJEB29999).

Received 15 May 2020; Accepted 14 September 2020; Published 09 October 2020

Author affiliations: [1]Laboratoire Génétique Moléculaire, Génomique et Microbiologie, UMR7156, Institut de Botanique, CNRS – Université de Strasbourg, Strasbourg, France; [2]Génomique Métabolique, Genoscope, Institut de Biologie François Jacob, Commissariat à l'Energie Atomique (CEA), CNRS, Université Evry, Université Paris-Saclay, Evry, France; [3]Plateforme Transcriptome et Epigenome, BioMics, Centre de Ressources et Recherches Technologiques, Institut Pasteur, Paris, France; [4]Hub Bioinformatique et Biostatistique, Centre de Bioinformatique, Biostatistique et Biologie Intégrative (C3BI, USR 3756, IP CNRS), Institut Pasteur, Paris, France; [5]Micalis Institute, INRA, AgroParisTech, Université Paris-Saclay, Jouy-en-Josas, France.
***Correspondence:** Florence Arsène-Ploetze, florence.ploetze@ibmp-cnrs.unistra.fr

Keywords: acid mine drainage (AMD); comparative genomics; genome evolution; adaptation; genomic islands; arsenic.

Abbreviations: AMD, acid mine drainage; dN, non-synonymous; dS, synonymous; GI, genomic island; HQ, high quality; ICE, integrative and conjugative element; MGE, mobile genetic element; MIC, minimum inhibitory concentration; OSR, oxidative stress response; SIM, stress-induced mutagenesis; SNV, single nucleotide variation.

†Present address: Hawai'i Institute of Marine Biology, University of Hawai'i at Mānoa, Kāne'ohe, HI, USA
‡Present address: Institut de Biologie Moléculaire des Plantes, CNRS, Université de Strasbourg, Strasbourg, France.

The accession number for the sequence data of the whole-population genome is ERR3040228 (BioProject number PRJEB29999). The GenBank/EMBL/DDBJ accession numbers for the genome sequences of Bio17B3, Sup16B3, OC7 and the new sequence of CB2 are LR131947–LR131949, LR131950–LR131952, LR131953–LR131955 and LR131956–LR131958, respectively (BioProject number PRJEB29999).
Data statement: All supporting data, code and protocols have been provided within the article or through supplementary data files. Ten supplementary figures and nine supplementary tables are available with the online version of this article.

INTRODUCTION

Understanding what processes have allowed bacteria to ultimately thrive in unique and challenging environments will allow for a better understanding of microbial genome evolution. Two processes likely responsible for bacterial adaptation include rapid changes in gene expression and the acquisition of beneficial mutations (point mutations, rearrangements or acquisition of beneficial genetic content via horizontal gene transfer) [1]. Exposure to stressful conditions affects cellular processes and can damage macromolecules, including DNA [1, 2]. Such damage can lead to the expression of proteins involved in DNA repair, specifically error-prone DNA repair machinery. This process was first studied in *Escherichia coli*: after stress-associated DNA damage, the SOS response induced the synthesis of the error-prone DNA polymerases IV and V, whereas the high-fidelity mismatch repair pathway was repressed [2]. Consequently, stressful conditions transiently increase the mutation rate, through a process called stress-induced mutagenesis (SIM) [3]. In addition, it was observed that SIM alters the relative frequency of different types of mutations. The range of this frequency is known as the mutation spectrum. Bacterial isolates with high mutation rates (known as mutators) and changes in the mutation spectrum have both been shown to be advantageous for some cells when it allows the population to acquire beneficial mutations over a short period of time [4–6]. SIM has been observed in various bacteria and has been shown to enable bacterial populations to acquire particular adaptive capacities in multiple extreme or stressful conditions [3, 7–10].

Among the myriad ecosystems on the planet, extreme environments provide interesting model systems to explore bacterial adaptation in relatively short timescales [10]. Acid mine drainage (AMD) sites are incredibly unique and challenging habitats for microbial communities, since several environmental factors, such as acidic pH or the presence of toxic compounds, interfere with cellular functions [11]. It has been proposed previously that microbes evolve faster in such environments compared to those inhabiting non-stressful environments [10–12]. It was also demonstrated that in AMD sites, multiple genotypic groups coexist within bacteria of the same genus or species, and fine variation in the genetic content of these strains results in different adaptive processes [12–15]. Heavy metals and metalloids found in AMD sites, such as arsenic, are particularly toxic, and may impact genome integrity and induce mutations [12, 16, 17]. Thus, such environments present an opportunity to determine whether bacteria in AMD sites undergo mutations more often than bacteria inhabiting environments without toxic compounds.

Various groups of prokaryotes are capable of colonizing AMD sites [11, 12, 18, 19]. These prokaryotes possess metabolic mechanisms to resist the hazardous effects of arsenic [20]. One essential method of resistance is to extrude arsenate via the transmembrane carrier protein ArsB (*ars* genes). Some prokaryotes are also able to oxidize arsenite to arsenate (*aio* genes) that can then be extruded, in a process initially regarded strictly as a method of detoxification. However, in *Herminiimonas arsenicoxydans*, it was suggested that the strain may gain additional energy from the arsenic oxidation process [21]. Other microbes can actively use various arsenic compounds in their metabolism, either as an electron donor or as a terminal electron acceptor for anaerobic respiration [20]. It has even been demonstrated that select micro-organisms can methylate inorganic arsenic, probably as a resistance mechanism, or demethylate organic arsenic compounds [20].

Impact Statement

Deciphering mechanisms of bacterial adaptation is crucial to revealing how bacteria persist and remain active in challenging habitats. Among the myriad ecosystems on the planet, extreme environments, such as acid mine drainage (AMD) sites, provide interesting model systems to explore bacterial adaptation across relatively short timescales. Included in the organisms found ubiquitously in AMD sites are several strains of the genus *Thiomonas*. Our study investigated the effect of increasing arsenic concentration on the evolution of *Thiomonas* sp. strain CB2 when grown as a biofilm. We assessed the impact of arsenic on this strain by characterizing the phenotype, genotype of the global population, genotype of variants produced, and gene expression in the parent CB2 strain as well as variants. Our results provide a better understanding of the mechanisms driving the emergence of cells in extreme environments that are more resistant to this toxic metal and potentially other toxic compounds including antibiotics. Ultimately, this study enhances the understanding of how bacterial genomes evolve in extreme conditions.

Among the organisms found ubiquitously in AMD sites are several strains of the genus *Thiomonas*, including *Thiomonas* sp. CB2 isolated from a site in the Reigous Creek near Carnoulès, France. *Thiomonas* strains oxidize arsenite to arsenate, promoting the sorption of arsenic by iron oxides and their coprecipitation, resulting in a natural process of arsenic attenuation [19, 22–25]. At the Carnoulès site, considerable microdiversity has been observed at both the genus and species levels of *Thiomonas* [24, 26–29]. Comparative genomic analyses were previously performed on several genomes of *Thiomonas* strains to define the differences and similarities among strains both between and within species of this genus [24, 27–30]. This research established a set of genes only found in select strains (the 'dispensable genome') and those conserved across all strains (the 'core genome'). These studies also highlighted that the *Thiomonas* genomes examined contained multiple genomic islands (GIs) [24, 27–30]. These GIs were previously defined [24, 27–30], present only in some strains and found to confer unique abilities for survival in AMD sites. This previous research suggested that variability in adaptation observed among *Thiomonas* strains possibly represented different ecotypes, ultimately enhanced the capacity for strain adaptation, and might be

caused by differences in mutations, genomic rearrangements or horizontal gene transfer events.

Among other important adaptation capacities, *Thiomonas* isolates also form biofilms, likely essential for survival in AMD sites for two main reasons. First, biofilms create ideal microenvironments enhancing diversification and horizontal gene transfer, allowing populations or communities to adapt while in unstable environmental conditions; this follows what is known as the 'insurance hypothesis' [31–35]. The presence of diverse subpopulations essentially allows for the overall community to survive in a broader range of conditions. Second, the biofilm matrix protects cells against external toxic compounds, such as arsenic, and extreme physicochemical conditions, because the organic matrix acts as a barrier isolating the cells from many environmental stresses [32, 34, 36].

The process of biofilm formation has been studied in the absence and presence of arsenic in *Thiomonas* sp. strain CB2 (referred to here as CB2) isolated from the Carnoulès AMD site [37]. The genes expressed during biofilm development were previously identified using transcriptomics after exposure to arsenic [29]. Interestingly, CB2 formed a particularly complex biofilm structure in the presence of arsenite, and within the biofilm, cell death and localized cell lysis events were observed. During biofilm maturation, the emergence of cells (termed variants) with greater survival capabilities than the starting population has been documented [37]. These previous studies suggested that *Thiomonas* bacteria may adapt over a relatively short-term timescale (after less than a hundred generations), in the presence of arsenic, leading to the generation of variants. We hypothesized that the disparity in survival rates between the original CB2 strain and variants may be due to differences in gene expression, mutations, duplications or deletions after growth in a biofilm.

Here, we attempt to identify mechanisms driving the emergence of variants in biofilms in the presence of arsenite, in order to enhance the understanding of how bacterial genomes evolve in extreme conditions. We first assessed how arsenic is involved in the accumulation of variants, and analysed both transcription and genome diversity in the parent strain; we also characterized the phenotypes and genomes of the variants generated. The results obtained allow for a better understanding of how *Thiomonas* spp. evolve and adapt to the particularly toxic environments found in AMD sites.

METHODS

Bacterial strains and growth conditions

Thiomonas sp. CB2 was isolated from the Carnoulès AMD site (Gard, France) [24, 26]. This strain and the variants were grown *in vitro* on modified 126 medium (M126), as previously described [26]. Arsenite was added at the concentration required (1.33, 2.67, 4, 5.33, 10.67, 13.25, 15.9, 21.2 or 26.5 mM, depending on the experiments detailed below) from a sterile stock of 667 mM obtained with $NaAsO_2$ salts (Prolabo). The concentrations were chosen according to ones used previously to study the effects of subinhibitory arsenite concentrations on biofilm formation and development in *Thiomonas* sp. CB2 [37].

Quantification of variants

To determine how effective arsenite exposure during growth under biofilm conditions was in generating variants, we quantified the number of variants generated from the original strain *Thiomonas* sp. CB2 using the approach illustrated in Fig. 1a. An initial culture of 5 ml was grown in a 15 ml centrifuge tube on a rotary shaker (300 r.p.m.) for 48 h in M126 medium lacking arsenite. This culture was then diluted to an OD_{600} of 0.002 in 20 ml, in the same medium, in a 150 ml Erlenmeyer flask that was also shaken during incubation (300 r.p.m.). After 40 h, the culture in the flask was then diluted to an OD_{600} of 0.002 and served as the inoculum for planktonic as well as biofilm cultures. Planktonic cultures were continually shaken in a culture tube or flask and incubated at 30 °C, while biofilm cultures were grown in a 12-well plate, not shaken, and incubated at 30 °C. Replicates of the planktonic and biofilm cultures were grown in M126 medium with the addition of one of three treatments: no arsenic, 2.67 mM arsenite or 5.33 mM arsenite. Three planktonic cultures and three wells (biofilm conditions) were maintained for each treatment for a total of 72 h. After 24 h, the medium overlying the biofilm in the 12-well plates was removed and fresh medium was added.

After 24, 48 or 72 h, the number of variants was evaluated by counting the number of colonies obtained on plates containing arsenite compared to the number on plates without arsenite (Fig. S1, available with the online version of this article). Briefly, cells obtained from biofilms were scraped with a small squeegee to remove them from the well bottom, resuspended by pipetting several times and finally vortexed to completely disperse them. The *Thiomonas* cells grown in biofilms were not strongly attached to each other and could be easily separated. Cell suspensions were serial 10-fold diluted and then 40 µl of these dilutions were spread on a small Petri dish containing solid M126 medium without arsenite or supplemented with 10.6 mM arsenite. This concentration corresponded to the minimum inhibitory concentration (MIC) of arsenite previously determined for *Thiomonas* sp. CB2 [24, 28]. In the case of 24 h cultures, specifically for the upper phase of biofilms grown with arsenite, the spread dilutions were 1/10000. In other cases, the dilutions were 1/100000. The plates were then incubated at 30 °C for 10 days. The percentage of arsenite MIC-resistant variants in the population was calculated by dividing the number of colonies counted on plates containing 10.6 mM arsenite by the number of colonies counted on plates without arsenite (Fig. S1) and then multiplying by 100. For the biofilm cultures, the percentage of variants was measured in both the phase above the biofilm (detached cells), as well as the cells fixed to the wall of the plate.

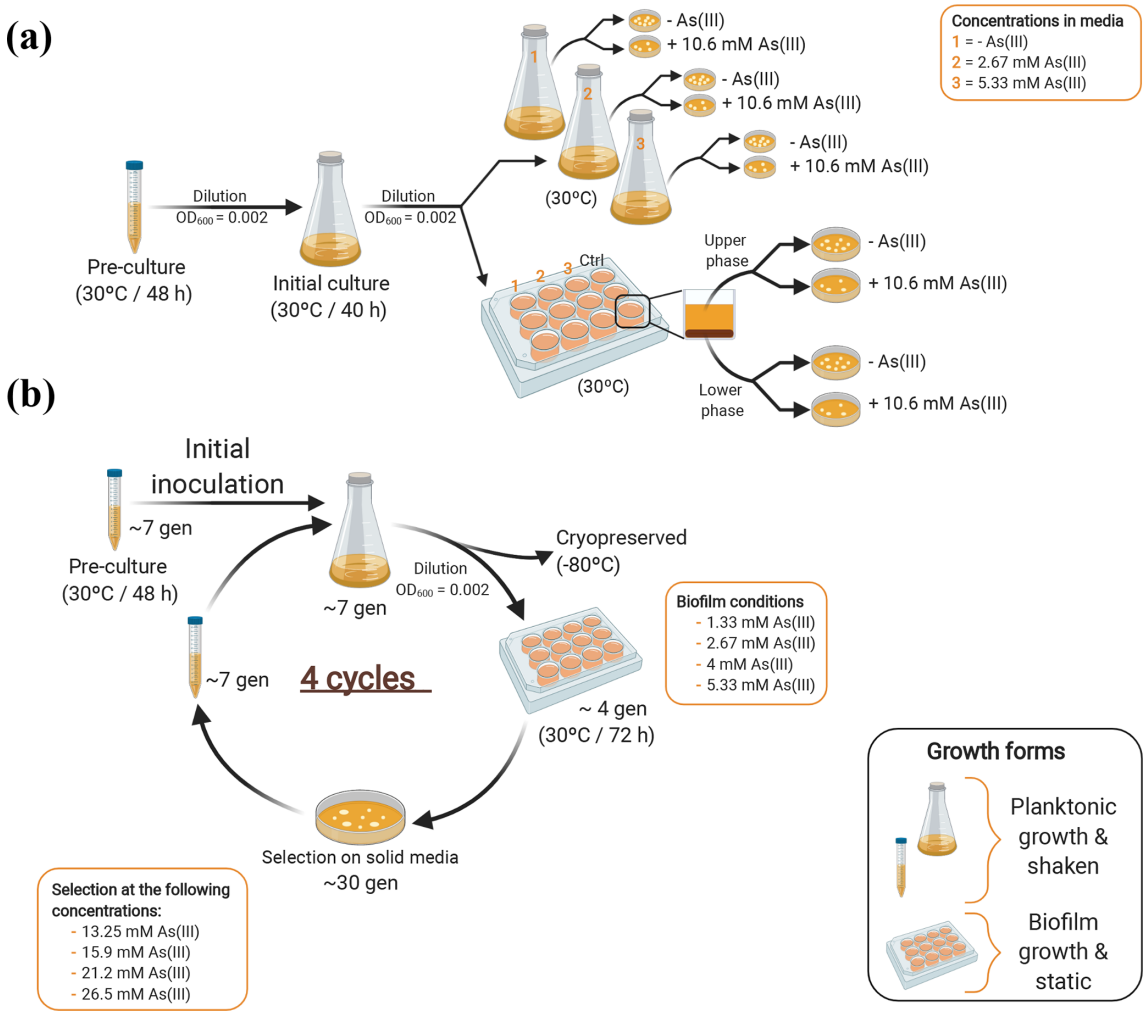


Fig. 1. Depictions of cultivation methods to obtain variants capable of survival at higher arsenite concentrations than the parent strain *Thiomonas* sp. CB2. (a) Determination of the number of variants generated after growth in planktonic or biofilm culture conditions. A planktonic culture was inoculated from a pre-culture, then used to start an initial culture to inoculate planktonic and biofilm cultures grown in (1) 0, (2) 2.67 or (3) 5.33 mM arsenite [As(III)]. Three replicates were used for biofilms (wells) and planktonic culture (flasks) per concentration. The percentage of variants was measured in the upper phase (detached cells) and the lower phase (attached cells) of the biofilm. Dilutions were made prior to spreading on solid M126 medium with 0 or 10.6 mM arsenite. (b) Generation of 'super variants' capable of growth at much higher levels than the parent strain. The biofilm was inoculated from a planktonic culture initially inoculated from a pre-culture. Biofilms were grown for 72 h with one change of medium at 24 h, grown with no arsenic or arsenite concentrations of 1.33, 2.67, 4 or 5.33 mM with two wells per concentration. The detached planktonic cells and the cells attached to the plate were separated prior to variant selection on solid medium containing 13.25, 15.9, 21.2 or 26.5 mM arsenite. This figure was created with bioRENDER (*https://biorender.com*).

Generation of variants

To obtain variants resistant to dramatically higher concentrations of arsenite compared to the parent strain CB2, we followed the scheme outlined in Fig. 1b. Briefly, a planktonic CB2 culture was first grown in 5 ml medium in the absence of arsenite in a centrifuge tube for 48 h (equivalent to seven generations). This initial culture was used to inoculate a larger flask (20 ml culture, in the absence of arsenite), which was then grown for an additional seven generations. At this point, an aliquot of the planktonic culture was stored at −80 °C (at a final concentration of 7% DMSO). The remainder of the culture was diluted to an OD_{600} of 0.002 and inoculated into a 12-well plate, with cultures grown in the absence or the presence of 1.33, 2.67, 4 or 5.33 mM arsenite. After four generations, the planktonic and fixed biofilm portion of each culture was plated on successively higher concentrations of arsenite than the previous culture's MIC, i.e. the first group of variants was plated at 10.6 mM. This corresponded to the first cycle of variant generation. Variants isolated after the first cycle from plates of solid medium were either prepared for cryopreservation or diluted to an OD_{600} of 0.002. To prepare cells for cryopreservation, they were grown up in a 20 ml culture flask

Table 1. MICs (denoted with an asterisk) and arsenite oxidation activity (denoted with #) of variants selected for their resistance to arsenic and select antibiotics

Strain/ variant ID	Arsenite (mM)*	Neomycin (mg l^{-1})*	Spectinomycin (mg l^{-1})*	Gentamicin (mg l^{-1})*	Streptomycin (mg l^{-1})*	Kanamycin (mg l^{-1})*	Arsenite oxidation activity #
CB2	12	16	8	4	8	2	++
Bio17A3	23.9	NT	NT	NT	NT	NT	NT
Bio17B3	23.9	16	8	4	8	2	–
Bio16B1	15.9	NT	NT	NT	NT	NT	NT
Sup16A1	23.9	NT	NT	NT	NT	NT	NT
Sup16B3	23.9	<16	8	4	8	2	–
Bio14B1	18.6	<16	8	4	8	2	NT
Sup14A1	21.3	NT	NT	NT	NT	NT	NT
Sup14A3	21.3	NT	NT	NT	NT	NT	NT
OC7	15.9	256	512	64	128	128	++

NT, Not tested.

for 40 h without arsenite in order to prevent any further effect of arsenite that could lead to additional mutations and, therefore, potential evolution, and frozen in 2 ml aliquots at a final concentration of 7% DMSO at −80 °C. The diluted culture was then inoculated into a 12-well plate, as described below, to obtain a second group of variants, more resistant compared to the first group. We performed a total of four cycles with selection at 13.25, 15.9, 21.2 and 26.5 mM arsenite, and a total of 124 variants were isolated. Among these variants, eight were selected based on their relatively higher resistance to arsenite for further experiments (explained in the following paragraph): Bio17A3, Bio17B3, Bio16B1, Sup16A1, Sup16B3, Bio14B1, Sup14A1 and Sup14A3. These variant were named according to the following scheme: 'Bio' indicated the variant was isolated from attached cells within biofilms grown in the static 12-well plate, whereas 'Sup' indicated the variant was isolated from detached cells inhabiting the upper phase of medium above the biofilm in the static 12-well plate. The number following Bio/Sup was assigned corresponding to the plate number or the experiment from which the variant was obtained.

In parallel, we also attempted to identify antibiotic-resistant variants. For this purpose, a different set of eight variants were isolated (OC1, OC2, OC3, OC4, OC5, OC6, OC7 and OC8) after incubation in the same growth conditions as above [in biofilm conditions grown in the absence of arsenic (OC1 and OC6), or the presence of 2.67 mM (OC2 and OC7) or 5.33 mM arsenite (OC3, OC4, OC5 and OC8)]. These isolates were either from attached (OC6, OC7 and OC8) or detached cells (OC1, OC2, OC3, OC4 and OC5) from the biofilm grown in the static 12-well plate. These eight variants were selected for their higher resistance to chloramphenicol (MIC >2 mg l^{-1}) (OC4 and OC5) or kanamycin (MIC>2 mg l^{-1}) (OC1, OC2, OC3, OC6, OC7 and OC8). From these variants, we selected OC7 (isolated from attached cells in the biofilm grown in the 12-well plate with 2.67 mM arsenite and kanamycin resistant).

Screening and assessment of variant phenotypes

We selected a subset of variants for the phenotype study, according to their arsenite or antibiotic MIC growth rate, and stability of resistance in the absence of selection pressure (i.e. in the absence of arsenite or antibiotic) (Table 1). Briefly, 18 variants were selected because they had been isolated on high concentrations of arsenite (>14 mM arsenite), suggesting increased resistance to arsenite, and 17 others by colony morphology, suggesting possible differences in ability to form biofilms. To carry out an initial screen and determine arsenite resistance, the 35 variants were first grown in multiple replicates directly on M126 medium without arsenite and on supplemented solid medium with arsenite concentrations ranging from 10.6 to 26.6 mM at intervals of 2.67 mM. The cultures were incubated at 30 °C for 12 days, after which the maximum concentration at which growth was visible was used to select eight variants: Bio14B1, Sup14A3, Sup14A1, Sup16B3, Bio17B3, Bio17A3, Sup16A1 and Bio16B1.

The MIC of arsenite for these eight selected variants was then measured more precisely and compared to CB2 (which was also re-assessed and determined to be 12 mM). For this purpose, exponential phase planktonic cultures (incubated at 30 °C for 40 h) were prepared and tenfold serial dilution were performed up to 10^{-5}. These dilutions were then replicated on solid M126 alone or supplemented with 13.3, 16, 18.6, 20, 21.3, 22.6 or 23.9 mM arsenite. The plates were incubated at 30 °C for 10 days and the MIC was determined based on the maximum concentration at which growth was clearly visible (Table 1).

To test the stability of arsenite resistance, the eight variants were grown in M126 liquid medium in the absence of arsenite for over 30 generations and then the MIC was remeasured using the same procedure described above. This experiment was done twice. The growth assessment was performed with 50 ml cultures inoculated with *Thiomonas* sp. CB2, as well as the eight chosen variants, from a starter culture at an OD_{600}0.001, incubated at 30 °C. For each strain, two cultures were grown and the OD_{600} was measured every 2 h until it reached 0.4.

The biofilm-forming capacities of these eight variants were tested using crystal violet staining, as previously described [29]. Three variants (Sup16B3, Bio17B3 and Bio14B1) were then selected on the basis of the biofilm-formation profiles observed. Confocal laser scanning microscopy was used to observe the biofilm architectures of these three variants, as described previously [29].

The antibiotic resistance of the variants was tested by determining the MIC using standard protocols as previously described [27], by inoculating 15 ml tubes containing M126 medium and a dilution series of the antibiotic being tested (including neomycin, spectinomycin, gentamicin, streptomycin or kanamycin). These tubes were inoculated with 5 µl planktonic cultures in exponential phase (previously incubated at 30 °C for 40 h). Growth was checked visually or by measuring the OD_{600} at 72 h after incubation at 30 °C with shaking.

We also tested the ability of variants to grow in the presence of H_2O_2 to determine whether the resistance to arsenite or the antibiotic was due to the development of a general oxidative stress resistance. For this purpose, appropriate dilutions were made from an overnight culture to obtain a concentration between 2×10^6 and 4×10^6 c.f.u. ml^{-1}. A volume of 2–4 ml per petri dish was inoculated by flooding on solid M126 medium. Discs containing different concentrations of hydrogen peroxide (H_2O_2) were prepared manually by soaking sterile discs of Whatman paper (6 mm in diameter) with different diluted solutions from a 30% stock solution (including 6, 3, 0.6 and 0.3%). These discs were then placed in different zones on the inoculated M126 agar plates and incubated at 30 °C. After 5 days, the diameters of any halos present around the discs containing H_2O_2, and indicating bacterial growth inhibition, were measured.

The ability of the three selected variants Bio17B3, Sup16B3 and OC7 and the original CB2 strain to oxidize arsenite was tested twice independently in liquid medium with 0.67 mM arsenite. Detection of arsenite and arsenate was performed by inductively coupled plasma-atomic emission spectrometry (ICP-AES), as described previously [38].

Whole-population genome sequencing to determine variation post-arsenic exposure

CB2 was grown in conditions favouring variant emergence, specifically in a biofilm over 72 h in the presence of 5.33 mM arsenite in a 12-well plate. After 24 h of biofilm growth, the overlying medium was replaced with fresh medium to remove planktonic cells that did not attach. This step ensured that the cells we categorized as growing as a biofilm were indeed involved in biofilm formation or generated from the biofilm. After 72 h, all cells in each well (attached or detached) were collected. DNA was extracted as previously described [28]. A mate-pair library with a 7708 bp insert size was constructed, this size corresponded to the 'layout nominal length' and was determined during the bank preparation by electrophoresis in an Agilent 2100 Bioanalyzer. The library was run on an Illumina MiSeq instrument (2×300 bp) yielding 2.4 Gb of data. Single nucleotide variations (SNVs) were identified using a bioinformatic pipeline, PALOMA (Polymorphism Analyses in Light Of MAssive DNA sequencing), which allowed comparison of the variants and the CB2 reference sequences. Based on the SSAHA2 package (https://www.sanger.ac.uk/science/tools), PALOMA takes into account raw sequence data and associated qualities to discriminate between true variations and sequencing errors. We used two parameters, a quality score and a minimum coverage threshold, to analyse sequences. The first score indicated the confidence in the prediction and was computed as follows: SNP score=$0.5\times S_{bio}+0.5\times S_{tech}$, where S_{bio}=allele rates and $S_{tech}=f$(quality, strand bias). This S_{tech} mathematical function, for which the maximum score (i.e. 1) corresponds to a well-balanced ratio of reads mapped on the forward and reverse strands (50/50), was evaluated with our data as the quality ratio and strand ratio are tightly linked to sequencing technology. A minimal SNP score was fixed at 0.35 according to the distribution of all the mutation scores (Fig. S2a), excluding the score of the first quartile. The second parameter was linked to the coverage threshold. The high-quality (HQ) reads correspond to the minimum number of reads at that position and had a quality at this position of Q=25, i.e. an accuracy of 99.7% at the level of base calls. The minimal HQ read value (number of HQ reads supporting mutation) was fixed at 10 regarding the distribution of all HQ reads across all mutations (Fig. S2b). The National Center for Biotechnology Information accession number for the sequences is ERR3040228 (BioProject number PRJEB29999).

Whole-genome sequencing of variants

DNA was extracted as previously described [28] and whole-genome sequencing was performed at Genoscope (Evry, France) using an Illumina MiSeq instrument. The CB2 genome was previously sequenced using Roche 454 shotgun sequencing and paired end (8 kb in size) libraries [28]. However, Roche 454 sequencing technology is prone to introducing sequencing errors and has distinct bias from that of Illumina sequencing technology. Therefore, the genome of the *Thiomonas* sp. CB2 strain was resequenced using the same technology used for the whole-population and variant genomes, to ensure that mutations found in the variants were not due to sequencing mistakes or mutations that already existed in the parental strain used in this study. For the *Thiomonas* strains Bio17B3, Sup16B3 and OC7, assemblies were obtained by combining paired-end and mate-pair data. First, multiplexed overlapping paired-end libraries

with a mean insert size of 560 bp were constructed and run on an Illumina MiSeq instrument (2×300 bp) generating 347 Mb (Bio17B3), 461 Mb (Sup16B3) and 377 Mb (OC7) of data. In parallel, mate-pair libraries with an insert size of approximately 8000 bp were constructed and loaded on the MiSeq (2×150 bp) generating 549 Mb (Bio17B3), 736 Mb (Sup16B3) and 578 Mb (OC7) of data. A subset (equivalent to approximately 120× coverage) of the paired-end and the mate-pair reads was assembled using Newbler version 2.9. Manual expertise was used to reconstruct the genome sequences. Automatic functional annotation was performed using the MicroScope platform (http://www.genoscope.cns.fr/agc/microscope). The GenBank/EMBL/DDBJ accession numbers for the genome sequences are LR131947–LR131949, LR131950–LR131952, LR131953–LR131955 and LR131956–LR131958 for Bio17B3, Sup16B3, OC7 and the new sequence of CB2, respectively (BioProject number PRJEB29999).

For the comparative genomic analysis, high-throughput sequencing (HTS) data were analysed using the PALOMA bioinformatic pipeline implemented in the MicroScope platform [39]. First, the HTS data were pre-processed to assess quality by completing the following steps: read trimming, merging and splitting of paired-end reads. Second, reads were mapped onto the *Thiomonas* sp. CB2 reference sequence [28] using the SSAHA2 package [40]. Only unique matches with an alignment score equal to at least half of their length were retained as seeds for full Smith–Waterman realignment [41] with a region extended on both sides by 5 nucleotides of the reference genome. All alignments generated then were screened for discrepancies between read and reference sequences, and finally, a score based on coverage, allele frequency, quality of bases and strand bias was computed for each detected event to assess relevance.

RNA sequencing and analysis

To determine which genes were possibly involved in biofilm formation and regulation in response to arsenite, RNA-seq had been used previously to determine which were expressed during biofilm development and/or regulated in response to arsenite [29]. These experiments were performed with *Thiomonas* sp. CB2 cultures maintained in biofilm growth conditions (grown in triplicate). Cells were incubated under stagnant conditions in the presence and in the absence of 5.33 mM arsenite for 24, 48 and 72 h [29]. The data obtained previously with this RNA-seq experiment were further analysed in the present study to identify genes involved in DNA repair.

Additionally, a second RNA sequencing experiment was performed in the present study with *Thiomonas* spp. cells grown in planktonic conditions. CB2, Bio17B3 and Sup16B3 were grown in liquid cultures with shaking as described above, in M126 with 5.33 mM arsenite or without arsenite. After 24 h, growth was halted by incubating cultures at 4 °C, the cultures were centrifuged at 8000 ***g*** for 15 min at 4 °C and stored at −80 °C. Three replicates from three independent cell cultures were sequenced. RNA extraction and sequencing were performed as described previously [29], using an Illumina HiSeq instrument (Illumina). Reads were cleaned to remove adapter sequences and low-quality sequences using an in-house program (https://github.com/baj12/clean_ngs). Only sequences at least 25 nt in length were considered for further analysis. Bowtie (v 0.12.7) using default parameters was used to align to the *Thiomonas* sp. CB2, Bio17B3 and Sup16B3 genomes [42]. For each gene, mapped reads were counted using the featureCounts program [43]. Counts were normalized according to the default method implemented in the DESeq2 R package [44].

RESULTS

Arsenite impacts the rate of resistant variant appearance in cells grown in a biofilm

To determine how arsenite and biofilm growth influence the emergence of higher arsenite resistance, we quantified the number of variants produced by CB2 when cultivated in the absence of arsenic or the presence of 2.67 or 5.33 mM arsenite (Figs 1a and Fig. S1) . In these experiments, we differentiated the planktonic cells, cells fixed within the biofilm and detached cells (see details in Methods), as follows: (i) planktonic cultures were grown in a flask incubated under agitation, (ii) attached cells corresponded to the cells within biofilms grown in the static 12-well plate, and finally (iii) detached cells corresponded to the cells inhabiting the upper phase of medium above the biofilm grown in the static 12-well plate. We measured variant emergence from replicate cultures at 24, 48 and 72 h. We observed that the numbers of cells in each population were similar between the populations with no arsenite compared to those grown in the presence of 2.67 and 5.33 mM arsenite (Fig. S1) .

In the absence of arsenite, at 24 h, the percentage of variants (number of resistant strains compared to the total number of cells) observed was <2% regardless of culture condition (Fig. 2a). For planktonic and fixed biofilm cells after 48 h, the proportion of variants obtained in the absence of arsenite increased to 25%, while the mean increased to 50% in detached biofilm cells; however, there was significant variation between detached cell replicates (Fig. 2b). At 72 h, the mean proportion of variants generated in the absence of arsenite increased to 30% in the planktonic cultures, but decreased to 15% after 72 h for both phases of biofilm culture (Fig. 2c). This decrease in the proportion of variants at 72 h could be due to a lower rate of variant emergence in biofilms without arsenite exposure, as well as unknown challenges to maintaining variant growth. Interestingly, after 72 h in the absence of arsenite, the proportion of variants was greater in the planktonic cells than in either part of the biofilm (fixed or detached cells). These results reveal that there is spontaneous emergence of variants in CB2 populations (between 15 and 30%) in the absence of arsenite regardless of culture type (e.g. planktonic or biofilm).

When grown in planktonic conditions, we found that the presence of arsenite did not have a predictable impact on the

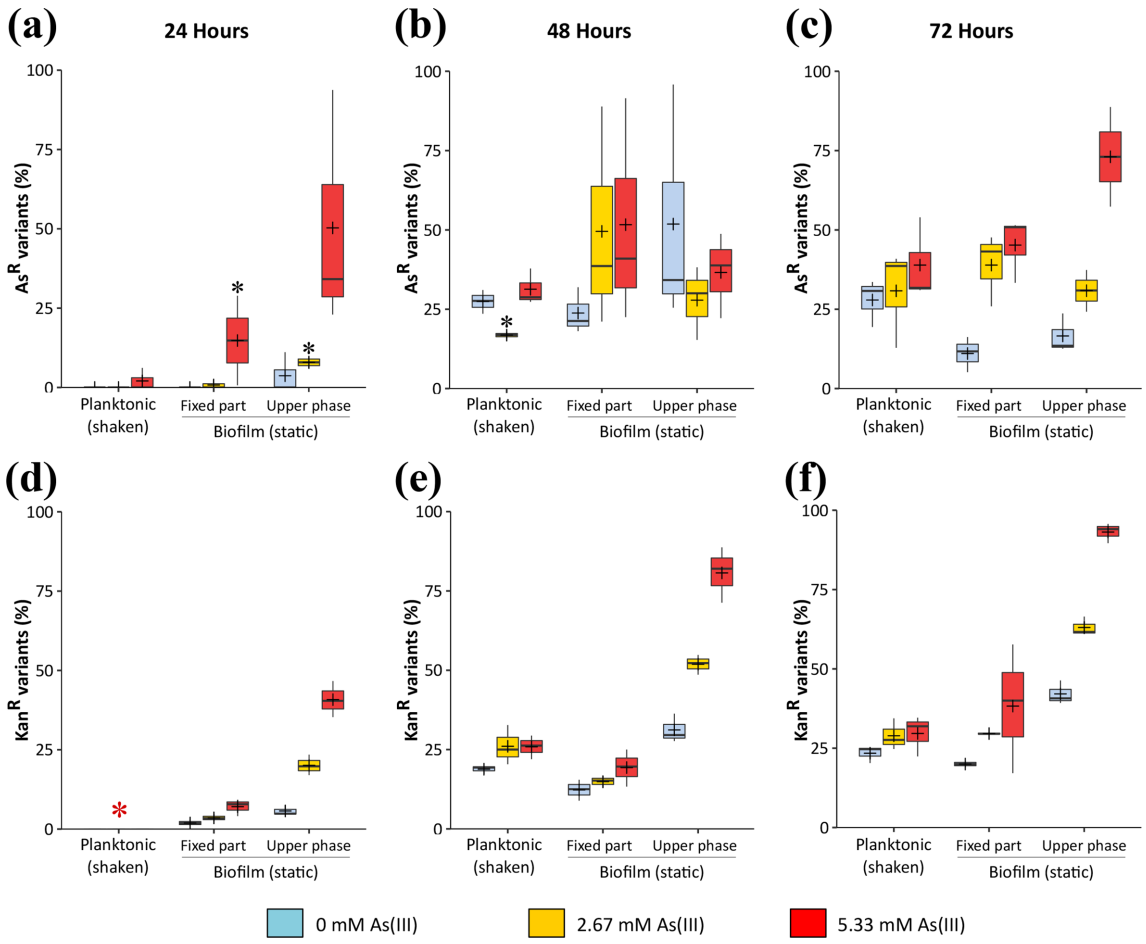


Fig. 2. Proportion (%) of bacterial cells that appear to be 'super resistant' to arsenite (a–c) or resistant to kanamycin (d–f) in strain *Thiomonas* sp. CB2 (variants). The percentage that represented variants was calculated by dividing the number of resistant cells by the total number of cells, with the boxplot showing median, quartiles and extremes; extreme values are indicated with bars. Crosses highlight the median values. The graphs are from samples grown as described in Fig. 1a for 24 h (a, d), 48 h (b, e) and 72 h (c, f). The red star indicates that the total number of cells was not evaluated for this condition. The data were calculated from three replicates except for samples indicated with the black star; in these cases, data were calculated from two replicates.

percentage of variants induced (Fig. 2a–c). In contrast, in the biofilm cultures, at 72 h, arsenite increased the percentage of variants present compared to controls (absence of arsenite), especially in the detached cells (Fig. 2c). In the presence of arsenite, after 24 and 48 h, in the fixed portion of the biofilm we observed a fluctuating number of variants between replicates generated (Fig. 2a, b). This observation suggests the fixed portion was highly heterogeneous in the presence of arsenite, especially after 48 h, while this population appeared to become more homogenous after 72 h, where less variability was observed. In the detached cells, a slight decrease in the number of variants was observed from 24 to 48 h in 5.33 mM arsenite; however, this was followed by an increase at 72 h. This could be due to the elimination of some cells during the renewal of culture medium at 24 h (see above and Methods). A previous study of CB2 biofilms found that cavities formed at the centre of the biofilms in the presence of arsenite were colonized by mobile cells [37]. Therefore, some of the cells supposedly collected from these detached cells could have originated from the fixed portion of the biofilm. Remarkably, the number of variants in the detached portion was surprisingly high (mean value >75%), when grown in the presence of 5.33 mM arsenite. In conclusion, although the appearance of variants was not unique to biofilm cultures, our results show that the percentage of variants generated increases in the presence of a high concentration of arsenite, especially in the detached cells after growth in biofilms (Fig. 2a–c).

The next step was to test whether the variants produced developed resistance to other toxic compounds in addition to arsenite. To test the specificity of the resistance developed, we quantified the frequency of kanamycin-resistant variant emergence when cells were grown in the presence of arsenite under the same conditions (Fig. 2d–f). Similar results were obtained compared to the generation of arsenic-resistant variants (Fig. 2a–c), suggesting that the variants generated did not

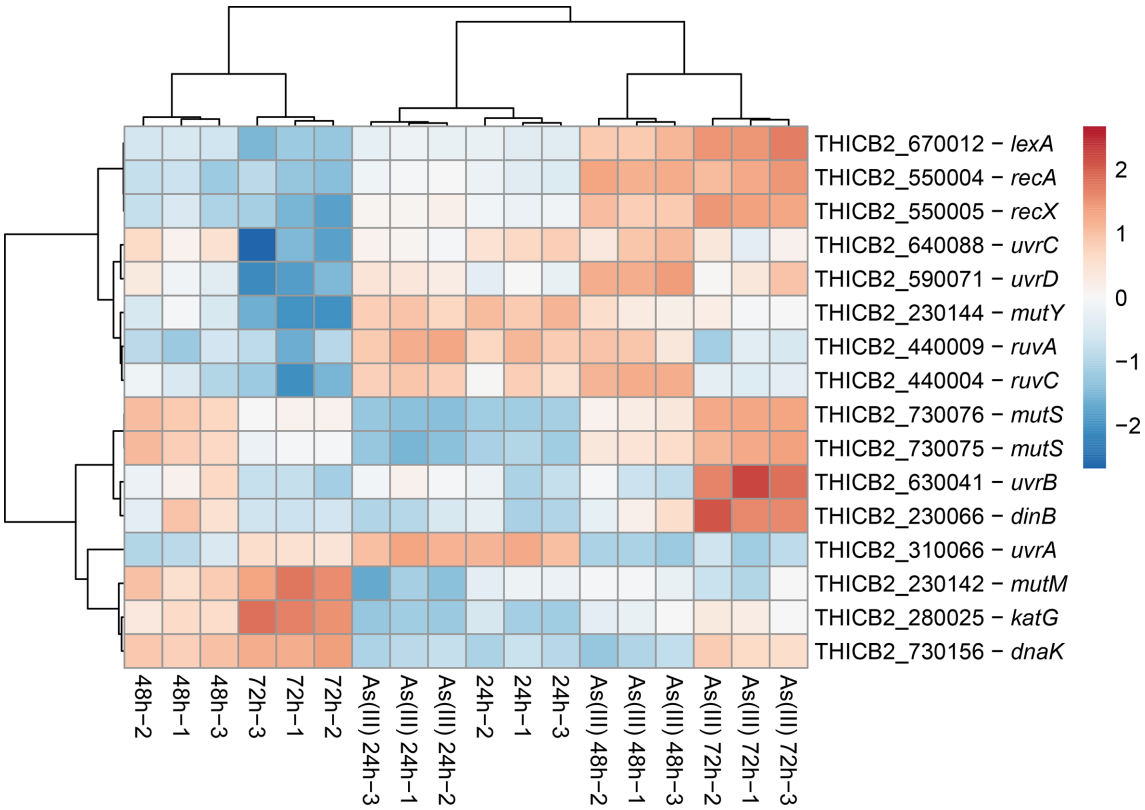


Fig. 3. Data analysis of previous RNA-seq experiments [29] revealing changes in expression of various genes in *Thiomonas* sp. CB2 grown in the presence of arsenite [5.33 mM] or absence of arsenite at 24, 48 and 72 h. Gene expression of 12 selected genes is shown. Each condition was replicated three times. The heatmap is based on the variance-stabilized transformed count matrix that has been adjusted for the replicate effect using the 'removeBatchEffect' function or the limma R package (version 3.44.3). Rows and columns have been re-ordered with hierarchical clustering (using the correlation and Euclidean distances, respectively, and the Ward aggregation criterion). The colour scale ranges from −2.5 to +2.5 as the rows of the matrix have been centred.

necessarily undergo genetic changes specific to enhancing arsenic resistance, but instead developed a general mechanism to survive exposure to toxic compounds. In a following experiment, we determined whether the cells underwent changes in gene expression that could lead to higher arsenite resistance or accumulated mutations in turn leading to resistance.

Arsenite impact on global gene expression and generation of mutations in cells grown in a biofilm

To better understand the mechanisms driving the appearance of resistant variants and variant phenotypes observed, we analysed data generated in a previous RNA-seq experiment [29] to identify genes expressed during biofilm development and/or regulated in response to exposure to arsenite. In this previous RNA-seq experiment, cells were incubated under biofilm conditions in the presence or absence of 5.33 mM arsenite and at 24, 48 and 72 h growth; RNA-seq was performed on fixed cells [29]. The conditions chosen for the RNA-seq experiment corresponded to those in which the rate of variant generation was highest (in the presence of 5.33 mM arsenite) compared to conditions where variant generation was lower (absence of arsenite) (Fig. 2). We previously observed that *aio* genes (involved in arsenite oxidation) were upregulated during biofilm formation in the presence of arsenic [29]. The ability to perform arsenite oxidation may help strains survive toxic environments; however, resistance also has been found to require a suite of *ars* genes [20]. Interestingly, expression of the *ars* genes involved in arsenic resistance was induced after 24 h, although their expression decreased after 48 h [29]. These previous observations indicated that cells within the biofilm probably oxidized arsenite, and both arsenite and arsenate were present and could be recognized by the cells. Here, we carefully re-examined the RNA-seq data to determine whether additional genes potentially involved in the appearance of variants were also upregulated.

We found that the expression of certain stress-response genes was induced during biofilm formation in the absence of arsenite but repressed in the presence of arsenite (Fig. 3), suggesting they may play different roles in biofilms depending on arsenite exposure. For example, genes encoding most chaperones and their expression regulators (*rpoH*, *dnaKJ*, *grpE*, *clp* genes, *groELS*, *htpG* and genes encoding small

Heat Schock Proteins (HSPs), as well as proteins involved in oxidative stress response (OSR) (*ahpCDF*, *katG*, *sodB* and *oxyR*) were downregulated in the presence (versus absence) of arsenite (Fig. 3). This suggests that cells grown in the presence of arsenite did not induce the expression of genes involved in general or oxidative stress, and that variant appearance was not due to a global stress response. In contrast, expression of genes involved in DNA repair (*uvrBCD*, *dinB*, *mutS* and to a lesser extent *mutY*) increased in the presence of arsenite at 72 h, whereas their expression decreased during biofilm formation in the absence of arsenite. Expression of genes involved in the SOS response or recombination (*recAX, ruvAC* and *lexA)* was induced in the presence of arsenite during biofilm development (after 48 or 72 h) (Fig. 3). However, expression of *mutM* and *uvrA* was repressed in the presence of arsenite, both known to prevent base pair damage due to either oxidative stress or mutations, or holiday junction resolution, respectively. In view of these RNA-seq data, we hypothesized that more variants appeared in the biofilm when in the presence of arsenite because genes involved in the error-prone DNA repair machinery are more highly expressed in such conditions. Therefore, we further tested the hypothesis that the rate of variant appearance was due (at least in part) to the accumulation of mutations when grown as a biofilm in the presence of arsenite.

A 'whole-population metagenomic sequencing' approach was used on the population grown in the conditions where we observed a higher number of variants, to determine whether mutations appeared in these conditions. Therefore, we grew CB2 in a biofilm for 72 h in the presence of 5.33 mM arsenite. To assess the mutations that were induced in these conditions, the DNA of the whole population was extracted, sequenced and compared to the CB2 genome. It is essential to note that the CB2 genome used for comparison was resequenced in this study (see Methods) and was obtained from a culture grown under planktonic conditions, in the absence of arsenite, which were the conditions we demonstrated led to a relatively low rate of variant appearance (Fig. 2). A total of 267 SNVs were identified; however, 232 were ultimately analysed post-read filtering, which involved setting a minimum threshold of coverage and quality score (see Methods). It has to be mentioned that these mutations were found only in a fraction of the population (see Tables S1–S9). Of all 232 SNVs, 39 were in non-coding regions (16.8%), 2 were found in the gene encoding the 23S rRNA (one substitution and one insertion) and 191 (82.3%) were in protein-encoding regions. Of the 191 mutations in protein-encoding regions, 41 (21.5%) corresponded to the insertion or deletion (indel) of 1 nucleotide, while 64 (33.5%) were synonymous (dS) and presumed neutral. The remaining 86 (45%) mutations were non-synonymous (dN). Ultimately, 127 mutations (including the 86 dN changes and 41 indels) likely affected protein sequence or length and could be considered to be changes at non-silent sites. A total of 39 mutations (representing 20.4% of the SNVs in coding regions) were found within coding sequences of either integrases or transposases. dN mutations or indels were identified in genes involved in various cellular process (Tables S1–S9) including lipid synthesis, general metabolism, protein folding (*hsp20*), replication (*dnaN*), cell division (*ftsW*), fimbrial biogenesis and cell-wall biosynthesis (*murG*). Interestingly, dN mutations or indels were found in *dinB* (involved in DNA repair, THICB2_v4_60517), in genes potentially involved in general toxic compound resistance [such as permeases of the drug/metabolite transporter superfamily (THICB2_v4_20175) with the EamA domain], in two genes involved in copper resistance – *copI* (THICB2_v4_10249) and *cueO* (THICB2_v4_70216), and in *aioA* (THICB2_v4_70221) involved in arsenite oxidation. A total of 31 mutations were found in the GI ICE19 (184721 kb), which is an integrative and conjugative element (ICE) carrying several genes involved in heavy-metal resistance [28]. This demonstrated that the frequency of mutations found in this ICE is higher than the rest of the genome (1 mutation per 5958 nucleotides in ICE19 compared to 1 mutation per 16664 nucleotides across the entire genome).

The frequencies of the mutation types are shown in Fig. S3. The percentages of mutation types observed were T/C (10.8%), C/A (9.5%), C/T (9.1%), A/G (9.1%), G/A (8.6%), G/T (11.6%) and insertion of 1 nucleotide (19%) (Fig. S3). These data revealed that point mutations appeared among the population within a biofilm grown in the presence of arsenite. However, this approach does not indicate whether there is a correlation between mutations and phenotypes nor reveal whether other mutations, such as rearrangements, took place in the genomes. To further explore genetic mutations and potential genomic rearrangement after biofilm exposure to arsenic, multiple variants were isolated and cryogenically preserved.

Selection and phenotype analysis of variants

We focused on 8 variants from the total of 124 arsenic-resistant and 8 antibiotic-resistant variants (see Methods), in order to study their phenotypes. According to the results of these phenotypic analyses, we selected three variants for genome sequencing and careful analysis. Of the 8 variants from the 124 arsenic-resistant variants, 3 were selected due to unique colony morphology compared to the parental strain (Bio14B1, Sup14A3 and Sup14A1), suggesting they possibly had impaired biofilm-formation capacities. Five others (Sup16B3, Bio17B3, Bio17A3, Sup16A1 and Bio16B1) were selected as they were generated after three or four biofilm cycles and were characterized by higher resistance to arsenite than other variants (MIC >14 mM) (Table 1). Four different approaches were undertaken with these eight variants to assess their phenotypes compared to the parental strain and investigate the mechanism leading to their enhanced resistance to arsenite.

First, to ensure that resistance to arsenite was not due to slower cell growth, the growth rates of the eight variants and original CB2 strain were measured. We determined that the same growth kinetics were observed for variants as well as CB2, all displaying exponential phrases between 24 and 40 h, reaching a final OD_{600} between 0.2 and 0.3

(see Fig. S4 for four of the growth curves obtained, additional data not shown). Variant Sup16B3 had a similar growth curve compared to CB2, whereas Bio17A3, Bio17B3 and Sup16A1 had longer lag phases. Growth rates were similar among all strains except for Bio17B3, which grew more rapidly after the lag phase.

Second, the ability to form biofilms was assessed with the eight selected variants at different concentrations of arsenite by crystal violet staining and compared to the original strain *Thiomonas* sp. CB2 (data not shown). Three general profiles were distinguished among these eight variants. The first profile was shared by variants that produced the most biofilms, notably in the presence of 2.67 mM arsenite. A second profile was characterized by Sup16B3, which produced less biofilm, compared to *Thiomonas* sp. CB2. A third profile was assigned to the remaining variants, that produced similar levels of biofilm as *Thiomonas* sp. CB2. We analysed in detail one variant representing each profile (Sup16B3, Bio17B3 and Bio14B1) using confocal analyses (Fig. S5) to determine whether resistance could be due to differential biofilm-formation capacities, since biofilms may enhance drug resistance [34]. The variants Bio14B1 and Bio17B3 had a marked increased capacity to form an extracellular matrix, whereas the number of fixed cells did not change compared to the parental strain CB2.

In parallel, eight other variants (OC1, OC2, OC3, OC4, OC5, OC6, OC7 and OC8) were generated and selected for their resistance to antibiotics, specifically kanamycin. Interestingly, OC7 was resistant to several other aminoglycosides. Therefore, we selected this multi-antibiotic-resistant variant and compared its drug-resistance capacity with that of Sup16B3 and Bio17B3 (Table 1). OC7 was more resistant to antibiotics than the two other variants and also slightly more resistant to arsenite than the original strain CB2.

Based on the phenotypic analyses outlined above, we continued analyses on three variants: OC7, Sup16B3 and Bio17B3. We tested the ability to oxidize arsenite in OC7, Sup16B3 and Bio17B3, as well as CB2, and found that while CB2 and OC7 could oxidize arsenite [26] (a potential mechanism of resistance [20]), Bio17B3 and Sup16B3 could not (Table 1, Fig. S6). This experiment was undertaken twice, and in both experiments little or no activity was detected in Sup16B3 or Bio17B3. These observations revealed that in these two variants, arsenite resistance is not due to an increase of arsenite oxidase activity. Additionally, while both H_2O_2 and arsenite induce oxidative stress, a disc diffusion assay revealed that no variants had a higher resistance to H_2O_2 than CB2. This suggested that the variants were not indirectly resistant to arsenic or kanamycin due to a higher resistance to oxidative stress. Since the arsenite resistance of these three variants was stable after 30 generations without exposure to arsenite, we hypothesized that resistance was likely due to mutations or genomic rearrangements. To test this, we sequenced the genomes of the three variants Bio17B3, Sup16B3 and OC7 in order to compare them to the parental strain CB2.

Table 2. Size (in nucleotides) of the chromosome and the chromosomal ICE19, as well as plasmids of CB2 and the variants

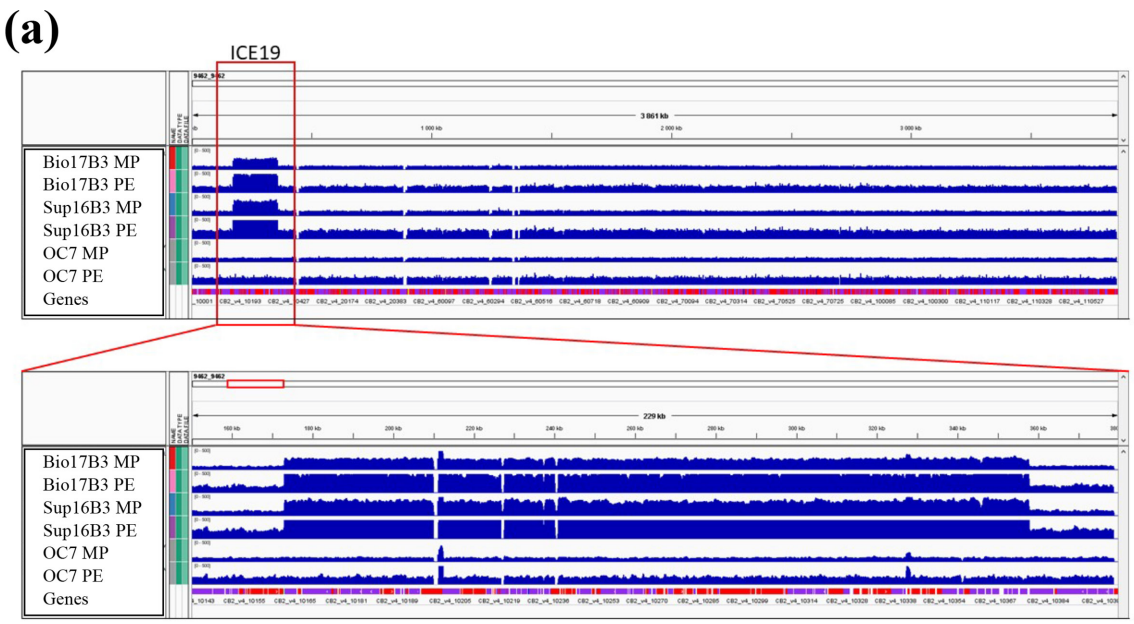
	Chromosome	ICE19	pA-	pB-	pC-
CB2	3866172	184821	39434	8340	0
OC7	3891617	184820	31833	9355	0
Bio17B3	3874064	184821	0	8714	204717
Sup16B3	3867933	184921	0	8082	184759

Genome analysis of three variants

The genomes of three variants (Bio17B3, Sup16B3 and OC7) were sequenced and compared to the genome of the parental strain, CB2 (which was resequenced using the same sequencing technology). The chromosome sizes were slightly different in Sup16B3 and Bio17B3, and marginally larger in OC7 (Table 2). However, these differences are almost completely due to a better quality assembly with the OC7 genome. The overall structures of the three variant chromosomes were compared to that of CB2 (Fig. S7) revealing that OC7 is very similar to CB2, since the synteny is conserved. The Bio17B3 and Sup16B3 genomes have highly similar structures, suggesting that no dramatic rearrangements occurred within the variant genomes compared to the original strain CB2.

OC7 harboured the same two plasmids as CB2 (pA-OC7 and pB-OC7), although they were slightly different in size (Table 2). In contrast, Bio17B3 and Sup16B3 only contained one plasmid each (pB-BIO17B3 and pB-SUP16B3, respectively), similar to the 8 kb plasmid of CB2 (pB-THICB2; Table 2). The synteny of these three plasmids was comparable, harbouring short inversions or insertions (Figs S8 and S9).

Remarkably, the two variants Bio17B3 and Sup16B3 each contained an additional circular piece of DNA (identified as pC-BIO17B3 and pC-SUP16B3, respectively) (Table 2, Fig. S10), which was highly similar to ICE19 of CB2. Interestingly, in the populations of these two variants, ICE19 was on the chromosome and/or found as a circular form such that on average, genes in this ICE were found in triplicate in these variants (Figs 4a and S10, Table 2). We determined that ICE19, as well as pC-BIO17B3 and pC-SUP16B3, carried genes involved in heavy-metal resistance. Specifically, genes for arsenic resistance (*ars* genes) and arsenite oxidation (*aio* genes) were carried on these plasmids and, therefore, present in triplicate in these variants. Expression of these genes was observed in the presence of arsenite in both variants (Tables S1–S9). However, our observations highlighted that arsenite oxidase activity is not the driving force of arsenite resistance in these two variants, since neither demonstrated arsenite oxidase activity (Fig. S6), even if their *aio* genes were expressed (Tables S1–S9) . In Bio17B3, one insertion sequence, BIO17B3–V2–50708 to 50712, carried on ICE19 was also duplicated in another region of the genome, BIO17B3–V2–10692 to 10696. This duplicate zone includes three genes potentially involved in cell adhesion (the first step in biofilm formation) which are:



(b)

ICE19 in CB2

```
attL (172891-172953)          AAAAAAGAAAAATACTAAGTGAATCAATCACGTGGGATTTTATTCTGGCGGAGAGAGGGGGAT
attR (357650-357712)          AAAATGAAAAAACCCTAAGTGAATCAATCACTTAGGGTCATGTTCTGGCGGAGAGAGGGGGAT
                              ****   *****  ***************** * ** *  * **********************

=> 49 nt; 6 mismatchs
```

ICE19 in Bio17 and Sup16

```
attL                          AAAAATGAAAAAACCCTAAGTGAATCAATCACTTAGGGTCATGTTCTGGCGGAGAGAGGGGGAT
attR                           AAAATGAAAAAACCCTAAGTGAATCAATCACTTAGGGTCATGTTCTGGCGGAGAGAGGGGGAT
                              ***********************************************************

=> 63 nt; 0 mismatchs
```

Fig. 4. ICE19 was found in triplicate in the Bio17B3 and the Sup16B3 variants, and 10 mutations were found in the variants in the *att* regions flanking this ICE19. (a) Read mapping profile obtained from sequencing data. This figure shows profiles for each set of sequencing data (PE, paired-end; MP, mate-pair data) and the three variant genomes, as well as the coverage of the reads mapped at each nucleotide of the CB2 genome. This figure was generated using IGV software. (b) Comparison of the *attL* and *attR* region of ICE19 in CB2 (at the top) and the *att* region in Bio17 and Sup16 (at the bottom).

(i) BIO17B3–V2–10692, only a fragment with homology to other genes known to play a role in adhesion and probably not functional; (ii) BIO17B3–V2–10693, homologous to *pilN*; and finally (iii) BIO17B3–V2–10694, homologous to *pilO*. Both *pilN* and *pilO* are required for fimbriae synthesis. The other two genes in this duplicated region encode the transposition machinery. It should be noted that three additional insertion sequences were found at other positions in the genomes: BIO17B3_v2_11743 to 11751, BIO17B3_v2_10704 to 10705 and BIO17B3_v2_12309. No such rearrangement was observed in the two other variant genomes.

We also carefully investigated the point mutations found in each genome using data from paired-end and mate-pair sequencing. A total of 55 mutations were observed in Bio17B3, 58 in Sup16B3 and 45 in OC7 (Tables S1–S9) . We found 37 mutations that were conserved across all three variants and in a population grown in the presence of arsenite in biofilm conditions (see above whole-population data). Such mutations may correspond to mutation hotspots. Transposases and integrases also appear to represent hotspots, since we observed 19 mutations in or near genes involved in transpositions. Moreover, 16 mutations were found within ICE19,

suggesting once more (see above) that ICE19 frequently experienced mutations. One gene (THICB2_v4_110298) encoding a methyl-accepting chemotaxis sensory transducer was mutated in all three variants as well as in the population (metagenomic). This gene may be involved in the regulation of motility and biofilm formation. Here, we observed that the variants have different capacities to form biofilms (Fig. S5). Mutations may concentrate at this locus that could, thus, correspond to mutation hotspots.

One deletion in a non-coding region was found in the pB-OC7 8 kb plasmid compared to pB-THICB2, and six mutations were unique to the OC7 chromosome. One of these unique mutations was in a gene encoding an alkyl hydroperoxide reductase, while four were in the genes encoding two putative prophage integrases, although these mutations proved to be neutral (dS). The only unique dN mutation was found in the gene encoding the formate dehydrogenase alpha subunit, FdhA. None of these mutations in OC7 clearly explained the antibiotic-resistance capability of this variant.

Fifty-five mutations were found both in the Bio17B3 and Sup16B3 variants but were not present in OC7 (Tables S1–S9). Two dS and one dN (insertion) mutations were found in Sup16B3 but not in the two other variants, the dN mutation was in one gene encoding a protein of unknown function (THICB2_v4_110244). A total of 37 of the mutations found in both Bio17B3 and Sup16B3 were in protein-encoding sequences. However, only 16 of these were indels or dN mutations. Among these dN mutations, four were found in integrase- or transposase-encoding regions, and one in a gene encoding a DinB-family protein. One nonsense mutation was found in THICB2_v4_20216, encoding a protein with an antisigma factor antagonist (ASA) domain. Such proteins (e.g. SpoIIAA in *Bacillus subtilis*) are known to inactivate antisigma via this ASA domain. The mutation found in Bio17B3 and Sup16B3 led to truncation of this protein (the serine residue was changed to a stop codon at position 1781). Additionally, four dN mutations were found in a putative signal transduction histidine kinase (THICB2_v4_20293). However, six dN mutations were found in other genes encoding proteins of unknown function and it is too speculative to state that one or several of these mutations might explain the loss of the arsenite activity in these two variants. Remarkably, we found 10 common mutations in both Sup16B3 and Bio17B3 in the region flanking ICE19 (*attL* and *attR*). These regions are known to be involved in the excision/integration of such genomic elements [45, 46]. In CB2, these two *att* regions are similar but not identical (6 mismatches in 49 nt). In both the Bio17B3 and Sup16B3 variants, the two *attL* and *attR* regions are composed of two regions of 63 nt long that are 100% identical (Fig. 4b). Thus, we observed that the *attL and attR* regions of ICE19 are longer and more similar between the variants Bio17B3 and Sup16B3, than in CB2.

DISCUSSION

The goal of our study was to understand how arsenic and/or biofilm conditions influence the emergence of variants, in order to ultimately determine how *Thiomonas* spp. evolve and adapt to the particularly toxic environments found at AMD sites. Arsenic affects the metabolism of most organisms and is thought to be mutagenic, carcinogenic and teratogenic [47, 48]. The toxicity of arsenic depends on its actual chemical state, and since arsenate is a structural analogue of phosphate, it competes with that essential ion in many enzymatic reactions. Arsenite has a very high affinity for thiol groups and may inhibit a variety of enzymes, and favour mutations [47, 48]. Due to the higher bioavailability in aqueous environments at neutral or acidic pH, arsenite is considered to be more toxic and dangerous than arsenate [47]. Extreme conditions create ideal environments to study the evolution of bacterial genomes. Furthermore, *Thiomonas* spp. have the capability of forming biofilms, likely a mechanism that allows for their survival in arsenic-contaminated sites. To understand how populations of *Thiomonas* spp. can survive in AMD sites, we explored how arsenic and growth in a biofilm affected gene expression, phenotype and mutation rates among *Thiomonas* sp. CB2 variants generated after growth in a biofilm exposed to toxic levels of arsenite.

We first aimed to determine how exposure to arsenite affected subpopulations in biofilms and observed that the rate of variant appearance increased in biofilms grown in arsenite (Fig. 2). Under biofilm conditions, in the upper phase, arsenite clearly had an effect at 24 and 72 h (Fig. 2c, e). For the fixed portion of cells, we also observed a less pronounced (but still present) effect (Fig. 2a, d). The 48 h time point (Fig. 2b) was taken carefully, since medium was renewed at 24 h and it is possible that the variants were more mobile or 'less attached' at 24 h and consequently preferentially eliminated when the medium was changed. Nevertheless, this study suggests that biofilm conditions alone are not enough to generate a significant increase of variants and arsenite did not enhance the generation of variants under planktonic culture conditions. To understand why growth in a biofilm and exposure to arsenite were required to induce variant generation, we undertook several approaches, including examining differences in transcription, changes to toxic compound resistance, and finally comparative genomics between three selected variants and the original parental strain, CB2.

Arsenite causes oxidative stress [20], and consequently may induce the expression of genes involved in the OSR and SOS response [2]. Interestingly, within the biofilm, we observed that the genes involved in the OSR (*sodB*, *katG* and *ahp*) were downregulated in the presence of arsenite, the same conditions where the number of variants had increased. Two explanations could be proposed to account for this downregulation. First, the biofilm matrix could have protected the cells, essentially stopping arsenite from affecting cells within the biofilm and inhibiting OSR induction. However, this is unlikely as the induction of *ars* and *aio* gene expression was observed in these conditions [28, 29]. Another explanation could be that arsenite is rapidly oxidized to arsenate. In planktonic conditions, it took over 35 h to completely oxidize 0.67 mM arsenite (Fig. S6). In our experiments, we added 2.67 or 5.33 mM arsenite (fourfold and eightfold more), in these

conditions we estimate that it took over 48 h to completely oxidize arsenite. There was probably a mixture of arsenite and arsenate in the medium, even after 72 h, as suggested by the induction of *ars* and *aio* genes in CB2 under the same planktonic or biofilm conditions as those used in our experiment (Tables S1–S9 and in published data [29]). Alternatively, the expression of these genes involved in the OSR could have been downregulated by an unknown mechanism. In a previous study of *Pseudomonas aeruginosa*, several genes encoding enzymes conferring protection against oxidative stress were downregulated in biofilms compared to planktonic conditions [49]. The authors also observed an increased frequency of variants resistant to antibiotics in the biofilms and proposed that the downregulation of the genes identified may enhance the rate of mutagenic events due to the accumulation of DNA damage. The downregulation we observed of *Thiomonas* genes involved in the OSR may similarly enhance the mutagenic events. In addition, we observed that genes involved in the error-prone DNA repair system (*dinB*) were upregulated in biofilms exposed to arsenic. Genes involved in error-prone DNA repair are regulated in other bacteria by the RecA–LexA-mediated SOS response and are induced when DNA is damaged [50]. Our transcriptomic approach revealed that in the conditions favouring emergence of variants, the *recA* gene and others involved in the SOS response (*recX, ruvAC* and *lexA*) were overexpressed, suggesting that at least a portion of the population was experiencing stressful conditions and, thus, DNA damage.

It is now recognized that bacterial mutation rates can increase dramatically in stressful conditions, a phenomenon known as SIM. SIM has been observed in several bacteria in response to different stresses and recently it was proposed that even a decrease in growth rate could lead to such an event [7, 51, 52]. The role of biofilms in the emergence of variants has been studied in pathogens [35, 53, 54]. For example, it was observed that subpopulations of *P. aeruginosa* generated extensive genetic diversity in biofilms, which has been shown to require *recA* [31]. Several studies have suggested that biofilms can serve as hotbeds of diversity that promote adaptation to harsh conditions [8, 31, 35, 49]. Indeed, the high diversity generated within biofilms offers protection against fluctuating challenging conditions, a principle that has been termed the insurance hypothesis, since it produces subpopulations with diverse functions, increasing the ability of the population to thrive as prevailing conditions change [35]. In the case of *Thiomonas* sp. CB2, our work revealed for what is believed to be the first time that both growth in biofilm and the presence of arsenite are necessary to generate high numbers of variants. Arsenite caused oxidative stress in CB2, likely producing reactive oxygen species, leading to breaks in the bacterial DNA [2, 20, 50]. Interestingly, in *E. coli*, it was proposed that both DNA breaks and the induction of *dinB* expression (via the SOS response and/or the *rpoS* regulon) are important factors leading to SIM [55].

Thus, in biofilms in the presence of arsenite, *Thiomonas* sp. CB2 could be subjected to SIM, since both DNA damage and *dinB* overexpression were observed. Our transcriptomic results revealed that *dinB* and *recA* expression (generally induced by the accumulation of DNA breaks [2, 50]), occur in the biofilm grown in the presence of arsenite. Furthermore, the whole-population metagenomic sequencing results revealed that many point mutations accumulated in the population, likely generated by error-prone DNA repair systems. After approximately 20 generations, over 200 mutations were acquired in this population. We would like to highlight that we resequenced the genome of the parent CB2 strain to ensure that the mutations observed were not due to sequencing errors. The mutations present in CB2 are only found in a fraction of this population (Tables S1–S9) and we cannot exclude the possibility that some may have appeared when the cells were grown on plates directly from the −80 °C stock (in the absence of arsenite) or during growth in the liquid pre-culture prior to inoculation in biofilm conditions. The majority of the mutations found in the whole-population genome (187 of 232) were not observed in the variants, and 21 mutations were found in one or two variants but not in the whole-population genome. These observations suggest that these mutations appeared independently, during growth under the various conditions tested. We observed the accumulation of certain point mutations (T/C, C/A, C/T, A/G, G/A, G/T and the insertion of 1 nucleotide). These mutations are possibly due to *dinB*, the error-prone DNA polymerase involved in translation repair. Alternatively, select mutations, for example G/T or G/A (20.3% of those observed), may be due to the reduced expression of genes involved in the high-fidelity DNA repair system (*mut* genes), which were downregulated at the beginning of biofilm formation (48 h) under our tested conditions. These genes are known to prevent base damage from oxidative stress. Altogether, these observations suggest that in *Thiomonas* sp. CB2 induction of genes involved in the SOS response (specifically *recA* and *dinB*) and downregulation of genes involved in the OSR partly accounted for the SIM (observed via the increased emergence of variants) in the biofilm when in the presence of arsenic.

Interestingly, *dinB* itself is involved in DNA repair and was shown to be mutated in our whole-genome-population sequencing experiment. Mutation in *dinB* may affect the mutation rate over a longer period, possibly generating a phenotype known as the 'hypermutator' phenotype. Indeed, populations have been found to occasionally evolve hypermutability due to mutations affecting the DNA repair system. In *E. coli*, long-term evolution experiments revealed that mutations arose in the population even when it was not subject to stress, possibly providing new capacities to the cells [6]. Hypermutability has been shown to affect the mutation rate, the activity of transposable elements and the rate of adaptation [6, 56]. Furthermore, hypermutability may be advantageous by allowing for the accumulation of beneficial mutations in populations over a short timescale, but deleterious over a long-term scale [4–6]. However, if the environment is sufficiently variable, and different stressful conditions occur frequently enough, SIM can be maintained [3]. The presence of several genotoxic compounds, such as metals, subjects bacterial populations in AMD sites

to several stresses and potentially to DNA damage. Single- or double-strand breaks followed by RecA repair in AMD conditions may result in higher mutation and recombination rates, possibly contributing to population diversification [31, 57–59]. Interestingly, it was observed that genomes of bacteria from AMD sites, including those of *Thiomonas*, have more genes involved in recombination and repair compared to other prokaryotes, likely enabling the rapid adaptation of these micro-organisms [10, 24, 28]. In the AMD site from which it was isolated, *Thiomonas* sp. CB2 would have experienced fluctuations in high concentrations of heavy metals and acidic pH, possibly maintaining SIM, driving the high diversity of *Thiomonas* genomes, and enhancing their adaptation capacities [24, 27–29].

In this study, we aimed to determine whether certain stressful conditions found in AMD sites, such as the presence of arsenite, could induce SIM in *Thiomonas* sp. CB2. The natural environment from which *Thiomonas* sp. CB2 was originally isolated is characterized by relatively high concentrations of dissolved arsenic ranging from 1 to 3.5 mM, with arsenite dominating as the primary form [16, 60]. In our experiments, we utilized similar (2.67 mM) as well as relatively higher (5.33 mM) concentrations of arsenite compared to the original isolation site. We did not test all of the variable conditions found across AMD sites, many of which have lower concentrations of dissolved arsenic; therefore, the conditions used here were quite different to those naturally found in many AMD sites. Moreover, we did not study the evolution of CB2 or its variants over a long time course. To further understand genome evolution in *Thiomonas*, it would be interesting to determine whether other stressful conditions found in AMD sites induce SIM in *Thiomonas* sp. CB2. It is likely that SIM led to the emergence of the variants Sup16B3, Bio17B3 and OC7, each have several mutations in their genomes. These variants, selected after CB2 was grown as a biofilm in the presence of arsenite, were more resistant to arsenic and antibiotics than the original strain, and acquired other capacities, possibly advantageous in certain circumstances and neutral or detrimental in others. Interestingly, the strains harboured differential abilities to form biofilms compared to the original CB2 (Fig. S5). Genome analysis of these three variants highlighted several mutations; however, there was no clear correlation between phenotype and the mutations observed, partly because several of these mutations were in genes of unknown function. To test the effect of each mutation on the phenotype, transformation, targeted mutagenesis or functional complementation must be optimized in *Thiomonas* sp. CB2, although attempts to obtain mutant *Thiomonas* strains have thus far failed.

It has been shown that mutational hotspots can occur in stressful conditions due to DNA breaks [9, 35, 55]. In our experiments, we observed that the frequency of mutations found in the ICE19, carrying several genes involved in arsenic resistance, is higher compared to the rest of the genome (1 mutation per 5958 nt in ICE19 compared to 1 mutation per 16 664 nt across the entire genome). The fact that this GI is subject to mutation in *Thiomonas* may be important for the evolution of resistance among these bacteria. ICE19, a previously described GI in CB2 [28], has characteristics of an ICE, essentially a mobile DNA element encoding fully functioning conjugative machinery present in two states (either excised or integrated) [45, 46]. Generally, ICEs are silent, integrated in the genome and vertically transmitted. In some conditions and select 'dedicated' cells, the ICE is excised after site-specific recombination between duplicated extremities (*attL* and *attR*), which requires an integrase and sometimes auxiliary proteins known as excisionases, recombination directionality factors or integration host factors. In some cases, a DDE transposase is involved in the replicative excision process. At low frequency, additional chromosomal regions can be added into the excised ICEs, as a result of defective excision [46]. The circular form of an ICE can be potentially transferred horizontally by conjugation, involving an *oriT* sequence and a relaxase-encoding gene. The transferred circular ICE can then integrate in the new host genome at specific recombination sites involving the *attI* (also known as *attP*) region of the ICE and a similar site in the genome known as *attB*. This site-specific integration is catalysed by an integrase, and results in the formation of the *attL* and *attR* duplicated extremities [45, 46]. The excision, transfer and integration processes are regulated in almost all bacteria where such ICEs have been studied in detail; however, the regulation process may be very different from one ICE to another [46]. ICEs often carry cargo genes, such as antibiotic or virulence genes, that can confer the new host selective advantages, possibly playing a major role in bacterial adaptation and evolution [45, 46]. ICE19 of *Thiomonas* sp. CB2 carries genes implicated in resistance to various heavy metals (arsenic, mercury, copper, zinc and cadmium), sulfate assimilation (*cys* genes) and biotin synthesis (*bio* genes). It also harbours many genes encoding integrases, as well as genes involved in conjugation [28]. At the two extremities of ICE19, 49 bp nearly perfect direct repeats called *attL* and *attR* have been identified, and the region between these two repeats contains the *tra* and *trb* genes potentially involved in conjugation [28]. The amplification of the junction boundary formed by this ICE excision using PCR was possible, suggesting that it could be excised in some cells and, therefore, was functional [28].

The two *att* regions of ICE19 were more similar to each other in the two arsenic-resistant variants (Sup16B3 and Bio17B3) than in CB2, due to the presence of several shared mutations. As mentioned above, in each of these variants the region was 63 nt long, while in CB2 it was only 49 nt and only shared 88% identity with the region found in the variants. These two regions are probably important for the excision/integration of this ICE, as described elsewhere [46]. Interestingly, this ICE was present in the population of the two variants both in the integrated form and the stable excised form. Using our data, we were not able to determine whether select cells had one copy integrated in the genome (as in CB2), while others had excised forms, or whether

the same cell contained both the integrated and excised forms. Delavat and collaborators have suggested that in the case of the *Pseudomonas* ICE*clc*, the excised and integrated forms of ICEs do not coexist in an individual cell, because the excision process is tightly regulated, and the two states have different requirements, a phenomenon referred to as bistability [46, 61]. Nevertheless, in some cases, such coexistence may occur when an element relies on replication for excision (as in *B. subtilis*) or if mutations could impair this bistability [46]. In our data, it appeared that the ICE19 was stable as an excised form at least in a portion of the population in Sup16B3 and Bio17B3. Once in the excised form, this ICE could potentially be transferred to other bacteria. We previously tried to transfer ICE19 by conjugation from *Thiomonas* sp. CB2, Sup16B3 or Bio17B3 to an *aio* mutant of *H. arsenicoxydans*, which had lost the capacity to oxidize arsenite [62, 63], in order to select transconjugants that exhibited arsenite oxidase activity. Unfortunately, this approach did not work, possibly because the transfer frequency was very low (as demonstrated in the case of other ICEs), if post-transfer ICE19 was not integrated, or if genes involved in arsenite oxidation were not expressed (or inactive). In the future, several other recipient cells must be tested to determine whether this ICE could be transferred by conjugation.

The mutations found in the *att* region in Sup16B3 and Bio17B3 could, in part, play a role in the increased proportion of the cells carrying the ICE19 in an excised form. Excision, which also requires the integrase, is performed via recombination between the *attR* and *attL* sites. We observed that whereas the two sites were not identical to CB2 (6 mismatches in 49 nt), in Bio17B3 and Sup16B3 both *attL* and *attR* are highly conserved (100% identity) over a longer region (63 nt). We hypothesize that the excision may be more efficient thanks to improved excisionase activity or perhaps RecA-dependent recombination. The shortest length of sequence homology necessary for efficient RecA-dependent recombination can vary greatly, depending on the bacteria, from 23 nt for *E. coli*, approximately 50 nt for *Ralstonia solanacearum*, to approximately 70 nt for *B. subtilis* [64–66]. A relationship between the flanking region length and recombination efficiency was demonstrated for several bacteria, and the fact that the *att* region size increased from 49 to 63 nt, while the identity also increased between both *att* regions, may favour the RecA-dependent mechanism as well as site-specific recombination and consequently the excision of the ICE19. Interestingly, it was demonstrated that the expression of SOS genes may influence the excision of such ICEs in some bacteria [46]. More generally, these genes might alter the role of LexA, the regulator of the SOS response in the regulation of genes from mobile genetic elements (MGEs) [67]. In our experiments, under the conditions when the SOS response was induced, variants appeared that were more resistant to arsenite, some of which accumulated mutations in one putative MGE (ICE19), leading to an increase in the excision efficiency of this ICE19. Finally, the variants we isolated bring new opportunities to decipher the regulation of excision and integration of ICEs and how such MGEs could evolve in response to heavy metals or growth within biofilm communities.

Funding information
This work was supported by the Université de Strasbourg, the Centre National de la Recherche Scientifique (CNRS) and the Region Alsace (J.F.). This study was also financed by THIOFILM (ANR-12-ADAP-0013) projects. K.C.F., O.C. and J.F. were supported by the Agence Nationale de la Recherche, ANR THIOFILM (ANR-12-ADAP-0013). The Transcriptome and EpiGenome Platform is a member of the France Génomique consortium (ANR10-NBS-09-08).

Author contributions
K. C. F., S. F., A. H, J. F., S. K., M. P., O. C., C. P., J. D., D. L. performed research; D. R., S. C., R. T. and H. V. contributed new reagents/analytical tools; K. C. F., S. F., J. F., H. V., R. B., J.-Y. C., V. B. and F. A.-P. analysed data; and K. C. F. and F. A.-P. wrote the paper.

Conflicts of interest
The authors declare that there are no conflicts of interest.

References

1. Brooks AN, Turkarslan S, Beer KD, Lo FY, Baliga NS. Adaptation of cells to new environments. *Wiley Interdiscip Rev Syst Biol Med* 2011;3:544–561.
2. Shinagawa H. SOS response as an adaptive response to DNA damage in prokaryotes. *EXS* 1996;77:221–235.
3. Lukačišinová M, Novak S, Paixão T. Stress-induced mutagenesis: stress diversity facilitates the persistence of mutator genes. *PLoS Comput Biol* 2017;13:e1005609.
4. Wielgoss S, Barrick JE, Tenaillon O, Wiser MJ, Dittmar WJ *et al*. Mutation rate dynamics in a bacterial population reflect tension between adaptation and genetic load. *Proc Natl Acad Sci USA* 2013;110:222–227.
5. Good BH, Desai MM. Evolution of mutation rates in rapidly adapting asexual populations. *Genetics* 2016;204:1249–1266.
6. Lenski RE. Experimental evolution and the dynamics of adaptation and genome evolution in microbial populations. *ISME J* 2017;11:2181–2194.
7. Agashe D. The road not taken: could stress-specific mutations lead to different evolutionary paths? *PLoS Biol* 2017;15:e2002862.
8. Rainey PB, Travisano M. Adaptive radiation in a heterogeneous environment. *Nature* 1998;394:69–72.
9. Rosenberg SM. Evolving responsively: adaptive mutation. *Nat Rev Genet* 2001;2:504–515.
10. Li S-J, Hua Z-S, Huang L-N, Li J, Shi S-H *et al*. Microbial communities evolve faster in extreme environments. *Sci Rep* 2014;4:6205.
11. Méndez-García C, Peláez AI, Mesa V, Sánchez J, Golyshina OV *et al*. Microbial diversity and metabolic networks in acid mine drainage habitats. *Front Microbiol* 2015;6:475.
12. Denef VJ, Mueller RS, Banfield JF. AMD biofilms: using model communities to study microbial evolution and ecological complexity in nature. *ISME J* 2010;4:599–610.
13. Denef VJ, VerBerkmoes NC, Shah MB, Abraham P, Lefsrud M *et al*. Proteomics-inferred genome typing (PIGT) demonstrates inter-population recombination as a strategy for environmental adaptation. *Environ Microbiol* 2009;11:313–325.
14. Denef VJ, Kalnejais LH, Mueller RS, Wilmes P, Baker BJ *et al*. Proteogenomic basis for ecological divergence of closely related bacteria in natural acidophilic microbial communities. *Proc Natl Acad Sci USA* 2010;107:2383–2390.
15. Denef VJ, Banfield JF. In situ evolutionary rate measurements show ecological success of recently emerged bacterial hybrids. *Science* 2012;336:462–466.

16. Egal M, Casiot C, Morin G, Elbaz-Poulichet F, Cordier MA *et al*. An updated insight into the natural attenuation of As concentrations in Reigous Creek (southern France). *Appl Geochem* 2010;25:1949–1957.
17. Singer PC, Stumm W. Acidic mine drainage: the rate-determining step. *Science* 1970;167:1121–1123.
18. Volant A, Bruneel O, Desoeuvre A, Héry M, Casiot C *et al*. Diversity and spatiotemporal dynamics of bacterial communities: physicochemical and other drivers along an acid mine drainage. *FEMS Microbiol Ecol* 2014;90:247–263.
19. Bertin PN, Heinrich-Salmeron A, Pelletier E, Goulhen-Chollet F, Arsène-Ploetze F *et al*. Metabolic diversity among main microorganisms inside an arsenic-rich ecosystem revealed by meta- and proteo-genomics. *ISME J* 2011;5:1735–1747.
20. Kruger MC, Bertin PN, Heipieper HJ, Arsène-Ploetze F. Bacterial metabolism of environmental arsenic-mechanisms and biotechnological applications. *Appl Microbiol Biotechnol* 2013;97:3827–3841.
21. Muller D, Médigue C, Koechler S, Barbe V, Barakat M *et al*. A tale of two oxidation states: bacterial colonization of arsenic-rich environments. *PLoS Genet* 2007;3:e53.
22. Duquesne K, Lieutaud A, Ratouchniak J, Muller D, Lett MC *et al*. Arsenite oxidation by a chemoautotrophic moderately acidophilic *Thiomonas* sp.: from the strain isolation to the gene study. *Environ Microbiol* 2008;10:228–237.
23. Bruneel O, Personné J-C, Casiot C, Leblanc M, Elbaz-Poulichet F *et al*. Mediation of arsenic oxidation by *Thiomonas* sp. in acid-mine drainage (Carnoulès, France). *J Appl Microbiol* 2003;95:492–499.
24. Arsène-Ploetze F, Koechler S, Marchal M, Coppée J-Y, Chandler M *et al*. Structure, function, and evolution of the *Thiomonas* spp. genome. *PLoS Genet* 2010;6:e1000859.
25. Battaglia-Brunet F, Joulian C, Garrido F, Dictor M-C, Morin D *et al*. Oxidation of arsenite by *Thiomonas* strains and characterization of *Thiomonas arsenivorans* sp. nov. *Antonie Van Leeuwenhoek* 2006;89:99–108.
26. Bryan CG, Marchal M, Battaglia-Brunet F, Kugler V, Lemaitre-Guillier C *et al*. Carbon and arsenic metabolism in *Thiomonas* strains: differences revealed diverse adaptation processes. *BMC Microbiol* 2009;9:127.
27. Arsène-Ploetze F, Chiboub O, Lièvremont D, Farasin J, Freel KC *et al*. Adaptation in toxic environments: comparative genomics of loci carrying antibiotic resistance genes derived from acid mine drainage waters. *Environ Sci Pollut Res Int* 2018;25:1470–1483.
28. Freel KC, Krueger MC, Farasin J, Brochier-Armanet C, Barbe V *et al*. Adaptation in toxic environments: arsenic genomic islands in the bacterial genus *Thiomonas*. *PLoS One* 2015;10:e0139011.
29. Farasin J, Koechler S, Varet H, Deschamps J, Dillies M-A *et al*. Comparison of biofilm formation and motility processes in arsenic-resistant *Thiomonas* spp. strains revealed divergent response to arsenite. *Microb Biotechnol* 2017;10:789–803.
30. Farasin J, Andres J, Casiot C, Barbe V, Faerber J *et al*. *Thiomonas* sp. CB2 is able to degrade urea and promote toxic metal precipitation in acid mine drainage waters supplemented with urea. *Front Microbiol* 2015;6:993.
31. Boles BR, Thoendel M, Singh PK. Self-generated diversity produces "insurance effects" in biofilm communities. *Proc Natl Acad Sci USA* 2004;101:16630–16635.
32. Aminov RI. Horizontal gene exchange in environmental microbiota. *Front Microbiol* 2011;2:158.
33. Madsen JS, Burmølle M, Hansen LH, Sørensen SJ. The interconnection between biofilm formation and horizontal gene transfer. *FEMS Immunol Med Microbiol* 2012;65:183–195.
34. Koechler S, Farasin J, Cleiss-Arnold J, Arsène-Ploetze F. Toxic metal resistance in biofilms: diversity of microbial responses and their evolution. *Res Microbiol* 2015;166:764–773.
35. Steenackers HP, Parijs I, Dubey A, Foster KR, Vanderleyden J. Experimental evolution in biofilm populations. *FEMS Microbiol Rev* 2016;40:373–397.
36. Stoodley P, Sauer K, Davies DG, Costerton JW. Biofilms as complex differentiated communities. *Annu Rev Microbiol* 2002;56:187–209.
37. Marchal M, Briandet R, Halter D, Koechler S, DuBow MS *et al*. Subinhibitory arsenite concentrations lead to population dispersal in *Thiomonas* sp. *PLoS One* 2011;6:e23181.
38. Weeger W, Lièvremont D, Perret M, Lagarde F, Hubert JC *et al*. Oxidation of arsenite to arsenate by a bacterium isolated from an aquatic environment. *Biometals* 1999;12:141–149.
39. Vallenet D, Calteau A, Cruveiller S, Gachet M, Lajus A *et al*. Microscope in 2017: an expanding and evolving integrated resource for community expertise of microbial genomes. *Nucleic Acids Res* 2017;45:D517–D528.
40. Ning Z, Cox AJ, Mullikin JC. SSAHA: a fast search method for large DNA databases. *Genome Res* 2001;11:1725–1729.
41. Smith TF, Waterman MS. Identification of common molecular subsequences. *J Mol Biol* 1981;147:195–197.
42. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 2009;10:R25.
43. Liao Y, Smyth GK, Shi W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res* 2013;41:e108.
44. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014;15:550.
45. Bellanger X, Payot S, Leblond-Bourget N, Guédon G. Conjugative and mobilizable genomic islands in bacteria: evolution and diversity. *FEMS Microbiol Rev* 2014;38:720–760.
46. Delavat F, Miyazaki R, Carraro N, Pradervand N, van der Meer JR. The hidden life of integrative and conjugative elements. *FEMS Microbiol Rev* 2017;41:512–537.
47. Kruger MC, Bertin PN, Heipieper HJ, Arsène-Ploetze F. Bacterial metabolism of environmental arsenic-mechanisms and biotechnological applications. *Appl Microbiol Biotechnol* 2013;97:3827–3841.
48. Hughes MF. Arsenic toxicity and potential mechanisms of action. *Toxicol Lett* 2002;133:1–16.
49. Driffield K, Miller K, Bostock JM, O'Neill AJ, Chopra I. Increased mutability of *Pseudomonas aeruginosa* in biofilms. *J Antimicrob Chemother* 2008;61:1053–1056.
50. Baharoglu Z, Mazel D. Sos, the formidable strategy of bacteria against aggressions. *FEMS Microbiol Rev* 2014;38:1126–1145.
51. Maharjan RP, Ferenci T. A shifting mutational landscape in 6 nutritional states: stress-induced mutagenesis as a series of distinct stress input-mutation output relationships. *PLoS Biol* 2017;15:e2001477.
52. Maharjan RP, Ferenci T. The impact of growth rate and environmental factors on mutation rates and spectra in *Escherichia coli*. *Environ Microbiol Rep* 2018;10:626–633.
53. Churton NWV, Misra RV, Howlin RP, Allan RN, Jefferies J *et al*. Parallel evolution in *Streptococcus pneumoniae* biofilms. *Genome Biol Evol* 2016;8:1316–1326.
54. Esposito A, Pompilio A, Bettua C, Crocetta V, Giacobazzi E *et al*. Evolution of *Stenotrophomonas maltophilia* in cystic fibrosis lung over chronic infection: a genomic and phenotypic population study. *Front Microbiol* 2017;8:1590.
55. Galhardo RS, Do R, Yamada M, Friedberg EC, Hastings PJ *et al*. DinB upregulation is the sole role of the SOS response in stress-induced mutagenesis in *Escherichia coli*. *Genetics* 2009;182:55–68.
56. Tenaillon O, Barrick JE, Ribeck N, Deatherage DE, Blanchard JL *et al*. Tempo and mode of genome evolution in a 50,000-generation experiment. *Nature* 2016;536:165–170.
57. Allegrucci M, Sauer K. Formation of *Streptococcus pneumoniae* non-phase-variable colony variants is due to increased mutation frequency present under biofilm growth conditions. *J Bacteriol* 2008;190:6330–6339.

58. Boles BR, Singh PK. Endogenous oxidative stress produces diversity and adaptability in biofilm communities. *Proc Natl Acad Sci USA* 2008;105:12503–12508.
59. van der Veen S, Abee T. Generation of variants in *Listeria monocytogenes* continuous-flow biofilms is dependent on radical-induced DNA damage and RecA-mediated repair. *PLoS One* 2011;6:e28590.
60. Casiot C, Morin G, Juillot F, Bruneel O, Personné J-C *et al*. Bacterial immobilization and oxidation of arsenic in acid mine drainage (Carnoulès creek, France). *Water Res* 2003;37:2929–2936.
61. Delavat F, Mitri S, Pelet S, van der Meer JR. Highly variable individual donor cell fates characterize robust horizontal gene transfer of an integrative and conjugative element. *Proc Natl Acad Sci USA* 2016;113:E3375–E3383.
62. Muller D, Lièvremont D, Simeonova DD, Hubert JC, Lett MC. Arsenite oxidase *aox* genes from a metal-resistant beta-proteobacterium. *J Bacteriol* 2003;185:135–141.
63. Koechler S, Cleiss-Arnold J, Proux C, Sismeiro O, Dillies MA *et al*. Multiple controls affect arsenite oxidase gene expression in *Herminiimonas arsenicoxydans*. *BMC Microbiol* 2010;10:53.
64. Shen P, Huang HV. Homologous recombination in *Escherichia coli*: dependence on substrate length and homology. *Genetics* 1986;112:441–457.
65. Bertolla F, Van Gijsegem F, Nesme X, Simonet P. Conditions for natural transformation of *Ralstonia solanacearum*. *Appl Environ Microbiol* 1997;63:4965–4968.
66. Khasanov FK, Zvingila DJ, Zainullin AA, Prozorov AA, Bashkirov VI. Homologous recombination between plasmid and chromosomal DNA in *Bacillus subtilis* requires approximately 70 bp of homology. *Mol Gen Genet* 1992;234:494–497.
67. Fornelos N, Browning DF, Butala M. The use and abuse of LexA by mobile genetic elements. *Trends Microbiol* 2016;24:391–401.