UNIVERSITÉ DE STRASBOURG

ÉCOLE DOCTORALE ED414

UMR 7156 Génétique Moléculaire Génomique et Microbiologie

THÈSE présentée par :

Jérémy ANDRES

soutenue le : **15 octobre 2014**

pour obtenir le grade de : **Docteur de l'université de Strasbourg**

Discipline/ Spécialité : Sciences du vivant / Aspects moléculaires et cellulaires de la biologie

Mécanismes adaptatifs et interactions métaboliques au sein de communautés microbiennes soumises au stress arsénié

THÈSE dirigée par :

M BERTIN Philippe	Professeur, Université de Strasbourg

RAPPORTEURS :

M PEYRET Pierre	Professeur, Université d'Auvergne
M NORMAND Philippe	Directeur de Recherche, Université Claude Bernard Lyon I

AUTRES MEMBRES DU JURY :

Mme SCHALK Isabelle	Directeur de Recherche, Université de Strasbourg
M MULLER Daniel	Maître de Conférences, Université Claude Bernard Lyon I

REMERCIEMENTS

Cette thèse a été financée par le Ministère de l'Éducation nationale, de l'Enseignement supérieur et de la Recherche et a été réalisée au sein du laboratoire de Génétique Moléculaire, Génomique et Microbiologie de l'Université de Strasbourg (UMR7156) dirigé successivement par Serge Potier et Ivan Tarassov. Je les remercie de m'avoir accueilli au sein de leur unité.

Je tiens à remercier particulièrement Philippe Bertin pour avoir dirigé ce travail de thèse, Florence Arsène-Ploetze pour son implication et son étroite collaboration ainsi que Frédéric Plewniak pour ses remarques constructives dans les domaines de la bio-informatique et des statistiques.

Je remercie également les autres membres de l'équipe, Sandrine, Jessica, David, Kelle, Julien, Antoine, François, Didier et Marie-Claire pour leur disponibilité et leur soutien ainsi que toutes les personnes du laboratoire qui, d'une manière ou d'une autre, m'ont aidé dans ce travail.

Je remercie vivement Dimitri Heintz et Raphaël Lugan pour m'avoir initié à la métabolomique et s'être impliqués lors de collaborations scientifiques enrichissantes ainsi que Jacques Faerber pour son aide précieuse en microscopie électronique. Je remercie également Daniel Muller, Florent Lassalle et Céline Brochier-Armanet pour leurs conseils avisés et leur aide respective.

Un grand merci à mes parents et mon frère qui ont su m'encourager et me soutenir durant toutes ces années d'études universitaires. Je n'oublie pas Maud, qui à su me donner confiance et m'épauler durant cette thèse, ainsi que tous mes amis.

Enfin, je tiens à exprimer toute ma reconnaissance aux membres du jury pour avoir accepté d'évaluer la qualité de ce travail.

SOMMAIRE

LISTE DES TABLES ET FIGURES

AVANT-PROPOS

Dans la nature, les micro-organismes sont soumis à de nombreux stress. Physiques ou chimiques, ces derniers éprouvent continuellement les microbes dans leur capacité à survivre et à s'adapter. Ceux-ci sont cependant capables d'adopter différentes stratégies comportementales ou physiologiques afin de répondre aux contraintes de leur niche écologique. À cet égard, l'arsenic est un contaminant majeur de nombreux écosystèmes aquatiques, à l'origine de nombreux problèmes de santé publique. De manière intéressante, certains organismes ont su adapter leur physiologie à cet élément et sont capable de résister à sa toxicité ou de l'intégrer dans leur métabolisme énergétique. Ces dernières capacités font des micro-organismes des intervenants majeurs du cycle biogéochimique de l'arsenic en influençant sa chimie, sa mobilité et sa biodisponibilité. De plus, en conditions naturelles, les micro-organismes ne sont pas isolés mais se développent au sein de communautés microbiennes. Ces structures multi-spécifiques sont le théâtre de diverses interactions cruciales au fonctionnement de l'écosystème et pouvant apporter un avantage aux partenaires en favorisant leur survie et croissance en milieu toxique.

L'introduction de ce manuscrit est un article de revue scientifique soumis à *Environmental Microbiology Reports* à la demande de l'éditeur.

INTRODUCTION

Arsenic is a natural component of the Earth's crust and a constituent of many minerals, and is therefore ubiquitously present on the surface of the globe. Natural geological processes and anthropogenic industrial activities have led to the release and prevalence of this toxic element in many regions and ecosystems (Nordstrom, 2002). Chronic exposure to the resulting contaminated waters is responsible for several human diseases, including cardiovascular and neurological disorders, cancer and diabetes (Kapaj et al., 2006). This toxic element has therefore been ranked by the World Health Organization among the ten major public health threatening chemicals. In addition, depending on the physico-chemical conditions and the biological activities at work, several forms of arsenic can be retrieved from aquatic environments, showing various levels of toxicity and bioavailability (Fig. 1; Rahman and Hassler, 2014). Arsenic can occur in water in several different oxidation and methylation states: the most common forms are inorganic arsenite and arsenate, i.e. As(III) and As(V) (Fig. 1A; Cullen and Reimer, 1989).

Micro-organisms have evolved several ways of coping with arsenic. On the one hand, resistance mechanisms such as those based on active extrusion, extracellular precipitation and chelation, and intracellular transformation, compartmentalization and sequestration reduce the mobility and the bioavailability of this element (Slyemi and Bonnefoy, 2012). On the other hand, some micro-organisms use inorganic or methylated forms of arsenic as a substrate in their energy or carbon metabolism, respectively (Fig. 1B; Maki et al., 2004; Silver and Phung, 2005b; Stolz et al., 2006; Páez-Espino et al., 2009; Tsai et al., 2009; Yoshinaga et al., 2011). Microbial processes therefore play a key role in arsenic's biogeochemical cycle (Fig. 1B; Gadd, 2010) and are the main factors affecting its mobilization, since they influence its passage between the biotic and abiotic compartments (Oremland and Stolz, 2005; Lièvremont et al., 2009). In this context, studies on the various microbial mechanisms whereby micro-organisms either thrive in contaminated ecosystems or metabolize arsenic are of particular interest, from both the fundamental and applied biological points of view. On the one hand, these processes may have been of great importance in the early stages of life or may result in phenotypical benefits making the colonization of specific ecological niches possible (Lebrun et al., 2003; Kulp et al., 2008). On the other hand, studies on the mechanisms involved should also shed light on the mobility of arsenic in given environments and help to develop bioremediation strategies (Tsai et al., 2009). Although the reactions in which arsenic is involved have been fairly well documented, little is known so far about the cellular processes underlying the overall responses of cells to this toxic element.

The use of molecular approaches such as genome sequencing and transcriptome, proteome and metabolome profiling methods, which have also been applied to higher organisms such as plants

and animals, including humans (Tripathi et al., 2012; Moore et al., 2013), has helped considerably to elucidate the genetic basis of micro-organisms' responses to arsenic and yielded a broad general physiological picture of the processes involved. The term "arsenomics" was recently coined to denote approaches focusing on the changes in the transcriptome, proteome and metabolome occurring during exposure to arsenic (Tripathi et al., 2012; Sacheti et al., 2013). In this review, which will also include studies based on genome analysis, and will therefore cover the fields of descriptive, comparative and functional genomics, it is proposed to outline how these approaches have contributed to our knowledge of the processes at work in micro-organisms exposed to arsenic, from cells to communities in the three domains of life.

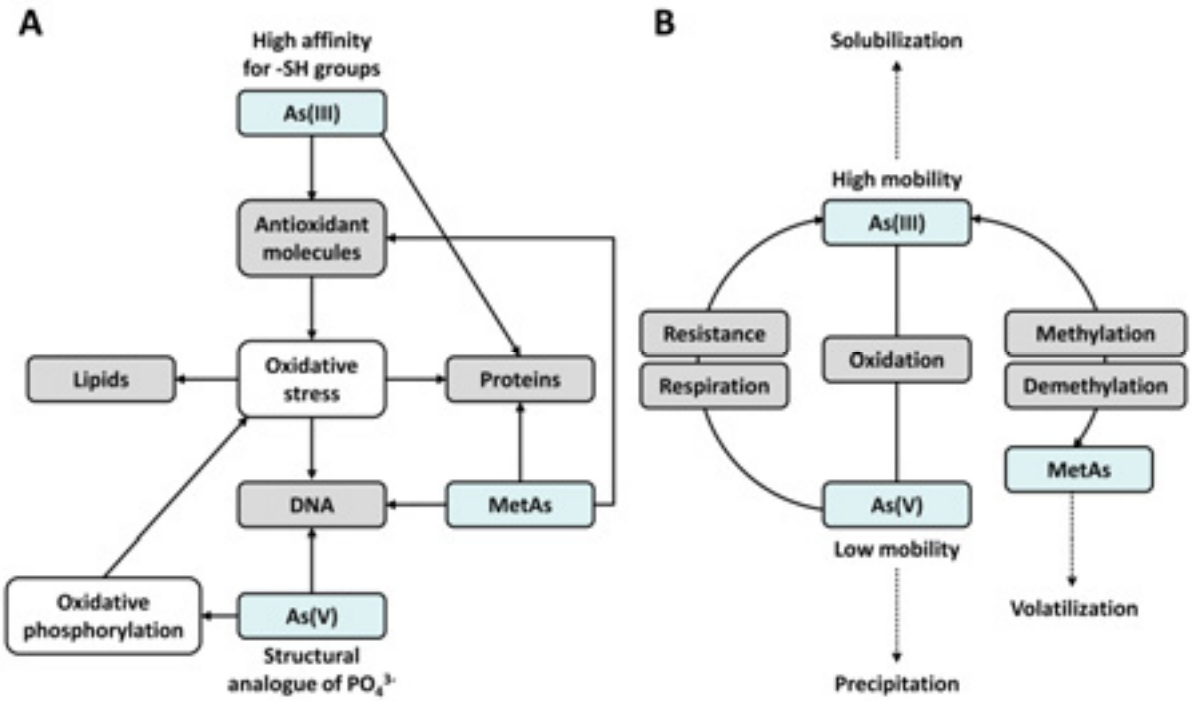


Figure 1. Toxicity and microbial cycle of arsenic. A. Arsenic species most commonly found in the environment and their cellular effects. B. Role of microbial processes in the biogeochemical cycle of arsenic and in the mobilization of this element. Arsenic forms are colored in light blue. -SH, PO_4^{3-} and MetAs stand for sulfhydryles groups, phosphate and methylated arsenic species.

1. UNVEILING MICRO-ORGANISMS' POTENTIAL RESPONSES TO ARSENIC USING DESCRIPTIVE AND COMPARATIVE GENOMIC APPROACHES

a) The use of genomic tools in studies on arsenic resistance and metabolism

The development of next-generation sequencing technologies, such as the Roche 454 and Illumina sequencing technologies which are now being widely used, has decreased the cost of sequencing and increased the throughputs and sequencing speeds (Parkhill, 2013). Thanks to these advances, genome sequencing has almost become a routine step in studies on micro-organisms, including those in the arsenic-related field. However, although the latest cutting-edge techniques have resulted in better genome coverage and given access to larger genomes and whole populations of communities, the increasingly large body of data available has raised new issues, such as the need for greater analytical and storage capacities. In addition, since the latest technologies yield shorter reads than the first-generation ones and less effort is spent on genome completion, the present genomic databases have moved from a few finished genomes to many permanent drafts (Parkhill, 2013, www.genomesonline.org/statistics). The resulting non-assembled sequences often include missing information, which is liable to make genetic analyses more complex: drafts may not make it possible to distinguish plasmidic and chromosomic sequences, for example. The combination of next generation sequencing with some interesting new alternative approaches such as that based on optical mapping may provide effective means of performing *de novo* sequencing and genome assembly (Levy-Sakin and Ebenstein, 2013). Once an organism's genome has been sequenced, it is possible to predict its gene content and protein functions and hence its metabolic potential, as well as to make comparisons with other organisms and environmental sequences. However, the increasing amount of data accumulating in databases has also led to a decrease in the proportion of genes with expert-validated annotations and experimentally established functions (Vallenet et al., 2013). This means that the interpretation of analyses based on sequence homologies, including those focusing on arsenic-related genes, must sometimes be handled with care (Schoepp-Cothenet et al., 2009). For all these reasons, the sequencing of the genomes of reference organisms and their physiological characterization are essential to be able to perform valid descriptive, comparative and functional analyses.

In this context, the heterotrophic *Betaproteobacterium Herminiimonas arsenicoxydans* was the first bacterium involved in arsenic redox reactions in which the arsenite oxidase operon and

genome were sequenced (Muller et al., 2003, 2007). Arsenic resisting and metabolizing microbes were subsequently isolated from diverse environments, including extreme ones such as soda lakes, hot springs and acid mine drainage. The genomes of more than 85 metabolizing micro-organisms have been sequenced to date (Fig. 2; Oremland and Stolz, 2003), including those of archaea, eukaryotes and bacteria, although the latter are mostly *Proteobacteria*, and these micro-organisms have various carbon and energy requirements and arsenic tolerance levels (Oremland and Stolz, 2003). With these genomic sequences, various bioinformatic tools can be used to perform prokaryotic genomic investigations. BacMet, for example, is a database containing experimentally confirmed and predicted information about biocide- and metal-resistance genes, such as their location and the nature and function of their products, and providing suitable tools for their identification in DNA sequences (Pal et al., 2014). The Microscope platform is also worth mentioning because it is one of the most complete and most user-friendly resources for the annotation, curation and comparative analysis of genomic data (Vallenet et al., 2013). The lattest arsenic-metabolizing prokaryotes' genomes available in databases have been added to the Arsenoscope project in order to provide a reference tool for genomic studies (https://www.genoscope.cns.fr/agc/mage/arsenoscope). Since arsenic resistance and methylation genes are too widely distributed and too complex to be definitely identified, respectively, the genomes bearing these genes have been included in this database only when they carry additional genes associated with arsenic metabolism, which make it possible to pursue further genomic studies.

The improvement of genome sequencing and analysis tools and the availabilty of the genomes of many arsenic metabolizing organisms have therefore contributed greatly to unveiling the genetic secrets of arsenic-related processes. Descriptive and comparative genomic tools have made it possible to describe the widespread patterns of distribution of arsenic-related genes throughout the tree of life, as well as to identify the genetic requirements of arsenic resistance and metabolism and determine the genetic context of the corresponding genes and their potential mobility, as well as the regulatory networks controlling their expression.

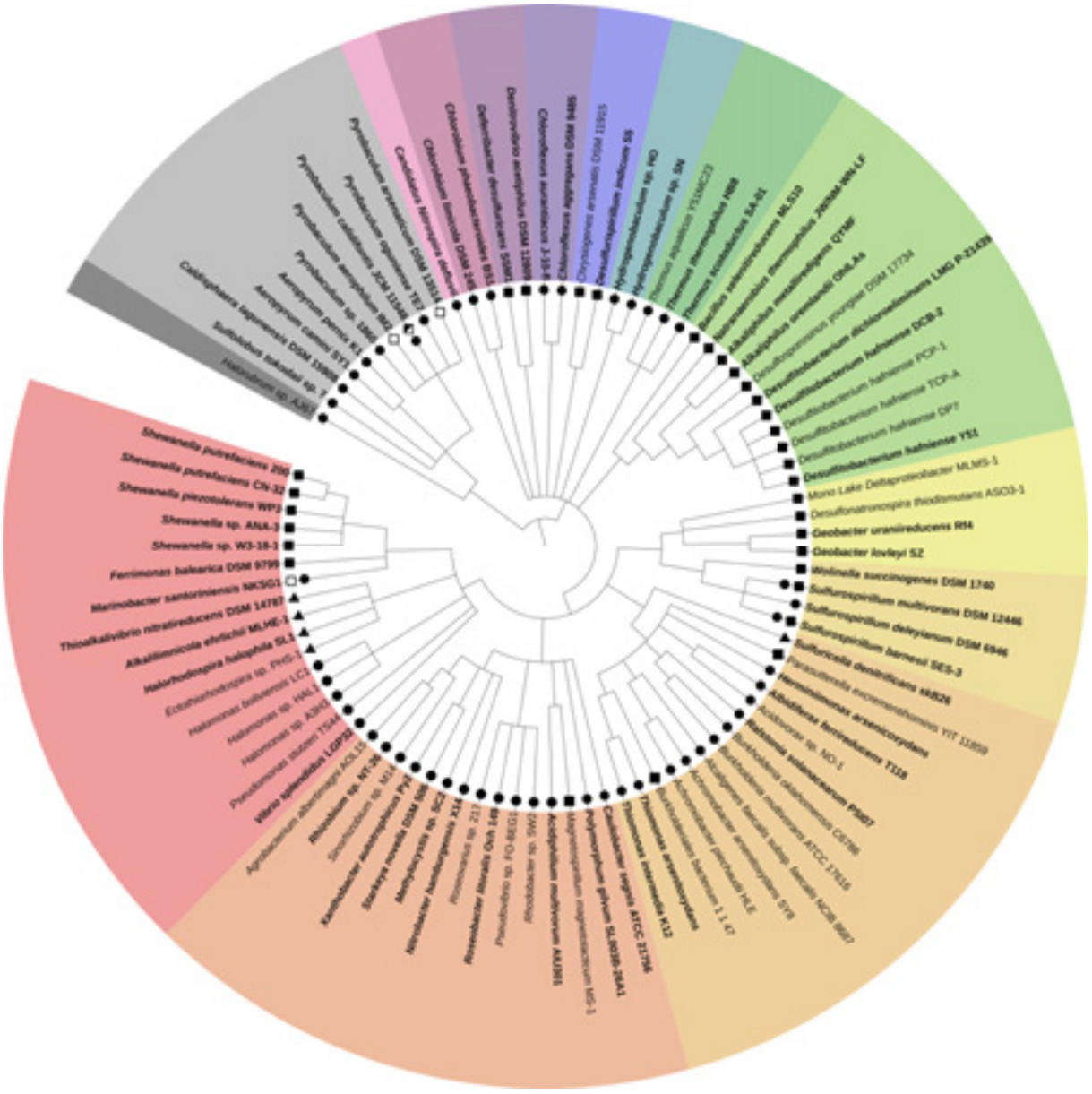


Figure 2. Cladogram of arsenic metabolizing prokaryotes whose genome has been sequenced. The Phyla of bacteria are highlighted in various colors, whereas those of archaea are presented in shades of grey. Complete and draft genomes are indicated by giving the names of organisms in bold and normal fonts, respectively. Filled circles, squares and triangles stand for organisms bearing *aio*, *arr* and *arx* genes, respectively. Open squares stand for organisms respiring As(V) in which the *arr* genes have not yet been clearly identified: in the case of *P. calidifontis*, see details in the text. Organisms with a completed genome sequence were retrieved by BLASTing reference arsenic gene sequences against the Uniprot complete microbial proteome database (The UniProt Consortium, 2014), whereas draft genomes were retrieved from the literature. The tree was based on NCBI taxonomy and constructed using the web-based iTOL tool (Wheeler et al., 2006; Letunic and Bork, 2011).

b) Widespread distribution of arsenic resistance and metabolism genes

Micro-organisms are capable of both resisting and metabolizing arsenic. Arsenic resistance is based on reactions which have been qualified as detoxification strategies. The *ars* or *ACR* system, which was the first and most thoroughly described arsenic transformation system, plays a purely protective role by conferring on the host organism resistance to both As(V) and As(III) by reducing As(V) into As(III) via a cytoplasmic arsenate reductase and extruding the latter from the cellular compartment by means of a membranous As(III) efflux pump (Silver and Phung, 2005a, 2005b). *ars* or *ACR* genes are present in bacteria, archaea and eukaryotes of various ecological origins (Mukhopadhyay and Rosen, 2002; Stolz et al., 2006). In addition, the As(III) oxidation process is thought to serve as a detoxifying mechanism in heterotrophic organisms such as *H. arsenicoxydans* (Muller et al., 2007). Thanks to the presence of *aio* gene products (Lett et al., 2012), As(III) is oxidized in the periplasm by an arsenite oxidase into the less bioavailable form As(V), which has less access to the cells.

On the other hand, arsenic metabolism involves several reactions contributing to the energy metabolism and, whereas arsenic resistance, seems to be a more species- or niche-specific process. First, in chemolithoautotrophic bacteria such as *Rhizobium* sp. NT-26, *Azoarcus* sp. DAO1 and *Dechloromonas* sp. ECC1-pb1, As(III) oxidation may provide the electrons for the aerobic and anaerobic respiration processes of oxygen, nitrate and chlorate (Santini and vanden Hoven, 2004; Rhine et al., 2007; Sun et al., 2010). Although genomic screening studies have shown that the *aio* system is phylogenetically and ecologically widespread, it seems to be restricted to bacteria and some archaea (Fig. 2, van Lis et al., 2013). Most of the bacterial genomes containing *aio* genes and most of the taxonomically assigned gene sequences present in databases belong to *Proteobacteria*, but other homologs have been found to exist in other groups such as *Deinococcus-Thermus*, green sulfur and green non-sulfur bacteria, along with a few members of the *Firmicutes* (Fig. 2; Stolz et al., 2006; van Lis et al., 2013). Although As(III) oxidation probably occurs in archaea, this assumption has yet to be confirmed as no *aio* gene-bearing strains have been proved so far to be physiologically capable of triggering this reaction. On the one hand, arsenic oxidation has been reported to occur in *Sulfolobus acidocaldarius*, but the genome of this organism has not yet been sequenced (Sehlin and Börje, 1992); and on the other hand, *aio* homologs are present in the genomes of its *Crenarchaeota* relatives *Sulfolobus tokodaii*, *Aeropyrum pernix*, *Pyrobaculum calidifontis* and *Pyrobaculum oguniense* as well as in the *Euryarchaeota Halorubrum* sp. AJ67 (Fig. 2; Lebrun et al., 2003; van Lis et al., 2013; Burguener et al., 2014), but the ability of these strains to oxidize As(III) has not yet been established. In addition, a study performed at a Yellowstone hot spring has shown that the part of the spring

where As(III) oxidation was observed was that where *Crenarchaeota* and *Euryarchaeota* 16S rDNA sequences predominated (Jackson et al., 2001).

Another oxidation system involving *arx* genes was recently identified using genomic tools (Kulp et al., 2008; Zargar et al., 2010). In this system, anaerobic As(III) oxidation is either combined with nitrate respiration or integrated into the electron transport chain of anoxygenic photosynthesis (Zargar et al., 2012). The *arx* system seems to be fairly species-specific, since it has only been described so far in three *Gammaproteobacteria* belonging to the *Ectothiorhodospiraceae* family (Fig. 2). *Alkalilimnicola ehrlichii* MLHE-1, which was the first organism bearing these genes whose genome was sequenced, has been experimentally proved to be able to oxidize As(III) (Zargar et al., 2010). The presence of *arx* homologs was subsequently reported in the As(III) oxidizers *Ectothiorhodospira* sp. PHS-1 and *Halorhodospira halophila* SL1, based on comparative genomic analyses (Zargar et al., 2012). A Blast search performed with the MLHE-1 ArxA sequence yielded two additional homologs, a formate dehydrogenase subunit and a molybdopterin oxidoreductase showing 75 % and 54 % amino acid identity with this sequence, respectively. The former homolog is also present in the genome of a fourth *Ectothiorhodospiraceae*, *Thioalkalivibrio nitratireducens*, whereas the second one was detected in a *Betaproteobacterium*, *Sulfuricella denitrificans*. In addition, the fact that other *arxA* sequences have been detected in samples originating from several arsenic-contaminated environments (Zargar et al., 2012) suggests that *arx* genes may be widespread. Further analyses (au pluriel) and experiments are required, however, before it is possible to put forward any plausible assumptions about the taxonomic or ecological distribution of these genes. To date, no biologically catalyzed As(III) oxidation processes have been definitely proved to occur in eukaryotic organisms. Although the red alga *Cyanidioschyzon* sp. 5508 is thought to be a likely candidate here (Lehr et al., 2007; Qin et al., 2009), no arsenite oxidase genes or enzymes have been identified so far in this organism, and the As(III) oxidation which was observed in this study may have resulted from the oxidative environment generated by the photosynthec activity of this alga.

The *arr* genes enable dissimilatory As(V) reduction to occur by encoding a membranous arsenate reductase which is responsible for the periplasmic anaerobic respiration of As(V), using various electron sources. Genes related to As(V) respiration have been identified in a large proportion of the bacterial domain, i.e. in the genomes of *Beta-*, *Gamma-*, *Delta*/*Epsilonproteobacteria*, *Firmicutes*, *Chrysiogenetes* and *Deferribacter* (Fig. 2, van Lis et al., 2013). By contrast, only two As(V) respirers have been detected in archaea to date, namely *Pyrobaculum arsenaticum* and *Pyrobaculum aerophilum*, both of which belong to the *Crenarchaeota* phylum (Huber et al., 2000). However, some degenerate primers reliably amplifying Arr genes in

bacteria failed to detect these genes in *P. arsenaticum*, and the *arr* annotated gene PAE2859 present in *P. aerophilum* was found to rather encode a polysulfide/thiosulfate reductase (Malasarn et al., 2004; Kulp et al., 2008, 2008; Schoepp-Cothenet et al., 2009). An exploration of the genome of *P. aerophilum* showed the presence of another gene, PAE1265, showing only a slight similarity with characterized *arr*, and based on transcriptomic results, it was established that PAE1265 is part of a three-gene operon that is upregulated in the presence of As(V), whereas the operon containing PAE2859 is not (Cozen et al., 2009). After analyzing the sequences of these genes, it was hypothesized that the three encoded proteins may be a molybdopterin oxidoreductase, an iron-sulfur-protein and a membrane anchoring subunit: these characteristics are typical of Arr, and the only orthologous operon is the operon present in the other archaeal As(V) respirer, *P. arsenaticum*. Lastly, a *P. calidifontis* variant containing a PAE1265 ortholog obtained after performing selection experiments showed stronger growth on As(V), which was found to coincide with the removal of a stop codon from this gene. All in all, these data support the existence of an archaeal Arr and although classical PCR-based methods have failed here, the latest genomic tools have shed interesting light on the mechanisms underlying As(V) respiration.

Arsenic methylation and demethylation are mediated by the products of the *arsM* and *arsI* genes, which encode an arsenite methyltransferase and a C·As lyase, respectively. These processes are considered to be part of the arsenic metabolism, but they are currently assumed to be simply detoxification processes since they confer higher arsenic tolerance, as shown by heterologous expression experiments (Qin et al., 2006, 2009; Yuan et al., 2008; Yoshinaga and Rosen, 2014). Although some methylated species can be more genotoxic than inorganic arsenic (Petrick et al., 2000; Dopp et al., 2010), others are almost innocuous and can be volatilized (Páez-Espino et al., 2009; Ye et al., 2012; Rahman and Hassler, 2014). However, although methylation is not involved in cellular energy pathways, demethylation may serve as a carbon or energy source in some bacteria (Maki et al., 2004). In a microbial genome survey using the sequence of the previously characterized arsenite methyltransferase rat enzyme, *arsM* genes were detected in a large number of bacteria and archaea, and further studies have suggested that orthologs are widespread in eukaryotes, from *Fungi* to protists and *Metazoa* (Qin et al., 2006; Bhattacharjee and Rosen, 2007; Yin et al., 2011; Ye et al., 2012). By contrast, no higher plant arsenite methyltransferase genes have been identified to date. Although methylated arsenic species have been detected in several plants such as *Fucus serratus* and *Oryza sativa*, they probably resulted from the activity of closely associated micro-organisms such as *Fungi* or bacteria (Zuccaro et al., 2008; Lomax et al., 2012). Lastly, sequence alignments have brought to light the existence of many putative *arsI* orthologs, which were all restricted to bacteria (Yoshinaga and Rosen, 2014).

Many micro-organisms are therefore capable of performing arsenic reduction, oxidation, methylation and demethylation although arsenotrophy, i.e. growth combined with As(III) oxidation or As(V) respiratory reduction, is restricted to prokaryotes (Zargar et al., 2012). Although methods based on random mutagenesis and heterologous expression have been used to identify the genes involved in these resistance and metabolic processes (Butcher et al., 2000; Kashyap et al., 2006; Qin et al., 2006; Andres et al., 2013), combining these approaches with genomic tools has greatly improved our knowledge of arsenic genetics.

c) Minimum and additional gene sets underlying arsenic resistance and metabolism processes

The increasingly large number of genomes of arsenic resisting and metabolizing organisms available has resulted in the identification of many arsenic-related clusters. The gene patterns which have consequently emerged have shown the heterogeneity of these clusters, in terms of both their genetic content and organization, and even that of closely related species, as recently observed in the case of *Burkholderiales* (Li et al., 2014). Comparisons between these clusters make it possible to distinguish a minimum set of common genes and additional genes that are more specific to each system.

Starting with the *ars* system, the *arsRBC* operon is the simplest set of cotranscribed genes that confers basal tolerance to arsenic in prokaryotes (Carlin et al., 1995; Silver and Phung, 2005a). In this system, the *arsRBC* operon codes for the regulator, the As(III) transporter and the arsenate reductase (Mukhopadhyay et al., 2002; Silver et al., 2002; Kruger et al., 2013). Supplementary genes may also be associated with this operon. In this context, *arsA* and *arsD* are occasionally found to occur and they code for an ATPase which binds to ArsB and makes the As(III) efflux more efficient, and for a protein which not only represses the operon but also acts as an As(III) chaperone, respectively (Silver and Phung, 2005a; Páez-Espino et al., 2009; Kruger et al., 2013). Although this enlarged *arsRDABC* operon has been found to occur in fewer bacterial genomes than *arsRBC*, it confers a higher level of arsenic resistance (Bhattacharjee and Rosen, 2007). Lastly, *arsH*, *arsN*, *arsP*, *arsTX*, *arsO* or *glo* have been found to occur in some clusters and to determine some accessory functions which are known to be or suspected of being involved in arsenic resistance (Páez-Espino et al., 2009). Genome explorations have shown the existence of some atypical *ars* clusters. First, *arsR* is not located in the *arsDABC* operon but still regulates its expression in *Shewanella* sp. ANA-3 (Murphy and Saltikov, 2009). Likewise, *arsC* is located distantly from the *ars* operon in *Lactobacillus plantarum*

in such a way that this gene is chromosomic whereas the operon is plasmidic, possibly reflecting the occurrence of a crosstalk between the two genetic determinants (van Kranenburg et al., 2005). Secondly, like some other archaea, *Ferroplasma acidarmanus* Fer1 carries only an *arsRB* gene subset and lacks *arsC* despite its high level of arsenic resistance, which suggests the occurrence of alternative coping strategies (Baker-Austin et al., 2007). Thirdly, the *Sinorhizobium meliloti* operon includes a gene coding for an aquaglyceroporin responsible for As(III) efflux instead of *arsB* (Yang et al., 2005; Rosen and Tamás, 2010). Fourthly, some *ars* operons have been found to show recent fusion events between genes encoding either an aquaglyceroporin and an ArsC reductase or an Acr3 and an ArsC in the case of the *Actinobacteria Frankia alni* and *Salinispora tropica* or that of *Mycobacterium tuberculosis*, respectively, resulting in a single protein mediating both reduction and efflux processes in each case (Wu et al., 2010). In eukaryotes such as *Saccharomyces cerevisiae* and other representatives, the *ACR1*, *ACR2*, *ACR3* gene cluster (which is sometimes denoted *YAP8* or *ARR1*, *ARR2* and *ARR3*) functions in a similar way to *arsRBC*, encoding the transcriptional activator, the arsenate reductase and the efflux pump, respectively (Ghosh et al., 1999). It is worth noting that while *arsB* is restricted to prokaryotes, *acr3* homologs are present throughout the tree of life, and there exist three classes of arsenate reductase, namely two bacterial ArsC and the eukaryotic Acr2p (Rosen, 2002). In addition, some genomes have been found to carry a gene encoding a Multidrug Resistance Protein involved in arsenic resistance. In *S. cerevisieae*, the ABC transporter Ycf1p belonging to this superfamilly is responsible for the translocation of the cytoplasmic As(III) into vacuoles after binding to glutathione (GSH) (Ghosh et al., 1999; Rosen, 1999).

To continue with the *aio* system, the *aioBA* operon is the only element which is consistently detected (Muller et al., 2003; van Lis et al., 2013). *aioB* and *aioA* encode the small and large subunits of the arsenite oxidase, respectively, and their order is conserved in all the As(III) oxidizers analyzed to date (Slyemi and Bonnefoy, 2012). As in the case of the *ars* cluster, the flanking genes and their pattern of organization differ from one organism to another. A *cytC* and sometimes one or two *moeA* encoding a *c*-type cytochrome and a molybdenum cofactor, respectively, are present downstream of *aioA* in most of the *Proteobacteria* genomes sequenced but not in other lineages (Slyemi and Bonnefoy, 2012; van Lis et al., 2013). Likewise, a periplasmic As(III) binding protein and a two-component regulatory system encoded by *aioXSR* are known to be part of the operon in several *Proteobacteria* but not in other phyla, which suggests that these genes may have appeared in this phylum. They are oriented either in the same direction as *aioBA* or in the opposite direction, depending on whether one is dealing with *Alpha*- or *Betaproteobacteria*, respectively (Li et al., 2013; van Lis et al., 2013). Moreover, bearing both *aio* and *ars* operons seems to confer higher average resistance on arsenic than bearing *ars* alone (Cai et al., 2009a).

To pursue this account, the *arr* and *arx* systems show several homologies despite the existence of some differences between their patterns of genetic organization. The sequence similarities observed between *aio*, *arr* and *arx* suggest that Arx may be a variant of Arr which performs the reverse reaction (van Lis et al., 2013). Although the *arrAB* operon encodes the respiratory arsenate reductase and is able to ensure As(V) respiration alone (Saltikov and Newman, 2003), *arx* As(III) oxidizing strains bear an *arxB'AB* operon. In this genetic arrangement, *arxA* and the downstream *arxB* are known to code for the enzyme subunits, whereas the exact function of the upstream *arxB* homolog *arxB'* has not yet been established (van Lis et al., 2013). Supplementary genes may also be found in the clusters of some strains. First, an *arr/arxC* gene may be present, which is probably involved in protein anchoring and electron transfer processes (Duval et al., 2008). When *arrC* is present, it is located upstream of *arrAB* in most cases, whereas *arxC* seems to be always located downstream of *arxB'AB*. Secondly, an *arr/arxD* gene coding for a chaperone is often present downstream of *arrAB* or *arxB'ABC*. Other elements possibly present in the *arr* cluster show greater variability from one strain to another (van Lis et al., 2013). In four of the five sequenced strains containing *arx*, namely *A. ehrlichii* MLHE-1, *Ectothiorhodospira* sp. PHS-1, *T. nitratireducens* and *S. denitrificans*, an *arxXSR* operon similar to *aioXSR* is present upstream of *arxB'AB* and is oriented in the opposite direction (Zargar et al., 2012; van Lis et al., 2013, this work). Nevertheless, since the *arx* system has only been quite recently discovered and only a few genomes containing *arx* are available so far, it may be risky to draw any definite conclusions about this system, since further studies may show the existence of greater diversity, as occurred in the case of the *ars* and *aio* clusters.

The last point worth mentioning here is that *arsM* is the only gene required for arsenic methylation, since it encodes an arsenite methyltransferase responsible for the three successive oxidative methylation and reduction steps starting from As(III) (Dombrowski et al., 2005; Qin et al., 2006; Rahman and Hassler, 2014). To date, *arsI* is the sole gene identified that confers demethylation activity on bacteria (Yoshinaga and Rosen, 2014). It encodes a C·As lyase responsible for methyl-arsonous acid demethylation and organo-arsenical degradation.

Studies based on the use of genomic methods combined with PCR-based technics have proved that most prokaryotic arsenic-related genes are part of operonic structures. The findings outlined above also show the considerable diversity of *ars*, *aio*, *arr* and *arx* clusters, which contain both common minimum sets of genes and specific additional ones. Studies based on the genomes available have yielded an overview of these clusters' genetic contexts, showing on the one hand the existence of larger genomic entities and on the other hand, the possible presence of several arsenic-related clusters in a single genome.

d) Genetic organisation of arsenic-related clusters

In a minireview describing the content of the 71 kb DNA sequence of *Alcaligenes faecalis* which contains the *aioBA* operon, the authors coined the concept of arsenic islands because of the clustering they observed beween *aioBA* and 21 genes potentially involved in arsenic resistance and metabolism processes, including *moeA*, *aioXSR*, *arsR*, *arsB* and several phosphate metabolism-related genes (*pho*, *pst*, *phn*) (Silver and Phung, 2005b). A genomic island of this kind was subsequently identified in *H. arsenicoxydans* (Muller et al., 2007), followed by the discovery, in an *in silico* analysis of a set of phylogenetically diverse bacteria, of 21 conserved arsenic islands ranging in size from 5 kb to 71 kb and showing significant synteny (Li et al., 2013). The concept of arsenic islands therefore suggested the possible existence of links between various cellular processes, since gene clusters often contain functionally related genes (Ettema et al., 2005). In this context, arsenic islands have been frequently found to cluster arsenic resistance and metabolism genes such as *ars* with *aio* as well as *ars* with *arx*, *arr* or *arsM*, as in *Ectothiorhodospira* sp. PHS-1, *Shewanella* species or *Alkaliphilus oremlandii*, respectively (Saltikov and Newman, 2003; Zargar et al., 2012; Li et al., 2013). Secondly, the highly conserved clustering of arsenic- and phosphate-related genes in the genomes available has given rise to the hypothesis that these genes may function as a group, enabling prokaryotes to use low concentrations of phosphorus under high arsenic conditions (Li et al., 2013). As(V) is known to be a structural analogue of phosphate which can interfere with its metabolism (Hughes, 2002). This interplay has been described in several studies, and a model for the coregulation of As(V) and phosphate metabolisms has been presented in *A. tumefaciens* 5A (see below and Kang et al., 2012b). Thirdly, in *Chloroflexus aurantiacus* and in *Deltaproteobacterium* MLMS-1 and *Desulfitobacterium hafniense*, a gene homologous to the *luxR* transcriptional regulator is located upstream of the *aioBA* and *arrAB* genes, respectively (van Lis et al., 2013), which suggests the possible existence of a link between cell to cell communications and arsenic-related reactions, as LuxR is involved in quorum sensing processes (Miller and Bassler, 2001). Lastly, other heavy metal resistance genes such as those involved in mercury, copper and cadmium resistance, have been detected in the vicinity of arsenic-related genes in *H. arsenicoxydans* and several *Thiomonas* species (Muller et al., 2007; Arsène-Ploetze et al., 2010). Although fewer data are available on eukaryotes and the arsenic island concept focuses on prokaryotes, the clustering of *ACR* genes in *S. cerevisiae* and of the genes encoding an ArsM homolog and arsenic resistance proteins observed in most of the fungal genomes available so far indicates that at least arsenic transformation-related genes might be clustered in this domain of life (Bobrowicz et al., 1997; Ye et al., 2012).

Further sequence explorations have shown in addition that several arsenic resistance- and metabolism-related clusters can coexist in a single genome. As far as prokaryotes and arsenic resistance reactions are concerned, the *H. arsenicoxydans* genome has been found to bear four *ars* loci for example, three of which are functional, and the *Anabaena* sp. PCC7120 genome has been found to carry three *arsC* copies, the expression of which is induced by arsenic (Muller et al., 2007; Pandey et al., 2013). One might therefore wonder what the purpose of bearing functionally redundant *ars* clusters might be. On the one hand, several *ars* clusters may confer greater tolerance on arsenic (Ordóñez et al., 2005; Kang et al., 2014); and on the other hand, multiple clusters may make it possible for a cellular activity to be expressed in a wider range of conditions, as in *P. putida* KT2440, which bears two virtually identical *arsRBCH* operons expressed at different temperatures (Páez-Espino et al., 2014). By contrast, as regards the metabolism of arsenic, although *arsM* may sometimes be duplicated, *aio*, *arr* and *arx* clusters seem to be restricted to a single copy wherever they occur, apart from a few exceptions such as those observed in *Acidiphilum multivorum* and some *Thiomonas* strains bearing two *aioBA* copies (Arsène-Ploetze et al., 2010; Li et al., 2013). Interestingly, the results of a genomic survey focusing on 188 *Burkholderiales* members and microarray-based comparisons between the genomes of some *Thiomonas* strains support the idea that the number of arsenic-related genes and their acquisition, respectively, may depend on the selective pressure exerted by this element itself (Arsène-Ploetze et al., 2010; Li et al., 2014). Multiple arsenic-related clusters can also occur in eukaryotic genomes. For example, the *Saccharomyces douglasii* genome includes two *acr* clusters, which is thought to explain why the tolerance of this strain to arsenic is greater than that of *S. cerevisiae* (Maciaszczyk et al., 2004). On similar lines to what occurs in prokaryotes, it has been established by performing selection experiments on *Cryptococcus neoformans* that greater resistance to arsenic is conferred on this strain by the massive duplication of a genomic region bearing *ARR3*, which further supports the hypothesis that the selective pressure exerted by arsenic may affect the number of gene copies (Chow et al., 2012).

In conclusion, the authors of genomic studies have described the clustering of genes responsible for arsenic resistance and metabolism along with the factors contributing to phosphate metabolism, quorum sensing and heavy metal resistance, and suggested the functional links possibly involved. It has also been observed that several clusters conferring specific traits can be present in a single genome. In prokaryotes, multiple copies of genes and metal resistance genes can be associated with extrachromosomal DNA such as plasmids and genomic islands (GEI) (Mergeay et al., 2003; Juhas et al., 2009; Janssen et al., 2010). As these two types of genetic elements can both be mobile, involved in horizontal gene transfer (HGT) and facilitate bacterial adaptation, several studies have been performed with a view to determining the genomic location of arsenic-related clusters.

e) Mobile genetic elements and adaptation to arsenic stress

Genome analyses, sometimes combined with electrophoretic methods, have made it possible to discriminate between sequences of plasmidic and chromosomic origin and to determine the locations of arsenic resistance and metabolism genes. First, *ars* can be chromosomic, plasmidic, or both when multiple copies are present. Interestingly, multiple chromosomic *ars* operons have also been reported to exist in *P. putida*, *Corynebacterium glutamicum*, *H. arsenicoxydans*, *Ochrobactrum tritici*, *Thiomonas* sp. 3As, *Leptospirillum ferriphilum*, *Geobacillus kaustophilus* and *A. tumefaciens* 5A (Ordóñez et al., 2005; Muller et al., 2007; Branco et al., 2008; Arsène-Ploetze et al., 2010; Li et al., 2010; Cuebas et al., 2011; Kang et al., 2014; Páez-Espino et al., 2014). Although *aio* clusters are generally chromosomic, they can also be plasmidic, as in *Rhizobium* sp. NT-26, *Sinorhizobium* sp. M14, *Nitrobacter hamburgensis*, *Ralstonia solanacearum*, *Halomonas* sp. A3H3 and *Thermus thermophilus* (Starkenburg et al., 2008; Remenant et al., 2010; Heinrich-Salmeron et al., 2011; Andres et al., 2013; Drewniak et al., 2013; Koechler et al., 2013). The results of a recent genomic survey, which are supported by experimental data, have suggested that *aio* clusters may more often be plasmidic than chromosomic in *Alphaproteobacteria* (Li et al., 2013). Little information is available so far in this connection on *arr*, but *arx* might be carried by either plasmidic or chromosomic elements. The *arx* cluster is located on a plasmid in the reconstructed *S. denitrificans* genome, whereas it is chromosomally encoded in *A. ehrlichii* MLHE-1, *H. halophila* SL1 and *T. nitratireducens*. However, the presence of arsenic-related genes in plasmids is not necessarily correlated with their lateral acquisition, since plasmids are not all conjugative.

Genomic sequence explorations, phylogenetic analyses and genome comparisons have yielded some interesting additional clues supporting some hypotheses about plasmid-associated arsenic-related gene transfers. On the one hand, an example focusing on the *ars* system was obtained thanks to the sequencing of the *L. plantarum* WCFS1 genome. The reconstruction of this genome has resulted in the identification of three plasmids, and sequence explorations have brought to light the location of an *arsRB* operon as well as that of a complete set of genes involved in mobilization processes in the largest plasmid, which suggests that the *ars* operon may be transferable. Further experiments showed that this large plasmid is a conjugative plasmid conferring arsenic resistance on another *Lactobacillus* strain carrying a chromosomic *arsC* gene (van Kranenburg et al., 2005). On the other hand, there exist several examples of plasmidic *aio* genes that may have been involved in HGTs. First, along with sequence explorations, phylogenetic analyses have often shown the possible occurrence of HGT by bringing to light discrepancies between the arsenic-related

genes and species phylogenies (Cai et al., 2009a; Heinrich-Salmeron et al., 2011; Li et al., 2013). The authors of a recent study on these lines have pointed out the existence of three inconsistencies, the *aioA* phylogeny including *Acidiphilum* strains along with *Chlorobi* and *Deinococcus-Thermus*, which all have a similar ecological origin, *Chloroflexus aggregans* along with *Deinococcus-Thermus* and *Ralstonia* sp. 22 along with *Achromobacter arsenitoxydans* SY8, where *aioA* is of plasmidic origin in each case (Li et al., 2013). Secondly, the presence of arsenic-related genes in only a few strains of a given genus is another clue suggesting the existence of HGT. In this respect, the authors of a study on 30 available *Rhizobium/Agrobacterium* genomes identified an *aio* cluster in only 3 strains, i.e. *Sinorhizobium* sp. M14, *A. tumefaciens* 5A and *Rhizobium* sp. NT-26, in which it is also plasmidic (Andres et al., 2013). Interestingly, the plasmid of *Sinorhizobium* sp. M14 harbours a complete set of arsenic oxidation genes and was recently found to be self-transferable, stable and to have a broad range of hosts (Drewniak et al., 2013). However, plasmids are not the sole mobile genetic elements liable to be involved in the acquisition of arsenic-related clusters.

De facto, clusters of this kind have been found to be located in other mobile elements such as GEIs and transposons, which can be identified by performing sequence explorations on the basis of their specific features (Langille et al., 2010). To begin with, a genomic study on *P. putida* KT2440, which bears two *arsRBCH*, has shown that one of them is the indigenous resistance cluster found to exist in almost all the members of this genus, while the other one is located in a 62 kb GEI which is lacking in the 5 most closely related strains: this suggests that the latter cluster may have been acquired by HGT (Páez-Espino et al., 2014). An exploration of the *Thiomonas* sp. 3As genome and its comparison with the genomes of several other members of this genus based on the use of CGH chips have shown that its *aioBA* operon is located on a GEI which is also specifically present in strains isolated from arsenic-rich environments: this finding again supports the possible occurrence of HGTs and suggests that they may contribute to the diversity of ecotypes (Arsène-Ploetze et al., 2010; Bertin et al., 2011). This island also contains *aio*, *ars* and several heavy metal-resistance genes, which fits the arsenic island concept, thus providing the bearing strains with a specific set of cellular functions facilitating their adaptation to their environment (Juhas et al., 2009; Arsène-Ploetze et al., 2010). An in-depth analysis of this GEI and a phylogenetic analysis of the *aio*, *ars*, *pst* and *phn* genes present in 21 arsenic islands detected in various bacteria have both underlined the composite nature of these islands and suggested that arsenic islands may have formed independently as the result of a series of DNA acquisitions, deletions and rearrangements rather than having evolved as a single unit (Arsène-Ploetze et al., 2010; Li et al., 2013). In this respect, the phylogenetic analysis of ArsB sequences from these same 21 arsenic islands showed the existence of three conflicts with species phylogenies, i.e. those of *Acidovorax* sp. NO1, *Thiomonas* sp. 3As and *A. faecalis* NCIB8687, in all of

which *arsB* is located on a transposon (Li et al., 2013). In addition, the *ars* cluster in *Acidithiobacillus caldus* is located in a functional Tn*21*-like transposon, which further supports the idea that such genetic elements may be involved in GEI formation and arsenic-related gene horizontal transfer processes (Tuffin et al., 2005).

Although studies on the lateral transfer of arsenic genes have focused on events between bacteria, the possibility that transfer may also occur between domains cannot be ruled out. In a recent study based on genome sequencing data, two *arsB* sequences were identified in the genome of the extremophilic metal resisting red alga *Galdieria sulphuraria* (Schönknecht et al., 2013). Not only are these sequences devoid of introns, but their closest relatives also belong to bacterial strains living in the same environment as this alga, which suggests that they may bave been acquired via interdomain HGT events. Interestingly, the number of events of this kind increase with the level of specialization of the organisms present in some ecological niches, thus reflecting the acquisition of adaptive abilities (Keeling and Palmer, 2008).

In conclusion, based on the findings obtained by exploring and comparing genomes and by performing phylogenetic analyses, it can be said that arsenic resistance and metabolism genes are frequently associated with mobile elements such as plasmids, GEIs and transposons, which suggests that these genes may have been involved in transfer processes, genome evolution and the adaptation of the host organisms to arsenic. Although these hypotheses still remain to be proved, they are often consistent with the relevant physico-chemical, ecological and geographical data. Despite the mobility of arsenic resistance and metabolism genes, the huge amount of genetic information obtained so far on a wide range of living organisms in all three domains of life has made it possible to address the following interesting evolutionary questions about arsenic-transforming systems.

f) Evolution of arsenic resistance and metabolism gene clusters

Several authors have used multiple sequence alignments, taxonomic patterns of distribution, paleogeochemical inferences and biochemical data to draw up hypotheses about the successive steps in the evolution of the various arsenic-related systems since the first signs of life appeared on primordial Earth (Mukhopadhyay et al., 2002; Lebrun et al., 2003; Oremland and Stolz, 2003; Duval et al., 2008; Kulp et al., 2008; Schoepp-Cothenet et al., 2009; van Lis et al., 2013). The first study on these lines focused on the evolution of the *ars* operon (Rosen, 1999). The author suggested that the

efflux system may have first arisen in an environment dominated by As(III). Arsenate reductases may then have evolved independently from phosphatases with the increasing level of oxidation and evolved convergently in all three domains of life, which shows how efficient this system is (Rosen, 1999; Mukhopadhyay and Rosen, 2002; Páez-Espino et al., 2009).

Two possible scenarios accounting for the evolution of arsenic metabolism were then presented and discussed, and the following one was taken to be the most likely (van Lis et al., 2013). According to this scenario, Aio, the first existing arsenic-related enzyme, was involved in nitrogen oxides' anaerobic respiration. After the split between bacteria and archaea occurred, Aio may have been recruited by more recent pathways such as bacterial photosynthesis. With the increasing oxidizing level of the environment resulting from photosynthesis, the activity of the arsenite oxidase may then have been integrated into the aerobic respiration of dioxygen. The accumulation of As(V) also caused by the oxidizing environment would next have led to the evolution of an enzyme involved in the reduction of sulfur into the respiratory arsenate reductase Arr. Arx may have emerged at approximately the same time, reversing the Arr reaction.

Bacteria, archaea, and eukaryotes' arsenite methyltransferases harbour motifs which are conserved in all three domains of life, as shown by multiple sequence alignments: this suggests that this enzyme must have come into existence a very long time ago (Ye et al., 2012). More extensive studies are required, however, before more detailed evolutionary hypotheses can be put forward. All in all, although HGT events may have blurred some of the information and some of the assumptions cannot be proved experimentally, combining phylogenetic analyses with biochemical data has yielded some interesting theories. Last but not least, studies on genomic sequences have also made it possible to address some of the mechanisms underlying arsenic-related gene expression.

g) The regulation of the expression of the genes involved in arsenic resistance and metabolism

The regulation of arsenic-related gene expression has been extensively studied and several transcriptional regulators have been identified and characterized to date. In short, ArsR and ArsD negatively control the expression of *ars* and *arr* operons as well as *arsM.* The two-component regulatory system AioSR as well as AioF positively regulate *aio* operon transcription. And lastly, Acr1 positively controls *ACR2* and *ACR3* expression (Ghosh et al., 1999; Slyemi and Bonnefoy, 2012).

Genomic tools have provided useful means of identifying genes and characterizing the molecular mechanisms involved in the regulation of most of these systems.

On the one hand, genomic analyses have helped to identify the genes involved in the regulation of several arsenic-related systems. In *Shewanella* sp. ANA-3, neither the *arsDABC* nor the *arrAB* operons, both of which are present in the arsenic island, are preceded by an upstream regulator (Murphy and Saltikov, 2009). Interestingly, genome explorations led to the identification of six genes coding for ArsR family proteins, three of which are present in the latter island and may therefore be involved in controlling *ars* and *arr* operon expression. The results of mutagenesis experiments, transcription profiling and DNA binding assays then brought to light the effective *arsR* gene, the product of which regulates the expression of both operons. Likewise, genomic tools have highlighted the existence of a gene involved in the control of *aioBA* expression via AioSR. Although the mechanism involved in the sensing of As(III) by AioSR was not clearly established at the time of these analyses, sequence comparisons have shown that an open reading frame was recurrently present directly upstream of AioSR in several clusters (Muller et al., 2007; Cai et al., 2009b). RT-PCR experiments then showed that this gene is cotranscribed with *aioSR*, and it has been predicted on the basis of sequence analyses that it codes for a periplasmic oxyanion binding protein (Cai et al., 2009b). It was thereofore suggested that its product, named AioX, may be involved along with AioSR itself in *aioBA* regulation by sensing As(III) and conveying the information to AioS (Cai et al., 2009b). Additional biochemical and mutagenesis experiments eventually confirmed that AioX binds to As(III) and is required for *aioBA* expression (Liu et al., 2012).

On the other hand, genomic tools have contributed to unraveling the molecular mechanisms involved in the AioSR-mediated regulation of *aioBA* expression. A study in which random mutagenesis was combined with qRT-PCR tests first established that the expression of arsenite oxidase in *H. arsenicoxydans* requires RpoN, an alternative sigma factor (σ^{54}) (Koechler et al., 2010). It was suggested in addition that RpoN may be involved in the initiation of *aioBA* transcription. The sequences containing the experimentally determined *aioBA* transcriptional start site were then analyzed, and the presence of a putative -12/-24 σ^{54}-dependent promoter motif was detected in *H. arsenicoxydans* and *Rhizobium* sp. NT-26, which supports the above hypothesis (Koechler et al., 2010; Sardiwal et al., 2010). Lastly, the RpoN regulation of *aioBA* expression was unambiguously confirmed in *A. tumefaciens* 5A by exactly deleting its binding site, which is located upstream of *aioB*, and performing RT-PCR tests (Kang et al., 2012a). In addition to these findings, multiple sequence alignments indicated that AioR shows sequence homologies with known σ^{54} RNA polymerase transcriptional activators and harbours a response regulator receiver domain, a σ^{54} interacting

domain and a DNA binding domain, which are shared by most enhancer-binding proteins (Koechler et al., 2010; Sardiwal et al., 2010). AioR is therefore thought to interact with RpoN, thus initiating *aioBA* transcription in both *Alpha-* and *Betaproteobacteria* bearing *aioXSR*, although these operons are differently orientated in these groups. Lastly, a multiple sequence alignment performed with all the *aioBA* regulatory sequences available in databases recently resulted in the identification of two conserved motifs located upstream of the *aioBA* transcriptional start site, namely GT[CT]CGN(6)CG[GA]AC in *Rhizobiales* and GTTNCN(6)GNAAC in *Burkholderiales*, which are thought to be likely candidate AioR binding sites (Andres et al., 2013). The putative RpoN and AioR binding sites and their concomitant presence have only been observed in strains bearing *aioXSR*, which further supports the idea that these putative upstream activating sequences may be involved in *aioBA* expression (Koechler et al., 2010; Andres et al., 2013; Li et al., 2013).

Furthermore, genomic findings have suggested that additional elements may be involved in the regulatory network controlling arsenic resistance and metabolism gene expression. First, although AioX, AioS and AioR seem to be essential to the expression of arsenite oxidase when the corresponding genes are present in a given strain (Kashyap et al., 2006; Koechler et al., 2010; Sardiwal et al., 2010), strains bearing only the *aioBA* operon, such as *Halomonas* sp. HAL1 and *Pseudomonas stutzeri* TS44, are able to oxidize As(III) (Cai et al., 2009b; Lin et al., 2012). In addition, the induction of *aioBA* expression by As(III) observed in the latter strain despite its lack of *aioXSR* (Cai et al., 2009b) suggests the existence of an alternative regulatory mechanism. Secondly, several RpoN binding sites were identified in an exploration of the *A. tumefaciens* 5A genome, one of which is associated with *dctA*, which codes for a malate/phosphate antiporter known to readily replace phosphate by As(V) (Kang et al., 2014). Since the expression of *dctA* is upregulated by As(III), RpoN is required for the expression of this gene, and the induction of *aioBA* is delayed in a *dctA* deletion mutant, this transporter may be involved in the regulatory network controlling arsenite oxidase expression. Thirdly, as mentioned above, a *luxR* homolog is sometimes present in *aio* or *arr* clusters, possibly reflecting the existence of a regulatory link between arsenic and quorum sensing processes (van Lis et al., 2013). Interestingly, although there are no *luxR* homologs in the vicinity of *aioBA* in *A. tumefaciens* 5A, further sequence explorations have shown the existence of several copies in the genome, and quorum sensing is indeed involved in controlling the expression of *aioBA* in this strain, as shown by cross culture experiments or addition of *N*-acylhomoserine lactone extracts to bacterial cultures (Kashyap et al., 2006). In heterotrophic bacteria, quorum sensing signals, which may be related to fixed communities, may promote the induction of detoxification strategies such as arsenite oxidation in cells unable to escape the stresses to which they are exposed. Alternatively, in chemoautotrophic bacteria, while isolated cells may use only one energy substrate, whole

populations of cells may rapidly lead to its depletion. Therefore, the induction of a more diversified metabolism by quorum sensing signals would serve to prevent starvation thanks to the use of various molecules, including As(III), as sources of energy. No studies have focused so far on the regulation of *arx* expression, but the results of comparisons between various arsenic islands have been taken to suggest that a separate operon may be located next to *arxB'ABCDE* and transcribed in the opposite direction, as occurs in *A. ehrlichii* MLHE-1 and other *arx* harboring strains (Zargar et al., 2010; van Lis et al., 2013). This operon is predicted to comprise regulatory components resembling *aioXSR*, which suggests that the processes regulating the expression of both types of arsenite oxidase may be similar.

All in all, the genomes currently available have shed interesting light on the regulatory mechanisms controlling the expression of arsenic-related genes, and the results thus obtained have suggested a model for the molecular mechanisms possibly involved in the regulation of *aioBA* expression (Li et al., 2013). However, although findings in the fields of descriptive and comparative genomics have greatly contributed to identifying the genetic determinants underlying the processes of arsenic resistance and metabolism as well as determining the functioning of these systems, these approaches have mainly served to define these very functions, and therefore justify the use of complementary approaches as a means of enlarging our picture of micro-organims' responses to arsenic.

2. DECIPHERING CELLS' AND MICROBIAL COMMUNITIES' RESPONSES TO ARSENIC USING FUNCTIONAL GENOMIC APPROACHES

a) Functional genomics and arsenic responses

In the field of functional genomics, several methods are used to adress cell physiology on the overall scale and at various levels from genome expression to the metabolome. The deletome approach which has been developed in the case of *S. cerevisiaea* involves the systematic deletion of all the coding DNA sequences longer than 100 bp and has enabled the subsequent screening of a 4.700-strain library to identify the genes involved in arsenic resistance processes (Haugen et al., 2004). Transcriptome profiling methods have been used to address gene expression and regulation topics in several studies. Although microarrays have been widely used in studies on these lines, High Throughput Sequencing-based RNA-sequencing is now becoming the standard method in the field of

transcriptomics (Coppée, 2008; Wang et al., 2009; McGettigan, 2013). Proteome profiling tools have yielded additional information, as the abundance of the proteins under investigation depends not only on the transcription rates but also on whether post-translational modifications or proteolysis have occurred (Bertin et al., 2008). Metabolome profiling studies have focused on the last and most dynamic level of biological objects, the metabolites. Although this approach has been applied only once in arsenomics and full analysis of metabolomes is currently impossible, it can be said that in view of the fast development of analytical chemistry techniques and database completion, this method promises to yield interesting findings (Weckwerth, 2010).

These functional genomic tools have been applied to determine the responses to arsenic of various micro-organisms belonging to several phyla of bacteria, archaea and eukaryotes, namely *Proteobacteria*, *Firmicutes*, *Nitrospirae*, *Cyanobacteria*, *Chrenarchaeota*, *Euryarchaeota*, *Ciliophora*, *Ascomycota* and *Euglenozoa*. The organisms studied were of various ecological origins, such as arsenic-contaminated waters, soils and sludges as well as clinical and laboratory environments, and had different carbon, energy and oxygen requirements, arsenic metabolism and tolerance levels. The genomes of some of these organisms have not yet been sequenced. Various functional genomic approaches have been used and sometimes combined in these studies, and the hypotheses to which they have given rise have been systematically checked by performing additional tests. The comparisons made in most of these studies have focused on the profiles obtained in the presence and absence of either As(III) or As(V), but some authors have also compared the responses observed between closely related species, ecotypes or exposure lengths (Bryan et al., 2009; Cleiss-Arnold et al., 2010; Pandey et al., 2012; Halter et al., 2014). A study on these lines was also performed in which genetic variants were compared. Two As(III) resistant genetic variants of *Leishmania amazonensis*, A and A', were obtained by performing selection experiments, and the transcriptomes of these variants were compared with the wild-type strain (Lin et al., 2008). When referring to this study in what follows, the terms up- and down-regulation will therefore be used when comparing these variants with the wild-type strain.

On the whole, these studies in which functional genomic tools have been applied to all these micro-organisms under various conditions have shed light on the main cellular processes affected by arsenic and those potentially involved in the responses of prokaryotes and eukaryotes to this toxic element. The recent development of novel technologies and data processing tools has made it possible to compare more complex data sets such as those bearing on whole microbial communities, including non-culturable micro-organisms, and thus to determine additional mechanisms triggered in the presence of arsenic.

b) Cellular processes underlying micro-organisms' responses to arsenic

i. *Colonization processes*

Micro-organisms are able to move towards more favourable conditions and away from deleterious ones, as well as to protect themselves from environmental stresses thanks to various processes. Several functional studies have shed interesting light on the effects of arsenic on these abilities. Chemotactism is a sensing mechanism which enables bacteria to detect changes in their environment and to adapt their motility accordingly. Proteins involved in this process are up- and downregulated in the presence of As(III) in *H. arsenicoxydans* and *Thiomonas* sp. 3As, respectively, and the former was shown to exhibit a positive chemotaxic response towards As(III) (Muller et al., 2007; Bryan et al., 2009; Weiss et al., 2009). The flagellum, which is closely related to chemotactic processes in bacteria, is the main microbial locomotor appendix, although there exist structural and functional differences between the three domains of life. Proteins involved in the biosynthesis and functioning of bacterial flagella accumulate preferentially in the presence of As(III) in *H. arsenicoxydans* and *Rhizobium* sp. NT-26, concomitantly showing greater motility (Muller et al., 2007; Weiss et al., 2009; Cleiss-Arnold et al., 2010; Andres et al., 2013). As(III) also increases the motility of *Euglena mutabilis*, a photosynthetic protist which was isolated from an arsenic-contaminated acid mine drainage (AMD), and transcripts encoding flagellar proteins were found to accumulate more conspicuously in *L. amazonensis* A than in the wild type strain (Lin et al., 2008; Halter et al., 2012a).

Secondly, flagella are involved in various stages of biofilm development. This extracellular matrix, which is composed of proteins, lipids, nucleic acids and polysaccharides, increases bacterial resistance to various physico-chemical stresses. Studies based on functional tools have shown that proteins and transcripts associated with exopolysaccharide synthesis are upregulated in *H. arsenicoxydans* and *Rhizobium* sp. NT-26, respectively, when grown in the presence of As(III) (Weiss et al., 2009; Andres et al., 2013). Likewise, a cell surface nucleic acid binding protein that may be involved in biofilm development is upregulated in the presence of As(V) in *Staphylococcus* sp. NBRIEAG-8 (Mackey-Lawrence et al., 2009; Srivastava et al., 2012). Quorum sensing processes are known to influence the switch between planktonic and sessile lifestyles. High levels of expression of a gene involved in this pathway, which has been reported to induce cellular motility, have been observed in *Rhizobium* sp. NT-26 in the presence of As(III): this finding supports the links previously suggested to exist between arsenic and quorum sensing processes (Andres et al., 2013).

Functional genomic studies have therefore clearly established that arsenic affects chemotactism, motility, biofilm synthesis and quorum sensing in bacteria and at least motility in eukaryotes. These findings may reflect micro-organisms' ability to sense some arsenic-related environmental changes and adopt appropriate protective or escape strategies. However, despite the existence of the latter strategies, cells may still have to cope with the presence of arsenic and its toxicity.

ii. *Arsenic permeability and cell envelope processes*

One of the strategies most commonly involved in resistance to heavy metals consists in preventing the entry of these toxic substances into the cellular compartment by modifying the permeability of the membrane. This is achieved in some cases by changing the composition of the cell envelope. Interestingly, in some bacteria, genes involved in the synthesis of the peptidoglycan, of the lipopolysaccharide and that of the cell envelope in general are differentially expressed in the presence of arsenic, i.e. in *Thiomonas arsenivorans*, *H. arsenicoxydans*, *L. ferriphilum*, *Staphylococcus* sp. NBRIEAG-8 and *Rhizobium* sp. NT-26, (Carapito et al., 2006; Bryan et al., 2009; Cleiss-Arnold et al., 2010; Li et al., 2010; Srivastava et al., 2012; Andres et al., 2013). σ^E, which is involved in the transcriptional control of several pathways contributing to phospholipid and lipopolysaccharide synthesis, is repressed in *H. arsenicoxydans* (Cleiss-Arnold et al., 2010). In addition, transcripts and proteins involved in the early steps of fatty acid biogenesis tend to accumulate preferentially in the presence of arsenic in *Pseudomonas aeruginosa*, *Cupriavidus metallidurans* and *Rhizobium* sp. NT-26 (Parvatiyar et al., 2005; Zhang et al., 2009; Andres et al., 2013). Lastly, an enzyme involved in the synthesis and maintenance of membrane fluidity is repressed in *Exiguobacterium* sp. PS in the presence of arsenic (Li et al., 2010). In the case of eukaryotes, a study on the *S. cerevisiae* deletome has pinpointed the role of genes involved in the biosynthesis of ergosterol, phospholipids and sphingolipids, which are components of the lipid bilayer, in arsenic tolerance processes (Thorsen et al., 2009). The yeast and *L. amazonensis* A transcriptomes show a differential expression of the genes involved in the synthesis of a glycoconjugate anchored to the cell surface and in fatty acid metabolism (Haugen et al., 2004; Jin et al., 2008; Lin et al., 2008). In agreement with these findings, biochemical tests have shown that the fatty acid composition of the cell membrane undergoes a change in the presence of As(III) in *E. mutabilis* and *Euglena gracilis* (Halter et al., 2012a). All in all, these data concordantly show that in the presence of arsenic, changes in the cell membrane composition occur in both bacteria and eukaryotes, possibly reflecting changes in the permeability of

the envelope and its reorganisation. These restructuring events may also involve components which are embedded in the membrane, such as transporters and signal receptors involved in transduction pathways.

As(V) and As(III) enter cells via phosphate transporters and via sugar uptake systems and aquaglyceroporins, respectively, and several functional studies have shown that arsenic perturbs the accumulation of these very proteins or their respective transcripts in all three domains of life (Páez-Espino et al., 2009; Rosen and Liu, 2009). First, arsenic affects phosphate transport. In *H. arsenicoxydans*, data obtained using a transcriptome approach have shown that the expression of *pit*, a non-specific inorganic phosphate transporter, is induced in the presence of As(III) (Cleiss-Arnold et al., 2010). More interestingly, arsenic induces the accumulation of phosphate-specific transporters called Pst in *Pseudomonas* sp. As-1, *Comamonas testosteroni* sp. CNB-1, *C. metallidurans* CH34, *H. arsenicoxydans*, *Staphylococcus* sp. NBRIEAG-8, *Rhizobium* sp. NT-26, *Exiguobacterium* sp. PS and in the eukaryote *E. mutabilis* (Patel et al., 2007; Zhang et al., 2007, 2009; Cleiss-Arnold et al., 2010; Srivastava et al., 2012; Andres et al., 2013; Sacheti et al., 2013; Halter et al., 2014). The selectivity of these phosphate-binding proteins is essential to enable sufficiently large amounts of phosphate to be incorporated by these organisms at high As(V) concentrations, thus ensuring their survival. Interestingly, this selectivity is especially pronounced in the highly As(V) resistant *Halomonas* strain GFAJ-1 (Elias et al., 2012). It has also been observed that transcripts or corresponding proteins resulting in the use of alternative sources of phosphorus such as phosphonate transporters accumulate preferentially in the presence of As(III) in *H. arsenicoxydans*, *Rhizobium* sp. NT-26 and *E. mutabilis* (Cleiss-Arnold et al., 2010; Andres et al., 2013; Halter et al., 2014). Arsenic also affects sugar transport: the fact that the expression of the genes coding for hexose transporters is repressed in the presence of As(III) in *Klebsiella pneumoniae* and *S. cerevisiae* reduces the incorporation of As(III) (Haugen et al., 2004; Jin et al., 2008; Daware et al., 2012). Lastly, the authors of functional genomic studies have reported that the levels of expression of porines vary in the presence of arsenic. The downregulation of non-specific diffusion channels of this kind, which are sometimes involved in resistance to antibiotics and metals has been observed in *K. pneumonia* and *Yersinia enterolitica* 1A (Daware et al., 2012; Mallik et al., 2012). The repression of σ^E observed in *H. arsenicoxydans* may also be associated with a downregulation of porin maturation mechanisms (Cleiss-Arnold et al., 2010). Surprisingly, the expression of a gene coding for a homologous protein is upregulated in *L. amazonensis* A', where it is thought to be possibly involved in As(III) efflux, as occurs in *S. meliloti* (Lin et al., 2008).

The expression or accumulation of diverse additional molecular transporters is also affected by arsenic in some organisms. The expression of amino acid transporters is repressed in *K. pneumoniae* and *Y. enterolitica* 1A, and the accumulation of transporters of unidentified molecules is affected in *Y. enterolitica* 1A, *C. testosteroni* sp. CNB-1, *C. metallidurans* and *S. cerevisiae* (Haugen et al., 2004; Zhang et al., 2007, 2009; Daware et al., 2012; Mallik et al., 2012, 2012; Andres et al., 2013). In the latter eukaryote, some metallic ion extrusion transporters are upregulated in the presence of As(III), whereas some transporters involved in metal uptake are repressed (Jin et al., 2008). The fact that several genes involved in iron, cation and transition metal transport are differentially expressed in the presence of As(III) may be attributable to a disruption of metal homeostasis (Jin et al., 2008).

The presence of arsenic also affects the expression of various signal transduction pathways. Chemotactism-related functions may be affected in bacteria, for example, as mentioned above. In addition, the two-component regulatory system AioSR controlling the expression of arsenite oxidase is upregulated in *H. arsenicoxydans* (Cleiss-Arnold et al., 2010), and a putative signal transduction protein and a histidine kinase are upregulated in the presence of arsenic in *L. ferriphilum* and *Anabaena* sp. PCC7120, respectively (Li et al., 2010; Pandey et al., 2012). In the case of eukaryotes, yeast deletome and transcriptome analyses have revealed the role of different signal transduction pathways in the responses of this organism to arsenic such as those involving cAMP dependant protein kinase A, protein kinase CK2, and MAPK, which are known to be involved in the transcriptional control of ribosome synthesis and stress responses (Haugen et al., 2004; Jin et al., 2008). Likewise, the MAPK pathway seems to play an important role in *L. amazonensis* A' tolerance to arsenic (Haugen et al., 2004; Lin et al., 2008). Interestingly, in *S. cerevisiae*, the microtubule associated protein kinase Hog1p controls As(III) uptake through the aquaglyceroporin Fps1p and is essential to arsenic tolerance, as shown by deletome experiments (Jin et al., 2008; Thorsen et al., 2009). Diverse signal transduction mechanisms are therefore affected by arsenic, starting with the envelope receptors, and these mechanisms may trigger further downstream processes contributing to prokaryotes' and eukaryotes' responses to arsenic.

All the above studies in the field of functional arsenomics have therefore taught us that arsenic affects processes related to the cell envelope's composition and functions in both prokaryotes and eukaryotes. These mechanisms may enable micro-organisms to acquire essential miconutrients while limiting the amount of arsenic incorporated, as well as controlling their responses to this toxic element.

iii. *Arsenic resistance and metabolism*

Not surprisingly, many functional studies have shown that the expression of arsenic-related genes is induced by arsenic stress. The expression of arsenic resistance genes is known to be induced by arsenic in micro-organisms belonging to all the domains of life: it has been clearly established by now that ArsR/Acr1p, ArsB/Acr3/Acr3p, ArsC/Acr2p, ArsA, ArsH, ArsP and ArsA encoding transcripts or proteins are preferentially accumulated and involved in the responses of bacteria, archaea and eukaryotes to arsenic (Haugen et al., 2004; Carapito et al., 2006; Thorsen et al., 2007, 2009; Zhang et al., 2007, 2009; Bryan et al., 2009; Cozen et al., 2009; Weiss et al., 2009; Cleiss-Arnold et al., 2010; Li et al., 2010; Pandey et al., 2012; Andres et al., 2013). In addition, the expression of *YCF1* and its homolog *PGPA*, both of which are responsible for As(III) vacuolar sequestration, is induced by As(III) and upregulated in comparison with the wild-type strain in *S. cerevisiae* and *L. amazonensis* A, respectively (Lin et al., 2008; Thorsen et al., 2009). Based on data obtained using the deletome approach, Ycf1p has also been found to play a crucial role in the resistance of yeast to arsenic (Jin et al., 2008).

On the other hand, while most functional studies have focused on arsenic resistant micro-organisms, some arsenic metabolizing microbes have also been investigated and the corresponding processes documented. In bacteria, studies using proteomic and transcriptomic approaches have shown that the expression of genes coding for arsenite oxidase and its two-component regulatory system AioSR is induced by As(III) in *Rhizobium* sp. NT-26 and *H. arsenicoxydans*, respectively (Cleiss-Arnold et al., 2010; Andres et al., 2013). The respiratory arsenate reductase operon identified in the archaea *P. aerophilum* is also induced in the presence of As(V), as mentioned above (Cozen et al., 2009). In eukaryotes, a study based on RNA sequencing has shown the constitutive expression of *arsM*, *arsA* and *acr3* homologs in *E. mutabilis* (Halter et al., 2014).

These data therefore amply confirm that arsenic triggers processes whereby micro-organisms either reduce the toxicity of this chemical via detoxifying reactions or metabolize it. However, despite the occurrence of all these arsenic resistance and envelope-related processes preventing it from penetrating into the cells, arsenic still often contaminates the intracellular compartment.

iv. *Oxidative stress responses*

Arsenic toxicity is partly due to oxidative stress, since it can generate reactive oxygen species such as hydrogen peroxide, superoxide anions, hydroxyl radicals and organic hyperoxides and deplete the cellular antioxidant pool (Bernstam and Nriagu, 2000; Flora, 2011). Arsenic therefore causes damage to proteins, lipids and DNA (Valko et al., 2005; Flora, 2011). Interestingly, some micro-organisms have evolved mechanisms limiting the oxidative stress and maintaining the reducing intracellular environment required by many cellular activities.

On the one hand, prokaryotes and eukaryotes produce similar responses to the oxidative stress caused by arsenic. In the first place, Fe- and Mn-superoxide dismutase homologue transcripts and proteins responsible for the process of superoxide anion dismutation into the less toxic hydrogen peroxide and oxygen accumulate preferentially in the presence of As(III) in *H. arsenicoxydans*, *K. pneumonia*, *Rhizobium* sp. NT-26 and *Exiguobacterium* sp. S17 (Weiss et al., 2009; Cleiss-Arnold et al., 2010; Daware et al., 2012; Andres et al., 2013; Belfiore et al., 2013). Although lower levels of accumulation of these proteins have been observed in *Anabaena* sp. PCC7120, which shows alternative responses, corresponding transcripts are upregulated in *L. amazonensis* A (Lin et al., 2008). Secondly, thiol peroxidases, thioredoxin reductases, thioredoxins and glutaredoxins are upregulated by arsenic in *H. arsenicoxydans*, *K. pneumoniae*, *Anabaena* sp. PCC7120, *F. acidarmanus* Fer1, *Tetrahymena pyriformis* and *S. cerevisiae* (Baker-Austin et al., 2007; Thorsen et al., 2007; Cleiss-Arnold et al., 2010; Daware et al., 2012; Pandey et al., 2012; Zhang et al., 2012). These enzymes are involved in thioredoxins' and glutaredoxins' redox pathways, which serve mainly to control the intracellular redox potential by providing antioxidant enzymes with reducing equivalents, thus maintaining a reducing thiol- rich state and decreasing the oxidative stress (Meyer et al., 2009). Thiol peroxidases are also involved in the reductive detoxification of hydrogen peroxide and organic hydroperoxides, and thioredoxins and glutaredoxins in the reduction of As(V) (Mukhopadhyay and Rosen, 2002; Meyer et al., 2009), which shows that these molecules contribute importantly to these organisms' responses to arsenic. Thirdly, while arsenic induces GSH biosynthesis and regeneration in only a few bacteria such as *C. testosteroni* sp. CNB-1, *L. ferriphilum*, *H. arsenicoxydans* and *Anabaena* sp. PCC7120 (Zhang et al., 2007; Cleiss-Arnold et al., 2010; Li et al., 2010; Pandey et al., 2012), all the functional genomic studies performed so far on eukaryotic organisms have established that GSH-related functions such as glutaredoxin, GSH peroxidase and reductase and gamma-glutamylcysteine synthetase are upregulated during exposure to arsenic stress in *S. cerevisiae* and *T. pyriformis* (Haugen et al., 2004; Thorsen et al., 2007, 2009; Jin et al., 2008; Zhang et al., 2012). In addition, a

study on *L. amazonensis* A has shown that all the components of the trypanothione dependent tryparedoxin cascade (trypanothione is a GSH derivative) are upregulated in comparison with the wild-type strain (Lin et al., 2008; Fahey, 2013). GSH, which is involved in glutaredoxin regeneration, takes part in arsenate reduction and antioxidant responses via its ability to trap radicals, reduce peroxides and maintain the cell redox state. GSH preserves the functional integrity of cells by chelating metal ions and mediating their vacuolar sequestration, by protecting proteins from oxidation and reparing oxidized ones, and by eliminating lipid and fatty acid peroxides (Grant, 2001; Pompella et al., 2003). Interestingly, the fact that GSH-related processes are induced in only a few bacteria, which also upregulate other antioxidant enzymes, suggests that GSH may play only an accessory role in bacterial resistance to oxidative stress. In agreement with this hypothesis, it has been reported that GSH is used in bacteria at low hydrogen peroxide concentrations, whereas alternative antioxidant mechanisms involving enzymes such as catalases are favoured at higher concentrations (see below, Vergauwen et al., 2003). By contrast, the induction of GSH metabolism observed in all the functional studies performed so far on eukaryotes supports the central role it has been assumed to play in these organisms (Vergauwen et al., 2003). Lastly, the authors of studies on transcriptomes and deletomes have described the upregulation of the Yap1p transcription factor in *S. cerevisiae* and its involvement in the response of this organism to As(III) (Haugen et al., 2004; Thorsen et al., 2007). Yap1p is known to control the expression of several targets, including *YCF1*, involved in oxidative stress resistance mechanisms, which were among the most highly upregulated processes, i.e. thioredoxins and GSH metabolisms, as shown by transcriptomic data.

On the other hand, some other specific processes have been found to contribute to micro-organisms' responses to oxidative stress. In the case of bacteria, organic hydroperoxide resistance proteins are upregulated in *Thiomonas* sp. 3As and *Rhizobium* sp. NT-26 (Bryan et al., 2009; Andres et al., 2013). Homologues of catalases, which are responsible for the scavenging and decomposition of high concentrations of hydrogen peroxide, are then upregulated in *C. metallidurans*, *H. arsenicoxydans*, *Anabaena* sp. PCC7120 and *Rhizobium* sp. NT-26 in the presence of arsenic (Vergauwen et al., 2003; Weiss et al., 2009; Zhang et al., 2009; Cleiss-Arnold et al., 2010; Pandey et al., 2012; Andres et al., 2013). Vitamin B6 metabolism is also upregulated in *C. metallidurans*, *H. arsenicoxydans*, *Anabaena* sp. PCC7120, *Exiguobacterium* sp. PS and *Rhizobium* sp. NT-26 (Zhang et al., 2009; Cleiss-Arnold et al., 2010; Pandey et al., 2012; Andres et al., 2013; Sacheti et al., 2013). Vitamin B6 has been found to contribute to several metabolic processes, including those involved in resistance to oxidative stress (Mukherjee et al., 2011). Other proteins, or their corresponding transcripts, which may also play a protective role against oxidative stress, are upregulated in *C. metallidurans*, *K. pneumoniae* and *Exiguobacterium* sp. PS, namely thioesterase, flavodoxin

reductase, cystine binding protein, arginine deaminase and pyridine nucleotide-disulfide reductase (Zhang et al., 2009; Daware et al., 2012; Sacheti et al., 2013). In eukaryotes, As(III) exposure induces the expression of polyamine transporters in *S. cerevisiae* and *E. mutabilis* (Jin et al., 2008; Halter et al., 2014). Polyamines such as spermidine and putrescine are involved in protecting organisms from reactive oxygen species (Chattopadhyay et al., 2006).

Functional genomics have therefore brought to light some common and more specific processes triggered by the presence of arsenic, whereby micro-organisms reduce the oxidative stress generated by this element and thus protect the precious intracellular medium. However, cell components such as DNA, proteins and lipids still run the risk of being damaged by processes such as mutagenesis, misfolding and peroxidation, respectively (Flora, 2011), and therefore require additional mechanisms in order to remain functional.

v. *Maintaining the functional integrity of cells*

One of the processes whereby micro-organisms keep their functional integrity under stressfull conditions consists in preserving existing molecules. This mechanism often involves general stress response proteins such as heat shock proteins (HSPs) or other chaperones. HSPs are involved in the tolerance of cells to various stresses by promoting proper protein folding, assembly, renaturation and degradation: the latter role prevents the accumulation of misfolded proteins, which are liable to cause cellular damage (Parsell and Lindquist, 1993; Hartl, 1996). Some HSPs are also involved in mRNA stability (Yoon et al., 2008). The Hsp70 system, consisting of DnaK (Hsp70), DnaJ (Hsp40) and GrpE, and the Hsp60 system, consisting of GroEL (Hsp60) and GroES (Hsp10) (Houry, 2001), are the two systems which have been the most commonly found in functional genomic studies to be induced by arsenic in all three domains of life. DnaK has been identified in almost half of these studies and found to be upregulated in bacteria, archaea and eukaryotes such as *C. testosteroni* sp. CNB-1, *H. arsenicoxydans*, *K. pneumoniae*, *Anabaena* sp. PCC7120, *Exiguobacterium* sp. S17 and PS, *Rhizobium* sp. NT-26, *F. acidarmanus* Fer1, *T. pyriformis* and *S. cerevisiae*, (Baker-Austin et al., 2007; Zhang et al., 2007, 2012; Jin et al., 2008; Weiss et al., 2009; Cleiss-Arnold et al., 2010; Daware et al., 2012; Pandey et al., 2012; Andres et al., 2013; Belfiore et al., 2013; Sacheti et al., 2013). Likewise, its co-chaperone and its nucleotide exchange factor, DnaJ and GrpE, are upregulated in *C. testosteroni* sp. CNB-1 and *P. aeruginosa*, respectively (Parvatiyar et al., 2005; Zhang et al., 2007). GroEL and GroES are upregulated by arsenic in *T. arsenivorans*, *K. pneumoniae*, *Exiguobacterium* sp. S17, *Rhizobium* sp. NT-26, *F. acidarmanus* Fer1 and *S. cerevisiae* (Baker-Austin et al., 2007; Thorsen et al., 2007; Bryan et

al., 2009; Daware et al., 2012; Andres et al., 2013; Belfiore et al., 2013). Other heat shock proteins are upregulated by arsenic in both bacteria and eukaryotes such as in *P. aeruginosa*, *C. testosteroni* sp. CNB-1, *H. arsenicoxydans*, *L. ferriphilum*, *K. pneumoniae* and *S. cerevisiae* (Haugen et al., 2004; Parvatiyar et al., 2005; Thorsen et al., 2007; Zhang et al., 2007; Jin et al., 2008; Cleiss-Arnold et al., 2010; Li et al., 2010; Daware et al., 2012). Other chaperones have only been identified in bacteria. Clp ATPases and the ribosome associated trigger factor, which are involved in the unfolding of proteins and aggregates prior to their degradation and the folding of newly-synthesized nascent polypeptide chains, respectively (Houry, 2001), are upregulated by As(III) in *K. pneumoniae* and *L. ferriphilum*, respectively (Li et al., 2010; Daware et al., 2012). Lastly, universal stress proteins are upregulated in the presence of arsenic in *C. metallidurans*, *H. arsenicoxydans*, and *Exiguobacterium* sp. S17 and PS (Zhang et al., 2009; Cleiss-Arnold et al., 2010; Belfiore et al., 2013; Sacheti et al., 2013). Interestingly, in the latter two strains, cold shock proteins are also upregulated in response to arsenic. These proteins corresponding to RNA chaperones are involved in several stress responses involving the rescue of RNA blocked in non-productive configurations (Belfiore et al., 2013). All in all, these functional genomic data converge to show that HSPs, along with other chaperones, universally play a central role in micro-organisms' responses to arsenic, preserving the RNA and protein pools. However, despite the protective effects of these general stress response proteins, nucleic acids, polypeptides and fatty acids are still liable to be affected by arsenic.

Arsenic species can indeed cause DNA damage such as single- or double-strand DNA breaks, base damage and the formation of alkali-labile sites (Ding et al., 2009; Flora, 2011), as well as inhibiting DNA repair processes and influencing epigenetic mechanisms by modifying histones chemically and changing the structure of chromatine (Dilda et al., 2008; Bustaffa et al., 2014). Functional studies have shown that various processes associated with these alterations are differentially expressed in the presence of arsenic. Several proteins and transcripts contributing to DNA repair mechanisms are upregulated in the presence of As(III) in *Thiomonas* sp. 3As, *H. arsenicoxydans*, *L. ferriphilum* and in many eukaryotes (Bryan et al., 2009; Weiss et al., 2009; Cleiss-Arnold et al., 2010; Li et al., 2010). The expression of genes coding for proteins involved in DNA repair systems such as mitochondrial DNA polymerase I and telomere maintenance protein, transcriptional activator of nucleotide and base excision repair, is either upregulated or induced by arsenic in *L. amazonensis* A and *S. cerevisiae* (Haugen et al., 2004; Jin et al., 2008; Lin et al., 2008). Likewise, *E. mutabilis* induces the expression of the DNA mismatch repair system in the presence of As(III), but also constitutively expresses genes involved in nucleotide excision repair and non-homologous end-joining systems, contrary to the laboratory strain *E. gracilis* (Halter et al., 2014). Interestingly, the cellular cycle of yeasts has been found to slow down to adapt to high intracellular

As(III) levels by triggering a G(1) or G(2) checkpoint delay which depends on Hog1p (Migdal et al., 2008). In agreement with these results, the expression of genes coding for cell cycle progression proteins is induced by arsenic in *S. cerevisiae*, and a study on the deletome profile has confirmed the involvement of these genes, including *HOG1*, in arsenic tolerance mechanisms (Haugen et al., 2004; Thorsen et al., 2009). These cell cycle perturbations may be associated with the control of DNA damage. Several functional studies have shown the existence of a differential expression of proteins involved in epigenetic modifications in the presence of arsenic. In this context, functions associated with DNA stability and chromatin structure such as histones, DNA gyrase and histone-like DNA-binding protein are upregulated in *T. arsenivorans*, *H. arsenicoxydans* and *L. amazonensis* A (Lin et al., 2008; Bryan et al., 2009; Cleiss-Arnold et al., 2010). In addition, although both hypo- and hyperacetylation of histones have been reported to occur in the presence of arsenic (Ren et al., 2011), a histone deacetylase was found to be upregulated in *L. amazonensis* A in comparison with the wild-type strain (Lin et al., 2008). DNA repair processes therefore seem to be induced in prokaryotes and eukaryotes in the presence of arsenic, which preserve the integrity of their genomes. Although the processes associated to the epigenetic modifications may result from the damage caused by arsenic, they may also reflect changes in the patterns of gene expression which are part of organisms' own protective responses to this toxic element.

Some functions associated with the control of gene expression and transcriptional processes are affected by arsenic in both prokaryotes and eukaryotes. Several transcriptional regulators are expressed differentially in the presence of this toxic element. Contrary to the repression of σ^E observed in *H. arsenicoxydans* mentioned above, σ^{54} expression is induced in *Rhizobium* sp. NT-26 (Cleiss-Arnold et al., 2010; Andres et al., 2013). Transcription regulators with a differential level of expression have been detected in *C. testosteroni* sp. CNB-1, *Anabaena* sp. PCC7120, *Exiguobacterium* sp. S17, *P. aerophilum*, *T. pyriformis* and *S. cerevisiae* (Haugen et al., 2004; Zhang et al., 2007, 2012; Cozen et al., 2009; Pandey et al., 2012; Belfiore et al., 2013). Studies on the deletome of *S. cerevisiae* have indicated that some stress-related transcription factors are involved in As(III) tolerance (Jin et al., 2008; Thorsen et al., 2009). The changes in the expression of these regulators induced by arsenic may reflect the setting up of arsenic responses. However, most of the changes in transcription-related functions identified in arsenomic studies have turned out surprisingly to point a decrease in the rates of mRNA synthesis in both bacteria and eukaryotes: downregulation of RNA polymerase β and α subunits and a transcription elongation factor was observed in the presence of arsenic in *Rhizobium* sp. NT-26, *Exiguobacterium* sp. PS and *L. ferriphilum*, respectively (Li et al., 2010; Andres et al., 2013; Sacheti et al., 2013). In addition, a polyribonucleotide nucleotidyltransferase involved in mRNA degradation is upregulated in *T. arsenivorans* and *H. arsenicoxydans* (Bryan et al., 2009; Weiss

et al., 2009). Likewise, the expression of genes encoding RNA helicase and decapping proteins potentially involved in mRNA degradation as well as negative regulators of transcription from RNA polymerase II promoters is induced in the presence of arsenic in *S. cerevisiae*, whereas the expression of genes encoding proteins involved in mRNA 3' end maturation is repressed (Haugen et al., 2004; Jin et al., 2008). However, mRNA synthesis is still an important cellular process contributing to the tolerance of the latter organism to arsenic, as established by authors using the deletome approach (Thorsen et al., 2009). The above findings may therefore be attributable to the development of a defense phenotype which redirects the energy available towards resistance mechanisms. In other words, cells may slow down their transcriptional activity and stimulate the expression of specific genes. In addition, micro-organisms may use nucleotides resulting from mRNA recycling processes to preserve the integrity of the DNA rather than for transcription or even replication purposes, as no transcripts or proteins specifically involved in the latter process have been identified in functional studies. In this respect, the expression of some genes involved in nucleotide metabolism, including nucleotide salvage pathways which enable energy savings, as well as their *de novo* synthesis, is upregulated in both bacteria and eukaryotes: those involved in the synthesis and transport of purines, pyrimidines and their precursors are upregulated in *L. ferriphilum*, *K. pneumoniae*, *Rhizobium* sp. NT-26, *Exiguobacterium* sp. S17 and PS, *S. cerevisiae* and *L. amazonensis* A (Haugen et al., 2004; Lin et al., 2008; Li et al., 2010; Daware et al., 2012; Andres et al., 2013; Belfiore et al., 2013; Sacheti et al., 2013). The overall results obtained using functional tools therefore show that changes in transcriptional activity occur in the presence of arsenic in both prokaryotes and eukaryotes, possibly enabling micro-organisms to favour specific processes, including those underlying DNA repair systems.

Arsenic can also cause damage to proteins involved in either the metabolism or resistance mechanisms (Shen et al., 2013). Several processes have been found to occur which enable micro-organisms to maintain the functional efficiency of their protein pool by fixing, suppressing, and replacing damaged polypeptides. Although chaperones usually help to retrieve misfolded proteins, some of the latter proteins may remain non functional and possibly even toxic and have to be eliminated. For this purpose, micro-organisms have evolved processes such as proteolysis, based on digestive proteases or the ubiquitin-proteasome system. However, the possible link between arsenic and degradation processes of this kind has not yet been clearly established in bacteria. While some proteases are upregulated in *Anabaena* sp. PCC7120 and *Exiguobacterium* sp. PS in the presence of arsenic (Pandey et al., 2012; Sacheti et al., 2013), others, as well as some proteasome-like degradation complex constituents are downregulated in *H. arsenicoxydans* and *Rhizobium* sp. NT-26 (Carapito et al., 2006; Andres et al., 2013). By contrast, all the functional analyses performed so far

on eukaryotes have shown that protein degradation activities are induced in the presence of arsenic. Some of the components of the ubiquitin-proteasome system, such as the ubiquitin carrier, ubiquitin ligase and some proteasome subunits and corresponding transcriptional activators, are upregulated in *S. cerevisiae*, *L. amazonensis* A, *T. pyriformis* and *E. mutabilis*, as shown by transcriptomic and proteomic studies (Haugen et al., 2004; Thorsen et al., 2007; Jin et al., 2008; Lin et al., 2008; Zhang et al., 2012; Halter et al., 2014). Although the enhanced expression of genes coding for proteasome-related proteins may be attributable to the renewal of some arsenic sensitive components (Haugen et al., 2004), the additional upregulation of ubiquitin-related proteins observed in the presence of arsenic seems more likely to reflect the induction of protein degradation. Concomitantly with this process removing the damaged polypeptides, the occurrence of protein synthesis would replace these cellular components and preserve the protein pool. In this connection, several molecules involved in translation processes, such as rRNA, ribosomal proteins, tRNA and elongation factors, are affected by arsenic in all three domains of life. In prokaryotes, *C. testosteroni* CNB-1, *Staphylococcus* sp. NBRIEAG-8 and *Exiguobacterium* sp. PS constitute one group in which ribosomal biogenesis and translation are repressed in the presence of arsenic (Zhang et al., 2007; Srivastava et al., 2012; Sacheti et al., 2013). By contrast, these activities are upregulated in the second group, which is composed of *P. aeruginosa*, *H. arsenicoxydans*, *T. arsenivorans*, *L. ferriphilum*, *Rhizobium* sp. NT-26, *Y. enterolitica* 1A, *K. pneumoniae*, *Anabaena* sp. PCC7120, *Exiguobacterium* sp. S17 and *F. acidarmanus* Fer1 (Parvatiyar et al., 2005; Carapito et al., 2006; Baker-Austin et al., 2007; Bryan et al., 2009; Weiss et al., 2009; Cleiss-Arnold et al., 2010; Li et al., 2010; Daware et al., 2012; Mallik et al., 2012; Pandey et al., 2012; Andres et al., 2013; Belfiore et al., 2013). In eukaryotes, on the other hand, in *T. pyriformis*, *S. cerevisiae*, *E. gracilis* and *L. amazonensis* wild-type strain, which were isolated in non-contaminated environments and do not show a high level of arsenic tolerance, the identified components involved in translation processes were all downregulated in the presence of arsenic or in comparison with the A variant in the case of *L. amazonensis* (Haugen et al., 2004; Thorsen et al., 2007; Jin et al., 2008; Lin et al., 2008; Zhang et al., 2012; Halter et al., 2014). The downregulation of the protein kinase A and CK2 observed in yeast may also be involved in a process of this kind (Jin et al., 2008). By contrast, in the arsenic-resistant strains *L. amazonensis* A and *E. mutabilis*, several rRNA, ribosomal proteins and transcripts involved in the initiation and termination of translation and post-translational modifications were upregulated in comparison with the wild type strain and in the presence of As(III), respectively (Lin et al., 2008; Halter et al., 2014). Since ribosome production may account for more than 50 % of the work of synthesis occurring in eukaryotic cells, organisms in which translation activities are repressed, such as arsenic sensitive eukaryotes and the first group of bacteria mentioned above, will presumably redirect their cellular resources towards other resistance processes such as those involving chaperones (Jin et al., 2008). However, the translation rates seem

to be a key to arsenic tolerance, as illustrated by the upregulation of protein synthesis observed in arsenic tolerant eukaryotes. Since the latter organisms are closely related to arsenic-sensitive ones such as *L. amazonensis* A, an arsenic-selected genetic variant of *L. amazonensis*, for example, the higher rates of protein synthesis observed may well result from the adaptation of these organisms to arsenic. Although this increase in translation activity may be linked to the protein turnover despite the fact that no degradation processes were detected in some cases, it may also enable the production of proteins specifically involved in these organisms' responses to arsenic. In line with the possible induction of translational processes, functional studies subsequently showed that genes involved in the metabolism of protein synthesis building blocks, i.e. amino acids, are differentially expressed in the presence of arsenic in both prokaryotes and eukaryotes. With the exception of *L. ferriphilum* and *Y. enterolitica* 1A, all the studies in which proteins or transcripts involved in amino acids' metabolism have been identified, i.e. in the case of *P. aeruginosa*, *H. arsenicoxydans*, *K. pneumoniae*, *Anabaena* sp. PCC7120, *Exiguobacterium* strains S17 and PS, *P. aerophilum*, *S. cerevisiae*, *L. amazonensis* A and *E. mutabilis*, have shown that either their biosynthesis, interconversion and transport are upregulated or their catabolism is downregulated (Haugen et al., 2004; Parvatiyar et al., 2005; Carapito et al., 2006; Thorsen et al., 2007; Jin et al., 2008; Lin et al., 2008; Cozen et al., 2009; Daware et al., 2012; Pandey et al., 2012; Belfiore et al., 2013; Sacheti et al., 2013). All these data based on the use of functional genomic tools show that the changes in protein degradation and synthesis which occur in response to the presence of arsenic tend to differ from one domain or one organism to another.

Micro-organisms therefore dispose of several maintenance and renewal mechanisms counteracting the toxic effects of arsenic on their DNA, RNA and protein pools. As we have seen, the upregulation of GSH biosynthesis and enzymes involved in fatty acid metabolism can also protect lipids from peroxidation and help with their recycling, respectively, thus stabilizing the cell envelope in the presence of arsenic. However, the efficiency of each of these processes depends on the occurrence of elementary reactions supplying either the precursors or the energy required.

vi. *Central metabolism: sulfur, nitrogen, carbon and energy*

The pool of GSH molecules is closely linked to the sulfur metabolism. Functional approaches have shed interesting light on the differential expression of several pathways involving sulfur which occurs in response to arsenic in both eukaryotes and bacteria. First, microarray data combined with deletion experiments suggested that Acr1p may be involved in the transcriptional control of sulfur

assimilation in *S. cerevisiae*, possibly highlighting the existence of a direct link between the sulfur and arsenic metabolisms (Haugen et al., 2004). In addition, transcriptome, proteome and deletome studies have all shown that the assimilation and metabolism of sulfur are upregulated and that they play an important role in the presence of As(III), from its incorporation to the biosynthesis of sulfur containing amino acids and GSH in *S. cerevisiae*, *E. mutabilis* and *E. gracilis* (Haugen et al., 2004; Thorsen et al., 2007; Jin et al., 2008; Halter et al., 2014). In agreement with these findings, a metabolite profiling study has shown that GSH precursors accumulate preferentially in the presence of As(III) in *S. cerevisiae* (Thorsen et al., 2007). Lastly, a bioinformatic analysis of promoter sequences combined with a microarray approach has shown the existence of common motifs in groups of genes showing a similar pattern of expression in *S. cerevisiae*. Assuming that these sequences play the role of transcription factor binding sites, their respective regulators are involved in sulfur assimilation, GSH metabolism and ribosomal protein synthesis (Thorsen et al., 2007). All these data further support the idea that GSH plays a central role in eukaryotes' responses to arsenic. In addition, the increase observed in the rates of GSH biosynthesis might explain the downregulation of protein synthesis observed in *S. cerevisiae* in the presence of As(III): the sulfur assimilated may be conveyed towards the synthesis of GSH rather than that of proteins (Haugen et al., 2004). By contrast, the authors of functional studies on bacteria have established that sulfur oxidation and transport are inhibited by As(III) in *Thiomonas* sp. 3As, in the late response of *H. arsenicoxydans* and in *Rhizobium* sp. NT-26 (Bryan et al., 2009; Cleiss-Arnold et al., 2010; Andres et al., 2013). However, it has recently been suggested that this inhibition may be linked to another regulatory mechanism (Andres et al., 2013). These three organisms express an arsenite oxidase, the activity of which is inhibited by some of the reaction intermediates of the sulfur metabolism (Lieutaud et al., 2010). In line with this assumption, the inhibition of the sulfur metabolism observed in *H. arsenicoxydans* only occurs in the late response, concomitantly with the induction of arsenite oxidase expression, as established on the basis of transcriptomic data (Cleiss-Arnold et al., 2010). These data therefore show that the changes in sulfur metabolism which occur in response to arsenic differ between eukaryotes and prokaryotes.

Nucleic and amino acid biosynthesis, which are affected by arsenic, are both involved in the metabolism of nitrogen. In *E. mutabilis*, these two pathways are both induced and associated with the active DNA repair and protein turnover processes in the presence of As(III), and the transcriptome is also enriched in transcripts associated with nitrate transport processes, which increases the nitrogen inputs with which these pathways are supplied (Halter et al., 2014). By contrast, based on the results of biochemical tests, *Anabaena* sp. PCC7120 has been found to show a lower nitrogen fixation activity in the presence of As(V). However, the results obtained using a proteomic approach have suggested that an asparinage may be upregulated, yielding the ammonia

required by the process of amino acid biosynthesis (Pandey et al., 2012). In other bacteria and one member of the archaea, namely *Rhizobium* sp. NT-26, *C. metallidurans* and *P. aerophilum*, the metabolism of nitrogen is induced in the presence of arsenic, but the corresponding identified transcripts are rather involved in the reduction of compounds containing nitrogen, and therefore in respiratory processes (Cozen et al., 2009; Zhang et al., 2009; Andres et al., 2013). However, the possible coupling between arsenite oxidation and nitrate reduction processes does not support the growth of *Rhizobium* sp. NT-26 (Andres et al., 2013).

Nucleic, amino and fatty acids are alternatively involved in the central metabolism which generates precursor metabolites as well as reducing equivalents maintaining the intracellular redox potential. Studies based on the use of functional genomic tools have shown that arsenic affects the expression of proteins involved in inorganic carbon fixation and organic carbon metabolism, which are mediated by the Calvin cylcle, pentose phosphate pathway, gluconeogenesis, glycolysis and the tricarboxilic acid (TCA) cycle. Unfortunately, however, because of the diversity of the metabolisms and the fact that only a few proteins were identified in these pathways in all these proteomic studies, it would be risky to put forward any overall hypotheses about these processes. In addition, some studies have shown that a given enzyme can be both up- and downregulated in different organisms, and the same applies to different enzymes involved in a given pathway in a single organism. However, various trends have emerged upon examining data independently of the trophic requirements. In one group of organisms composed of *C. testosteroni* CNB-1, *T. arsenivorans*, *H. arsenicoxydans*, *Y. enterolitica* 1A, *Anabaena* sp. PCC7120, *Exiguobacterium* strains S17 and PS, *F. acidarmanus* Fer1, *S.cerevisiae*, *L. amazonensis* A and *T. pyriformis*, glycolysis and TCA cycle enzymes identified using functional approaches were all found to be upregulated under arsenic stress conditions (Baker-Austin et al., 2007; Zhang et al., 2007, 2012; Jin et al., 2008; Lin et al., 2008; Bryan et al., 2009; Weiss et al., 2009; Mallik et al., 2012; Pandey et al., 2012; Belfiore et al., 2013; Sacheti et al., 2013). In the second group, which is composed of *T. arsenivorans* and *Rhizobium* sp. NT-26, although some enzymes involved in glycolysis and the TCA cycle were found to be downregulated only in the latter organism, enzymes known to be involved in the Calvin cycle were upregulated by As(III) in both organisms (Bryan et al., 2009; Andres et al., 2013). These findings are supported by the ability of both strains to grow autotrophically, using As(III) as the electron donor (Santini et al., 2000; Bryan et al., 2009). Lastly, identified enzymes involved in glycolysis and the TCA cycle show a mixed pattern of regulation in *K. pneumoniae*, whereas the pentose phosphate pathway seems to be upregulated. This may lead to the accumulation of reducing equivalents and amino and nucleic acids, which, along with the downregulation of the amino acid catabolism observed, supports the idea that protein synthesis is upregulated in response to As(III) in this organism (Daware et al., 2012). All these

results suggest that both prokaryotes and eukaryotes might favour reducing equivalents and energy productive pathways to provide the energy requirements for the functioning of arsenic tolerance mechanisms.

In agreement with this hypothesis, gluconeogenesis is inhibited in both eukaryotes and bacteria in the presence of arsenic, as established in studies using functional genomic tools. Proteins and transcripts encoding proteins involved in this pathway accumulate less strongly in the presence of arsenic in *S. cerevisiae*, *Thiomonas* sp. 3As and *Anabaena* sp. PCC7120 (Haugen et al., 2004; Thorsen et al., 2007; Bryan et al., 2009; Pandey et al., 2012). Therefore, rather than consuming energy to produce reserves, the cells may give priority to generating more energy. In addition, several metabolic pathways ending in the TCA cycle are induced in response to arsenic. First, amino acids generated by protein degradation activities may either serve as energy substrates which are fed into the TCA cycle or as nitrogen and carbon sources. Secondly, studies on the deletome and transcriptomes of *S. cerevisiae* and *L. amazonensis* A have shown that the fatty acid metabolism is involved in the responses of these organisms to arsenic (Jin et al., 2008; Lin et al., 2008): enzymes involved in fatty acid oxidation are upregulated in both organisms, resulting in the degradation of long chains into acetyl-CoA, which is able to enter the TCA cycle (Haugen et al., 2004; Jin et al., 2008; Lin et al., 2008). In addition, some enzymes involved in the first few steps of fatty acid biogenesis catalyze reversible reactions, and the fact that they are upregulated in *P. aeruginosa*, *C. metallidurans* and *Rhizobium* sp. NT-26 may also be due to these molecules being preferntially used as substrates (Parvatiyar et al., 2005; Zhang et al., 2009; Andres et al., 2013). Lastly, a study on the deletome of *S. cerevisiae* has shown that an alcool dehydrogenase plays an essential role in As(III) tolerance (Jin et al., 2008). Interestingly, the substrates in question, namely fatty acids and alcool, may enter the TCA cycle via the glyoxylate shunt, some components of which are upregulated in the presence of arsenic in *Thiomonas* sp. 3As, *T. arsenivorans* and *C. testosteroni* (Zhang et al., 2007; Bryan et al., 2009).

In agreement with the induction of these metabolic pathways and the above energy production hypothesis, an upregulation of transcripts or proteins involved in all the steps of oxidative phosphorylation has been observed in most of the studies performed on representatives of all three domains of life. Apart from *Rhizobium* sp. NT-26 and *L. ferriphilum*, in which proteins are regulated in both ways, some of the components of NADH dehydrogenase, succinate dehydrogenase, cytochrome bc_1 complex, cytochrome *c* oxydase and ATP synthase are upregulated in the presence of arsenic in *C. testosteroni* CNB-1, *H. arsenicoxydans*, *Y. enterolitica* 1A, *K. pneumoniae*, *Anabaena* sp. PCC7120, *Exiguobacterium* sp. PS, *F. acidarmanus* Fer1, *S. cerevisiae*, *L. amazonensis* A and *T. pyriformis* (Baker-

Austin et al., 2007; Zhang et al., 2007, 2012; Jin et al., 2008; Lin et al., 2008; Weiss et al., 2009; Daware et al., 2012; Mallik et al., 2012; Pandey et al., 2012; Sacheti et al., 2013). In addition, the study of the *S. cerevisiae* deletome has shown the importance of mitochondrial functions and biogenesis in arsenic tolerance (Thorsen et al., 2009). Lastly, proteins involved in photosynthesis are either upregulated or their normal rates of accumulation are recovered in an acclimatation phase to As(V) in *Anabaena* sp. PCC7120 (Pandey et al., 2012). The induction of these various processes might allow micro-organisms to maintain sufficiently high intracellular energy levels during their exposure to arsenic stress.

All these data obtained using functional approaches have brought to light various processes enabling micro-organisms to cope with arsenic, therefore describing at least some of the overall physiological effects of this toxin on cells. Some of the mechanisms described are common to either closely or distantly related organisms, whereas others are specific to a given organism. These data therefore give a picture of the multiple cellular processes involved in microbial responses to arsenic (Fig. 3). There are still many gaps to be filled, however, and further studies will have to be performed on more phylogenetically diverse bacteria, as well as micro-eukaryotes and archaea, on which few studies have focused so far although these organisms are thought to play a key role in arsenic cycling processes. Functional studies on higher eukaryotes are likely to provide useful information about the former group, but archaea may deploy some rather unusual strategies, as suggested above in the case of *F. acidarmanus* Fer1. It is also possible that some atypical processes may be at work in non-culturable organisms belonging to the three domains of life. Culture independent approaches such as metagenomic and metaproteomic approaches have therefore been used in order to understand in greater detail how micro-organisms and microbial communities thrive in arsenic contaminated environments.

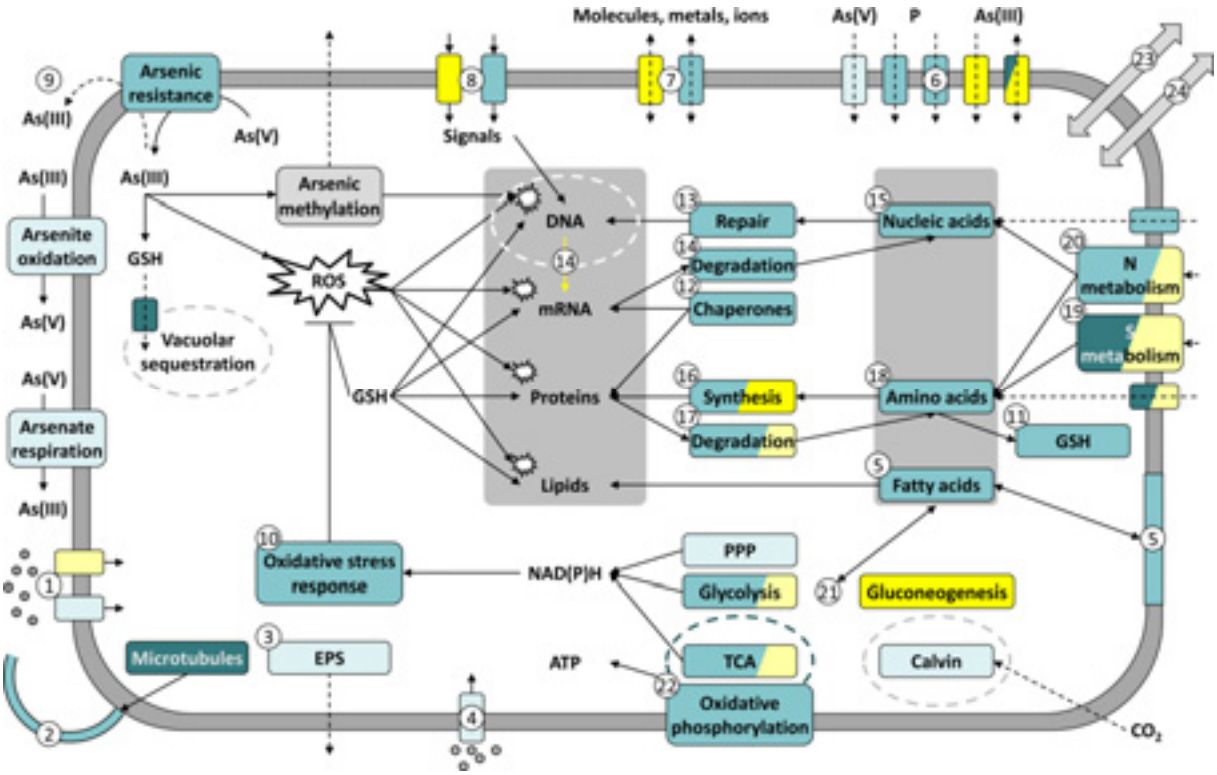


Figure 3. Diagram of the multiple cellular processes involved in the microbial responses to arsenic determined using functional genomic approaches. Mechanisms colored light or dark yellow have been found to be downregulated in prokaryotes or in both prokaryotes and eukaryotes, respectively, whereas those colored light, dark or medium blue are upregulated in prokaryotes, eukaryotes or both, respectively. Boxes with two colors mean that a different type of regulation (up- or down-) has been observed, depending on the organisms. Dashed and standard arrows stand for molecular transport and signaling/regulatory or metabolic pathways, respectively. Dashed circles stand for internal eukaryotic membranes. Because of their structural differences, the prokaryotic and eukaryotic cellular envelope and flagellum have been simplified. Numbers 1 to 22 refer to biological functions in their order of appearance in the text. 1: chemotactism; 2: flagellum synthesis; 3: biofilm synthesis; 4: quorum sensing; 5: envelope composition and fatty acid metabolism; 6: arsenic entry routes such as those involving phosphorus, hexose transporters and porins; 7: diverse molecular transporters; 8: signaling and regulatory pathways; 9: arsenic resistance and metabolism; 10: oxidative stress responses; 11: glutathione metabolism; 12: molecular chaperones; 13: DNA repair systems; 14: mRNA synthesis and degradation; 15 and 18: nucleic and amino acid synthesis and transport, respectively; 16 and 17: protein degradation and synthesis, respectively; 19: sulfur metabolism, transport or oxidation; 20: nitrogen metabolism, transport, fixation or respiration; 21: central metabolism; 22: oxidative phosphorylation; 23 and 24: genetic and metabolic interactions, respectively. ROS, GSH, EPS, TCA and PPP are acronyms which stand for reactive oxygen species, glutathione, extracellular polymeric substances, tricarboxylic acid cycle and the pentose phosphate pathway, respectively.

c) Accessing non-culturable micro-organisms and microbial communities using genomic approaches

i. *The metagenomics of arsenic contaminated ecosystems*

Since most micro-organisms cannot be isolated using culture methods (Amann et al., 1995), these approaches are not suitable for studying some specific organisms and microbial communities. The development of sequencing methods has made it possible to circumvent this problem by directly analyzing the environmental DNA. The so-called metagenomic approach gives access to the whole pool of genes of environmental samples, and therefore yields useful information about the composition and the metabolic potential of the microbial communities they include. Methods of this kind have recently been applied in studies on the microbial communities inhabiting AMD environments, which are characterized by low pH values and high metal concentrations, including arsenic.

Along with PCR-based approaches, metagenome sequencing methods can be used to study the diversity of micro-organisms in arsenic contaminated environments by identifying and comparing taxonomic biomarkers. In this context, the latter approaches have been used to explore the genetic diversity of microbial communities inhabiting arsenic contaminated sites, focusing on 16S rDNA, 18S rDNA and concatenated marker sequences, (Baker and Banfield, 2003; Aguilera et al., 2007; Baker et al., 2009; Bertin et al., 2011; Cai et al., 2013; Plewniak et al., 2013; Mendez-Garcia et al., 2014; Volant et al., 2014). Studies on these lines have shown that the Richmond Mine and Carnoulès AMD are both characterized by a relatively low level of prokaryotic diversity, i.e. they show low species richness or only a few predominant populations in comparison with other similar environments such as the Los Rueldos AMD and arsenic contaminated marine sediments (Baker and Banfield, 2003; Bertin et al., 2011; Plewniak et al., 2013; Mendez-Garcia et al., 2014). This difference is thought to be due to the physico-chemical conditions, which are less extreme in the latter two ecosystems. Secondly, metagenome sequencing methods can be used to study the distribution, diversity and abundance of functional markers such as the genes involved in arsenic metabolism (Cai et al., 2013). Although these genes do not always reflect the taxonomic diversity or cannot be assigned to their respective species because of the possible occurrence of HGT (Heinrich-Salmeron et al., 2011; Li et al., 2013), studies on the presence of arsenic reduction, oxidation and methylation genes and their abundance in a given environment should help to understand or predict the fate of arsenic, i.e. whether it has been or is likely to be mobilized, immobilized or volatilized. In this connection, the

authors of a recent differential metagenomic study have identified sulfate reduction, arsenate reduction and fermentation as major functions in a microbial community inhabiting arsenic contaminated marine sediments, which explains why arsenic was found to show greater mobility here than at other sites (Plewniak et al., 2013). Thirdly, metagenomic DNA can be used to screen environmental samples for unknown genes possibly involved in arsenic resistance. For example, the community DNA obtained from the sludge of an effluent treatment plant and more recently, that of the Rio Tinto AMD have been used to construct DNA fragment libraries (Chauhan et al., 2009; Morgante et al., 2014). *E. coli* transformations performed with these libraries have made it possible to look for genes conferring higher arsenic tolerance. The first study on these lines has led to the identification of *arsN*, which encodes a protein similar to acetyltransferases, homologues of which are closely associated with arsenic resistance genes in the genomes of many other bacteria (Chauhan et al., 2009). In the second study of this kind, genes involved in transport processes, stress responses, DNA repair and amino acid and phospholipid synthesis were identified (Morgante et al., 2014), which confirms previous assumptions about the role of these processes in various organisms' responses to arsenic. A similar approach has also been recently used to identify novel eukaryotic genes conferring resistance on heavy metals in yeast (Lehembre et al., 2013). Lastly, the introduction of molecular methods has also yielded information about less abundant micro-organisms, which are sometimes referred to as cryptic. For example, the often underestimated importance of archaea and their role in ecosystems have been highlighted in the case of the Richmond mine AMD, where these organisms are thought to be involved in carbon cycling processes (Tyson et al., 2004; Baker et al., 2006; Justice et al., 2012).

Rather than focusing on taxonomic markers or single biological functions, the metagenomic approach also helps to unravel the metabolic potential of whole communities exposed to arsenic and to describe the complete genome of non-culturable organisms. However, eukaryotic organisms have often been omitted from these studies, and have only been listed in 18S rDNA inventories despite their high level of diversity and the substantial ecological role they play (Aguilera et al., 2007; Baker et al., 2009). Four metagenome studies have been conducted so far on the communities inhabiting arsenic contaminated ecosystems (Tyson et al., 2004; Bertin et al., 2011; Plewniak et al., 2013; Mendez-Garcia et al., 2014), but we will focus here on the Richmond mine and Carnoulès AMD because these sites present the highest arsenic concentrations of all those studied so far. The first study, which focused on the biofilm formed at the Richmond mine, resulted in the assembly of the near-complete genomes of two predominant organisms belonging to *Leptospirillum* group II and *Ferroplasma* type II, as well as the partial genomes of three other organisms, including a representative of *Leptospirillum* group III (Tyson et al., 2004). Based on their genome assembly, the

predominant organisms are thought to be responsible for iron oxidation and therefore to be involved in the generation of AMD. *Leptospirillum* group II is also liable to be responible for primary colonization because of its motility and its ability to synthesize biofilm but also for carbon fixation and for the supply of its heterotrophic partners with organic matter (Tyson et al., 2004). In addition, the metagenome study has suggested the presence of several conspecific strains belonging to *Leptospirillum* group II and *Ferroplasma* type II, which differ in terms of their nucleotide sequences by less than 0.1 % and 2.2 % (Tyson et al., 2004). Data analyses have shown that the population of archaea consists of multiple strains bearing mosaic genomes resulting from extensive homologous recombinations (Tyson et al., 2004). An analysis performed on data obtained on the same site, in which some Ferroplasma type I sequences were compared, i.e. the genome of *F. acidarmanus* fer1 genome and environmental sequences corresponding to a population of the same species, showed the presence of phage, plasmid and transposase insertions and deletions. These processes may serve to maintain diversity and confer greater stability on ecosystems by increasing the microbial populations' resilience to environmental stresses (Allen et al., 2007). In addition, various ecotypes will presumably show some functional diversity and therefore play different roles in a given ecosystem, which would explain their coexistence (Simmons et al., 2008; Denef et al., 2010a; Wilmes et al., 2010). The second study on these lines focused on the microbial community inhabiting the Carnoulès AMD, and led to the description and reconstruction of the genomes of seven predominant bacteria (named CARN1 to CARN7) of different taxonomic origins, five of which are non-culturable bacteria (Bertin et al., 2011). Based on the identification of *aio* and *rus* genes in species of *Thiomonas* (CARN2) and *Acidithiobacillus* (CARN5), respectively, these seven bacteria are thought to be involved in an arsenic attenuation process involving the oxidation and co-precipitation of arsenic with iron and sulfur. The authors of this study also characterized two genetic variants of a novel phylum, *Candidatus* Fodinabacter communificans (CARN1 and CARN4), which possibly assist the other community members by recycling organic matter (Bertin et al., 2011). It is worth noting that both of the latter metagenomic studies were facilitated by the relatively low complexity of the respective communities (Denef et al., 2010b; Bertin et al., 2011), whereas *de novo* genome assembly was not possible in two other metagenomic studies because of the greater microbial diversity (Plewniak et al., 2013; Mendez-Garcia et al., 2014). The results of these two studies nevertheless suggested that two important biological processes, arsenic mobilization and metabolic stratification, might be at work in the respective communities.

In order to investigate organisms more closely, however, it is often necessary to isolate them and maintain them in culture. By reconstructing the genome of non-culturable organisms, it is possible to identify the existing and missing pathways with a view to developing cell culture or

organism selection strategie, defining the inverse metagenomic approach. In this context, metagenome analyses performed on the biofilm of the Richmond mine AMD have shown that *Leptospirillum ferrodiazotrophum* is the sole organism capable of nitrogen fixation which has enabled its selection and isolatation in culture (Tyson et al., 2004, 2005). However, even with methods of this kind, some micro-organisms might still not lend themselves to culture procedures. Although metagenomics have proved to be a useful tool for investigating novel species and functions, care should be taken when drawing conclusions, as with classical PCR-based approaches, because of the possible bias due to the sampling, DNA extraction and sequencing methods used. In addition, it can be difficult to annotate the genes detected and attribute them to specific organisms, which sometimes makes it even harder to determine the underlying ecological or evolutionary processes. Since metagenomic studies yield information about genetic potential but not about the expression and functioning of cellular processes, they have to be combined with other methods. In recent studies on arsenic impacted communities, for instance, metagenomic approaches have been usefully combined with functional tools in order to identify the active micro-organisms in the corresponding communities, and the results obtained have made it possible to put forward some hypotheses as to how whole communities may function (Ram et al., 2005; Denef et al., 2010b; Bertin et al., 2011; Mendez-Garcia et al., 2014).

ii. *Functional studies on arsenic exposed microbial communities*

Metagenomic data on the Richmond mine and Carnoulès AMDs have provided useful bases for downstream metaproteomic analyses (Ram et al., 2005; Bertin et al., 2011). Thanks to the genomic dataset obtained, databases of predicted proteins were newly created and have helped to identify some of the proteins present in the samples and to assign them to specific organisms. Metaproteomics can therefore be a worthwhile means of addressing the actual expression of cellular functions identified on the basis of metagenomic findings, and thus make it possible to draw a more accurate picture of the role of each organism present in communities and describe the metabolic networks and evolutionary processes possibly ensuring the survival of these communities inhabiting arsenic contaminated ecosystems.

Despite the availability of the metagenomes, most of the proteins detected in the two metaproteomic studies cited above could not be assigned with specific functions because of their low homology with the proteins having determined functions which are listed in databases. Since some proteins are conserved in several organisms, they may contribute to AMD adaptation processes (Ram

et al., 2005; Bertin et al., 2011). The Richmond mine metaproteome analysis has shown that proteins involved in cell envelope biogenesis, oxidative stress responses, translation and protein pool maintenance, amino and nucleic acid metabolism and energy production were among the most abundant proteins detected in the community in question (Ram et al., 2005), whereas the Carnoulès metaproteome showed the presence of proteins involved in colonisation processes, transport and regulatory systems including phosphate-specific transporters, oxidative stress responses and DNA repair (Bertin et al., 2011). Although these cellular mechanisms may also be triggered by other stresses, they are all involved in the responses to arsenic described above. In addition, the findings made using the metaproteomic approach have confirmed the validity of several hypotheses suggested by the results of metagenome analyses. First, the protein richness described in both studies was consistent with the number of assigned sequencing reads, possibly reflecting the abundance of the respective organisms (Ram et al., 2005; Bertin et al., 2011). Secondly, by establishing the presence of cytochromes involved in iron oxidation processes and proteins involved in nitrogen fixation and colonisation processes, the findings made on the Richmond mine metaproteome confirm that some micro-organisms may contribute to the process of AMD generation, and that some of them are responsible for nitrogen fixation or primary colonization (Ram et al., 2005). Thirdly, the identification of significant levels of arsenite methyltransferases, rustycianins and arsenite oxidases in the Carnoulès metaproteome is consistent with the presence of methylated arsenic species and the involvement of organisms in the natural arsenic attenuation process (Bertin et al., 2011).

Metagenomic and metaproteomic studies have also brought to light some other processes possibly contributing to the survival of organisms in these ecosystems. On the one hand, the findings obtained on the Richmond mine metaproteome support the idea that genetic transfers and microbial diversity enhance the resilience of species by promoting their evolution and the acquisition of new functions. Several detected proteins with unknown functions are encoded by genes located in genomic regions associated with mobile elements (Ram et al., 2005; Denef et al., 2010b), and several transposases, integrases and recombinases have also been identified. In addition, some regions of single proteins were found to match various genomes, providing further evidence for the occurrence of recombination events (Ram et al., 2005; Denef et al., 2010b). The finding that some proteins which are present in several members of *Leptospirillum* group II showed different levels of expression depending on the genotype involved supports the idea that each organism may play a specific ecological role in the ecosystem (Denef et al., 2010a). Results obtained using an alternative approach in which the genomes of some *Thiomonas* strains were compared using a microarray-based comparative genomic hybridization method showed the specific presence of some GEI in AMD-

isolated strains (Arsène-Ploetze et al., 2010). Evidence is also available that some of these genomic islands are acquired via HGT, the components of which enable the host species to cope with AMD conditions, including the presence of arsenic, which further supports the idea that genetic transfers contribute to species' survival and adaptation. On the other hand, the specialization which has been suggested to occur among the members of *Leptospirillum* group II is thought to play an important role in the Carnoulès microbial community. The data obtained in the latter study suggest that a a syntrophic partnership may be formed between the main members, which enables the whole community to thrive in this toxic environment. Each member of the community was found to express specific functions. *Thiomonas* sp. (CARN2) may be involved in inorganic nutrient metabolism and arsenic detoxification functions and along with *Acidithiobacillus* sp. (CARN5), in carbon fixation and urea degradation pathways. The CARN6 bin may play a role in cellulose metabolism, and the ecotypes *Candidatus* Fodinabacter communificans (CARN1 and CARN4) may be involved in nucleic and amino acid degradation and cofactor synthesis (Bertin et al., 2011). Some organisms may therefore be primary producers, while others may be involved in the recycling of organic matter: they therefore all participate in the overall functioning of the ecosystem and in the arsenic attenuation process (Bertin 2011). In addition, cells of *E. mutabilis*, which form a mat contacting the latter bacterial community, may participate in the functioning of the community by acting as a primary producer (Bertin et al., 2011; Halter et al., 2012b). Some interesting additional clues supporting the importance of these trophic relationships in AMDs were presented in a recent metabolomic study performed on the AMD from the Richmond mine. Three metabolites possibly involved in the adaptation of mirco-organisms to their environment, namely phosphatidylethanolamine lipids, taurine and hydroxyectoine, were identified in that study. Taurine is known to play several physiological roles, which include protecting proteins, nucleic acids and membranes from the free radicals that may be generated upon exposure to arsenic, or alternatively, it may be used as a source of carbon, nitrogen and sulfur (Mosier et al., 2013). Based on these findings and metagenomic data, the authors of the latter study suggested that taurine may be produced by the fungus *Acidomyces richmondensis* and consumed by the *Sulfobacillus* members of the community. Lastly, in view of the differences observed between closely neighbouring microbial communities, the authors of a metaproteogenomic study performed on the AMD from Los Rueldos also suggested that even the development of microbial communities in environments of this kind may depend on trophic relationships (Mendez-Garcia et al., 2014).

Studies using the metagenomic approaches described above have therefore shed interesting light on some of the genetic and metabolic interactions occurring between members of microbial communities, which help them to thrive in arsenic contaminated environments. These two processes

have therefore been added to the overall repertoire of micro-organisms' responses to arsenic (Fig. 3). Although studies based on the use of genomic techniques may have only revealed the role of a set of metabollicaly dominant organisms, these studies have made a considerable step towards understanding how microbial communities inhabiting AMD environments function.

Descriptive and comparative genomics have certainly helped to elucidate micro-organisms' genetic potential regarding arsenic. The use of these tools has contributed to the present picture of the widespread and more specific patterns of occurrence of arsenic-related genes in micro-organisms, and has helped to determine the genetic bases and the regulatory mechanisms involved in both resistance to arsenic and its metabolism. In addition, findings in the field of functional genomics have shown the great diversity of the cellular processes involved in micro-organisms' responses to arsenic, limiting its entry in the cells and preventing the occurrence of cellular damage as well as maintaining the functionality of the cells' components. In addition, these tools have yielded new insights into the microbial communities, including non-culturable organisms, which thrive in arsenic contaminated environments, and shed new light on the genetic and trophic interactions involved in species' adaptation to extreme conditions and their ability to survive. The present account of the various findings made in the field of arsenomics clearly shows that the survival of micro-organisms in arsenic contaminated environments depends on their adaptation, resistance, metabolism and interaction abilities.

Approaches in which genomics have been combined with classical methods as well as physico-chemical data have yielded an interesting but still partial picture of how the organisms under investigation function. Ecosystems are composed of both biotic and abiotic components, which are involved in a complex network of interactions. It would certainly now be worth conducting further investigations on how these communities function and how their members interact, especially by studying the role of micro-eukaryotic organisms, archaea and cryptic organisms in various arsenic contaminated niches. Elucidating the functional specificities of the large proportion of the molecules present in environmental samples which have not yet been characterized will probably require the use of other methods. The currently emerging single cell genomics approach seems to be a promising means of overcoming some technical bottlenecks. Thanks to the recent development of molecular and microfluidic tools, it is now possible to address biological issues at the single cell level (Blainey, 2013). In the field of arsenomics, it may be possible to combine flow cytometry approaches, which enables cell sorting, selection and isolation, with whole genome amplification methods in order to sequence the genomes of single cells obtained from contaminated environments, without having to perform culture procedures. The use of mass spectrometry-based proteomics and metabolomics on sorted microbial subpopulations is also likely to yield some interesting new insights into micro-organisms' responses to arsenic (Jahn et al., 2013).

Aguilera, A., Zettler, E., Gómez, F., Amaral-Zettler, L., Rodríguez, N., and Amils, R. (2007) Distribution and seasonal variability in the benthic eukaryotic community of Río Tinto (SW, Spain), an acidic, high metal extreme environment. *Syst. Appl. Microbiol.* **30**: 531–546.

Allen, E.E., Tyson, G.W., Whitaker, R.J., Detter, J.C., Richardson, P.M., and Banfield, J.F. (2007) Genome dynamics in a natural archaeal population. *Proc. Natl. Acad. Sci. U. S. A.* **104**: 1883–1888.

Amann, R.I., Ludwig, W., and Schleifer, K.H. (1995) Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. *Microbiol. Rev.* **59**: 143–169.

Andres, J., Arsène-Ploetze, F., Barbe, V., Brochier-Armanet, C., Cleiss-Arnold, J., Coppée, J.-Y., et al. (2013) Life in an arsenic-containing gold mine: genome and physiology of the autotrophic arsenite-oxidizing bacterium *Rhizobium* sp. NT-26. *Genome Biol. Evol.* **5**: 934–953.

Arsène-Ploetze, F., Koechler, S., Marchal, M., Coppée, J.-Y., Chandler, M., Bonnefoy, V., et al. (2010) Structure, function, and evolution of the *Thiomonas* spp. genome. *PLoS Genet.* **6**: e1000859.

Baker-Austin, C., Dopson, M., Wexler, M., Sawers, R.G., Stemmler, A., Rosen, B.P., and Bond, P.L. (2007) Extreme arsenic resistance by the acidophilic archaeon "*Ferroplasma acidarmanus*" Fer1. *Extrem. Life Extreme Cond.* **11**: 425–434.

Baker, B.J. and Banfield, J.F. (2003) Microbial communities in acid mine drainage. *FEMS Microbiol. Ecol.* **44**: 139–152.

Baker, B.J., Tyson, G.W., Goosherst, L., and Banfield, J.F. (2009) Insights into the diversity of eukaryotes in acid mine drainage biofilm communities. *Appl. Environ. Microbiol.* **75**: 2192–2199.

Baker, B.J., Tyson, G.W., Webb, R.I., Flanagan, J., Hugenholtz, P., Allen, E.E., and Banfield, J.F. (2006) Lineages of acidophilic archaea revealed by community genomic analysis. *Science* **314**: 1933–1935.

Belfiore, C., Ordoñez, O.F., and Farías, M.E. (2013) Proteomic approach of adaptive response to arsenic stress in *Exiguobacterium* sp. S17, an extremophile strain isolated from a high-altitude Andean Lake stromatolite. *Extrem. Life Extreme Cond.* **17**: 421–431.

Bernstam, L. and Nriagu, J. (2000) Molecular aspects of arsenic stress. *J. Toxicol. Environ. Health B Crit. Rev.* **3**: 293–322.

Bertin, P.N., Heinrich-Salmeron, A., Pelletier, E., Goulhen-Chollet, F., Arsène-Ploetze, F., Gallien, S., et al. (2011) Metabolic diversity among main micro-organisms inside an arsenic-rich ecosystem revealed by meta- and proteo-genomics. *ISME J.* **5**: 1735–1747.

Bertin, P.N., Médigue, C., and Normand, P. (2008) Advances in environmental genomics: towards an integrated view of micro-organisms and ecosystems. *Microbiol. Read. Engl.* **154**: 347–359.

Bhattacharjee, H. and Rosen, B.P. (2007) Arsenic metabolism in prokaryotic and eukaryotic microbes. In, Nies,D.H. and Silver,S. (eds), *Molecular Microbiology of Heavy Metals*, Microbiology Monographs. Springer Berlin Heidelberg, pp. 371–406.

Blainey, P.C. (2013) The future is now: single-cell genomics of bacteria and archaea. *FEMS Microbiol. Rev.* **37**: 407–427.

Bobrowicz, P., Wysocki, R., Owsianik, G., Goffeau, A., and Ułaszewski, S. (1997) Isolation of three contiguous genes, *ACR1*, *ACR2* and *ACR3*, involved in resistance to arsenic compounds in the yeast *Saccharomyces cerevisiae*. *Yeast Chichester Engl.* **13**: 819–828.

Branco, R., Chung, A.-P., and Morais, P.V. (2008) Sequencing and expression of two arsenic resistance operons with different functions in the highly arsenic-resistant strain *Ochrobactrum tritici* SCII24T. *BMC Microbiol.* **8**: 95.

Bryan, C.G., Marchal, M., Battaglia-Brunet, F., Kugler, V., Lemaitre-Guillier, C., Lièvremont, D., et al. (2009) Carbon and arsenic metabolism in *Thiomonas* strains: differences revealed diverse adaptation processes. *BMC Microbiol.* **9**: 127.

Burguener, G.F., Maldonado, M.J., Revale, S., Fernández Do Porto, D., Rascován, N., Vázquez, M., et al. (2014) Draft genome sequence of the polyextremophilic *Halorubrum* sp. Strain AJ67, Isolated from hyperarsenic lakes in the argentinian puna. *Genome Announc.* **2**:

Bustaffa, E., Stoccoro, A., Bianchi, F., and Migliore, L. (2014) Genotoxic and epigenetic mechanisms in arsenic carcinogenicity. *Arch. Toxicol.* **88**: 1043–1067.

Butcher, B.G., Deane, S.M., and Rawlings, D.E. (2000) The chromosomal arsenic resistance genes of *Thiobacillus ferrooxidans* have an unusual arrangement and confer increased arsenic and antimony resistance to *Escherichia coli*. *Appl. Environ. Microbiol.* **66**: 1826–1833.

Cai, L., Liu, G., Rensing, C., and Wang, G. (2009a) Genes involved in arsenic transformation and resistance associated with different levels of arsenic-contaminated soils. *BMC Microbiol.* **9**: 4.

Cai, L., Rensing, C., Li, X., and Wang, G. (2009b) Novel gene clusters involved in arsenite oxidation and resistance in two arsenite oxidizers: *Achromobacter* sp. SY8 and *Pseudomonas* sp. TS44. *Appl. Microbiol. Biotechnol.* **83**: 715–725.

Cai, L., Yu, K., Yang, Y., Chen, B.-W., Li, X.-D., and Zhang, T. (2013) Metagenomic exploration reveals high levels of microbial arsenic metabolism genes in activated sludge and coastal sediments. *Appl. Microbiol. Biotechnol.* **97**: 9579–9588.

Carapito, C., Muller, D., Turlin, E., Koechler, S., Danchin, A., Van Dorsselaer, A., et al. (2006) Identification of genes and proteins involved in the pleiotropic response to arsenic stress in *Caenibacter arsenoxydans*, a metalloresistant *Betaproteobacterium* with an unsequenced genome. *Biochimie* **88**: 595–606.

Carlin, A., Shi, W., Dey, S., and Rosen, B.P. (1995) The *ars* operon of *Escherichia coli* confers arsenical and antimonial resistance. *J. Bacteriol.* **177**: 981–986.

Chattopadhyay, M.K., Tabor, C.W., and Tabor, H. (2006) Polyamine deficiency leads to accumulation of reactive oxygen species in a *spe2*Δ mutant of *Saccharomyces cerevisiae*. *Yeast Chichester Engl.* **23**: 751–761.

Chauhan, N.S., Ranjan, R., Purohit, H.J., Kalia, V.C., and Sharma, R. (2009) Identification of genes conferring arsenic resistance to *Escherichia coli* from an effluent treatment plant sludge metagenomic library. *FEMS Microbiol. Ecol.* **67**: 130–139.

Chow, E.W.L., Morrow, C.A., Djordjevic, J.T., Wood, I.A., and Fraser, J.A. (2012) Microevolution of *Cryptococcus neoformans* driven by massive tandem gene amplification. *Mol. Biol. Evol.* **29**: 1987–2000.

Cleiss-Arnold, J., Koechler, S., Proux, C., Fardeau, M.-L., Dillies, M.-A., Coppee, J.-Y., et al. (2010) Temporal transcriptomic response during arsenic stress in *Herminiimonas arsenicoxydans*. *BMC Genomics* **11**: 709.

Coppée, J.-Y. (2008) Do DNA microarrays have their future behind them? *Microbes Infect. Inst. Pasteur* **10**: 1067–1071.

Cozen, A.E., Weirauch, M.T., Pollard, K.S., Bernick, D.L., Stuart, J.M., and Lowe, T.M. (2009) Transcriptional map of respiratory versatility in the hyperthermophilic crenarchaeon *Pyrobaculum aerophilum*. *J. Bacteriol.* **191**: 782–794.

Cuebas, M., Villafane, A., McBride, M., Yee, N., and Bini, E. (2011) Arsenate reduction and expression of multiple chromosomal *ars* operons in *Geobacillus kaustophilus* A1. *Microbiol. Read. Engl.* **157**: 2004–2011.

Cullen, W.R. and Reimer, K.J. (1989) Arsenic speciation in the environment. *Chem. Rev.* **89**: 713–764.

Daware, V., Kesavan, S., Patil, R., Natu, A., Kumar, A., Kulkarni, M., and Gade, W. (2012) Effects of arsenite stress on growth and proteome of *Klebsiella pneumoniae*. *J. Biotechnol.* **158**: 8–16.

Denef, V.J., Kalnejais, L.H., Mueller, R.S., Wilmes, P., Baker, B.J., Thomas, B.C., et al. (2010a) Proteogenomic basis for ecological divergence of closely related bacteria in natural acidophilic microbial communities. *Proc. Natl. Acad. Sci. U. S. A.* **107**: 2383–2390.

Denef, V.J., Mueller, R.S., and Banfield, J.F. (2010b) AMD biofilms: using model communities to study microbial evolution and ecological complexity in nature. *ISME J.* **4**: 599–610.

Dilda, P.J., Perrone, G.G., Philp, A., Lock, R.B., Dawes, I.W., and Hogg, P.J. (2008) Insight into the selectivity of arsenic trioxide for acute promyelocytic leukemia cells by characterizing *Saccharomyces cerevisiae* deletion strains that are sensitive or resistant to the metalloid. *Int. J. Biochem. Cell Biol.* **40**: 1016–1029.

Ding, W., Liu, W., Cooper, K.L., Qin, X.-J., de Souza Bergo, P.L., Hudson, L.G., and Liu, K.J. (2009) Inhibition of poly(ADP-ribose) polymerase-1 by arsenite interferes with repair of oxidative DNA damage. *J. Biol. Chem.* **284**: 6809–6817.

Dombrowski, P.M., Long, W., Farley, K.J., Mahony, J.D., Capitani, J.F., and Di Toro, D.M. (2005) Thermodynamic analysis of arsenic methylation. *Environ. Sci. Technol.* **39**: 2169–2176.

Dopp, E., Kligerman, A.D., and Diaz-Bone, R.A. (2010) Organoarsenicals. Uptake, metabolism, and toxicity. *Met. Ions Life Sci.* **7**: 231–265.

Drewniak, L., Dziewit, L., Ciezkowska, M., Gawor, J., Gromadka, R., and Sklodowska, A. (2013) Structural and functional genomics of plasmid pSinA of *Sinorhizobium* sp. M14 encoding genes for the arsenite oxidation and arsenic resistance. *J. Biotechnol.* **164**: 479–488.

Duval, S., Ducluzeau, A.-L., Nitschke, W., and Schoepp-Cothenet, B. (2008) Enzyme phylogenies as markers for the oxidation state of the environment: the case of respiratory arsenate reductase and related enzymes. *BMC Evol. Biol.* **8**: 206.

Elias, M., Wellner, A., Goldin-Azulay, K., Chabriere, E., Vorholt, J.A., Erb, T.J., and Tawfik, D.S. (2012) The molecular basis of phosphate discrimination in arsenate-rich environments. *Nature* **491**: 134–137.

Ettema, T.J.G., de Vos, W.M., and van der Oost, J. (2005) Discovering novel biology by *in silico* archaeology. *Nat. Rev. Microbiol.* **3**: 859–869.

Fahey, R.C. (2013) Glutathione analogs in prokaryotes. *Biochim. Biophys. Acta* **1830**: 3182–3198.

Flora, S.J.S. (2011) Arsenic-induced oxidative stress and its reversibility. *Free Radic. Biol. Med.* **51**: 257–281.

Gadd, G.M. (2010) Metals, minerals and microbes: geomicrobiology and bioremediation. *Microbiol. Read. Engl.* **156**: 609–643.

Ghosh, M., Shen, J., and Rosen, B.P. (1999) Pathways of As(III) detoxification in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U. S. A.* **96**: 5001–5006.

Grant, C.M. (2001) Role of the glutathione/glutaredoxin and thioredoxin systems in yeast growth and response to stress conditions. *Mol. Microbiol.* **39**: 533–541.

Halter, D., Andres, J., Plewniak, F., Poulain, J., Da Silva, C., Arsène-Ploetze, F., and Bertin, P.N. (2014) Arsenic hypertolerance in the protist *Euglena mutabilis* is mediated by specific transporters and functional integrity maintenance mechanisms. *Environ. Microbiol.*

Halter, D., Casiot, C., Heipieper, H.J., Plewniak, F., Marchal, M., Simon, S., et al. (2012a) Surface properties and intracellular speciation revealed an original adaptive mechanism to arsenic in the acid mine drainage bio-indicator *Euglena mutabilis*. *Appl. Microbiol. Biotechnol.* **93**: 1735–1744.

Halter, D., Goulhen-Chollet, F., Gallien, S., Casiot, C., Hamelin, J., Gilard, F., et al. (2012b) *In situ* proteo-metabolomics reveals metabolite secretion by the acid mine drainage bio-indicator, *Euglena mutabilis*. *ISME J.* **6**: 1391–1402.

Hartl, F.U. (1996) Molecular chaperones in cellular protein folding. *Nature* **381**: 571–579.

Haugen, A.C., Kelley, R., Collins, J.B., Tucker, C.J., Deng, C., Afshari, C.A., et al. (2004) Integrating phenotypic and expression profiles to map arsenic-response networks. *Genome Biol.* **5**: R95.

Heinrich-Salmeron, A., Cordi, A., Brochier-Armanet, C., Halter, D., Pagnout, C., Abbaszadeh-fard, E., et al. (2011) Unsuspected diversity of arsenite-oxidizing bacteria as revealed by widespread distribution of the *aoxB* gene in prokaryotes. *Appl. Environ. Microbiol.* **77**: 4685–4692.

Houry, W.A. (2001) Chaperone-assisted protein folding in the cell cytoplasm. *Curr. Protein Pept. Sci.* **2**: 227–244.

Huber, R., Sacher, M., Vollmann, A., Huber, H., and Rose, D. (2000) Respiration of arsenate and selenate by hyperthermophilic archaea. *Syst. Appl. Microbiol.* **23**: 305–314.

Hughes, M.F. (2002) Arsenic toxicity and potential mechanisms of action. *Toxicol. Lett.* **133**: 1–16.

Jackson, C.R., Langner, H.W., Donahoe-Christiansen, J., Inskeep, W.P., and McDermott, T.R. (2001) Molecular analysis of microbial community structure in an arsenite-oxidizing acidic thermal spring. *Environ. Microbiol.* **3**: 532–542.

Jahn, M., Seifert, J., von Bergen, M., Schmid, A., Bühler, B., and Müller, S. (2013) Subpopulation-proteomics in prokaryotic populations. *Curr. Opin. Biotechnol.* **24**: 79–87.

Janssen, P.J., Van Houdt, R., Moors, H., Monsieurs, P., Morin, N., Michaux, A., et al. (2010) The complete genome sequence of *Cupriavidus metallidurans* strain CH34, a master survivalist in harsh and anthropogenic environments. *PloS One* **5**: e10433.

Jin, Y.H., Dunlap, P.E., McBride, S.J., Al-Refai, H., Bushel, P.R., and Freedman, J.H. (2008) Global transcriptome and deletome profiles of yeast exposed to transition metals. *PLoS Genet.* **4**: e1000053.

Juhas, M., van der Meer, J.R., Gaillard, M., Harding, R.M., Hood, D.W., and Crook, D.W. (2009) Genomic islands: tools of bacterial horizontal gene transfer and evolution. *FEMS Microbiol. Rev.* **33**: 376–393.

Justice, N.B., Pan, C., Mueller, R., Spaulding, S.E., Shah, V., Sun, C.L., et al. (2012) Heterotrophic archaea contribute to carbon cycling in low-pH, suboxic biofilm communities. *Appl. Environ. Microbiol.* **78**: 8321–8330.

Kang, Y.-S., Bothner, B., Rensing, C., and McDermott, T.R. (2012a) Involvement of RpoN in regulating bacterial arsenite oxidation. *Appl. Environ. Microbiol.* **78**: 5638–5645.

Kang, Y.-S., Heinemann, J., Bothner, B., Rensing, C., and McDermott, T.R. (2012b) Integrated co-regulation of bacterial arsenic and phosphorus metabolisms. *Environ. Microbiol.* **14**: 3097–3109.

Kang, Y.-S., Shi, Z., Bothner, B., Wang, G., and McDermott, T.R. (2014) Involvement of the Acr3 and DctA anti-porters in arsenite oxidation in *Agrobacterium tumefaciens* 5A. *Environ. Microbiol.*

Kapaj, S., Peterson, H., Liber, K., and Bhattacharya, P. (2006) Human health effects from chronic arsenic poisoning - a review. *J. Environ. Sci. Health Part A Tox. Hazard. Subst. Environ. Eng.* **41**: 2399–2428.

Kashyap, D.R., Botero, L.M., Franck, W.L., Hassett, D.J., and McDermott, T.R. (2006) Complex regulation of arsenite oxidation in *Agrobacterium tumefaciens*. *J. Bacteriol.* **188**: 1081–1088.

Keeling, P.J. and Palmer, J.D. (2008) Horizontal gene transfer in eukaryotic evolution. *Nat. Rev. Genet.* **9**: 605–618.

Koechler, S., Cleiss-Arnold, J., Proux, C., Sismeiro, O., Dillies, M.-A., Goulhen-Chollet, F., et al. (2010) Multiple controls affect arsenite oxidase gene expression in *Herminiimonas arsenicoxydans*. *BMC Microbiol.* **10**: 53.

Koechler, S., Plewniak, F., Barbe, V., Battaglia-Brunet, F., Jost, B., Joulian, C., et al. (2013) Genome sequence of *Halomonas* sp. Strain A3H3, isolated from arsenic-rich marine sediments. *Genome Announc.* **1**:

van Kranenburg, R., Golic, N., Bongers, R., Leer, R.J., de Vos, W.M., Siezen, R.J., and Kleerebezem, M. (2005) Functional analysis of three plasmids from *Lactobacillus plantarum*. *Appl. Environ. Microbiol.* **71**: 1223–1230.

Kruger, M.C., Bertin, P.N., Heipieper, H.J., and Arsène-Ploetze, F. (2013) Bacterial metabolism of environmental arsenic--mechanisms and biotechnological applications. *Appl. Microbiol. Biotechnol.* **97**: 3827–3841.

Kulp, T.R., Hoeft, S.E., Asao, M., Madigan, M.T., Hollibaugh, J.T., Fisher, J.C., et al. (2008) Arsenic(III) fuels anoxygenic photosynthesis in hot spring biofilms from Mono Lake, California. *Science* **321**: 967–970.

Langille, M.G.I., Hsiao, W.W.L., and Brinkman, F.S.L. (2010) Detecting genomic islands using bioinformatics approaches. *Nat. Rev. Microbiol.* **8**: 373–382.

Lebrun, E., Brugna, M., Baymann, F., Muller, D., Lièvremont, D., Lett, M.-C., and Nitschke, W. (2003) Arsenite oxidase, an ancient bioenergetic enzyme. *Mol. Biol. Evol.* **20**: 686–693.

Lehembre, F., Doillon, D., David, E., Perrotto, S., Baude, J., Foulon, J., et al. (2013) Soil metatranscriptomics for mining eukaryotic heavy metal resistance genes. *Environ. Microbiol.* **15**: 2829–2840.

Lehr, C.R., Kashyap, D.R., and McDermott, T.R. (2007) New insights into microbial oxidation of antimony and arsenic. *Appl. Environ. Microbiol.* **73**: 2386–2389.

Lett, M.-C., Muller, D., Lievremont, D., Silver, S., and Santini, J. (2012) Unified nomenclature for genes involved in prokaryotic aerobic arsenite oxidation. *J. Bacteriol.* **194**: 207–208.

Letunic, I. and Bork, P. (2011) Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Res.* **39**: W475–478.

Levy-Sakin, M. and Ebenstein, Y. (2013) Beyond sequencing: optical mapping of DNA in the age of nanotechnology and nanoscopy. *Curr. Opin. Biotechnol.* **24**: 690–698.

Li, B., Lin, J., Mi, S., and Lin, J. (2010) Arsenic resistance operon structure in *Leptospirillum ferriphilum* and proteomic response to arsenic stress. *Bioresour. Technol.* **101**: 9811–9814.

Lieutaud, A., van Lis, R., Duval, S., Capowiez, L., Muller, D., Lebrun, R., et al. (2010) Arsenite oxidase from *Ralstonia* sp. 22: characterization of the enzyme and its interaction with soluble cytochromes. *J. Biol. Chem.* **285**: 20433–20441.

Lièvremont, D., Bertin, P.N., and Lett, M.-C. (2009) Arsenic in contaminated waters: biogeochemical cycle, microbial metabolism and biotreatment processes. *Biochimie* **91**: 1229–1237.

Li, H., Li, M., Huang, Y., Rensing, C., and Wang, G. (2013) *In silico* analysis of bacterial arsenic islands reveals remarkable synteny and functional relatedness between arsenate and phosphate. *Front. Microbiol.* **4**: 347.

Lin, Y.-C., Hsu, J.-Y., Shu, J.-H., Chi, Y., Chiang, S.-C., and Lee, S.T. (2008) Two distinct arsenite-resistant variants of *Leishmania amazonensis* take different routes to achieve resistance as revealed by comparative transcriptomics. *Mol. Biochem. Parasitol.* **162**: 16–31.

Lin, Y., Fan, H., Hao, X., Johnstone, L., Hu, Y., Wei, G., et al. (2012) Draft genome sequence of *Halomonas* sp. strain HAL1, a moderately halophilic arsenite-oxidizing bacterium isolated from gold-mine soil. *J. Bacteriol.* **194**: 199–200.

van Lis, R., Nitschke, W., Duval, S., and Schoepp-Cothenet, B. (2013) Arsenics as bioenergetic substrates. *Biochim. Biophys. Acta* **1827**: 176–188.

Liu, G., Liu, M., Kim, E.-H., Maaty, W.S., Bothner, B., Lei, B., et al. (2012) A periplasmic arsenite-binding protein involved in regulating arsenite oxidation. *Environ. Microbiol.* **14**: 1624–1634.

Li, X., Zhang, L., and Wang, G. (2014) Genomic evidence reveals the extreme diversity and wide distribution of the arsenic-related genes in *Burkholderiales*. *PloS One* **9**: e92236.

Lomax, C., Liu, W.-J., Wu, L., Xue, K., Xiong, J., Zhou, J., et al. (2012) Methylated arsenic species in plants originate from soil micro-organisms. *New Phytol.* **193**: 665–672.

Maciaszczyk, E., Wysocki, R., Golik, P., Lazowska, J., and Ulaszewski, S. (2004) Arsenical resistance genes in *Saccharomyces douglasii* and other yeast species undergo rapid evolution involving genomic rearrangements and duplications. *FEMS Yeast Res.* **4**: 821–832.

Mackey-Lawrence, N.M., Potter, D.E., Cerca, N., and Jefferson, K.K. (2009) *Staphylococcus aureus* immunodominant surface antigen B is a cell-surface associated nucleic acid binding protein. *BMC Microbiol.* **9**: 61.

Maki, T., Hasegawa, H., Watarai, H., and Ueda, K. (2004) Classification for dimethylarsenate-decomposing bacteria using a restrict fragment length polymorphism analysis of 16S rRNA genes. *Anal. Sci. Int. J. Jpn. Soc. Anal. Chem.* **20**: 61–68.

Malasarn, D., Saltikov, C.W., Campbell, K.M., Santini, J.M., Hering, J.G., and Newman, D.K. (2004) *arrA* is a reliable marker for As(V) respiration. *Science* **306**: 455.

Mallik, S., Virdi, J.S., and Johri, A.K. (2012) Proteomic analysis of arsenite - mediated multiple antibiotic resistance in *Yersinia enterocolitica* biovar 1A. *J. Basic Microbiol.* **52**: 306–313.

McGettigan, P.A. (2013) Transcriptomics in the RNA-seq era. *Curr. Opin. Chem. Biol.* **17**: 4–11.

Mendez-Garcia, C., Mesa, V., Sprenger, R.R., Richter, M., Diez, M.S., Solano, J., et al. (2014) Microbial stratification in low pH oxic and suboxic macroscopic growths along an acid mine drainage. *ISME J.* **8**: 1259–1274.

Mergeay, M., Monchy, S., Vallaeys, T., Auquier, V., Benotmane, A., Bertin, P., et al. (2003) *Ralstonia metallidurans*, a bacterium specifically adapted to toxic metals: towards a catalogue of metal-responsive genes. *FEMS Microbiol. Rev.* **27**: 385–410.

Meyer, Y., Buchanan, B.B., Vignols, F., and Reichheld, J.-P. (2009) Thioredoxins and glutaredoxins: unifying elements in redox biology. *Annu. Rev. Genet.* **43**: 335–367.

Migdal, I., Ilina, Y., Tamás, M.J., and Wysocki, R. (2008) Mitogen-activated protein kinase Hog1 mediates adaptation to G1 checkpoint arrest during arsenite and hyperosmotic stress. *Eukaryot. Cell* **7**: 1309–1317.

Mikael Sehlin, H. and Börje Lindström, E. (1992) Oxidation and reduction of arsenic by *Sulfolobus acidocaldarius* strain BC. *FEMS Microbiol. Lett.* **93**: 87–92.

Miller, M.B. and Bassler, B.L. (2001) Quorum sensing in bacteria. *Annu. Rev. Microbiol.* **55**: 165–199.

Moore, L.E., Karami, S., Steinmaus, C., and Cantor, K.P. (2013) Use of OMIC technologies to study arsenic exposure in human populations. *Environ. Mol. Mutagen.* **54**: 589–595.

Morgante, V., Mirete, S., de Figueras, C.G., Postigo Cacho, M., and González-Pastor, J.E. (2014) Exploring the diversity of arsenic resistance genes from acid mine drainage micro-organisms. *Environ. Microbiol.* n/a–n/a.

Mosier, A.C., Justice, N.B., Bowen, B.P., Baran, R., Thomas, B.C., Northen, T.R., and Banfield, J.F. (2013) Metabolites associated with adaptation of micro-organisms to an acidophilic, metal-rich environment identified by stable-isotope-enabled metabolomics. *mBio* **4**: e00484–00412.

Mukherjee, T., Hanes, J., Tews, I., Ealick, S.E., and Begley, T.P. (2011) Pyridoxal phosphate: biosynthesis and catabolism. *Biochim. Biophys. Acta* **1814**: 1585–1596.

Mukhopadhyay, R. and Rosen, B.P. (2002) Arsenate reductases in prokaryotes and eukaryotes. *Environ. Health Perspect.* **110 Suppl 5**: 745–748.

Mukhopadhyay, R., Rosen, B.P., Phung, L.T., and Silver, S. (2002) Microbial arsenic: from geocycles to genes and enzymes. *FEMS Microbiol. Rev.* **26**: 311–325.

Muller, D., Lièvremont, D., Simeonova, D.D., Hubert, J.-C., and Lett, M.-C. (2003) Arsenite oxidase *aox* genes from a metal-resistant *Betaproteobacterium*. *J. Bacteriol.* **185**: 135–141.

Muller, D., Médigue, C., Koechler, S., Barbe, V., Barakat, M., Talla, E., et al. (2007) A tale of two oxidation states: bacterial colonization of arsenic-rich environments. *PLoS Genet.* **3**: e53.

Murphy, J.N. and Saltikov, C.W. (2009) The ArsR repressor mediates arsenite-dependent regulation of arsenate respiration and detoxification operons of *Shewanella* sp. strain ANA-3. *J. Bacteriol.* **191**: 6722–6731.

Nordstrom, D.K. (2002) Public health. Worldwide occurrences of arsenic in ground water. *Science* **296**: 2143–2145.

Ordóñez, E., Letek, M., Valbuena, N., Gil, J.A., and Mateos, L.M. (2005) Analysis of genes involved in arsenic resistance in *Corynebacterium glutamicum* ATCC 13032. *Appl. Environ. Microbiol.* **71**: 6206–6215.

Oremland, R.S. and Stolz, J.F. (2005) Arsenic, microbes and contaminated aquifers. *Trends Microbiol.* **13**: 45–49.

Oremland, R.S. and Stolz, J.F. (2003) The ecology of arsenic. *Science* **300**: 939–944.

Páez-Espino, A.D., Durante-Rodríguez, G., and de Lorenzo, V. (2014) Functional coexistence of twin arsenic resistance systems in *Pseudomonas putida* KT2440. *Environ. Microbiol.*

Páez-Espino, D., Tamames, J., de Lorenzo, V., and Cánovas, D. (2009) Microbial responses to environmental arsenic. *Biometals Int. J. Role Met. Ions Biol. Biochem. Med.* **22**: 117–130.

Pal, C., Bengtsson-Palme, J., Rensing, C., Kristiansson, E., and Larsson, D.G.J. (2014) BacMet: antibacterial biocide and metal resistance genes database. *Nucleic Acids Res.* **42**: D737–743.

Pandey, S., Rai, R., and Rai, L.C. (2012) Proteomics combines morphological, physiological and biochemical attributes to unravel the survival strategy of *Anabaena* sp. PCC7120 under arsenic stress. *J. Proteomics* **75**: 921–937.

Pandey, S., Shrivastava, A.K., Singh, V.K., Rai, R., Singh, P.K., Rai, S., and Rai, L.C. (2013) A new arsenate reductase involved in arsenic detoxification in *Anabaena* sp. PCC7120. *Funct. Integr. Genomics* **13**: 43–55.

Parkhill, J. (2013) What has high-throughput sequencing ever done for us? *Nat. Rev. Microbiol.* **11**: 664–665.

Parsell, D.A. and Lindquist, S. (1993) The function of heat-shock proteins in stress tolerance: degradation and reactivation of damaged proteins. *Annu. Rev. Genet.* **27**: 437–496.

Parvatiyar, K., Alsabbagh, E.M., Ochsner, U.A., Stegemeyer, M.A., Smulian, A.G., Hwang, S.H., et al. (2005) Global analysis of cellular factors and responses involved in *Pseudomonas aeruginosa* resistance to arsenite. *J. Bacteriol.* **187**: 4853–4864.

Patel, P.C., Goulhen, F., Boothman, C., Gault, A.G., Charnock, J.M., Kalia, K., and Lloyd, J.R. (2007) Arsenate detoxification in a *Pseudomonas* hypertolerant to arsenic. *Arch. Microbiol.* **187**: 171–183.

Petrick, J.S., Ayala-Fierro, F., Cullen, W.R., Carter, D.E., and Vasken Aposhian, H. (2000) Monomethylarsonous acid [MMA(III)] is more toxic than arsenite in Chang human hepatocytes. *Toxicol. Appl. Pharmacol.* **163**: 203–207.

Plewniak, F., Koechler, S., Navet, B., Dugat-Bony, E., Bouchez, O., Peyret, P., et al. (2013) Metagenomic insights into microbial metabolism affecting arsenic dispersion in Mediterranean marine sediments. *Mol. Ecol.* **22**: 4870–4883.

Pompella, A., Visvikis, A., Paolicchi, A., De Tata, V., and Casini, A.F. (2003) The changing faces of glutathione, a cellular protagonist. *Biochem. Pharmacol.* **66**: 1499–1503.

Qin, J., Lehr, C.R., Yuan, C., Le, X.C., McDermott, T.R., and Rosen, B.P. (2009) Biotransformation of arsenic by a Yellowstone thermoacidophilic eukaryotic alga. *Proc. Natl. Acad. Sci. U. S. A.* **106**: 5213–5217.

Qin, J., Rosen, B.P., Zhang, Y., Wang, G., Franke, S., and Rensing, C. (2006) Arsenic detoxification and evolution of trimethylarsine gas by a microbial arsenite S-adenosylmethionine methyltransferase. *Proc. Natl. Acad. Sci. U. S. A.* **103**: 2075–2080.

Rahman, M.A. and Hassler, C. (2014) Is arsenic biotransformation a detoxification mechanism for micro-organisms? *Aquat. Toxicol. Amst. Neth.* **146**: 212–219.

Ram, R.J., Verberkmoes, N.C., Thelen, M.P., Tyson, G.W., Baker, B.J., Blake, R.C., et al. (2005) Community proteomics of a natural microbial biofilm. *Science* **308**: 1915–1920.

Remenant, B., Coupat-Goutaland, B., Guidot, A., Cellier, G., Wicker, E., Allen, C., et al. (2010) Genomes of three tomato pathogens within the *Ralstonia solanacearum* species complex reveal significant evolutionary divergence. *BMC Genomics* **11**: 379.

Ren, X., McHale, C.M., Skibola, C.F., Smith, A.H., Smith, M.T., and Zhang, L. (2011) An emerging role for epigenetic dysregulation in arsenic toxicity and carcinogenesis. *Environ. Health Perspect.* **119**: 11–19.

Rhine, E.D., Ní Chadhain, S.M., Zylstra, G.J., and Young, L.Y. (2007) The arsenite oxidase genes (*aroAB*) in novel chemoautotrophic arsenite oxidizers. *Biochem. Biophys. Res. Commun.* **354**: 662–667.

Rosen, B.P. (2002) Biochemistry of arsenic detoxification. *FEBS Lett.* **529**: 86–92.

Rosen, B.P. (1999) Families of arsenic transporters. *Trends Microbiol.* **7**: 207–212.

Rosen, B.P. and Liu, Z. (2009) Transport pathways for arsenic and selenium: a minireview. *Environ. Int.* **35**: 512–515.

Rosen, B.P. and Tamás, M.J. (2010) Arsenic transport in prokaryotes and eukaryotic microbes. *Adv. Exp. Med. Biol.* **679**: 47–55.

Sacheti, P., Bhonsle, H., Patil, R., Kulkarni, M.J., Srikanth, R., and Gade, W. (2013) Arsenomics of *Exiguobacterium* sp. PS (NCIM 5463). *RSC Adv.* **3**: 9705–9713.

Saltikov, C.W. and Newman, D.K. (2003) Genetic identification of a respiratory arsenate reductase. *Proc. Natl. Acad. Sci. U. S. A.* **100**: 10983–10988.

Santini, J.M. and vanden Hoven, R.N. (2004) Molybdenum-containing arsenite oxidase of the chemolithoautotrophic arsenite oxidizer NT-26. *J. Bacteriol.* **186**: 1614–1619.

Santini, J.M., Sly, L.I., Schnagl, R.D., and Macy, J.M. (2000) A new chemolithoautotrophic arsenite-oxidizing bacterium isolated from a gold mine: phylogenetic, physiological, and preliminary biochemical studies. *Appl. Environ. Microbiol.* **66**: 92–97.

Sardiwal, S., Santini, J.M., Osborne, T.H., and Djordjevic, S. (2010) Characterization of a two-component signal transduction system that controls arsenite oxidation in the chemolithoautotroph NT-26. *FEMS Microbiol. Lett.* **313**: 20–28.

Schoepp-Cothenet, B., Duval, S., Santini, J.M., and Nitschke, W. (2009) Comment on "Arsenic (III) fuels anoxygenic photosynthesis in hot spring biofilms from Mono Lake, California." *Science* **323**: 583; author reply 583.

Schönknecht, G., Chen, W.-H., Ternes, C.M., Barbier, G.G., Shrestha, R.P., Stanke, M., et al. (2013) Gene transfer from bacteria and archaea facilitated evolution of an extremophilic eukaryote. *Science* **339**: 1207–1210.

Shen, S., Li, X.-F., Cullen, W.R., Weinfeld, M., and Le, X.C. (2013) Arsenic binding to proteins. *Chem. Rev.* **113**: 7769–7792.

Silver, S. and Phung, L.T. (2005a) A bacterial view of the periodic table: genes and proteins for toxic inorganic ions. *J. Ind. Microbiol. Biotechnol.* **32**: 587–605.

Silver, S. and Phung, L.T. (2005b) Genes and enzymes Involved in bacterial oxidation and reduction of inorganic arsenic. *Appl. Environ. Microbiol.* **71**: 599–608.

Silver, S., Phung, L.T., and Rosen, B.P. (2002) Arsenic metabolism: resistance, reduction and oxidation W. T. Frankenberger, Jr. CRC Press, New York, N. Y.

Simmons, S.L., Dibartolo, G., Denef, V.J., Goltsman, D.S.A., Thelen, M.P., and Banfield, J.F. (2008) Population genomic analysis of strain variation in *Leptospirillum* group II bacteria involved in acid mine drainage formation. *PLoS Biol.* **6**: e177.

Slyemi, D. and Bonnefoy, V. (2012) How prokaryotes deal with arsenic. *Environ. Microbiol. Rep.* **4**: 571–586.

Srivastava, S., Verma, P.C., Singh, A., Mishra, M., Singh, N., Sharma, N., and Singh, N. (2012) Isolation and characterization of *Staphylococcus* sp. strain NBRIEAG-8 from arsenic contaminated site of West Bengal. *Appl. Microbiol. Biotechnol.* **95**: 1275–1291.

Starkenburg, S.R., Larimer, F.W., Stein, L.Y., Klotz, M.G., Chain, P.S.G., Sayavedra-Soto, L.A., et al. (2008) Complete genome sequence of *Nitrobacter hamburgensis* X14 and comparative genomic analysis of species within the genus *Nitrobacter*. *Appl. Environ. Microbiol.* **74**: 2852–2863.

Stolz, J.F., Basu, P., Santini, J.M., and Oremland, R.S. (2006) Arsenic and selenium in microbial metabolism. *Annu. Rev. Microbiol.* **60**: 107–130.

Sun, W., Sierra-Alvarez, R., Milner, L., and Field, J.A. (2010) Anaerobic oxidation of arsenite linked to chlorate reduction. *Appl. Environ. Microbiol.* **76**: 6804–6811.

The UniProt Consortium (2014) Activities at the Universal Protein Resource (UniProt). *Nucleic Acids Res.* **42**: D191–D198.

Thorsen, M., Lagniel, G., Kristiansson, E., Junot, C., Nerman, O., Labarre, J., and Tamás, M.J. (2007) Quantitative transcriptome, proteome, and sulfur metabolite profiling of the *Saccharomyces cerevisiae* response to arsenite. *Physiol. Genomics* **30**: 35–43.

Thorsen, M., Perrone, G.G., Kristiansson, E., Traini, M., Ye, T., Dawes, I.W., et al. (2009) Genetic basis of arsenite and cadmium tolerance in *Saccharomyces cerevisiae*. *BMC Genomics* **10**: 105.

Tripathi, R.D., Tripathi, P., Dwivedi, S., Dubey, S., Chatterjee, S., Chakrabarty, D., and Trivedi, P.K. (2012) Arsenomics: omics of arsenic metabolism in plants. *Front. Physiol.* **3**: 275.

Tsai, S.-L., Singh, S., and Chen, W. (2009) Arsenic metabolism by microbes in nature and the impact on arsenic remediation. *Curr. Opin. Biotechnol.* **20**: 659–667.

Tuffin, I.M., de Groot, P., Deane, S.M., and Rawlings, D.E. (2005) An unusual Tn*21*-like transposon containing an *ars* operon is present in highly arsenic-resistant strains of the biomining bacterium *Acidithiobacillus caldus*. *Microbiol. Read. Engl.* **151**: 3027–3039.

Tyson, G.W., Chapman, J., Hugenholtz, P., Allen, E.E., Ram, R.J., Richardson, P.M., et al. (2004) Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **428**: 37–43.

Tyson, G.W., Lo, I., Baker, B.J., Allen, E.E., Hugenholtz, P., and Banfield, J.F. (2005) Genome-directed isolation of the key nitrogen fixer *Leptospirillum ferrodiazotrophum* sp. nov. from an acidophilic microbial community. *Appl. Environ. Microbiol.* **71**: 6319–6324.

Valko, M., Morris, H., and Cronin, M.T.D. (2005) Metals, toxicity and oxidative stress. *Curr. Med. Chem.* **12**: 1161–1208.

Vallenet, D., Belda, E., Calteau, A., Cruveiller, S., Engelen, S., Lajus, A., et al. (2013) MicroScope - an integrated microbial resource for the curation and comparative analysis of genomic and metabolic data. *Nucleic Acids Res.* **41**: D636–647.

Vergauwen, B., Pauwels, F., and Van Beeumen, J.J. (2003) Glutathione and catalase provide overlapping defenses for protection against respiration-generated hydrogen peroxide in *Haemophilus influenzae*. *J. Bacteriol.* **185**: 5555–5562.

Volant, A., Bruneel, O., Desoeuvre, A., Héry, M., Casiot, C., Bru, N., et al. (2014) Diversity and spatiotemporal dynamics of bacterial communities: physicochemical and others drivers along an acid mine drainage. *FEMS Microbiol. Ecol.*

Wang, Z., Gerstein, M., and Snyder, M. (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **10**: 57–63.

Weckwerth, W. (2010) Metabolomics: an integral technique in systems biology. *Bioanalysis* **2**: 829–836.

Weiss, S., Carapito, C., Cleiss, J., Koechler, S., Turlin, E., Coppee, J.-Y., et al. (2009) Enhanced structural and functional genome elucidation of the arsenite-oxidizing strain *Herminiimonas arsenicoxydans* by proteomics data. *Biochimie* **91**: 192–203.

Wheeler, D.L., Barrett, T., Benson, D.A., Bryant, S.H., Canese, K., Chetvernin, V., et al. (2006) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **34**: D173–D180.

Wilmes, P., Bowen, B.P., Thomas, B.C., Mueller, R.S., Denef, V.J., VerBerkmoes, N.C., et al. (2010) Metabolome-proteome differentiation coupled to microbial divergence. *mBio* **1**:

Wu, B., Song, J., and Beitz, E. (2010) Novel channel enzyme fusion proteins confer arsenate resistance. *J. Biol. Chem.* **285**: 40081–40087.

Yang, H.-C., Cheng, J., Finan, T.M., Rosen, B.P., and Bhattacharjee, H. (2005) Novel pathway for arsenic detoxification in the legume symbiont *Sinorhizobium meliloti*. *J. Bacteriol.* **187**: 6991–6997.

Ye, J., Rensing, C., Rosen, B.P., and Zhu, Y.-G. (2012) Arsenic biomethylation by photosynthetic organisms. *Trends Plant Sci.* **17**: 155–162.

Yin, X.-X., Zhang, Y.-Y., Yang, J., and Zhu, Y.-G. (2011) Rapid biotransformation of arsenic by a model protozoan *Tetrahymena pyriformis* GL-C. *Environ. Pollut. Barking Essex 1987* **159**: 837–840.

Yoon, H., Hong, J., and Ryu, S. (2008) Effects of chaperones on mRNA stability and gene expression in *Escherichia coli*. *J. Microbiol. Biotechnol.* **18**: 228–233.

Yoshinaga, M., Cai, Y., and Rosen, B.P. (2011) Demethylation of methylarsonic acid by a microbial community. *Environ. Microbiol.* **13**: 1205–1215.

Yoshinaga, M. and Rosen, B.P. (2014) A C·As lyase for degradation of environmental organoarsenical herbicides and animal husbandry growth promoters. *Proc. Natl. Acad. Sci. U. S. A.* **111**: 7701–7706.

Yuan, C., Lu, X., Qin, J., Rosen, B.P., and Le, X.C. (2008) Volatile arsenic species released from *Escherichia coli* expressing the AsIII S-adenosylmethionine methyltransferase gene. *Environ. Sci. Technol.* **42**: 3201–3206.

Zargar, K., Conrad, A., Bernick, D.L., Lowe, T.M., Stolc, V., Hoeft, S., et al. (2012) ArxA, a new clade of arsenite oxidase within the DMSO reductase family of molybdenum oxidoreductases. *Environ. Microbiol.* **14**: 1635–1645.

Zargar, K., Hoeft, S., Oremland, R., and Saltikov, C.W. (2010) Identification of a novel arsenite oxidase gene, *arxA*, in the haloalkaliphilic, arsenite-oxidizing bacterium *Alkalilimnicola ehrlichii* strain MLHE-1. *J. Bacteriol.* **192**: 3755–3762.

Zhang, Y.-B., Monchy, S., Greenberg, B., Mergeay, M., Gang, O., Taghavi, S., and van der Lelie, D. (2009) ArsR arsenic-resistance regulatory protein from *Cupriavidus metallidurans* CH34. *Antonie Van Leeuwenhoek* **96**: 161–170.

Zhang, Y., Ma, Y.-F., Qi, S.-W., Meng, B., Chaudhry, M.T., Liu, S.-Q., and Liu, S.-J. (2007) Responses to arsenate stress by *Comamonas* sp. strain CNB-1 at genetic and proteomic levels. *Microbiol. Read. Engl.* **153**: 3713–3721.

Zhang, Y.-Y., Yang, J., Yin, X.-X., Yang, S.-P., and Zhu, Y.-G. (2012) Arsenate toxicity and stress responses in the freshwater ciliate *Tetrahymena pyriformis*. *Eur. J. Protistol.* **48**: 227–236.

Zuccaro, A., Schoch, C.L., Spatafora, J.W., Kohlmeyer, J., Draeger, S., and Mitchell, J.I. (2008) Detection and identification of fungi intimately associated with the brown seaweed *Fucus serratus*. *Appl. Environ. Microbiol.* **74**: 931–941.

OBJECTIFS DE LA THÈSE

Ce travail de thèse a pour objectif une meilleure compréhension de la réponse adaptative des micro-organismes au stress arsénié et du rôle des interactions procaryote-eucaryote dans un tel contexte toxique. En effet, si les mécanismes biologiques permettant la transformation des espèces arséniées sont aujourd'hui bien étudiés, l'éventail des processus physiologiques permettant le maintien des micro-organismes en environnement contaminé est encore mal connu.

Pour répondre à cette problématique, l'intérêt s'est tout d'abord porté sur *Rhizobium* sp. NT-26. L'isolement de cette souche bactérienne au contact d'un minéral riche en arsenic ainsi que son appartenance à la famille des *Rhizobiaceae*, comprenant de nombreuses espèces vivant en association avec des plantes, laissaient initialement penser qu'elle serait un modèle propice à l'atteinte des objectifs fixés. Dans un second temps, ces travaux ont conduit à s'intéresser à la réponse adaptative mise en place par *Euglena mutabilis*, un protiste photosynthétique. De plus, son isolement au contact de la communauté microbienne du Drainage Minier Acide de Carnoulès, caractérisé par de fortes concentrations en arsenic, suggérait l'existence d'interactions microbiennes dont l'étude a constitué la troisième partie de ce travail.

De manière générale, les outils de génomique descriptive, comparative et fonctionnelle ont été utiles tout au long de ce travail, permettant de dévoiler d'importantes informations génomiques, transcriptomiques, protéomiques ou métabolomiques. De plus, leur couplage à des techniques plus conventionnelles de mutagénèse, à des tests physiologiques ou de biologie moléculaire ainsi qu'à des observations microscopiques a permis de préciser la physiologie de ces micro-organismes soumis au stress arsénié.

Enfin, ces différentes approches ont mené à la rédaction de plusieurs articles scientifiques, acceptés, soumis ou en cours de rédaction et qui sont utilisés tout au long de ce manuscrit dans le but de présenter ces travaux de thèse.

CHAPITRE I

LA RÉPONSE AU STRESS ARSENIÉ DE *RHIZOBIUM* SP. NT-26

INTRODUCTION

La première partie de ce travail a concerné l'étude des mécanismes adaptatifs de la souche *Rhizobium* sp. NT-26 au stress arsénié et visait à mettre en évidence de potentielles interactions procaryote-eucaryote dans un tel contexte.

La souche bactérienne *Rhizobium* sp. NT-26 a été isolée d'une mine d'or, au contact d'échantillons d'arsénopyrite, un minéral composé de fer, d'arsenic et de soufre (Santini et al., 2000). La caractérisation de son métabolisme a permis de montrer sa capacité à croître de manière chimiolithoautotrophe en utilisant l'As(III) comme source d'électron, l'oxygène comme accepteur terminal d'électron et le dioxyde de carbone, CO_2, ou bicarbonate, HCO_3^-, comme source de carbone (Santini et al., 2000). De manière intéressante, cette souche est également capable d'utiliser l'As(III) en conditions hétérotrophiques et montre un temps de génération court en comparaison avec la plupart des bactéries arsénite-oxydantes. De plus, elle est capable de tolérer d'importantes concentrations en arsenic, allant jusqu'à 0,5 M d'As(V) et plus de 18,5 mM d'As(III) (Andres et al., 2013). Ces données préliminaires suggèrent ainsi que la physiologie de *Rhizobium* sp. NT-26 est particulièrement adaptée à cet élément toxique, et justifient le choix de cette souche comme modèle d'étude.

Ensuite, l'analyse phylogénétique de la séquence de l'ARN ribosomique 16S de *Rhizobium* sp. NT-26 classe cette souche au sein des *Alphaproteobacteria* et plus particulièrement dans la famille des *Rhizobiaceae* (Santini et al., 2000). Tandis que ces bactéries sont très répandues, leur physiologie en regard de l'arsenic est peu décrite. De plus, cette famille regroupe de nombreuses espèces bactériennes vivant en association avec les plantes. Les bactéries des genres *Rhizobium* et *Ensifer* sont par exemple connues pour leur implication dans des relations mutualistes avec des plantes légumineuses (Bromfield et al., 2010). Cette dernière caractéristique a conforté le choix de *Rhizobium* sp. NT-26, permettant potentiellement d'examiner le rôle des interactions procaryote-eucaryote dans le maintien des organismes en contexte arsénié.

Le génome de *Rhizobium* sp. NT-26 a tout d'abord été séquencé et analysé dans le but de mettre en évidence les mécanismes impliqués dans l'adaptation de cette souche à son environnement. La disponibilité du génome a ensuite rendu possible des analyses transcriptomiques et protéomiques différentielles. Celles-ci ont été développées afin d'examiner les fonctions cellulaires affectées par l'arsenic et potentiellement impliquées dans la réponse à cet élément. Enfin,

des résultats obtenus chez *H. arsenicoxydans* et *Thiomonas* sp. CB2 suggéraient un lien entre les processus de colonisation et le métabolisme de l'arsenic (Muller et al., 2007; Marchal et al., 2010, 2011). Cette relation a été étudiée par des approches de mutagénèse aléatoire, des tests physiologiques et des observations microscopiques.

L'intégralité de ces résultats a permis la publication de l'article scientifique présenté ici dans *Genome, Biology and Evolution*. J'ai regroupé les données phylogénétiques, d'interactions, transcriptomiques et protéomiques et contribué à l'écriture du manuscrit. Du point de vue expérimental, j'ai réalisé certaines analyses génomiques, complété la banque de mutants, caractérisé ces derniers et enfin effectué les tests physiologiques de mobilité et de synthèse de biofilm.

MANUSCRIT

GBE

Life in an Arsenic-Containing Gold Mine: Genome and Physiology of the Autotrophic Arsenite-Oxidizing Bacterium *Rhizobium* sp. NT-26

Jérémy Andres[1], Florence Arsène-Ploetze[1], Valérie Barbe[2], Céline Brochier-Armanet[3], Jessica Cleiss-Arnold[1], Jean-Yves Coppée[4], Marie-Agnès Dillies[4], Lucie Geist[1], Aurélie Joublin[1], Sandrine Koechler[1], Florent Lassalle[3,5,6], Marie Marchal[1], Claudine Médigue[7], Daniel Muller[6], Xavier Nesme[6], Frédéric Plewniak[1], Caroline Proux[4], Martha Helena Ramírez-Bahena[6,8], Chantal Schenowitz[2], Odile Sismeiro[4], David Vallenet[7], Joanne M. Santini[5,*], and Philippe N. Bertin[1,*]

[1]Laboratoire Génétique Moléculaire, Génomique et Microbiologie, UMR7156 CNRS Université de Strasbourg, Strasbourg, France

[2]Laboratoire de Finition, CEA-IG-Genoscope, Evry, France

[3]Université de Lyon, Université Lyon 1, CNRS, Laboratoire de Biométrie et Biologie Evolutive, UMR5558, Villeurbanne, France

[4]Plate-forme Technologique Transcriptome et Epigénome, Institut Pasteur, Paris, France

[5]Institute of Structural and Molecular Biology, University College London, United Kingdom

[6]Université de Lyon, Université Lyon 1, CNRS, INRA, Laboratoire Ecologie Microbienne Lyon, UMR5557, USC1193, Villeurbanne, France

[7]Laboratoire Analyses Bioinformatiques pour la Génomique et le Métabolisme, Genoscope-IG-CEA, Evry, France

[8]Instituto de Recursos Naturales y Agrobiologia, IRNASA-CSIC, Salamanca, Spain

***Corresponding authors:** E-mail: philippe.bertin@unistra.fr; j.santini@ucl.ac.uk.

Accepted: April 9, 2013

Data deposition: This project has been deposited at NCBI and ArrayExpress under the accession numbers 1125847 and E-MEXP-3021, respectively.

Abstract

Arsenic is widespread in the environment and its presence is a result of natural or anthropogenic activities. Microbes have developed different mechanisms to deal with toxic compounds such as arsenic and this is to resist or metabolize the compound. Here, we present the first reference set of genomic, transcriptomic and proteomic data of an *Alphaproteobacterium* isolated from an arsenic-containing goldmine: *Rhizobium* sp. NT-26. Although phylogenetically related to the plant-associated bacteria, this organism has lost the major colonizing capabilities needed for symbiosis with legumes. In contrast, the genome of *Rhizobium* sp. NT-26 comprises a megaplasmid containing the various genes, which enable it to metabolize arsenite. Remarkably, although the genes required for arsenite oxidation and flagellar motility/biofilm formation are carried by the megaplasmid and the chromosome, respectively, a coordinate regulation of these two mechanisms was observed. Taken together, these processes illustrate the impact environmental pressure can have on the evolution of bacterial genomes, improving the fitness of bacterial strains by the acquisition of novel functions.

Key words: arsenic metabolism, motility/biofilm, *Rhizobium/Agrobacterium*, transcriptomics/proteomics, phylogeny, rhizosphere.

Introduction

To deal with high concentrations of toxic metals, micro-organisms have evolved various strategies, which enable them to detoxify their environment. These processes involve physicochemical reactions, for example, precipitation or solubilization, adsorption or desorption (Borch et al. 2010), and metabolic oxido-reduction reactions (Gadd 2010). In addition, most of the metallic elements found in the periodic table may play a crucial role in microbial physiology, for example, as components of metalloproteins, or as electron

donors or acceptors in energy metabolism (Stolz 2011). Such a metabolism may have been important in the early stages of life, due to a high concentration of metals, including arsenic, in the primordial planet (reviewed in van Lis et al. 2012).

In recent years, the various "omics" methods, which include genome sequencing, comparative genomics, and transcriptome or proteome analysis, have allowed to address the physiology of organisms in a global way. Such approaches have therefore greatly improved the understanding of microbial metabolism (Bertin et al. 2008; Holmes et al. 2009; Wilkins et al. 2009), including the global functioning of ecosystems, as recently demonstrated for an arsenic-rich microbial community (Bertin et al. 2011). To date, the genomes of more than 20 arsenic-metabolizing strains have been sequenced. They originate from various environments, belong to unrelated taxonomic groups, and have different carbon and energy requirements (reviewed in Bertin et al. 2012; van Lis et al. 2013).

Regarding arsenic, which is mainly present in two oxidation states in aquatic environments, that is, arsenite [As(III)] and arsenate [As(V)], microorganisms have acquired various metabolic capacities. These include As(V) reduction, which is usually part of the resistance mechanism, but also functions involved in As(III) oxidation or methylation (reviewed in Stolz 2011). Unlike arsenite methyltransferase genes, which are not often found in bacterial genomes, genes encoding arsenite oxidase are widespread in Bacteria and Archaea (Heinrich-Salmeron et al. 2011 and reviewed in Osborne and Santini 2012). In *Herminiimonas arsenicoxidans*, the arsenite oxidase *aioBA* genes are located in an arsenic genomic island, which also contains genes involved in arsenic resistance and biosynthesis of a molybdenum cofactor of the Aio enzyme (Muller et al. 2007). Such a genetic organization has also been observed in *Thiomonas arsenitoxydans* (Arsène-Ploetze et al. 2010; Bertin et al. 2012) and the presence of *aio* genes on a plasmid has been reported in *Nitrobacter hamburgensis* and *Thermus thermophilus* str. HB8 (Bertin et al. 2012). These observations suggest that *aioBA* genes may be acquired by horizontal gene transfer.

The adaptative response to arsenic has been recently shown as occurring in two steps (Cleiss-Arnold et al. 2010). First, bacterial cells express various genes involved in defence mechanisms, for example, oxidative stress and arsenic efflux. Next, several metabolic activities are induced, including arsenite oxidation which, in heterotrophic bacteria like *H. arsenicoxydans*, may be principally considered as a detoxification mechanism (Muller et al. 2007). In contrast, in bacteria that can grow autotrophically such as *T. arsenitoxydans* (Arsène-Ploetze et al. 2010) arsenite oxidation is part of a bioenergetic mechanism involved in energy generation. Despite some similarities, the genome organization of these two bacteria and their arsenic response, including biofilm formation, have been shown to differ markedly (Marchal et al. 2010, 2011).

In their natural environment, bacteria usually grow in biofilms, which are structured microbial communities embedded in extracellular polymeric substances (EPS) composed of sugars, proteins, and DNA (Hall-Stoodley et al. 2004; McDougald et al. 2012). Even though biofilm formation can be a problem in the field of human health, it allows bacteria to survive and thrive in highly toxic environments, including those characterized by high concentrations of heavy metals or metalloids such as arsenic (Guibaud et al. 2006; Muller et al. 2007). Unlike *H. arsenicoxydans* (Marchal et al. 2010), *T. arsenitoxydans* has been shown to induce biofilm formation in the presence of As(III) (Marchal et al. 2011). In addition, after biofilm development, the induction of cell motility has led to accelerated cell dispersion, an important process in the colonization of alternative ecological niches.

To gain a better understanding of the genetic determinants involved in the metabolism of arsenic, we have investigated the response to As(III) in *Rhizobium* sp. NT-26, a motile, chemolithoautotrophic arsenite oxidizer isolated from a gold mine in Australia (Santini et al. 2000). This strain belongs to the *Rhizobiaceae* family of the *Alphaproteobacteria*, which includes many species living in association with plants, such as plant mutualists of the *Rhizobium* and *Ensifer* (formerly *Sinorhizobium*) genera (Martens et al. 2007) and plant pathogens or plant growth-promoting rhizobacteria of the *Agrobacterium* genus (Hao et al. 2011). The *Rhizobium* sp. NT-26 genome was sequenced and annotated, and its physiology was investigated using differential transcriptomics and proteomics, and random mutagenesis. Remarkably, the synthesis of flagella was shown to be controlled by arsenite, suggesting a possible coordinate regulation between clusters located on two genetic elements. Indeed, proteins involved in As(III) oxidation were shown to be encoded by genes present on a megaplasmid, whereas flagellar genes are located on the chromosome.

Materials and Methods

Bacterial Strains, Plasmid, and Growth Conditions

Rhizobium sp. NT-26 and its mutant strains were cultivated at 28 °C in minimal salts medium (MSM) containing 0.04% yeast extract (Santini et al. 2000) and supplemented with As(III) and agar when required. *Escherichia coli* S17.1 λ*pir* was cultivated at 28 °C in Luria–Bertani (LB) (MP Biomedicals) medium supplemented with 20 mg/l kanamycin (Sigma) for the maintenance of the pTGN/mini-Tn*5 gfp-km* plasmid (Tang et al. 1999).

Random Mutagenesis and Screening

Using the suicide vector pTGN carrying the mini-Tn*5* transposon, random mutagenesis was performed to construct a mutant library and to identify genes involved either in arsenite oxidation or in motility. Mobilization of the plasmid was performed using *E. coli* S17.1 λ*pir* carrying plasmid pTGN as the

donor and *Rhizobium* sp. NT-26 as the recipient. For conjugation, both strains in exponential phase, respectively, corresponding to an optical density (OD) of 0.6 and 0.135 at 600 nm, were superposed on LB plates at 28 °C for 24 h. As the *Rhizobium* sp. NT-26 strain used in this study is rifampicin resistant (Santini and vanden Hoven 2004), mutants were then selected on LB plates supplemented with 20 mg/l kanamycin and rifampicin.

Colonies from the library were screened for the loss of arsenite oxidation or the loss of motility. Briefly, mutants were individually inoculated into 96-well microtiter plates containing MSM with 0.04% yeast extract and 8 mM As(III) and incubated at 28 °C in 1.5% agar for 48 h or 0.3% agar for 24 h, respectively. The library was screened in the following two ways: 1) the silver nitrate method was used to detect arsenite oxidation (Lett et al. 2001; Muller et al. 2003) and 2) the diameter of the swarming ring was used to determine whether the cells were motile (Muller et al. 2007). Each phenotype was subsequently confirmed on Petri dishes in the corresponding medium. Mutants unable to oxidize arsenite were also tested for motility and vice versa.

To identify the disrupted gene in each mutant, the genomic region close to the mini-Tn*5* insertion site was amplified by inverse polymerase chain reaction (PCR). Total DNA was extracted with the Wizard Genomic DNA purification kit according to the manufacturer's instructions (Promega). One microgram of DNA was digested with 50 U of restriction enzymes that do not cut the transposon sequence (ClaI or PstI) in a 50 µl reaction volume at 37 °C for 2 h. After precipitation by ethanol and sodium acetate, digested DNA was ligated with 10 U of DNA ligase (Fermentas) in a volume of 20 µl overnight at 16 °C. PCR was carried out on 25 ng of this template in a 25 µl volume reaction with iProof DNA Polymerase (Bio-Rad) and Oend (ACTTGTGTATAAGAGTCAG) and Iend (AGATCTGATCAAGAGACAG) primers. The program used involved a denaturation step at 98 °C for 30 s, followed by 35 cycles of denaturation at 98 °C for 10 s, annealing at 52 °C for 30 s and elongation at 72 °C for 3 min, and a final elongation step at 72 °C for 10 min. Amplification products were checked on an agarose gel and sequenced with Oend by MilleGen (http://www.millegen.com/, last accessed April 30, 2013). The Blastn tool on the MaGe interface (Vallenet et al. 2006) was used to align the sequences with that of the *Rhizobium* sp. NT-26 genome allowing for identification of the disrupted gene. For each mutant, the precise insertion site and orientation of the mini-Tn*5* was determined by PCR, combining the Oend and Iend primers with new specific primers (supplementary table S1, Supplementary Material online) designed around each probable insertion site.

Biofilm Quantification

Biofilm formation by *Rhizobium* sp. NT-26 wild-type and mutant strains grown in the presence or absence of arsenite was measured by the crystal violet method. Cultures were grown in MSM medium containing 0.04% yeast extract supplemented with and without 8 mM As(III) and incubated at 28 °C overnight with shaking (120 rpm). The cultures were then diluted with fresh medium to an OD of 0.1 at 600 nm. Each strain was tested with six replicates of 200 µl in two flat-bottomed polystyrene 96-well microtiter plates (Nunc). Cultures were incubated at 28 °C for 24 h and 48 h without agitation. Biofilm formation was quantified using crystal violet, as previously described (Hommais et al. 2002). Briefly, after removing the culture medium, wells were gently rinsed three times with 0.1 M phosphate-buffered saline (PBS). To fix biofilms, plates were dried at 55 °C for 25 min, then 200 µl of 0.1% [w/v] crystal violet solution (Merck) was added to the wells and the plates were incubated at 30 °C for 30 min. Free crystal violet was removed and wells were washed three times with PBS. Plates were dried at room temperature and the biofilm was subsequently dissolved in 200 µl of 95% [v/v] ethanol over 30 min. Finally, the absorbance was read at 595 nm with a microplate reader (Synergy HT).

Pulsed-Field Gel Electrophoresis

Plasmid profiles were determined by a modified Eckhardt agarose gel electrophoresis technique, as described previously (Hynes and McGregor 1990). *Rhizobium* sp. NT-26 was grown in LB until an OD of 0.5 at 600 nm was reached, and 150 µl of culture were used per well. Electrophoresis was carried out at 4 °C, 5 V for 30 min and 85 V for 7 h on a 0.7% agarose gel containing 1% [w/v] sodium dodecyl sulfate (SDS) (Ramírez-Bahena et al. 2012). Plasmid size was estimated by comparison with those from *Agrobacterium tumefaciens* C58 (Wood et al. 2001).

Plant Trapping Tests

Nodulation experiments were performed under gnotobiotic conditions. Seeds of *Macroptilium atropurpureum, Vicia faba, Phaseolus vulgaris*, and *Pisum sativum* were surface sterilized for 2 min in 95% ethyl alcohol and then three times for 3 min in 1% sodium hypochlorite, each time washed with sterile water. Germination was carried out at 28 °C in dark conditions on glass plates covered with sterile filter paper moistened with sterile water. Pots with a capacity of 1.5 l were filled with sterile vermiculite, and 200 ml of nutrient sterile solution (Rigaud and Puppo 1975) was added per pot. Two seedlings were sown in each pot and plants were inoculated with a suspension of 10^5 CFU/ml 1 week after their transfer to hydroponic growth. *Rhizobium* sp. NT-26 was grown on YMB medium (Mannitol 0.7%, Yeast extract 0.2%, KH_2PO_4 0.02%, $MgSO_4$ 0.02%) and 1 ml of inoculum was applied to each seedling. Plants were regularly observed for nodule formation, and nodulation was quantified after inoculation as described in Gremaud and Harper (1989).

Genome Sequencing

The complete genome sequence of *Rhizobium* sp. NT-26 was obtained by combining Sanger and 454 sequencing methods. Sanger reads were obtained from a 10 kb insert library constructed after mechanical shearing of the genomic DNA and cloning of the generated inserts into the plasmid pCNS, as described previously (Muller et al. 2007). Plasmid DNA was purified and end-sequenced (26,888 reads) by dye-terminator chemistry with ABI3730 sequencers (Applied Biosystems) leading approximately to a 4× coverage. Reads were assembled by Newbler with around 20× coverage of 454 GS FLX reads (Roche) and validated via the Consed interface. Finishing steps were performed using primer walking of clones, PCR and in vitro transposition technology with the Template Generation System II Kit (Finnzyme), corresponding to 252, 32 and 8,404 additional reads, respectively. Approximately 70× coverage of 36 bp Illumina reads were mapped in the polishing phase, using SOAP (http://soap.genomics.org.cn/, last accessed April 30, 2013), as previously described (Aury et al. 2008).

Comparative Analysis of 24 *Rhizobiaceae* Genomes

The 23 genomes of *Rhizobiaceae* publicly available at the time of experiments (supplementary table S2, Supplementary Material online) were retrieved from ENA database (http://www.ebi.ac.uk/ena/, last accessed April 30, 2013). A homologous gene family database was built under the HOGENOM procedure (Penel et al. 2009) based on these 23 genomes and the one of *Rhizobium* sp. NT-26. Homologous protein sequences were aligned using MUSCLE (v.3.8.31, default parameters) (Edgar 2004) and then retro-translated with the pal2nal program (v.14) (Suyama et al. 2006). Nucleic acid alignments were restricted to conserved blocks with Gblocks (v.0.91b, minimum 50% of sequences in conserved and flank positions and all gaps allowed, codon mode) (Castresana 2000) and gene trees were computed from these alignments with PhyML (v.3.0, GTR + G8 + I model of evolution, best of SPR and NNI moves, SH-like branch supports) (Guindon and Gascuel 2003). All alignments and phylogenetic trees are shown in supplementary methods S1 and S2, Supplementary Material online. Replicon mapping and gene content comparison were done with custom Python scripts.

Species Phylogenies

The "core" set contained 822 gene families present in every 24 strains in only one copy. The "ribosomal" set contained 51 gene families whose products were annotated as "ribosomal protein" or related terms in at least one genome and were present in at least 22 strains. Full alignments of both family sets, and third codon-removed version of core family set were concatenated and used for species tree construction with RaxML (version 7.2.8-ALPHA, GTRCAT model with 50 categories, branch supports from 200 and 1,000 rapid bootstrap trees for "core" and "ribosomal" alignments, respectively) (Stamatakis 2006) (species trees are stored in supplementary methods S3, Supplementary Material online).

Tree Pattern Matching

Phylogenetic trees of gene families were searched for particular phylogenetic patterns, that is, subtrees with specific arrangement of relative branching leaves representing taxa, with TPMS software (Bigot et al. 2012): "NT26outAgro", that is *Rhizobium* sp. NT-26 as a direct outgroup of *Agrobacterium* genus; "NT26inAgro", that is *Rhizobium* sp. NT-26 as an ingroup of *Agrobacterium* and sister group of *A. tumefaciens*. Both searches were made first without considering branch support and then matching only with >0.9 SH-like branch support at nodes of interest (supplementary methods S4, Supplementary Material online).

Aio Phylogenies

Homologs of AioA, AioB, AioR, AioS, and AioX were retrieved from the nr database at the NCBI (http://www.ncbi.nlm.nih.gov/, last accessed April 30, 2013) using the BlastP program (Altschul et al. 1997) with the protein sequences of *Rhizobium* sp. NT-26 as queries and default parameters except the "Max target sequences" parameter which was set to 1,000. For each Aio protein, the 500 homologs displaying the highest similarity with the sequence of *Rhizobium* sp. NT-26 were retrieved and aligned using MAFFT (version 6, default parameters) (Katoh and Toh 2008). The resulting alignments were trimmed using BMGE (default parameters) (Criscuolo and Gribaldo 2010). Preliminary phylogenies were inferred using the Neighbor-Joining method implemented in SeaView (Poisson evolutionary distance) (Gouy et al. 2010). The robustness of the resulting trees was estimated with the nonparametric bootstrap procedure implemented in SeaView (100 replicates of the original alignments). Based on the resulting trees, the closest relatives of *Rhizobium* sp. NT-26 sequences were identified and used for more detailed phylogenetic analyses. The corresponding sequences were realigned and the resulting alignments trimmed using the same procedure. Final phylogenetic analyses were performed using the maximum likelihood and Bayesian approaches implemented in PhyML (version 3) (Guindon et al. 2009) and MrBayes (version 3.2) (Ronquist et al. 2012), respectively. PhyML was run with the LG evolutionary model (Le and Gascuel 2008) and a gamma distribution with four categories of substitution rates (Γ4) and an estimated alpha parameter. The robustness of the maximum likelihood trees was estimated by the nonparametric procedure implemented in PhyML (100 replicates of the original alignments). MrBayes was run with a mixed substitution model and a Γ4 distribution. Four chains were run in parallel for 1,000,000 generations. The first 2,000 generations were discarded as "burnin." The remaining trees were sampled every 100 generations to build the consensus tree.

Total RNA Extraction, Microarrays, and Data Analysis

A custom 15 K microarray with a probe length of 60 mer was manufactured by Agilent Technologies following the protocol used for *H. arsenicoxydans* (Weiss et al. 2009). Total RNA was extracted from *Rhizobium* sp. NT-26 strain grown heterotrophically in MSM containing 0.04% yeast extract in the absence and presence of 5.3 mM As(III) until late log phase (OD at 600 nm of 0.115 and 0.152, respectively) as described previously (Santini et al. 2007). RNA quality was checked using an Agilent Bioanalyzer. Ten micrograms of total RNA was reverse transcribed using the Fairplay III Microarray labeling kit (Agilent Technologies) and cDNA were indirectly labeled using Cy3 or Cy5 Mono reactive dyes (GE Healthcare). Labeled cDNA quality and quantity were determined by spectroscopy at 260, 280, 550, and 650 nm. The labeled Cy3 and Cy5 target quantities were adjusted to 250 pmol, mixed together and concentrated with Microcon YM-30 (Millipore). Hybridization was performed for 17 h at 65 °C. Three distinct biological RNA samples as well as dye swap experiments were performed for each culture condition. Arrays were scanned as described previously (Weiss et al. 2009). Data were acquired by Genepix Pro 6.0 (Axon Instrument) and statistically analyzed as described previously (Koechler et al. 2010). Genes having a BH adjusted *P* value lower than 0.05 were considered as differentially expressed between the two conditions and were retained for further study. Microarray data were deposited in ArrayExpress (E-MEXP-3021).

Preparation of Proteins Extracts and 2D Gel Electrophoresis

Experiments were performed with four protein extracts from four replicates for each growth condition. Strain NT-26 was grown heterotrophically in MSM containing 0.04% yeast extract in the absence or presence of 5.3 mM As(III). Exponential phase cultures were harvested by centrifugation at 6,000 × g for 10 min at 4 °C. Pellets were suspended in 400 µl of distilled water supplemented with 1 µl Benzonase Nuclease (Sigma) and 4 µl of Protease Inhibitor Mix (GE Healthcare). Cell suspensions were sonicated on ice with 10 pulses of 30 s at 28% of amplitude with 30 s intervals using a VC 750 sonicator (Bioblock Scientific). Cellular debris were removed by two centrifugations, the first at 6,000 × g for 5 min and the second at 16,000 × g for 90 min. Protein concentrations were measured using the Bradford method (Bradford 1976).

Differential accumulation of proteins was either monitored by Colloidal Brilliant Blue staining or DIGE (Marouga et al. 2005). For Colloidal Brilliant Blue staining experiments, 300 µg of protein extract were diluted to a final volume of 350 µl with rehydration buffer (8 M urea, 2% [w/v] CHAPS, 0.5% [v/v] IPG buffer pH 3-10, 40 mM DTT, and 0.01% [w/v] bromophenol blue). For DIGE experiments, 50 µg of protein was adjusted to pH 8.8 by adding 50 mM Tris–HCl final concentration, and either stained with 400 pmol of Cy3 or Cy5 (GE Healthcare). In addition, 25 µg of each 8 extracts (4 replicates for 2 conditions) were pooled and stained with 1,600 pmol of Cy2 to serve as an internal standard. For the staining, each CyDye DIGE fluor stock solution was diluted in high grade dimethylformamide to a final concentration of 400 pmol/µl. One microliter of the dilution was added to 50 µg of protein (Cy3 and Cy5) or 4 µl to 200 µg of the internal standard pool (Cy2) and kept in the dark and on ice for 30 min. The reaction was stopped by adding 1 µl of a 10 mM solution of lysine to 50 µg of protein (4 µl to 200 µg) and then 1 volume of 2× sample buffer (9 M urea, 3 M thiourea, 130 mM DTT, 4% [w/v] CHAPS, 2% [v/v] IPG buffer Pharmalyte 3-10) was added prior to incubation on ice for 10 min. One Cy3-labeled sample (condition 1) and one Cy5-labeled sample (condition 2) were mixed with one-fourth of the Cy2-labeled pool and rehydration buffer was added to reach a volume of 350 µl. Dye swap experiments were performed for each culture condition.

For protein separation, samples were first loaded onto an 18 cm pH 4–7 IPG strip. IEF was conducted using the Ettan IPGphor system (GE Healthcare), as previously described (Weiss et al. 2009). The strips were equilibrated in SDS equilibration buffer (30 mM Tris–HCl pH 8.8, 6 M urea, 34.5% [v/v] glycerol, 2% [w/v] SDS, 0.01% [w/v] bromophenol blue) supplemented with 1% [w/v] DTT for 15 min and then with 2.5% [w/v] iodoacetamide for 15 min. SDS polyacrylamide gel electrophoresis was subsequently performed using 11.5% SDS gels, using the Ettan DAlt*six* system (GE Healthcare) with the following steps: 1 h at 60 mA, 80 V, 4 W and 1 h at 240 mA, 500 V, 52 W. Gels were stained with Colloidal Brilliant Blue or digitized using a Typhoon Scanner (GE Healthcare).

Differential protein expression analysis was performed as previously described (Bryan et al. 2009; Weiss et al. 2009). Spots were selected and identified by MALDI-TOF and Nano LC-MS/MS, and data analysis were performed with Mascot (Matrix Science Ltd.) as described previously (Bryan et al. 2009) against a *Rhizobium* sp. NT-26 protein database. All identifications were incorporated into the "InPact" proteomic database developed previously (http://inpact.u-strasbg.fr/~db/, last accessed April 30, 2013) (Bertin et al. 2008).

Transmission Electron Microscopy

Rhizobium sp. NT-26 or the *aioR* mutant were grown in MSM containing 0.04% yeast extract in the presence or absence of 8 mM As(III) for 24 h. A drop of culture was deposited onto Formvar-coated nickel grids and after cell decantation, the liquid excess was removed. Uranyl acetate 2% was added to negatively stain bacteria and flagella and these samples were dried. Grids were observed with a Hitachi H-600 transmission electron microscope (TEM) at 75 kV and photographed with a Hamamatsu ORCA-HR camera using the AMT software (Advanced Microscopy Techniques).

Results and Discussion

General Genome Features

Chromosome, Plasmids, and Genomic Plasticity

The *Rhizobium* sp. NT-26 genome includes a single 4.2 Mbp chromosome and two plasmids. The circular chromosome consists of 4,239,731 bp with 4,380 coding sequences, including 4,303 coding DNA sequences and 59 RNA genes, and representing 90.28% of the whole genome (fig. 1). Among these coding sequences (CDS), 34.40% are of unknown function.

The mean G + C content of the chromosome is 61.97% but its distribution is not homogenous (fig. 1), and the *Rhizobium* sp. NT-26 chromosome exhibits 65 regions of genomic plasticity (RGP, supplementary table S3, Supplementary Material online) (Vallenet et al. 2009) in comparison with that of *A. tumefaciens* 5A. The G + C content of these regions, their size (5–207 kb) and the codon adaptation index lower than the average are characteristic of genomic islands (GEI) (Juhas et al. 2009). Moreover, transposable elements and tRNA encoding genes are present in several of these regions, which further support the lateral transfer of these potential genomic islands (Daubin et al. 2003). Such genetic events are known to promote bacterial adaptation under environmental stresses by the acquisition of various capacities through horizontal gene transfer, an important mechanism of microbial genome evolution (Juhas et al. 2009). In agreement with this, more than 15 loci coding for metabolic functions that may improve the fitness of the strain to its environment were found among the 65 RGP identified in the genome of *Rhizobium* sp. NT-26. These include amino acids and carbon sources transport, inorganic carbon fixation, nitrogen metabolism, and sulfur oxidation (supplementary table S3, Supplementary Material online).

The genomes of bacteria in the *Agrobacterium* and *Rhizobium* genera are known to include several extrachromosomal replicons, which encode various functions required for the adaptation to specific niches (López-Guerrero et al. 2012). The *Rhizobium* sp. NT-26 genome comprises two plasmids, including a megaplasmid of 322,264 bp containing 367 coding sequences (CDS) (fig. 1). The presence and the size of the megaplasmid were confirmed experimentally by a modified Eckhardt gel electrophoresis method (supplementary fig. S1, Supplementary Material online). The second plasmid is 15,430 bp and more than half of its CDS encode proteins with unknown functions.

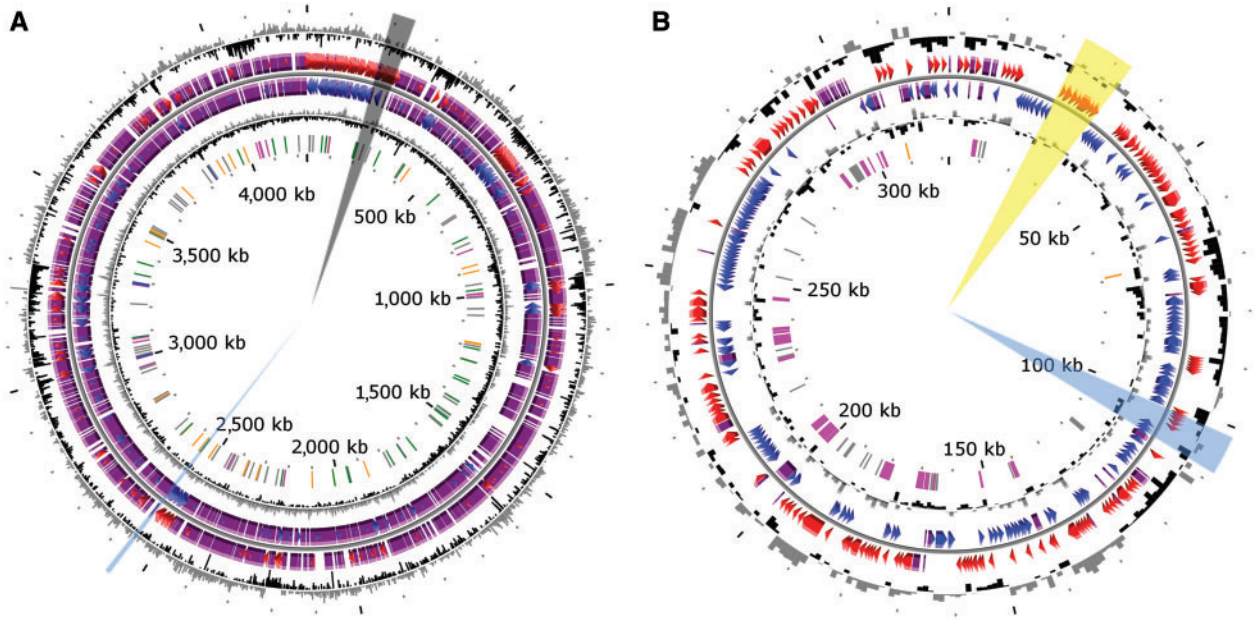


Fig. 1.—Circular representation of the *Rhizobium* sp. NT-26 genome. The chromosomal (*A*) and plasmidic (*B*) characteristics are 4,239-Mb long, 61.97% GC, 9 16S-23S-5S rRNA, 50 tRNA, and 4,294 CDSs; and 322-kb long, 60.19% GC, 0 16S-23S-5S rRNA, 0 tRNA, and 367 CDSs, respectively. From outside, circles display 1) the GC percent deviation in a 1,000 bp window (GC window – mean GC); 2) and 3) predicted CDSs transcribed in the clockwise and counterclockwise direction, respectively; red and blue colors correspond to validated annotations, orange to automatic annotation and purple to primary automatic annotation; 4) GC skew (G + C/G – C) in a 1,000 bp window; 5) rRNA are shown in blue, tRNA in green, miscRNA in orange, transposable elements in pink, and pseudogenes in gray. The regions with the genes coding for proteins involved in motility, reduction of arsenate or oxidation of arsenite are highlighted in black, blue, or yellow, respectively. The figure does not represent the p2 plasmid and the scale between the two genetic determinants is not respected (https://www.genoscope.cns.fr/agc/microscope/home/index.php, last accessed April 30, 2013).

The large plasmid harbors four different replication systems, in particular three *repABC* operons. The first one is duplicated. These two *repABC* operons are related to *repABC* of *Dinoroseobacter* (>90% identity) of the *Rhodobacteraceae,* a family of the *Rhodobacterales*. The third *repABC* operon is related to *repABC* of *Rhizobiales*. The fourth replication system is composed of a ParB/RepC replication system homolog of *A. tumefaciens* NCPPB925 plasmid origin of replication. Such a redundancy is not rare in *Rhizobiales*. For example, two replicons in *R. etli* CFN42, one in *R. leguminosarum* 3841 and one in *Ruegeria* sp. PR1b contain two *repABC* operons (Zhong et al. 2003; González et al. 2006; Young et al. 2006).

The smaller plasmid replication system is different from the canonical *repABC* replication system. It is constructed as the replication system described in pTAR of *A. vitis* (Gallie and Kado 1988), that is, the origin region carries a *repA*-like gene, a *parA* gene and a putative regulator locus coding for a putative segregation protein. Besides these replication transfer genes, the 15 kb plasmid harbors a toxin antitoxin system, which may explain its maintenance in *Rhizobium* sp. NT-26.

Plasmidic Adaptive Traits

In symbiotic bacteria, plasmids are known to play a role in their interaction with plants (López-Guerrero et al. 2012). In addition to multiple transposases and insertion sequences, the megaplasmid identified in *Rhizobium* sp. NT-26 encodes a putative type IV secretion system known to be involved in conjugal DNA transfer, including between bacteria and plants. Indeed, two complete *tra* clusters were found on the 322 kb plasmid: the first one is related to the type IV secretion system found in the *Rhizobium/Agrobacterium* genus, whereas the second one is related to the type IV secretion system of marine bacteria members of the *Rhodobacteraceae* family, that is, *Oceanibulbus indolifex* or *Ruegeria* sp. PR1b plasmid pSD25. However, canonical *nodABC* genes coding for proteins NodA (acetyl transferase), NodC (oligomerization of *N*-acetyl-glucosamine), and NodB (chitooligosaccharide deacatylase) that are required for the synthesis of the core structure of lipo-chitooligosaccharide (i.e., Nod factor) (Dénarié et al. 1996) were not identified in the genome of *Rhizobium* sp. NT-26. CDS displaying some similarities with other *nod* genes, that is, encoding enzymes that control specific substitutions on the chitooligosaccharide backbone, or the *fix* operon are present (supplementary table S4, Supplementary Material online), but these genes are also well conserved in nonsymbiotic prokaryotes. Nevertheless, it has been demonstrated that the nodulation of some *Fabaceae* by rhizobia occurs in the absence of the *nodABC* genes and lipochitooligosaccharidic Nod factors (Giraud et al. 2007). This indicates that other signaling strategies can trigger nodule organogenesis in some legumes. Nodulation assays on various *Fabaceae* plants were therefore performed as previously described (Gremaud and Harper 1989), but no nodules were observed at 3 weeks after inoculation or later at 4 weeks (data not shown).

Despite a lack of plant nodulation, root inoculation by *Rhizobium* sp. NT-26 suggested a potential phytobeneficial effect (fig. 2). Direct plant-growth promotion can be derived from phosphorus solubilization (Richardson et al. 2009), production of plant growth regulators (phytohormones) such as auxins, gibberellins, and cytokinins (Spaepen et al. 2009), NO production and/or by supplying biologically fixed nitrogen (Creus et al. 2005). Increasing the bioavailability of phosphate as micronutrient is mediated by bacterial phosphatase activity, and a phosphatase homolog, that is, NT26v4_0651, is present in the *Rhizobium* sp. NT-26 genome. Moreover, two main classes of dissimilatory nitrite reductase (Nir) involved in NO production exist among denitrifying bacteria: the heme-cytochrome cd_1 type encoded by *nirS* genes and the copper-containing type encoded by *nirK* genes (Zumft 1997). A *nirK* homolog but no *nirS* homolog was identified in the *Rhizobium* sp. NT-26 genome. Finally, no other homolog of classical phytobeneficial functions was identified when analyzing the genome, for example, phytohormone synthesis such as auxin by *ipdC/ppdC* or acetoin-2,3-butanediol by *budABC*, or nitrogen fixation by nitrogenase *nifHDK*.

Indirect plant growth-promoting mechanisms used by plant-growth-promoting rhizobacteria (PGPR) include induced systemic resistance, antibiotic protection against pathogens, reduction of iron availability in the rhizosphere by

Fig. 2.—Phytobeneficial effect of *Rhizobium* sp. NT-26 on *Phaseolus vulgaris*. Erlenmeyer flasks of *P. vulgaris* were inoculated with *Rhizobium* sp. NT-26 on the left, and with water on the right.

sequestration with siderophores, synthesis of fungal cell wall-lysing or lytic enzymes, and competition for nutrients and colonization sites with pathogens (Dobbelaere and Okon 2007). *Rhizobium* sp. NT-26 contains loci coding for polyketide synthases (NT26v4_3331, NT26v4_3332, and NT26v4_3333), which are involved in nonribosomal synthesis of antibiotics, or coding for proteins involved in siderophore transport (NT26v4_2008, NT26v4_4195, and NT26v4_4199). Taken together, these observations suggest that *Rhizobium* sp. NT-26 does not exert any direct interaction with plants but it may have an indirect role in plant growth and protection by its metabolic activities in the rhizosphere.

Unlike *H. arsenicoxydans* (Muller et al. 2007) and *T. arsenitoxydans* (Arsène-Ploetze et al. 2010), which metabolize and provide resistance to arsenic using proteins encoded by chromosomally borne genes, proteins involved in arsenic resistance in *Rhizobium* sp. NT-26 are encoded by *ars* genes present on both the chromosome and the megaplasmid. The *aio* genes involved in arsenite oxidation are present only on the megaplasmid (fig. 1*B*), as shown in *The. thermophilus* str. HB8. The *Rhizobium* sp. NT-26 *aio* cluster also contains genes coding for phosphate transport and molybdenum cofactor biosynthesis, as previously observed in other arsenite-oxidizing bacteria (Arsène-Ploetze et al. 2010; Bertin et al. 2011). In addition, like the metallo-resistant strain *Cupriavidus metallidurans* (Janssen et al. 2010), the *Rhizobium* sp. NT-26 megaplasmid contains numerous genes involved in resistance to heavy metals such as chromium, cadmium, and mercury. These observations suggest a loss of most plasmid-encoded functions known to be involved in bacteria–plant interactions and an acquisition of multiple genes allowing the organism to grow in its natural habitat, a goldmine known to contain toxic metals and metalloids.

The gene cluster coding for arsenite oxidase contains 5 *aio* genes in *Rhizobium* sp. NT-26. The survey of the nr database revealed the existence of numerous homologs of AioA, AioB, AioR, AioS, and AioX. Preliminary phylogenetic analyses of AioA homologs showed that the sequence from *Rhizobium* sp. NT-26 belongs to a well-supported clade of proteobacterial sequences corresponding to the groups I and II, which were recently described (Heinrich-Salmeron et al. 2011). Subsequent phylogenetic analyses revealed that the *Rhizobium* sp. NT-26 AioA branched among alphaproteobacterial sequences (group I), within a strongly supported clade composed of sequences from various *Agrobacterium* species and *Sinorhizobium* sp. M14 (Rhizobiaceae), from *Ochrobacterium tritici* (Brucellaceae), and uncultured organisms (bootstrap value [BV] = 98% and posterior probability [PP] = 1.00, supplementary fig. S2*A*, Supplementary Material online). Phylogenetic analyses of other Aio proteins showed similar branching patterns (supplementary fig. S2*B–E*, Supplementary Material online). This suggests that the five *aio* genes have co-evolved, which is not entirely surprising given that they are functionally related and clustered together when present in a genome.

A careful examination of the taxonomic distribution within the subgroups (supplementary fig. S2, Supplementary Material online) revealed that only 2 *Rhizobium/Agrobacterium* complete genomes contain the *aio* genes although nearly 30 genome sequences are available at NCBI. In addition, the relationships among the sequences were not always in agreement within the phylogeny of species, for example, the grouping of a member of *Brucellaceae* within *Rhizobiaceae*. This strongly suggests that horizontal gene transfers may have played a role in the spread of *aio* genes among these species. The alternative hypothesis, that is the presence of *aio* genes in the common ancestor of the *Rhizobium/Agrobacterium* group followed by multiple independent gene losses during the diversification of this lineage, appears less likely and does not explain the discrepancies among the Aio phylogenies and the taxonomy. On the contrary, these inconsistencies may be easily explained by the initial acquisition of the *aio* genes by one member of the *Rhizobium/Agrobacterium* group followed by a few horizontal gene transfers to related species or strains. Such transfers may have been favored by the colocation of *aio* genes on genomes and their location on plasmids in a few strains (e.g., pSinA in *Sinorhizobium* sp. M14 and pAt5A for *A. tumefaciens* 5A).

Taxonomic Relationship of *Rhizobium* sp. NT-26 with Other *Rhizobiaceae*

Rhizobium sp. NT-26 strain has been previously assigned to the *Rhizobium* genus on the basis of 16S RNA sequence (Santini et al. 2000), but its phylogenetic relationship with other *Rhizobiaceae* remains quite unclear. Therefore, the sequence of the *Rhizobium* sp. NT-26 chromosome has been compared with the sequences present in the "Prokaryotic Genome DataBase" (PkGDB) (Vallenet et al. 2006) and RefSeq (NCBI Reference Sequences) data banks. The highest synteny conservation was observed with the *A. tumefaciens* C58 circular chromosome, which is 67.85% of the CDS in this genome share synteny with the chromosome of *Rhizobium* sp. NT-26 and the average size of the syntons is 8.4 CDS. This gene order conservation is higher than that observed with *Rhizobium* spp. strains (6.9–7.4), suggesting a closer evolutionary relationship of strain *Rhizobium* sp. NT-26 with the *Agrobacterium* lineage. To determine more precisely the taxonomic position of *Rhizobium* sp. NT-26 among *Rhizobiaceae*, we compared its genome with a set of 23 other sequenced genomes of this family in a phylogenetic framework. We computed maximum-likelihood (ML) trees based on the concatenated alignments of sequences of either all-homologous genes that are common to and unique in all strains (822 "core" genes) or genes of ribosomal proteins (51 "ribosomal" genes). Intriguingly, both data sets yielded phylogenies that agree on all major splits in the taxon, but not on the position

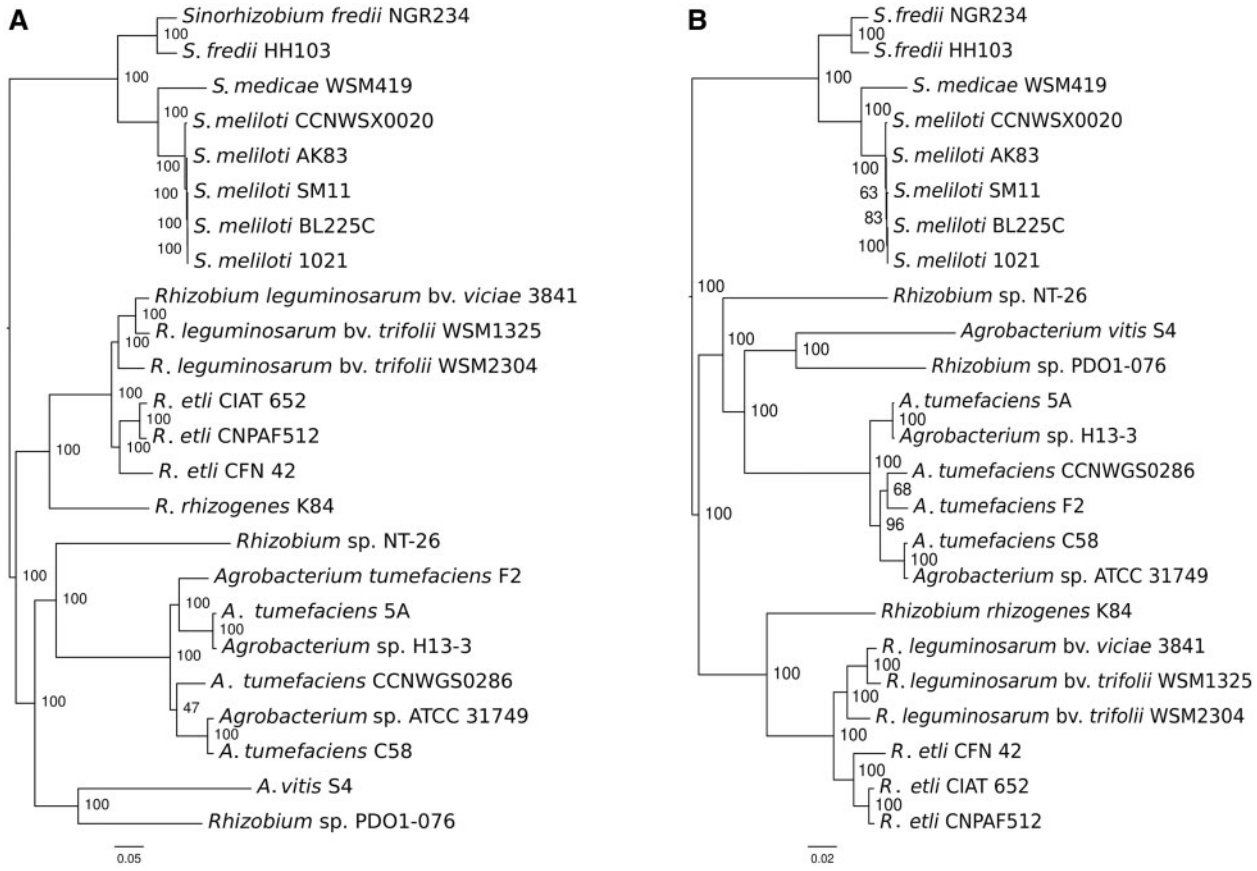


Fig. 3.—Phylogeny of *Rhizobium* sp. NT-26 among *Rhizobiaceae*. ML phylogenies of 24 *Rhizobiaceae* including strain NT-26 were built from concatenated alignments of (*A*) 822 core genes and (*B*) 51 ribosomal genes. Branch supports are percentage of the 200 and 1,000 bootstrap trees having the bipartition, respectively.

of *Rhizobium* sp. NT-26. According to core genes, this strain branched as an "in-group" of the *Agrobacterium* subgroup, being the brother clade of *A. tumefaciens* after the split with *A. vitis* (fig. 3*A*). Instead, according to ribosomal genes, strain NT-26 branched as an "out-group" of the *Agrobacterium* clade that encompasses *A. vitis* and *A. tumefaciens* (fig. 3*B*). In both cases, the conflicting bipartitions are well supported, and removing third codon positions in the alignment of core genes, because of the possible saturation of the substitution signal in non-housekeeping genes, did not change the observed pattern (supplementary methods S3, Supplementary Material online). This suggests that among the core gene set, which include the majority of the ribosomal gene set, different genes have different histories, causing the average history (core phylogeny) to be different from that of a subset (ribosomal phylogeny). This may have been caused by horizontal gene transfer to and from *A. tumefaciens*, *A. vitis*, *Rhizobium*, or other more phylogenetically distant taxon that would blur the signal for vertical inheritance.

We therefore computed individual phylogenies for all homologous families to determine what scenario each gene supported. On a set of 2,878 homologous gene family trees containing 3,537 *Rhizobium* sp. NT-26 genes, we searched for subtrees displaying the unambiguous patterns of either *Rhizobium* sp. NT-26 as a direct out-group of the *Agrobacterium* clade ("NT26outAgro") or as an in-group of the *Agrobacterium* clade and a brother group of *A. tumefaciens* ("NT26inAgro"). "NT26inAgro" was prevalent with 338 genes (268 considering only high branch support) versus 255 (146) for "NT26outAgro," and even though the amount of genes displaying such unambiguous patterns was relatively low, "NT26inAgro" was significantly more frequent (χ^2 test, P value $< 10^{-3}$). The location of those genes along the chromosome of *Rhizobium* sp. NT-26 showed no grouping with a particular pattern that could support a large-scale transfer event (Mann–Whitney–Wilcoxon test, P value >0.9; supplementary fig. S3, Supplementary Material online). The homogeneous dispersal of phylogenetic signatures rather suggests that numerous small-scale transfer events occurred, as observed in the case of frequent homologous recombination with partners of different taxa (Didelot et al. 2010). Alternatively, our observations may be the consequence of a

poor resolution of phylogenies. This may be due to the frequent artifacts in the phylogenetic reconstruction of the relationship of strain NT-26 to *Agrobacterium*, such as those caused by the long branch leading to *Rhizobium* sp. NT-26. In the future, more phylogenetic information might be provided by sampling strains branching at the base of the *Rhizobium/Agrobacterium* group.

Most of *Rhizobiaceae* contain a secondary chromosome or megaplasmids, generally referred to as chromids (Harrison et al. 2010), that are members of a same family of large replicons derived from a plasmid (Slater et al. 2009). Although *Rhizobium* sp. NT-26 genome contains a megaplasmid with chromid characteristics (Harrison et al. 2010), this megaplasmid show very limited homology with chromids of this family (supplementary table S5, Supplementary Material online). The absence of a typical *Rhizobiaceae* secondary replicon makes the genome structure of *Rhizobium* sp. NT-26 unusual when compared with other members of the family. Comparison of the homologous gene location in *Rhizobium* sp. NT-26 and related organisms may help with the understanding of its evolution history. With this aim, the closest homolog of each of its genes present in several strains of *Rhizobium* and *Agrobacterium* were mapped along the *Rhizobium* sp. NT-26 chromosome (supplementary fig. S4, Supplementary Material online). It appeared that the vast majority of strain NT-26 chromosomal genes map to the principal chromosome in *Rhizobium* and *A. vitis*, suggesting that the *Rhizobium* sp. NT-26 lineage has completely lost the secondary chromosome of the *Rhizobium/Agrobacterium* ancestor. The history of intragenomic translocations have been documented in this taxon (Slater et al. 2009) and locating *Rhizobium* sp. NT-26 genes whose homologs have migrated at a specific divergence time would help to date the age of the divergence of the *Rhizobium* sp. NT-26 lineage. In this respect, its chromosome possesses a large chromosomal fragment (spanning from 2.55 to 3.40 Mb), which is specifically present on the secondary (linear) chromosome in *A. tumefaciens* (supplementary fig. 3, Supplementary Material online), further supporting a divergence of the *Rhizobium* sp. NT-26 lineage predating the *A. tumefaciens* speciation and synapomorphic translocation events. Similarly, strain NT-26 appears to have conserved the majority of the genes that are specifically borne by the secondary chromosome of *R. rhizogenes* (119 homologs in *Rhizobium* sp. NT-26 over 129 specific translocated genes). If strain NT-26 belonged to the *Rhizobium* lineage, the majority of these genes would probably have been lost with this whole chromid, although we cannot rule out potential translocations of those genes back to the main chromosome along with the chromid loss. Taken together, and even though our current data do not allow a more accurate classification, our observations support the inclusion of *Rhizobium* sp. NT-26 in the *Agrobacterium* subgroup. Nevertheless, according to the current nomenclature (Young et al. 2001), *Rhizobium* is still a valid genus name for strain NT-26.

Finally, although chromosomes are mainly dedicated to housekeeping functions, chromids carry genes involved in specific ecological functions, that is, legume symbiosis enabled by symbiotic plasmids in *Rhizobium* and *Sinorhizobium* (Harrison et al. 2010) and plant-related functions in *A. tumefaciens* C58 (Lassalle et al. 2011). All these functions relate to the interactions inside the rhizosphere and soil that represent the canonical habitat of *Rhizobiaceae*. The loss by *Rhizobium* sp. NT-26 of this large replicon housing rhizosphere-associated functions may be related to the drastic shift in environment the lineage has experienced. Indeed (discussed earlier), this loss coincides with the gain of genes enabling resistance to heavy metals, and arsenic and sulfur metabolisms, both traits with a potentially great adaptive value in sustaining life on an arsenopyrite (FeAsS)-containing rock.

Functional Approaches to Investigate Arsenic Metabolism and Resistance

Proteomic and Transcriptomic Profiling

Genomic tools have been used to study the bacterial response to arsenic mainly in arsenite-oxidizing *Betaproteobacteria* such as *H. arsenicoxydans*, a chemoorganotroph (Carapito et al. 2006; Muller et al. 2007; Weiss et al. 2009), and *T. arsenitoxydans*, a chemolithoautotroph (Bryan et al. 2009; Arsène-Ploetze et al. 2010). Arsenic metabolism was investigated in the chemolithoautotrophic *Alphaprotebacterium Rhizobium* sp. NT-26 by two complementary approaches: protein and RNA profiling using 2D gel electrophoresis and DNA microarrays, respectively. The comparisons of expression were done on proteins and RNA isolated from strain NT-26 grown heterotrophically with and without arsenite. The main results are summarized in table 1 and a complete list of the data is presented in supplementary table S6, Supplementary Material online.

The 2D gel proteomic profile of *Rhizobium* sp. NT-26 was quite similar to those previously obtained for *H. arsenicoxydans* (Carapito et al. 2006) and *T. arsenitoxydans* (Bryan et al. 2009), which are also neutrophilic bacteria. Sixty-three spots showed a significant difference in their accumulation pattern in strain NT-26 grown with and without As(III). Their analysis by mass spectrometry led to the identification of 141 proteins (supplementary table S6a, Supplementary Material online), including arsenite oxidase, which was identified for the first time on 2D gels. Like membrane proteins, such periplasmic proteins are often eliminated with cell debris before their solubilization during sample preparation. Proteins up- or downregulated with a fold-change ranging from +39.6 to −5.8 when *Rhizobium* sp. NT-26 was grown in the presence of As(III) had a molecular mass ranging from 15 to 109 kDa and a pI value from 4.2 to 7.9. Among them, 24% were involved in cell envelope and cellular processes, 12% in

Table 1

Major Arsenic-Regulated Functional Categories Identified in Transcriptomics and Proteomics Experiments

Functional Category	MaGe ID	Gene	Function	FC	
				RNA[a]	Protein[a]
Oxidative stress	NT26v4_0103	*rpoN*	RNA polymerase σ^{54} factor	1.32	
	NT26v4_0389	*katA*	Catalase A	1.37	
	NT26v4_0773	*ohr*	Organic hyperoxide resistance	1.41	
	NT26v4_0799	*sodB*	Superoxide dismutase		3
Carbon metabolism	NT26v4_0674	*cbbF*	Fructose-1.6-bisphosphatase	1.3	
	NT26v4_0670	*cbbL*	RuBisCo large subunit	1.37	
	NT26v4_2684	*cbbT*	Transketolase		3.4
	NT26v4_0667	*cbbE*	Ribulose-phosphate 3-epimerase	1.35	
Nitrogen metabolism	NT26v4_3645	*norQ*	Putative NorD protein	1.52	
	NT26v4_3643	*norC*	Nitric oxide reductase subunit C	1.42	
	NT26v4_3641	*norE*	Involved in nitric oxide reduction	1.60	
	NT26v4_3654	*nirV*	Involved in nitrite reduction	1.51	
	NT26v4_3653	*nirK*	Cu-containing nitrite reductase		39.6
Arsenic metabolism	NT26v4_p10030	*aioA*	Arsenite oxidase large subunit	3.89	10
	NT26v4_p10029	*aioB*	Arsenite oxidase small subunit	4.27	
	NT26v4_p10118	*arsC1b*	Arsenate reductase ArsC		22
	NT26v4_p10122	*arsH1*	Arsenical resistance protein		2.9
Sulfur metabolism	NT26v4_2623	*soxG*	Sulfur oxidation protein	−1.37	
	NT26v4_2619	*soxV*	Sulfur oxidation protein	−1.39	
	NT26v4_2618	*soxW*	Thioredoxin	−1.52	
	NT26v4_2617	*soxX*	Sulfur oxidizing protein	−1.68	
	NT26v4_2616	*soxY*	Sulfur oxidation protein	−1.45	
	NT26v4_2615	*soxZ*	Sulfur oxidation protein	−1.51	
	NT26v4_2882	*cysT*	Sulfate/thiosulfate transport protein	−1.72	
Phosphate metabolism	NT26v4_1226	*phoE1*	Phosphonate ABC transporter	1.45	
	NT26v4_0079	*phoR*	Phosphate regulon kinase	1.32	
	NT26v4_p10016	*phoE2*	Phosphonate ABC transporter subunit	1.61	
	NT26v4_p10017	*phoT2*	Phosphonate ABC transporter subunit	1.42	
	NT26v4_p10024	*pstS2*	High-affinity phosphate transporter		3.7
Motility/biofilm	NT26v4_0204	*fliF*	Flagellar M-ring protein	1.44	
	NT26v4_0227	*flaA*	Flagellin A		7.2
	NT26v4_0228	*fla*	Flagellin		9.7
	NT26v4_0655	*qseB*	Quorum sensing regulator QseB	1.39	
	NT26v4_2748	*noeJ*	Mannose-1-P guanylyltransferase	1.41	
	NT26v4_1615	*kdsA*	KDO 8-P synthase	1.40	
Plant/bacteria interactions	NT26v4_1705	*cgmA*	Beta-1.2-glucan modification protein	1.7	
	NT26v4_p10302	*avhB10*	Type IV system transglycosylase	1.41	

Note.—Induced and repressed functions are shown in blue and black, respectively. No value is indicated in the FC column if the gene is not statistically differentially expressed in transcriptomics or if the protein has not been identified in proteomics. Complete data are presented in supplementary table S6, Supplementary Material online.

[a]Fold-change observed in transcriptomics and proteomics data, respectively.

transport and binding proteins, 11% in information and regulation pathways, 42% in metabolism, 1% in transcription, and 10% were of unknown function.

The second approach used whole genome microarrays to perform a differential expression profiling experiment. Under As(III) stress, the transcript level of 199 genes, that is 4.5% of the whole genome, showed an increase of up to more than four times with a *P* value ≤ 0.05. At the same time, the expression of 416 genes, that is 9.5% of the whole genome, decreased by up to more than three times (supplementary table S6c, Supplementary Material online).

General Response to Arsenic Stress

Several proteins involved in arsenic resistance were shown to be accumulated on 2D gels when the organism was grown in the presence of As(III), for example, an ArsH1 NADPH-dependent FMN reductase and an ArsC1 arsenate

reductase (table 1) with fold changes of 2.9 and 22, respectively. These two proteins are encoded by an *ars* operon located on the megaplasmid, which also contains genes coding for an ArsB efflux pump and an ArsR regulator. Moreover, a second operon located on the chromosome contains an *arsA* gene coding for an ATPase associated with an ArsB arsenite efflux pump. ArsA enables the strain to increase arsenic resistance by ATP-dependent extrusion of the metalloid out of the cell (Branco et al. 2008).

Microarray experiments showed that genes encoding the two arsenite oxidase subunits, that is, the small subunit, AioB, which contains the Rieske 2Fe-2S cluster and the catalytic subunit, AioA, which contains a molybdopterin guanine dinucleotide at the active site and a 3Fe–4S cluster (Santini and vanden Hoven 2004), were about 4-fold induced in the presence of arsenite (table 1). In contrast, the expression of two genes located downstream of the *aioBA* operon, that is, *cytC* encoding the periplasmic cytochrome c_{552}, which can serve as an electron acceptor to the arsenite oxidase (Santini et al. 2007) and *moeA1* encoding a molybdenum cofactor biosynthesis gene, was not significantly affected under arsenite stress (supplementary table S6c, Supplementary Material online). These observations are further supported by proteomic experiments showing that, among these proteins, AioA was found to be preferentially accumulated in the presence of As(III). All these results are consistent with previous data (Santini et al. 2007), which suggests that the expression of *aioBA* genes is induced by arsenite while genes located downstream are constitutively expressed even though they may have a role in arsenic metabolism.

Located upstream of *aioBA* are two regulatory genes, *aioS* and *aioR*, which encode a sensor histidine kinase and a response regulator, respectively (Sardiwal et al. 2010). Both proteins have been shown to be required for the transcriptional regulation of the *aioBA* genes (Koechler et al. 2010; Sardiwal et al. 2010). Moreover, it has been demonstrated that the expression of the *aioBA* genes requires the RpoN alternative sigma factor (σ^{54}) in *H. arsenicoxydans* (Koechler et al. 2010). Similarly, a role for RpoN in arsenite oxidation has been recently highlighted in *A. tumefaciens* 5A (Kang et al. 2012). In this respect, a putative σ^{54}-dependent promoter region has been detected upstream of the *aioB* gene in *Rhizobium* sp. NT-26 (Santini et al. 2007), suggesting that it is also involved in the expression of the *aioBA* operon in strain NT-26 (Sardiwal et al. 2010). This hypothesis is supported by our transcriptomic data, which revealed an induced expression of *rpoN* in *Rhizobium* sp. NT-26 when it was grown in the presence of As(III) (table 1), in contrast to the constitutive expression recently observed in *A. tumefaciens* 5A (Kang et al. 2012). Similarly, microarray and 2D-gel data showed a 2-fold increase in the expression of genes coding for general chaperones, that is, DnaK and GroEL, in the presence of arsenite (supplementary table S6a and c, Supplementary Material online), which is in agreement with the role played by proteins of the heat-shock family in As(III) oxidation in *H. arsenicoxydans* (Koechler et al. 2010).

Rhizobium sp. NT-26 also tolerates arsenate concentration greater than 0.5 M (Clarke A. and Santini J.M., unpublished data), suggesting the existence of an alternative mode of resistance. The first one is an Ars-type arsenic resistance system, components of which were found to be upregulated when the strain was grown with arsenite (discussed earlier) and the second is the presence of a specific phosphate transport system which is thought to limit arsenate entry into the cell (Weiss et al. 2009). Indeed, in *Rhizobium* sp. NT-26, the arsenic genomic island contains a *pst* operon in the vicinity of the *aio* operon. The *pst* operon encodes proteins implicated in the specific transport of phosphate into the cell to maintain a sufficient level of this ion despite the presence of arsenate, a structural analog of phosphate (Muller et al. 2007; Cleiss-Arnold et al. 2010). PstS2, a periplasmic protein involved in phosphate transport and encoded by this operon, had a 3.7-fold increase in expression when strain NT-26 was grown in the presence of As(III) (table 1). The *pst* operon is regulated by *phoR*, which encodes a membrane-associated protein kinase that phosphorylates PhoB in response to environmental signals. Indeed, microarray data showed that the expression of *phoR* was also upregulated in the presence of As(III) (table 1). Moreover, the PhoR protein may be involved in biofilm formation as *phoB* overexpression has been shown to increase biofilm formation in *A. tumefaciens* (Danhorn et al. 2004). This is supported by the presence in the vicinity of the *pho* chromosomal operon of a cluster of genes involved in EPS biosynthesis (NT26v4_1233–NT26v4_1263).

Arsenic is known to induce oxidative stress by generating free radicals (Bernstam and Nriagu 2000). An induction of genes involved in the resistance to such a stress has been previously observed in *Pseudomonas aeruginosa* and in *H. arsenicoxydans* under arsenite exposure (Parvatiyar et al. 2005; Weiss et al. 2009; Cleiss-Arnold et al. 2010). In *Rhizobium* sp. NT-26, an increase in *katA* mRNA, which encodes a catalase involved in the protection against oxidative stress by scavenging endogenously produced H_2O_2, was observed in microarray experiments (table 1). Similarly, the expression of *ohr*, which promotes bacterial resistance to hydroperoxide, was also up-regulated in *Rhizobium* NT-26 (table 1). Finally, results of the proteomic experiments showed a 3-fold increase in the SodB superoxide dismutase accumulation when strain NT-26 was grown in the presence of As(III). These observations further support the strong link, which exists in bacteria between arsenic response and protection against oxidative stress (Bertin et al. 2012).

Carbon, Nitrogen, and Energy Metabolism

Rhizobium sp. NT-26 is able to use various carbon or electron sources for growth. Indeed, multiple carbohydrates such as acetate, succinate, fumarate, lactate, glucose, fructose,

xylose, and galactose are potential carbon sources for this bacterium (Santini et al. 2000). Alternatively, *Rhizobium* sp. NT-26 is able to grow chemolithoautotrophically in the presence of bicarbonate as a carbon source, oxygen as an electron acceptor and arsenite as an electron donor (Santini et al. 2000). Transcriptomics and proteomics experiments revealed that several genes and proteins involved in the fixation of CO_2 via the Calvin cycle were upregulated with a fold change ranging from 1.3 to 3.4 when strain NT-26 was grown in the presence of As(III) (table 1). This is in agreement with the proteomics results obtained in *T. arsenivorans,* where an accumulation of the ribulose-1,5-biphosphate carboxylase/oxygenase large subunit and of the fructose-1,6-biphosphate has been observed when the organism was grown in the presence of As(III) (Bryan et al. 2009). Both strains may thus improve their capacity to fix CO_2 when arsenite is present.

In addition, our microarrays data showed that the expression of *nirV* encoding a protein involved in nitrite reduction was induced. Moreover, 2D-gel data showed that NirK, which also participates in the reduction of nitrite to nitric oxide, was 39.6 times more accumulated when strain NT-26 was grown in the presence of arsenite. These experiments also showed an induction of several genes of the *norEFCBQD* nitric oxide reductase gene cluster when *Rhizobium* sp. NT-26 was grown in the presence of As(III) (table 1). These genes encode proteins that catalyze the reduction of nitric oxide to nitrous oxide, that is, *norQ, norE* and *norC* coding for a protein involved in nitric oxide reduction, a nitric oxide reductase activating protein, and the small subunit of the nitric oxide reductase, respectively. This suggests that the chemolithoautotrophic bacterium *Rhizobium* sp. NT-26 may fix CO_2 and couple nitrite reduction with As(III) oxidation. However, as no growth was observed with arsenite on either nitrate or nitrite, arsenite oxidation using nitrite as electron acceptor in autotrophic conditions seems unable to support sufficient energy (ATP) generation to sustain growth.

Rhizobium sp. NT-26 has been shown to grow with hydrogen sulfide, elemental sulfur, and thiosulfate (Santini J.M., unpublished data). In agreement with these observations, a *sox* cluster implicated in the oxidation of thiosulfate is present in the *Rhizobium* sp. NT-26 genome. Nevertheless, many genes involved in sulfur metabolism, that is, *soxGVWXYZ* and *cysT* were downregulated by up to 2-fold when the strain was grown in the presence of arsenite (table 1). Our results therefore suggest that, even though *Rhizobium* sp. NT-26 may be able to grow by using sulfur as an electron donor, the strain represses sulfur oxidation when grown in the presence of As(III). One hypothesis may be that, in such a case, the strain expresses a repressor of the *sox* genes. Their products serve for the oxidation of thiosulfate to sulfate and the reaction intermediates, that is, sulfite, sulfide, and hydrogen sulfide, have been shown to inhibit arsenite oxidase activity (Lieutaud et al. 2010).

Physiological and Genetic Approaches: Flagellar Motility and Biofilm Formation

Flagellum Cascade Features

Rhizobium sp. NT-26 is motile by the means of two subterminal flagella (Santini et al. 2000). Genes involved in their biosynthesis are organized in a large chromosomal cluster of 55 genes showing a perfect synteny with those of *S. meliloti*. In this flagellar regulon, *visN* and *visR* form part of the master operon and encode the proteins forming the VisNR heterodimer that acts as a global transcriptional regulator. This master regulator activates the expression of genes located in the cascade that encode flagella, motor, and chemotaxis proteins (Sourjik et al. 2000). In *Rhizobium* sp. NT-26, microarray data showed that the expression of *fliF*, coding for the flagellum M-ring protein, was induced when the organism was grown in the presence of As(III) (table 1). Furthermore, proteomic data showed a 9.7- and 7.2-fold-increase in the accumulation of flagellin proteins Fla and FlaA, respectively (table 1). The expression of *qseB* was also induced when *Rhizobium* sp. NT-26 was grown in the presence of As(III). QseB has been shown to participate in the flagellum and motility bacterial regulatory network via a quorum-sensing mechanism. Indeed, in *E. coli*, *qseBC* expression enhances the transcription of flagellar genes in response to the autoinducer by a direct binding of QseB to the *flhDC* master operon promoter (Clarke and Sperandio 2005). Finally, microarray data showed the induction of genes possibly involved in the synthesis of an exopolysaccharide matrix in *Rhizobium* sp. NT-26, that is, *noeJ*, coding for a mannose-1-phosphate guanylyltransferase and *kdsA*, coding for a 2-dehydro-3-deoxyphosphooctonate aldolase. These observations suggest that As(III) has an impact on flagellum synthesis, that is to say motility, and biofilm formation in *Rhizobium* sp. NT-26, as observed in *H. arsenicoxydans* (Muller et al. 2007; Marchal et al. 2010). To test this hypothesis, swarming assays were performed on 0.3% agar plates. The presence of As(III) was shown to increase the swarming ring by up to 2-fold in the presence of 8 mM As(III) (fig. 4*A*). Remarkably, cell observation under a TEM revealed that flagellum biosynthesis occurred immediately in the presence of 8 mM As(III) while more than two days were needed to observe flagella in the absence of arsenite (fig. 4*B* and *C*), providing evidence that arsenite promotes motility in *Rhizobium* sp. NT-26. In addition, a two-fold reduction in biofilm formation was observed in the first 24 h of growth in the presence of As(III) (fig. 4*D*), which suggests a preferential development as motile planktonic cells rather than as unflagellated sessile cells as in *H. arsenicoxydans* (Marchal et al. 2010).

Random Mutagenesis

The *Rhizobium* sp. NT-26 genome organization suggests that motility and arsenite oxidation depend on genes located on its

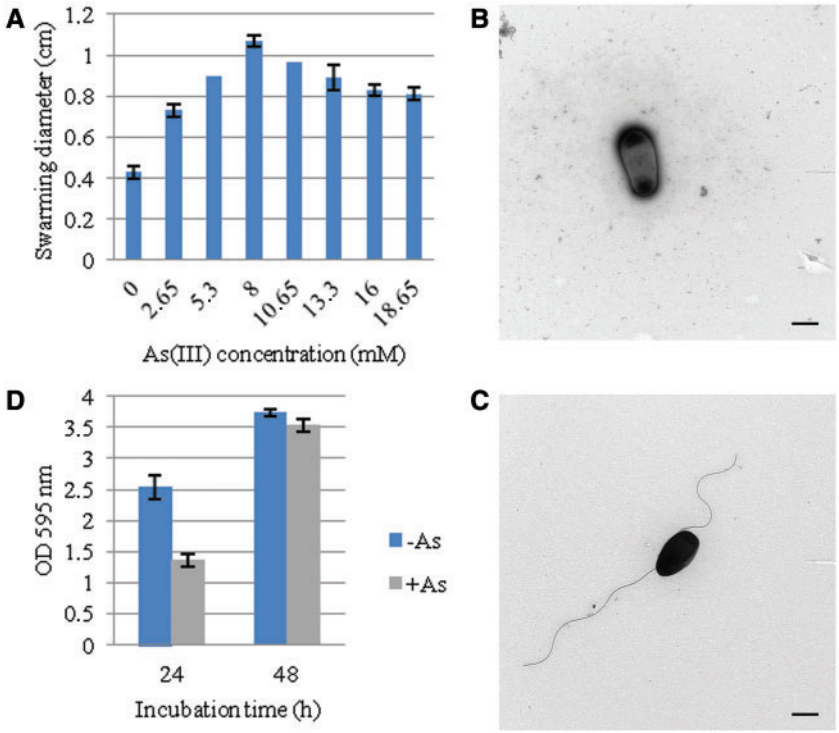


Fig. 4.—Motility phenotype of *Rhizobium* sp. NT-26 grown at different concentration of arsenite. (*A*) Swarming diameter measured after 48 h in MSM containing 0.04% yeast extract and supplemented by different concentrations of As(III). Results are the mean values of three independent experiments. (*B*) and (*C*) TEM observations of *Rhizobium* sp. NT-26 at 24 h of culture, without or with 8 mM As(III), respectively. The scale bar corresponds to 500 nm and the pictures are representative of 10 pictures. (*D*) Biofilm formation by strain NT-26, without or with 8 mM As(III) visualized by the crystal violet method. Results are the mean values of 24 replicates.

chromosome and on its megaplasmid, respectively. With the aim to analyze the possible link between these physiological processes, a mutant library was constructed by random transposon mutagenesis (Tang et al. 1999). The motility of 6,000 kanamycin-resistant transposition derivatives was tested on semisolid medium, which led to the isolation of 22 motility-deficient mutants. The mutations that resulted in a loss of motility were identified by sequencing the mini-Tn*5* transposon insertion sites. Fourteen mutations were shown to directly disrupt motility genes (table 2), and the proteins encoded by these genes are either structural or regulatory components of the flagellum cascade, that is, 5 Flg proteins (FlgE, FlgF, FlgG, FlgI, and FlgL), 5 Fli proteins (FliF, FliK, Flip, and FliR), 1 Flh protein (FlhA), and 1 Vis protein (VisR). No flagellin-defective mutant was obtained, which may be explained by the presence of four different flagellin-encoding genes on the chromosome.

Similarly, six mutations resulting in a lack of arsenite oxidation as compared with the wild-type and motility mutants were obtained after screening 6,000 kanamycin-resistant clones with the silver nitrate method (Muller et al. 2007) (table 2). First, two mutations were identified in the arsenite oxidase genes, that is, *aioA* and *aioB*. One mutation was also shown to affect *aioR,* which encodes the regulatory protein of the AioRS two-component system. No oxidation of As(III) to As(V) was detected by HPLC-ICP-AES (Muller et al. 2007) in this mutant, as compared with complete As(III) oxidation determined in the motility mutant deficient in the flagellum master regulator VisR. A fourth mutation was located in the *moeB* gene involved in the synthesis of the molybdopterin cofactor required for arsenite oxidase activity. Finally, the inactivation of the *aioX* gene, which is located upstream of *aioSR*, also resulted in a loss of arsenite oxidation in *Rhizobium* sp. NT-26. In *A. tumefaciens* 5A, the periplasmic AioX has been recently shown to be involved in the regulation of As(III) oxidation (Liu et al. 2012).

To determine the link between As(III) oxidation and colonization properties in *Rhizobium* sp. NT-26, the ability of various mutants to move and to form a biofilm was evaluated in the presence of arsenite (fig. 5). Mutations in flagellar genes resulted in a loss of motility and in a decrease in biofilm formation. Indeed, all the mutants we tested were nonmotile (fig. 5*A*) and lost between 12% and 45% of their ability to form a biofilm when compared with the wild-type strain (fig. 5*B*). This observation demonstrates that, although

Table 2

Rhizobium sp. NT-26 Mutants Isolated on the Basis of a Loss of Motility or Arsenite Oxidation

Mutant	MaGe ID[a]	Gene	Function	Gene Location[b]	Insertion[c]
Motility					
2B5	NT26v4_0222	*flgI*	Flagellar P-ring protein precursor	221267–222391	$258_{I\text{-}O}$
2D6	NT26v4_0245	*flhA*	Flagellar biosynthesis protein	243262–245349	$38_{I\text{-}O}$
4H7	NT26v4_0204	*fliF*	Flagellar M-ring protein	206351–208027	$477_{I\text{-}O}$
5B7	NT26v4_0226	*fliP*	Flagellar biosynthetic protein	224272–225009	$109_{O\text{-}I}$
6E4	NT26v4_0220	*flgG*	Flagellar basal-body rod protein	219981–220769	$182_{I\text{-}O}$
10A11	NT26v4_0248		Putative FlgJ-like protein	246658–247212	
10G2	NT26v4_2965		Conserved hypothetical protein	2880029–2881285	$44_{O\text{-}I}$
16B6	NT26v4_0245	*flhA*	Flagellar biosynthesis protein	243262–245349	$55_{I\text{-}O}$
16B7	NT26v4_0240	*flgL*	Flagellar hook-associated protein	240428–241549	$11_{I\text{-}O}$
18D11	NT26v4_0238	*flgE*	Flagellar hook protein	237398–238930	$363_{I\text{-}O}$
20E7	NT26v4_0246	*fliR*	Flagellar biosynthetic protein	245378–246130	$110_{O\text{-}I}$
23E5	NT26v4_0214	*flgF*	Flagellar basal-body rod protein	216054–216788	$44_{O\text{-}I}$
29G6	NT26v4_0247		Putative FliR/FliJ-like chaperone	246137–246553	$91_{I\text{-}O}$
35A8	NT26v4_2314		Putative two-component sensor histidine kinase	2263060–2264472	$113_{O\text{-}I}$
37C12	NT26v4_2314		Putative two-component sensor histidine kinase	2263060–2264472	$48_{O\text{-}I}$
37H1	NT26v4_2267		Putative ATP-dependent hydrolase protein	2217585–2219546	$337_{O\text{-}I}$
38B4	NT26v4_0206	*visR*	Master transcriptional regulator of flagellar regulon	209070–209798	$53_{O\text{-}I}$
38G3	NT26v4_0204	*fliF*	Flagellar M-ring protein	206351–208027	$235_{I\text{-}O}$
39G12	NT26v4_2671		Conserved protein of unknown function	2586119–2587117	$94_{O\text{-}I}$
40E5	NT26v4_0234	*fliK*	Flagellar hook-length regulator	234324–235835	$83_{I\text{-}O}$
50E11	NT26v4_0250		Conserved integral membrane protein of unknown function	247603–248142	$92_{O\text{-}I}$
61C2	NT26v4_3970		Conserved exported protein of unknown function	3918374–3918853	$196_{I\text{-}O}$
Arsenite oxidation					
8G1	NT26v4_p10026	*aioX*	Putative periplasmic phosphite-binding-like protein precursor; PtxB-like protein	23892–24806	$275_{I\text{-}O}$
11B3	NT26v4_4048	*moeB*	Putative molybdopterin biosynthesis protein MoeB	3998958–3999725	$245_{O\text{-}I}$
24B7	NT26v4_p10028	*aioR*	Two-component response regulator	26262–27584	$65_{I\text{-}O}$
37C3	NT26v4_p10026	*aioX*	Putative periplasmic phosphite-binding-like protein precursor; PtxB-like protein	23892–24806	$191_{I\text{-}O}$
55H7	NT26v4_p10029	*aioB*	Arsenite oxidase small subunit	27721–28248	
60E6	NT26v4_p10030	*aioA*	Arsenite oxidase large subunit	28261–30798	$818_{I\text{-}O}$

[a]Identification number of the gene in the MaGe interface.
[b]Position of the corresponding gene on the chromosome or the plasmid.
[c]Position of the codon immediately upstream of the transposon insertion site. Subscripts indicate the orientation of the insertion.

Rhizobium sp. NT-26 has a preferential motile life style in the presence of arsenite, flagella have a role as adhesive appendages in the first steps of biofilm formation, which has been shown previously in other studies (Kirov et al. 2004; Nejidat et al. 2008). This result is in agreement with those obtained with *H. arsenicoxydans*, where mutations resulting in nonfunctional flagella led to a more rapid adhesion as compared with the wild-type. Finally, *aioA* and *aioR* mutants were less motile and formed 30% and 45%, respectively, less biofilm than the wild-type (fig. 5), further supporting the role of motility and flagella in biofilm formation.

Regulation of Flagella Synthesis by AioR

The "omics" data showed that flagellar proteins and genes were upregulated when strain NT-26 was grown in the presence of arsenite (table 1). Remarkably, both *aioA* and *aioR* mutations resulted in a moderate reduction in motility (fig. 5*A*). This can be explained by the reduction in energy available to the cells as they are unable to metabolize arsenite (Santini et al. 2000). In addition, TEM observations of the *aioR* mutant, affected in the two-component signal transduction system, revealed the presence of flagella in the early log phase of growth even in the absence of As(III), which suggests that AioR may be involved in the repression of motility when no arsenite is present (fig. 6). AioR may thus interact, directly or indirectly, with components of the flagellar cascade.

To identify possible AioR-binding sites in the *Rhizobium* sp. NT-26 genome, multiple sequence alignments of all *aioBA* regulatory sequences available in databases were performed with fuzznuc (Rice et al. 2000). This enabled us to suggest the possible existence of two AioR putative binding sites upstream of the *aioBA* transcriptional start site, that is, GT[CT]CGN(6)CG[GA]AC in the *Rhizobiales* strains and GTTNCN(6)GNAAC in the *Burkholderiales*

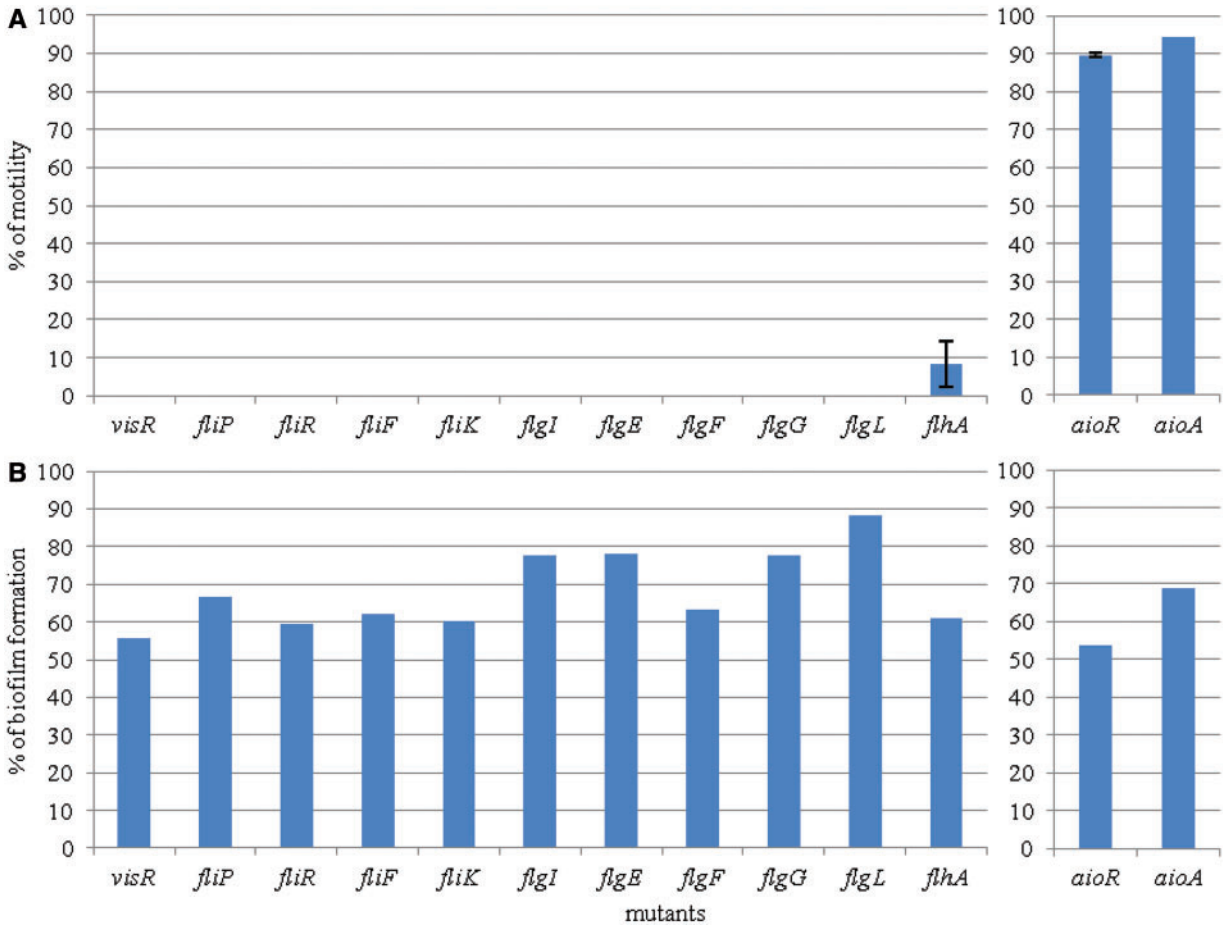


FIG. 5.—Percentage of motility and biofilm formation in various mutants as compared with *Rhizobium* sp. NT-26 wild-type strain. The left and the right panels show the results obtained in mutants affected in motility and As(III) oxidation, respectively. (*A*) % of swarming motility measured after 24 h. Results are the mean values calculated from the % of three independent experiments. In each experiment, mutant and wild-type strains were tested in triplicates. (*B*) % of biofilm formation visualized by crystal violet coloration. Results are the % calculated from the mean values of six replicates for each strain.

strains. In contrast, strains lacking the two-component system *aioSR* operon did not harbor any of these putative AioR-binding sites, which further supports a role for these motifs in the regulation of *aioBA* operon expression by AioR. Although the GT[CT]CGN(6)CG[GA]AC putative binding site was found at 49 locations on the *Rhizobium* sp. NT-26 chromosome, it is only in the upstream region of the *aioBA* operon that this motif was associated with the -12/-24 σ^{54}-dependent promoter sequence needed for the RpoN-dependent transcription initiation of arsenite oxidase genes (Koechler et al. 2010). Moreover, no clear connection could be observed between those putative binding sites and motility-related genes. Nevertheless, a search of the whole genome of strain NT-26 with a relaxed version of the pattern allowing any nucleotide at its degenerated positions, GTNCGN(6)CGNAC, yielded 39 more hits than with the canonical pattern. This low number of new hits suggests that the presence of this signature is not due to chance and this sequence may therefore have a potential regulatory role. None of the new hits was associated with a RpoN motif site although one hit was found within the coding sequence of the flagellar master regulatory gene *visN*. Therefore, although we cannot rule out an indirect effect of AioR by controlling another regulatory protein, we can hypothesize that the binding to this mildly degenerate motif of unphosphorylated AioR in the absence of arsenic would result in a *visN* repression and a delayed motility. Such a transcriptional repression via binding of the coding sequence of target genes has already been observed for other regulatory proteins and two-component system regulators, for example, OxyR (Zheng et al. 2001) and PrrA (Eraso et al. 2008). Taken together, our results demonstrate the importance of arsenite oxidation in the behavioral response of *Rhizobium* sp. NT-26, suggesting that arsenic metabolism enhances the ability of the organism to explore and colonize its environment.

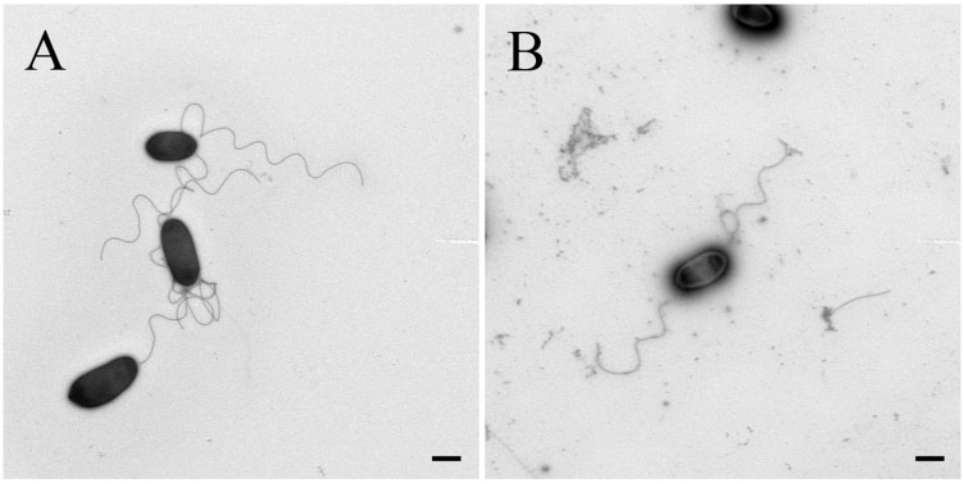


Fig. 6.—TEM observations of the *aioR* mutant. The *aioR* mutant was cultivated 24 h (*A*) in the absence of As(III) and (*B*) in the presence of 8 mM As(III). Pictures are representative of 10 photographs. The scale bar corresponds to 500 nm.

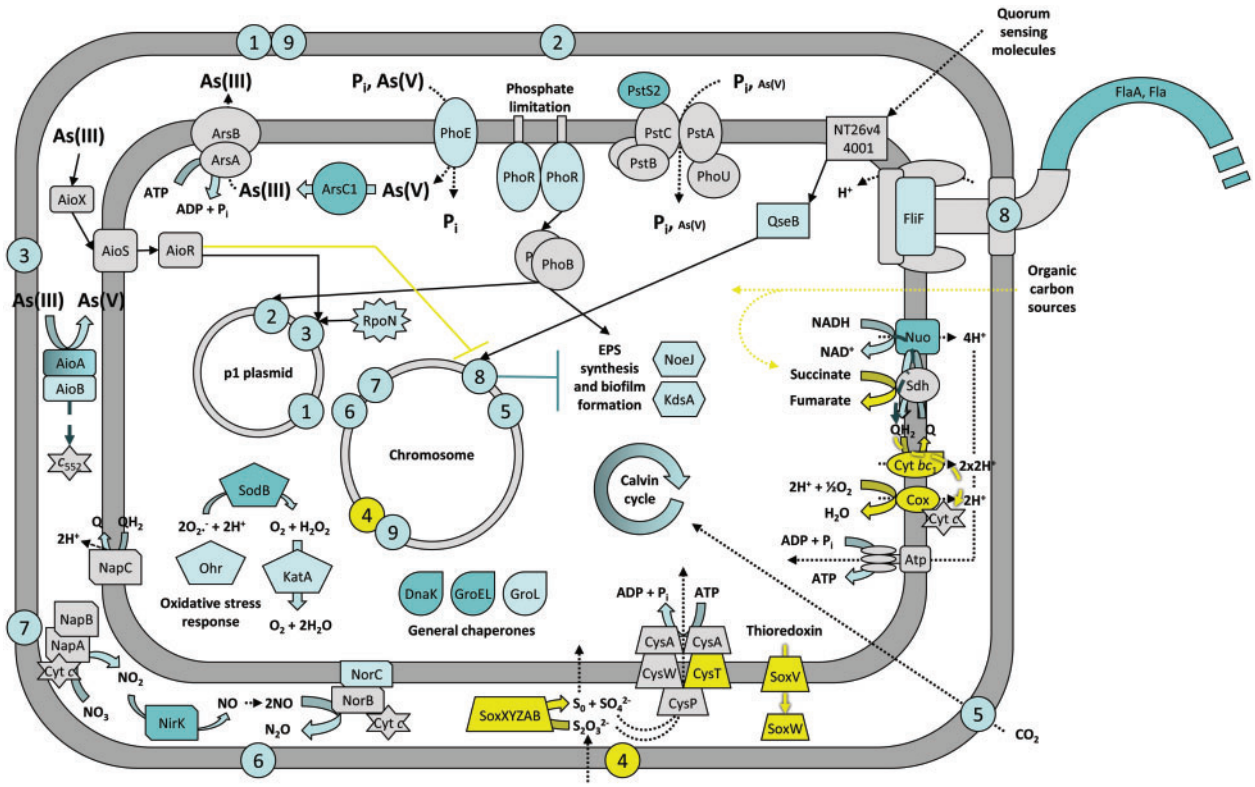


Fig. 7.—Conceptual representation of the *Rhizobium* sp. NT-26 response to arsenite exposure. This representation takes into account our genomic, transcriptomic, proteomic, and physiological results as well as data from the literature. Numbers 1 to 9 represent biological functions and the approximate genomic location of the gene clusters encoding their corresponding proteins. 1 and 9: arsenate reduction; 2: phosphate and arsenate transport; 3 and 4: arsenite and sulfur oxidation, respectively; 5: carbon fixation; 6 and 7: nitrate and nitrite reduction, respectively; 8: motility. Block, dashed, dotted, and standard arrows symbolize chemical reactions, electron flow, transport/utilization of molecules and signaling/regulatory pathways, respectively. When highlighted in dark blue, light blue, yellow or gray, elements have been identified as being induced by proteomic, induced by transcriptomic, repressed by transcriptomic, and present in the genome, respectively. For clarity reasons, proteins for which the exact function is still unknown but that are related to the different processes are not shown, the protein complexes of the respiratory chain, that is the NADH dehydrogenase, the fumarate reductase, the cytochrome bc_1 and the cytochrome c oxidase are designated by Nuo, Sdh, Cyt bc_1, and Cox, respectively, plasmid p2 is not shown and only one flagella is represented. Finally, Cyt c and c_{552} are for cytochrome c and cytochrome c_{552}, respectively, and NT26v4_4001 is a homolog of *qseC*.

Conclusion

This study extends our knowledge of the physiological response to arsenic in arsenite-oxidizing bacteria. Our results provide for the first time a reference set of genomic, transcriptomic, and proteomic data of an *Alphaproteobacterium* isolated from an arsenopyrite-containing goldmine, which allowed us to propose a model for the *Rhizobium* sp. NT-26 response to arsenite exposure (fig. 7). Although phylogenetically related to the plant-associated bacteria, strain sp. NT-26 has lost the major colonizing capabilities needed for symbiosis. Instead, this bacterium has acquired on a megaplasmid the various genes which allow it to metabolize arsenate. Remarkably, a link between flagellar motility/biofilm formation and arsenite oxidation was observed although the genes required for these physiological activities are carried by different genetic determinants, that is, the chromosome and the megaplasmid, respectively. This suggests the existence of a mechanism, probably indirect and which remains to be characterized at a molecular level, of a coordinate regulation of these two important biological processes. This underlines the importance of arsenite oxidation in the colonization of arsenic-rich ecosystems, a toxic element widespread on Earth. Importantly, our data also illustrate the major contribution of environmental pressure on the evolution of bacterial genomes, which results in a gain and loss of multiple functions, improving the fitness of the strains to extreme ecological niches.

Supplementary Material

Supplementary figures 1–4, tables S1–S6, and methods S1–S4 are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

This work was supported by the Université de Strasbourg (UdS), the Consortium National de Recherche en Génomique (CNRG), the Centre National de la Recherche Scientifique (CNRS), the Australian Research Council (DP034305), the French Ministère de l'Enseignement Supérieur et de la Recherche to J.A. and J.C.-A., the Direction Générale de l'Armement to L.G., a studentship from the Agence Nationale de la Recherche (ANR) COBIAS project (PRECODD 2007) to M.M., and the Ecole Normale Supérieure (ENS) of Lyon to F.L. The authors thank Mathieu Erhardt for the TEM observations and Ashleigh Clarke for her work on arsenic resistance in strain NT-26. This work was done in part in the frame of the Groupement de Recherche—Métabolisme de l'Arsenic chez les Microorganismes (GDR2909-CNRS) (http://gdr2909.alsace.cnrs.fr/, last accessed April 30, 2013).

Literature Cited

Altschul SF, et al. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25:3389–3402.

Arsène-Ploetze F, et al. 2010. Structure, function, and evolution of the *Thiomonas* spp. genome. PLoS Genet. 6:e1000859.

Aury J-M, et al. 2008. High quality draft sequences for prokaryotic genomes using a mix of new sequencing technologies. BMC Genomics 9:603.

Bernstam L, Nriagu J. 2000. Molecular aspects of arsenic stress. J Toxicol Environ Health B Crit Rev. 3:293–322.

Bertin PN, et al. 2011. Metabolic diversity among main microorganisms inside an arsenic-rich ecosystem revealed by meta- and proteogenomics. ISME J. 5:1735–1747.

Bertin PN, et al. 2012. Microbial arsenic response and metabolism in the genomics era. In: Santini JM, Ward SA, editors. The metabolism of arsenite. London: CRC Press. p. 99–114.

Bertin PN, Médigue C, Normand P. 2008. Advances in environmental genomics: towards an integrated view of micro-organisms and ecosystems. Microbiology 154:347–359.

Bigot T, Daubin V, Lassalle F, Perrière G. 2012. TPMS: a set of utilities for querying collections of gene trees. BMC Bioinformatics 14:109.

Borch T, et al. 2010. Biogeochemical redox processes and their impact on contaminant dynamics. Environ Sci Technol. 44:15–23.

Bradford MM. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem. 72:248–254.

Branco R, Chung A-P, Morais PV. 2008. Sequencing and expression of two arsenic resistance operons with different functions in the highly arsenic-resistant strain *Ochrobactrum tritici* SCII24T. BMC Microbiol. 8:95.

Bryan CG, et al. 2009. Carbon and arsenic metabolism in *Thiomonas* strains: differences revealed diverse adaptation processes. BMC Microbiol. 9:127.

Carapito C, et al. 2006. Identification of genes and proteins involved in the pleiotropic response to arsenic stress in *Caenibacter arsenoxydans*, a metalloresistant beta-proteobacterium with an unsequenced genome. Biochimie 88:595–606.

Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol. 17:540–552.

Clarke MB, Sperandio V. 2005. Transcriptional regulation of flhDC by QseBC and sigma (FliA) in enterohaemorrhagic *Escherichia coli*. Mol Microbiol. 57:1734–1749.

Cleiss-Arnold J, et al. 2010. Temporal transcriptomic response during arsenic stress in *Herminiimonas arsenicoxydans*. BMC Genomics 11:709.

Creus CM, et al. 2005. Nitric oxide is involved in the *Azospirillum brasilense*-induced lateral root formation in tomato. Planta 221:297–303.

Criscuolo A, Gribaldo S. 2010. BMGE (block mapping and gathering with entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. BMC Evol Biol. 10:210.

Danhorn T, Hentzer M, Givskov M, Parsek MR, Fuqua C. 2004. Phosphorus limitation enhances biofilm formation of the plant pathogen *Agrobacterium tumefaciens* through the PhoR-PhoB regulatory system. J Bacteriol. 186:4492–4501.

Daubin V, Lerat E, Perrière G. 2003. The source of laterally transferred genes in bacterial genomes. Genome Biol. 4:R57.

Dénarié J, Debellé F, Promé JC. 1996. *Rhizobium* lipo-chitooligosaccharide nodulation factors: signaling molecules mediating recognition and morphogenesis. Annu Rev Biochem. 65:503–535.

Didelot X, Lawson D, Darling A, Falush D. 2010. Inference of homologous recombination in bacteria using whole-genome sequences. Genetics 186:1435–1449.

Dobbelaere S, Okon Y. 2007. The plant growth-promoting effect and plant responses. In: Elmerich C, Newton WE, editors. Associative and endophytic nitrogen-fixing bacteria and cyanobacterial associations. Nitrogen fixation: origins, applications, and research progress V. Dordrecht (The Netherlands): Springer. p. 145–170.

Edgar RC. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5: 113.

Eraso JM, et al. 2008. Role of the global transcriptional regulator PrrA in *Rhodobacter sphaeroides* 2.4.1: combined transcriptome and proteome analysis. J Bacteriol. 190:4831–4848.

Gadd GM. 2010. Metals, minerals and microbes: geomicrobiology and bioremediation. Microbiology 156:609–643.

Gallie DR, Kado CI. 1988. Minimal region necessary for autonomous replication of pTAR. J Bacteriol. 170:3170–3176.

Giraud E, et al. 2007. Legumes symbioses: absence of *Nod* genes in photosynthetic bradyrhizobia. Science 316:1307–1312.

Gonzalez V, et al. 2006. The partitioned *Rhizobium etli* genome: genetic and metabolic redundancy in seven interacting replicons. Proc Natl Acad Sci U S A. 103:3834–3839.

Gouy M, Guindon S, Gascuel O. 2010. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol Biol Evol. 27:221–224.

Gremaud MF, Harper JE. 1989. Selection and initial characterization of partially nitrate tolerant nodulation mutants of soybean. Plant Physiol. 89:169–173.

Guibaud G, van Hullebusch E, Bordas F. 2006. Lead and cadmium biosorption by extracellular polymeric substances (EPS) extracted from activated sludges: pH-sorption edge tests and mathematical equilibrium modelling. Chemosphere 64:1955–1962.

Guindon S, Delsuc F, Dufayard J-F, Gascuel O. 2009. Estimating maximum likelihood phylogenies with PhyML. Methods Mol Biol. 537:113–137.

Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol. 52: 696–704.

Hall-Stoodley L, Costerton JW, Stoodley P. 2004. Bacterial biofilms: from the natural environment to infectious diseases. Nat Rev Microbiol. 2: 95–108.

Hao Y, Charles TC, Glick BR. 2011. ACC deaminase activity in avirulent *Agrobacterium tumefaciens* D3. Can J Microbiol. 57:278–286.

Harrison PW, Lower RPJ, Kim NKD, Young JPW. 2010. Introducing the bacterial 'chromid': not a chromosome, not a plasmid. Trends Microbiol. 18:141–148.

Heinrich-Salmeron A, et al. 2011. Unsuspected diversity of arsenite-oxidizing bacteria as revealed by widespread distribution of the aoxB gene in prokaryotes. Appl Environ Microbiol. 77:4685–4692.

Holmes A, et al. 2009. Comparison of two multimetal resistant bacterial strains: *Enterobacter* sp. YSU and *Stenotrophomonas maltophilia* ORO2. Curr Microbiol. 59:526–531.

Hommais F, et al. 2002. Effect of mild acid pH on the functioning of bacterial membranes in *Vibrio cholerae*. Proteomics 2:571–579.

Hynes MF, McGregor NF. 1990. Two plasmids other than the nodulation plasmid are necessary for formation of nitrogen-fixing nodules by *Rhizobium leguminosarum*. Mol Microbiol. 4:567–574.

Janssen PJ, et al. 2010. The complete genome sequence of *Cupriavidus metallidurans* strain CH34, a master survivalist in harsh and anthropogenic environments. PLoS One 5:e10433.

Juhas M, et al. 2009. Genomic islands: tools of bacterial horizontal gene transfer and evolution. FEMS Microbiol Rev. 33:376–393.

Kang Y-S, Bothner B, Rensing C, McDermott TR. 2012. Involvement of RpoN in regulating bacterial arsenite oxidation. Appl Environ Microbiol. 78:5638–5645.

Katoh K, Toh H. 2008. Recent developments in the MAFFT multiple sequence alignment program. Brief Bioinform. 9:286–298.

Kirov SM, Castrisios M, Shaw JG. 2004. *Aeromonas flagella* (polar and lateral) are enterocyte adhesins that contribute to biofilm formation on surfaces. Infect Immun. 72:1939–1945.

Koechler S, et al. 2010. Multiple controls affect arsenite oxidase gene expression in *Herminiimonas arsenicoxydans*. BMC Microbiol. 10:53.

Lassalle F, et al. 2011. Genomic species are ecological species as revealed by comparative genomics in *Agrobacterium tumefaciens*. Genome Biol Evol. 3:762–781.

Le SQ, Gascuel O. 2008. An improved general amino acid replacement matrix. Mol Biol Evol. 25:1307–1320.

Lett M-C, Paknikar KM, Lièvremont D. 2001. A simple and rapid method for arsenite and arsenate speciation. In: Ciminelli VST, Garcia O Jr, editors. Biohydrometallurgy: fundamentals, technology and sustainable development, Part B. New York: Elsevier Science. p. 541–546.

Lieutaud A, et al. 2010. Arsenite oxidase from *Ralstonia* sp. 22. J Biol Chem. 285:20433–20441.

Liu G, et al. 2012. A periplasmic arsenite-binding protein involved in regulating arsenite oxidation. Environ Microbiol. 14:1624–1634.

López-Guerrero MG, et al. 2012. Rhizobial extrachromosomal replicon variability, stability and expression in natural niches. Plasmid 68: 149–158.

Marchal M, Briandet R, Koechler S, Kammerer B, Bertin PN. 2010. Effect of arsenite on swimming motility delays surface colonization in *Herminiimonas arsenicoxydans*. Microbiology 156:2336–2342.

Marchal M, et al. 2011. Subinhibitory arsenite concentrations lead to population dispersal in *Thiomonas* sp. PLoS One 6:e23181.

Marouga R, David S, Hawkins E. 2005. The development of the DIGE system: 2D fluorescence difference gel analysis technology. Anal Bioanal Chem. 382:669–678.

Martens M, et al. 2007. Multilocus sequence analysis of Ensifer and related taxa. Int J Syst Evol Microbiol. 57:489–503.

McDougald D, Rice SA, Barraud N, Steinberg PD, Kjelleberg S. 2012. Should we stay or should we go: mechanisms and ecological consequences for biofilm dispersal. Nat Rev Microbiol. 10:39–50.

Muller D, et al. 2007. A tale of two oxidation states: bacterial colonization of arsenic-rich environments. PLoS Genet. 3:e53.

Muller D, Lièvremont D, Simeonova DD, Hubert J-C, Lett M-C. 2003. Arsenite oxidase aox genes from a metal-resistant β-proteobacterium. J Bacteriol. 185:135–141.

Nejidat A, Saadi I, Ronen Z. 2008. Effect of flagella expression on adhesion of *Achromobacter piechaudii* to chalk surfaces. J Appl Microbiol. 105: 2009–2014.

Osborne TH, Santini JM. 2012. Prokaryotic aerobic oxidation of arsenite. In: Santini JM, Ward SA, editors. The metabolism of arsenite. London: CRC Press. p. 61–72.

Parvatiyar K, et al. 2005. Global analysis of cellular factors and responses involved in *Pseudomonas aeruginosa* resistance to arsenite. J Bacteriol. 187:4853–4864.

Penel S, et al. 2009. Databases of homologous gene families for comparative genomics. BMC Bioinformatics 10(6 Suppl):S3.

Ramírez-Bahena MH, Nesme X, Muller D. 2012. Rapid and simultaneous detection of linear chromosome and large plasmids in proteobacteria. J Basic Microbiol. 52:736–739.

Rice P, Longden I, Bleasby A. 2000. EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet. 16:276–277.

Richardson AE, Barea J-M, McNeill AM, Prigent-Combaret C. 2009. Acquisition of phosphorus and nitrogen in the rhizosphere and plant growth promotion by microorganisms. Plant Soil. 321:305–339.

Rigaud J, Puppo A. 1975. Indole-3-acetic acid catabolism by soybean bacteroids. J Gen Microbiol. 88:223–228.

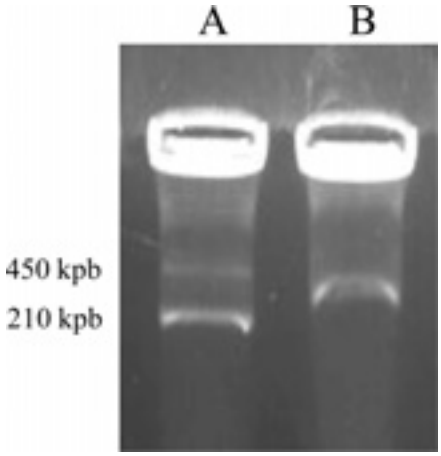
Ronquist F, et al. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol. 61:539–542.

Santini JM, et al. 2007. The NT-26 cytochrome c552 and its role in arsenite oxidation. Biochim Biophys Acta. 1767:189–196.

Santini JM, Sly LI, Schnagl RD, Macy JM. 2000. A new chemolithoautotrophic arsenite-oxidizing bacterium isolated from a gold mine: phylogenetic, physiological, and preliminary biochemical studies. Appl Environ Microbiol. 66:92–97.

Santini JM, vanden Hoven RN. 2004. Molybdenum-containing arsenite oxidase of the chemolithoautotrophic arsenite oxidizer NT-26. J Bacteriol. 186:1614–1619.
Sardiwal S, Santini JM, Osborne TH, Djordjevic S. 2010. Characterization of a two-component signal transduction system that controls arsenite oxidation in the chemolithoautotroph NT-26. FEMS Microbiol Lett. 313:20–28.
Slater SC, et al. 2009. Genome sequences of three *Agrobacterium biovars* help elucidate the evolution of multichromosome genomes in bacteria. J Bacteriol. 191:2501–2511.
Sourjik V, Muschler P, Scharf B, Schmitt R. 2000. VisN and VisR are global regulators of chemotaxis, flagellar, and motility genes in *Sinorhizobium (Rhizobium) meliloti*. J Bacteriol. 182:782–788.
Spaepen S, Vanderleyden J, Okon Y. 2009. Chapter 7: Plant growth-promoting actions of rhizobacteria. In: Van Loon LC, editor. Advances in botanical research, Vol. 51. Burlington: Academic Press. p. 283–320.
Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688–2690.
Stolz JF. 2011. Microbial metal and metalloid metabolism: advances and applications, 1st ed. Washington DC: ASM Press.
Suyama M, Torrents D, Bork P. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res. 34:W609–W612.
Tang X, Lu BF, Pan SQ. 1999. A bifunctional transposon mini-Tn5gfp-km which can be used to select for promoter fusions and report gene expression levels in *Agrobacterium tumefaciens*. FEMS Microbiol Lett. 179:37–42.
Vallenet D, et al. 2006. MaGe: a microbial genome annotation system supported by synteny results. Nucleic Acids Res. 34:53–65.
Vallenet D, et al. 2009. MicroScope: a platform for microbial genome annotation and comparative genomics. Database (Oxford) 2009: bap021.
van Lis R, Nitschke W, Duval S, Schoepp-Cothenet B. 2012. Evolution of arsenite oxidation. In: Santini JM, Ward SA, editors. The metabolism of arsenite. London: CRC Press. p. 125–144.
van Lis R, Nitschke W, Duval S, Schoepp-Cothenet B. 2013. Arsenics as bioenergetic substrates. Biochim Biophys Acta. 1827: 176–188.
Weiss S, et al. 2009. Enhanced structural and functional genome elucidation of the arsenite-oxidizing strain *Herminiimonas arsenicoxydans* by proteomics data. Biochimie 91:192–203.
Wilkins MJ, et al. 2009. Proteogenomic monitoring of Geobacter physiology during stimulated uranium bioremediation. Appl Environ Microbiol. 75:6591–6599.
Wood DW, et al. 2001. The genome of the natural genetic engineer *Agrobacterium tumefaciens* C58. Science 294:2317–2323.
Young JM, Kuykendall LD, Martínez-Romero E, Kerr A, Sawada H. 2001. A revision of *Rhizobium* Frank 1889, with an emended description of the genus, and the inclusion of all species of *Agrobacterium* Conn 1942 and *Allorhizobium* undicola de Lajudie et al. 1998 as new combinations: *Rhizobium radiobacter*, *R. rhizogenes*, *R. rubi*, *R. undicola* and *R. vitis*. Int J Syst Evol Microbiol. 51:89–103.
Young JPW, et al. 2006. The genome of *Rhizobium leguminosarum* has recognizable core and accessory components. Genome Biol. 7: R34.
Zheng M, et al. 2001. Computation-directed identification of OxyR DNA binding sites in *Escherichia coli*. J Bacteriol. 183:4571–4579.
Zhong Z, et al. 2003. Nucleotide sequence based characterizations of two cryptic plasmids from the marine bacterium *Ruegeria* isolate PR1b. Plasmid 49:233–252.
Zumft WG. 1997. Cell biology and molecular basis of denitrification. Microbiol Mol Biol Rev. 61:533–616.

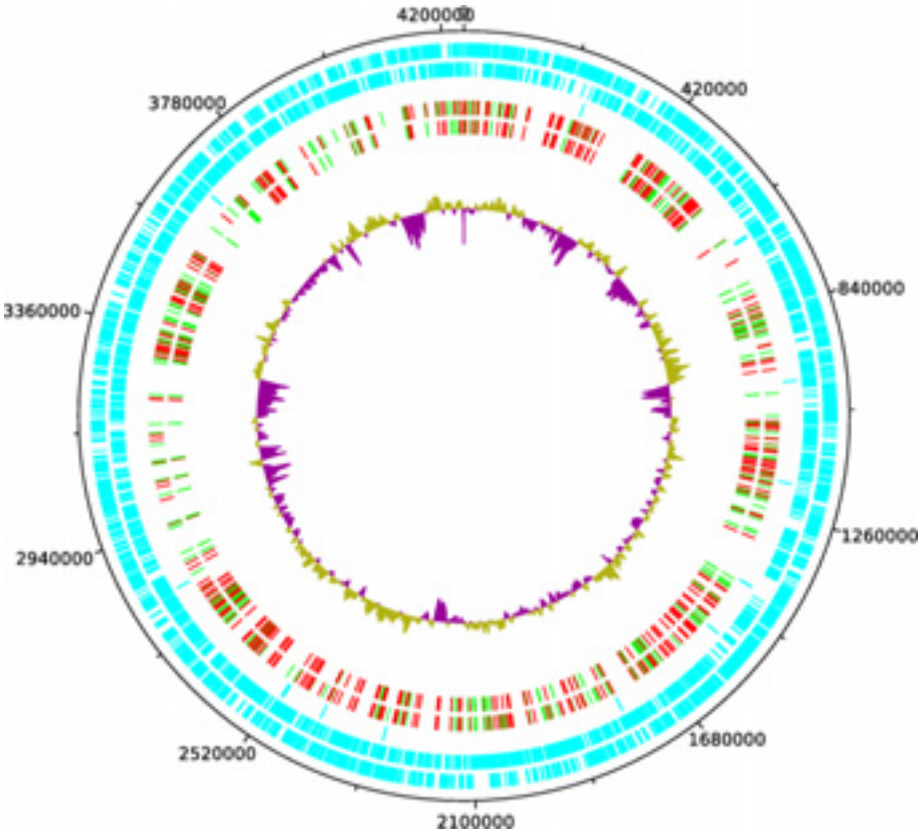
Associate editor: Purificación López-García



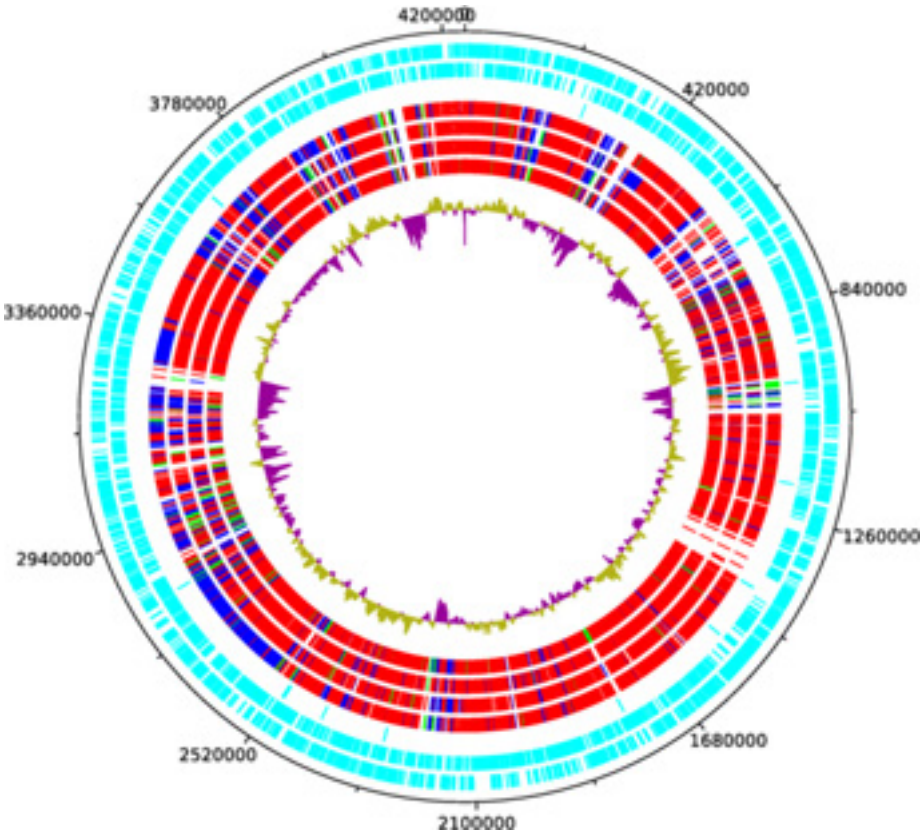
Supplementary Figure 1. Pulsed-Field Gel Electrophoresis analysis of *Rhizobium* sp. NT-26 genomic DNA. (*A*) *A. tumefaciens* C58 presenting 2 large plasmid: pAt (450 kbp) and pTi (210 kbp); (*B*) *Rhizobium* sp. NT-26 presenting the 322 kbp p1 plasmid. This electrophoretic profile was obtained by a modified Eckhardt procedure.


A
B
C
D
0.1

Supplementary Figure 2. Bayesian phylogenetic trees of Aio proteins. AioA (*A*) (109 sequences, 352 amino acid positions), AioB (*B*) (49 sequences, 153 amino acid positions), AioR (*C*) (46 sequences, 291 amino acid positions), AioS (*D*) (49 sequences, 214 amino acid positions) and AioX (*E*) (40 sequences, 201 amino acid positions) inferred with MrBayes. The sequences from *Alphaproteobacteria*, *Betaproteobacteria*, *Gammaproteobacteria* and *Firmicutes* are shown in orange, dark blue, light blue and pink, respectively, whereas the *Rhizobium* sp. NT-26 sequences are in red. Numbers at nodes represent posterior probabilities and bootstrap values computed with MrBayes and PhyML, respectively. For clarity reasons, only values greater than 0.50 and 50% are shown. Scale bars indicate the estimated average number of substitutions per site. The subgroups containing the *Rhizobium* sp. NT-26 sequences discussed in the paper are indicated by brackets.



Supplementary Figure 3. Distribution of genes matching "NT26inAgro" or "NT26outAgro" patterns around the *Rhizobium* sp. NT-26 chromosome. From the outside, circles display: (1) coordinates in bp; (2), (3) and (4) genes on direct strand, reverse strand, and pseudogenes, respectively; (5) and (6) genes matching the simple patterns and patterns with support higher than 90%, respectively; in red: "NT26inAgro", in green: "NT26outAgro"; (7) GC % deviation from the genomic mean, represented in yellow and purple for high and low GC %, respectively.



Supplementary Figure 4. Replicon location of *Rhizobium* sp. NT-26 gene homologs in related *Rhizhobiaceae* genomes. From the outside, circles display: (1) coordinates in bp; (2), (3) and (4) genes on direct strand, reverse strand, and pseudogenes, respectively; (5), (6), (7) and (8) represent replicon location of closest homolog in *A. tumefaciens* C58, *A. vitis* S4, *R. rhizogenes* K84, and *R. etli* CFN42, respectively. Represented in red: genes that are located on the primary chromosome, in blue: on the secondary chromosome/chromid (include p42b, p42e and p42f), and in green: on the plasmids; (9) GC % deviation from the genomic mean, represented in yellow and purple for high and low GC %, respectively.

DISCUSSION

Les résultats présentés dans cet article montrent comment *Rhizobium* sp. NT-26 s'est adaptée au stress arsénié et les données obtenues ont permis de reconstituer un modèle de la réponse globale de cette souche à l'arsenic.

Tout d'abord, le séquençage du génome de *Rhizobium* sp. NT-26 et son analyse semblent indiquer que cette souche a acquis des fonctions métaboliques et de résistance utiles à la colonisation de sa niche écologique et à son développement. En effet, même si ces hypothèses sont difficilement sujettes à l'expérimentation, l'exploration du génome a en particulier mis en évidence la présence de gènes liés aux métabolismes du carbone, de l'azote, du soufre et de l'arsenic mais également à la résistance à différents métaux lourds, sur des éléments génétiques pouvant avoir été acquis par transferts horizontaux.

Les approches de génomique fonctionnelle indiquent ensuite un impact de l'arsenic sur certaines de ces fonctions. *Rhizobium* sp. NT-26 semble ainsi privilégier une croissance chimiolithoautotrophe en induisant la fixation de carbone inorganique et l'oxydation d'As(III) en présence de cet élément. De plus, bien que ces outils ne permettent pas de mettre en évidence tout les mécanismes impliqués dans la réponse à l'arsenic, ils dévoilent ici le rôle de fonctions liées à la réduction et l'extrusion de cet élément, au transport du phosphate et à la résistance au stress oxydatif.

De manière intéressante, les données transcriptomiques et protéomiques soulignent également un lien entre les processus de colonisation et l'arsenic. Les différents tests physiologiques montrent que la présence d'arsenic favorise un mode de vie planctonique et mobile au détriment de la formation de biofilm chez *Rhizobium* sp. NT-26, ce qui peut avoir un rôle important dans la colonisation d'environnements riches en cet élément. Une régulation croisée des processus de colonisation et d'oxydation de l'arsenic a d'ailleurs été observée et l'hypothèse d'un contrôle de la mobilité par le biais du système de régulation de l'oxydation a été proposée grâce à des analyses de séquences.

Enfin, malgré la présence de certains éléments génétiques ainsi qu'un phytobénéfice observé, les tests n'ont pas permis de mettre en évidence un hôte potentiel pour cette *Rhizobiaceae* ou du moins l'existence d'interactions directes avec une plante. En effet, il semble que parallèlement

au gain de fonctions de résistance, cette bactérie ait perdu celles liées à la nodulation, soulignant l'impact des pressions environnementales sur l'évolution des génomes. Ainsi, pour pouvoir étudier le rôle des interactions procaryote-eucaryote en contexte arsénié, la suite de ces travaux a conduit à un changement de modèle d'étude, et plus précisément à s'intéresser à la communauté microbienne du Drainage Minier Acide de Carnoulès.

CHAPITRE II

LA RÉPONSE ADAPTATIVE D'*EUGLENA MUTABILIS* AU STRESS ARSENIÉ

INTRODUCTION

1. LES DRAINAGES MINIERS ACIDES

Une des principales causes anthropiques de contamination de l'environnement par l'arsenic est liée aux activités minières, qu'elles soient passées ou présentes. En effet, au cours de l'exploitation d'une mine, d'importants volumes de roches sont excavés et les stériles miniers résultants de l'extraction des éléments d'intérêt sont accumulés sous formes de terrils. Les roches sont ainsi exposées à l'oxygène atmosphérique et aux précipitations, ce qui favorise l'oxydation chimique et biologique des minéraux qu'elles contiennent. En particulier, l'oxydation des minéraux soufrés tels que la pyrite induit une production d'acide sulfurique et la solubilisation du fer (Johnson and Hallberg, 2003). Les conditions acides ainsi générées conduisent ensuite au lessivage des roches et à la libération d'autres métaux, métaux lourds ou métalloïdes. Ces effluents portent alors le nom de Drainage Minier Acide (DMA) et tandis que leur composition dépend des roches présentes, ils sont généralement caractérisés par un pH acide, entre 2 et 4, et d'importantes concentrations en fer. Il est à noter que dans certains cas, lorsque le fond géochimique d'une région est riche en minéraux soufrés, un phénomène naturel mais similaire peut se développer et l'on parle alors de Drainage Rocheux Acide.

Les DMA peuvent par la suite perturber de nombreux écosystèmes aquatiques situés en aval tels que des nappes phréatiques, des rivières, des fleuves, des lacs ou des estuaires et se poursuivent jusqu'à l'oxydation complète des roches. Ils sont ainsi à l'origine de pollutions étendues et persistantes. Cependant, tandis que les DMA sont qualifiés d'environnements extrêmes et sont toxiques pour la plupart des formes de vie, certains micro-organismes procaryotes et eucaryotes sont capables de s'y installer, de s'y maintenir et sont même trouvés de manière récurrente dans ces environnements (Johnson and Hallberg, 2005). De plus, par leur métabolisme, ces organismes influencent l'évolution des DMA. En effet, d'une part l'utilisation de l'oxydation du fer ou de composés soufrés dans le métabolisme énergétique joue un rôle clé dans la formation des DMA, et d'autre part, la sulfatoréduction, la modification de la spéciation des éléments métalliques ou la génération d'alcalinité par différentes voies métaboliques peut favoriser un processus d'atténuation naturelle en entraînant l'immobilisation des métaux (Johnson and Hallberg, 2005). Dans le but de mieux comprendre ces processus, nombre de DMA font l'objet d'un suivi physico-chimique et microbiologique régulier.

2. LE DRAINAGE MINIER ACIDE DE CARNOULÈS

L'ancienne exploitation d'une mine de plomb et de zinc à Carnoulès, située dans le Sud de la France, a conduit à l'accumulation de 1,5 MT de roches riches en arsénopyrite sous la forme d'un terril minier (Figure 2.1; Leblanc et al., 1996). L'oxydation abiotique et biotique de ces minéraux génère un DMA qui, en percolant à travers la roche, se mélange aux eaux de ruissellement d'une source localisée sous le terril, donnant naissance au Reigous (Figure 2.1). Le suivi régulier des caractéristiques physico-chimiques de ce ruisseau montre que malgré des variations saisonnières (Egal et al., 2010; Giloteaux et al., 2013), il présente de manière constante un pH acide (2,7-3,4) et de fortes concentrations en sulfates (2000-7700 $mg.L^{-1}$), arsenic (50-350 $mg.L^{-1}$) et fer (870-2700 $mg.L^{-1}$) (Leblanc et al., 1996; Bruneel et al., 2008). Des relevés réalisés le long du cours de ce ruisseau indiquent par ailleurs qu'au fur et à mesure de son écoulement, ces concentrations tendent à diminuer, témoignant d'un processus naturel d'atténuation (Table 2.1; Leblanc et al., 1996; Bruneel et al., 2008).

Malgré ces conditions toxiques, différentes études révèlent par des approches moléculaires ou culturales qu'un certain nombre de micro-organismes appartenant aux archées, bactéries et eucaryotes se développent dans cet écosystème (Bruneel et al., 2006, 2008; Delavat et al., 2013; Volant et al., 2014). Si ces travaux mettent en évidence une biodiversité faible en comparaison à d'autres écosystèmes non pollués, ils montrent également la stabilité de cette communauté microbienne dans le temps, témoignant de son adaptation à cette niche écologique particulière. De plus, la combinaison de ces résultats à des tests réalisés en laboratoire suggère que le phénomène d'atténuation observé à Carnoulès résulte au moins en partie de l'activité biologique de ces micro-organismes. D'une part des séquences d'ARNr 16S et des protéines GroEL attribuées à *Acidithiobacillus ferrooxydans* et *Thiomonas* sp. sont détectées dans les sédiments du site montrant la présence active de ces organismes. D'autre part des tests réalisés sur des isolats correspondant obtenus à partir des sédiments prouvent leur capacité respective à oxyder le Fe(II) en Fe(III) et l'As(III) en As(V), ces dernières réactions pouvant ensuite aboutir à la précipitation du fer, du sulfate et de l'arsenic sous forme d'un minéral insoluble, la tooéléite (Bruneel et al., 2003, 2006, 2011; Duquesne et al., 2003; Morin et al., 2003).

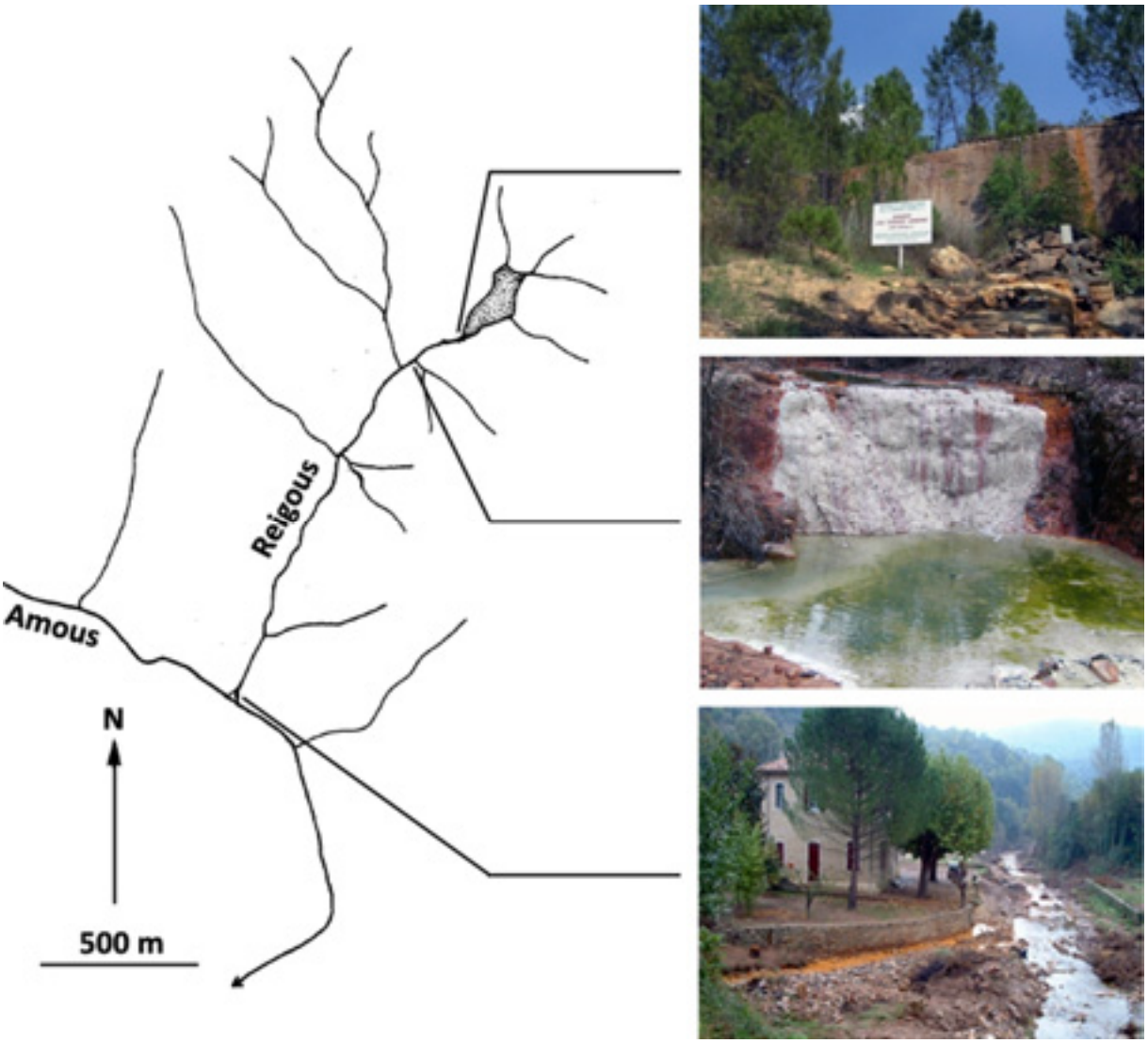


Figure 2.1. Plan schématique et photographies du DMA de Carnoulès. Les photographies indiquent la localisation de différents points d'étude sur le site. De haut en bas : le terril minier et le point S1 d'où jaillit le Reigous. Le point COWG correspondant à une retenue d'eau située 30 m en aval de S1. On peut y apercevoir la présence d'euglènes sous forme d'un tapis vert. Le point CONF localisé à la confluence du Reigous et de l'Amous 1500 m en aval de S1. Schéma adapté de (Leblanc et al., 1996), photographies Halter D. et Ploetze F.

Point	pH (±SD)	T (°C)	OD	As(III) (±SD)	As(V) (±SD)	Fe(II) (±SD)	SO_4^{2-} (±SD)
S1	2,50 (±0,05)	14,6	1	**30,0 (±0,8)**	**39,0 (±2)**	**879 (±70)**	**4388 (±441)**
COWG	2,74 (±0,05)	10,6	5-6	**22,0 (±0,08)**	**22,0 (±0,08)**	**501 (±40)**	**1785 (±182)**
CONF	3,25 (±0,05)	6,7	3-4	**0,53 (±0,02)**	**0,53 (±0,02)**	**25 (±2)**	**749 (±75)**

Table 2.1. Principales caractéristiques physico-chimiques du Reigous en novembre 2005. Ces données mettent en évidence les fortes concentrations en arsenic, fer et sulfates ainsi que le processus d'atténuation observé le long du ruisseau. Unités de concentration en $mg.L^{-1}$, SD : Standard Deviation, OD : oxygène dissout, Fe(III) non detecté. À titre de comparaison, la limite de concentration en arsenic dans l'eau de boisson fixée par l'OMS est de 10 µg.L-1. Adapté de (Bruneel et al., 2008)

3. LA COMMUNAUTÉ MICROBIENNE DE CARNOULÈS

Dans le but de mieux comprendre le fonctionnement de la communauté microbienne de Carnoulès et notamment les processus impliqués dans son maintien et dans le phénomène d'atténuation, une approche de métagénomique descriptive et fonctionnelle a été entreprise (Bertin et al., 2011). Cette étude dont les résultats principaux sont énoncés en introduction générale couple les approches de métagénomique et métaprotéomique. Brièvement, ces travaux ont conduit à la reconstruction des génomes des 7 espèces bactériennes majoritaires, nommées CARN1 à CARN7, présentes dans les sédiments du site. De plus, ces approches ont permis de proposer un modèle du fonctionnement de cette communauté où chacun des organismes joue un rôle spécifique, y compris dans le processus d'atténuation (Figure 2.2). Ces résultats suggèrent ainsi l'existence d'un réseau complexe de flux métaboliques et donc d'interactions trophiques entre ces sept organismes et leur écosystème, pouvant expliquer la stabilité de la communauté microbienne. Notamment, il est suggéré que le protiste *Euglena mutabilis*, isolé au contact de ces mêmes sédiments (Figure 2.1 et 2.2), est impliqué dans ces échanges, créant ainsi une opportunité d'étudier les interactions procaryote-eucaryote en contexte arsénié.

4. *EUGLENA MUTABILIS*

E. mutabilis est un protiste photosynthétique, retrouvé dans divers environnements aquatiques acides tels que des lacs acides, des tourbières, ou de nombreux sites contaminés par des DMA (Figure 2.3; Brake et al., 2001; Johnson and Hallberg, 2003; Aguilera et al., 2007), cette dernière caractéristique lui valant même d'être considéré comme un indicateur biologique de ces derniers écosystèmes (Valente and Gomes, 2007). A la différence des algues unicellulaires avec lesquelles elles sont souvent confondues, les euglènes ont la capacité d'alterner entre un métabolisme photo-autotrophe ou chimio-hétérotrophe, privilégiant par exemple le second en l'absence de lumière (Sumida et al., 2007). Malgré ce dernier mode de vie, elles ont gardé la capacité de synthétiser tout les composants indispensables à leur croissance en milieu minéral, à l'exception de la cobalamine, dont les voies de biosynthèse sont restreintes aux organismes procaryotes (Warren et al., 2002; Halter, 2011). Par son métabolisme photosynthétique *E. mutabilis* est considéré comme un producteur primaire, qui plus est particulièrement important dans les DMA présentant souvent une faible teneur en matière organique (Johnson and Hallberg, 2003; Das et al., 2009). Ces différentes caractéristiques soulignent ainsi à la fois l'adaptation d'*E. mutabilis* aux stress liés au DMA et

l'existence de potentielles interactions trophiques entre ce protiste et la communauté bactérienne de Carnoulès.

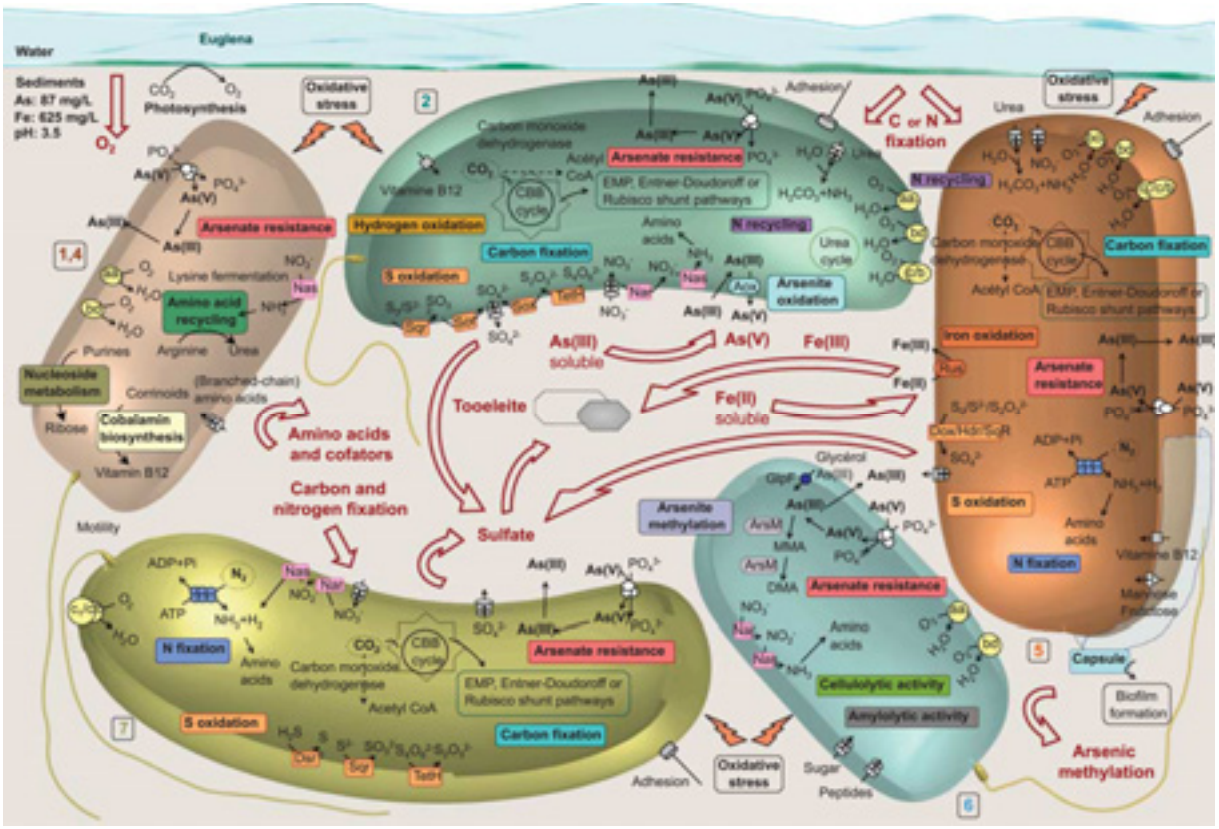


Figure 2.2. Représentation conceptuelle de la communauté bactérienne de Carnoulès indiquant les fonctions majeures identifiées par le séquençage du métagénome et la caractérisation du métaprotéome. Les CARNs sont numérotés de 1 à 7, en notant que le 3 est absent en raison de la faible quantité de données relative à cet organisme. Les flèches rouges représentent les interactions potentielles entre les bactéries majoritaires de la communauté ou avec leur environnement incluant des composés chimiques ou d'autres organismes. (Bertin et al., 2011)

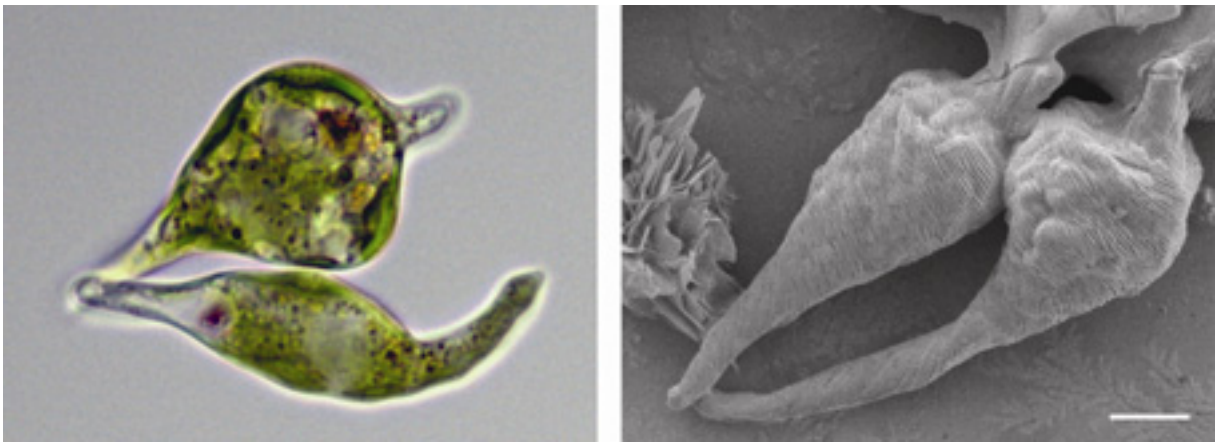


Figure 2.3. Images d'*E. mutabilis* obtenues en microscopie optique et microscopie électronique à balayage. La barre d'échelle représente 5 µm. Images personnelles avec la contribution de Faerber J.

5. ÉTUDE DE LA RÉPONSE ADAPTATIVE D'*E. MUTABILIS* AU STRESS ARSÉNIÉ

Avant de s'intéresser à ces potentielles relations procaryote-eucaryote, des approches ont été entreprises dans le but d'identifier les mécanismes ayant permis l'adaptation d'*E. mutabilis* à Carnoulès, et notamment ceux conférant une hypertolérance au stress arsénié. En effet, la résistance de ce protiste à des concentrations supérieures à 2500 $mg.L^{-1}$ d'As(III) laisse supposer un fonctionnement original (Halter et al., 2012a). Tout d'abord, une caractérisation du protéome exprimé *in situ* par *E. mutabilis* a mis en évidence l'expression de différentes protéines liées à la tolérance aux stress oxydatif et acide ainsi qu'à d'autres processus cellulaires (Halter et al., 2012b). Si les deux premières fonctions sont probablement impliquées dans la résistance aux conditions régnant à Carnoulès, il est difficile de juger du rôle potentiel des autres protéines dans la réponse à l'arsenic en l'absence de point de comparaison. De plus, la résistance au stress oxidatif ne justifie pas à elle-seule l'hypertolérance d'*E. mutabilis* vis-à vis de l'arsenic étant donné que la toxicité de cet élément est également lié à l'analogie structurale entre l'As(V) et le phosphate. Une approche comparative a ensuite été développée et compare les réponses au stress arsénié d'*E. mutabilis* et d'une souche phylogénétiquement proche mais non accoutumée à cet élément toxique, *Euglena gracilis* (Halter et al., 2012a). Cette étude a ainsi permis d'identifier des différences phénotypiques entre les deux protistes, pouvant expliquer en partie l'adaptation d'*E. mutabilis* à sa niche écologique. La surface cellulaire d'*E. mutabilis* présente notamment un caractère très hydrophobe en comparaison à *E. gracilis*, ce qui peut permettre de limiter l'exposition des cellules à des ions métalliques solubles tels que l'As(III). Ceci est par ailleurs conforté par une moindre accumulation intracellulaire d'As(III) observée chez *E. mutabilis*. Ces deux espèces montrent de plus un métabolisme de l'arsenic différent, des formes méthylées de l'arsenic, potentiellement génotoxiques, étant accumulées dans le compartiment intracellulaire d'E. *gracilis* mais pas d'*E. mutabilis*. Finalement, la présence d'arsenic semble influencer les processus de colonisation *d'E. mutabilis* en induisant sa mobilité, phénomène qui n'est pas observé chez *E. gracilis*.

Le travail présenté ici a été entrepris dans le but de compléter ces données en comparant les réponses à l'arsenic d'*E. mutabilis* et d'*E. gracilis* d'une manière plus globale, grâce à une approche de génomique fonctionnelle par séquençage des ARN messagers. Cette approche présente en effet l'avantage de mettre en évidence des processus potentiellement impliqués dans l'adaptation et la résistance sans *a priori*, et ce malgré l'absence de génome disponible. Les transcrits des souches *E. mutabilis* et *E. gracilis*, cultivées en présence et en absence d'arsenic, ont ainsi été extraits, séquencés, et les profils d'expression obtenus comparés. Ceci a permis d'émettre de premières

hypothèses qui ont ensuite été validées par différentes approches de biologie moléculaire et des tests physiologiques.

L'ensemble de ces travaux a conduit à la publication de l'article scientifique suivant dans *Environmental Microbiology*. J'ai contribué à ce travail en réalisant les différents tests physiologiques appuyant les hypothèses avancées par l'analyse des transcrits. J'ai notamment adapté les tests de comètes aux euglènes, mesuré l'activité du protéasome et réalisé les tests de culture sur boîtes de Petri.

MANUSCRIT

environmental microbiology

Environmental Microbiology (2014) doi:10.1111/1462-2920.12474

Arsenic hypertolerance in the protist *Euglena mutabilis* is mediated by specific transporters and functional integrity maintenance mechanisms

David Halter,[1] Jérémy Andres,[1] Frédéric Plewniak,[1] Julie Poulain,[2] Corinne Da Silva,[2] Florence Arsène-Ploetze[1] and Philippe N. Bertin[1]*

[1]*Génétique Moléculaire, Génomique et Microbiologie, Département Micro-organismes, Génomes, Environnement, UMR7156 Université de Strasbourg/CNRS, Strasbourg, France.*

[2]*Commissariat à l'Energie Atomique (CEA), Institut de Génomique (IG), Evry, France.*

Summary

Arsenic is a toxic metalloid known to cause multiple and severe cellular damages, including lipid peroxidation, protein misfolding, mutagenesis and double and single-stranded DNA breaks. Thus, exposure to this compound is lethal for most organisms but some species such as the photosynthetic protist *Euglena mutabilis* are able to cope with very high concentrations of this metalloid. Our comparative transcriptomic approaches performed on both an arsenic hypertolerant protist, i.e. *E. mutabilis*, and a more sensitive one, i.e. *E. gracilis*, revealed multiple mechanisms involved in arsenic tolerance. Indeed, *E. mutabilis* prevents efficiently the accumulation of arsenic in the cell through the expression of several transporters. More surprisingly, this protist induced the expression of active DNA reparation and protein turnover mechanisms, which allow *E. mutabilis* to maintain functional integrity of the cell under challenging conditions. Our observations suggest that this protist has acquired specific functions regarding arsenic and has developed an original metabolism to cope with acid mine drainages-related stresses.

Introduction

Acid mine drainages (AMDs) are important and persistent pollutions resulting from former mining activities. Such phenomena are mainly caused by biotic and abiotic oxidations of the sulfuric minerals associated with the excavated rocks accumulated as mine tailings. Previous analysis performed on AMD-exposed sediments revealed a relatively low microbial diversity in comparison to other sediments, highlighting that those ecosystems are strongly challenging for living organisms (Johnson and Hallberg, 2003; Denef *et al*., 2010; Bertin *et al*., 2011). Surprisingly, some microbial species, including bacteria (*Acidithiobacillus*, *Thiobacillus, Thiomonas sp*., . . .), algae (*Chlamydomonas acidophila*, . . .) or protists (*Euglena mutabilis*) are frequently found in AMDs, suggesting that they have developed specific mechanisms in order to cope with the toxicity associated to such environments (Valente and Gomes, 2007; Baker *et al*., 2009; Arsène-Ploetze *et al*., 2010; Halter *et al*., 2012a). Due to their direct involvement in bioremediation processes, several bacterial species, e.g. *Acidithiobacillus* or *Thiobacillus*, have been extensively studied but, in contrast, little is known about the adaptive mechanisms at work in the eukaryotic community in such ecosystems (Dasgupta *et al*., 2012).

The AMD of Carnoulès (Gard, France) originates from a mine tailing composed of 1.5 MT arsenopyrite-rich (FeAsS) rocks that were accumulated during the mining exploitation. These toxic effluents flow into a downstream creek, the Reigous, which is thereby characterized by a low pH and very high concentrations of arsenic, iron and sulfate, making these aquatic ecosystems very toxic for biota (Casiot *et al*., 2003). Nevertheless, some microorganisms are able to grow in this particular AMD impacted site (Bruneel *et al*., 2006; 2008; Bertin *et al*., 2011). Studies performed on the bacterial community from this site revealed its contribution to the observed attenuation process through the oxidation of iron, sulphuric molecules and arsenite. Indeed, the bacterial metabolism leads to the co-precipitation of these toxic compounds under their insoluble and less bio-available forms, mainly tooleite and amorph AsIII/FeIII and AsV/FeIII complexes (Duquesne *et al*., 2003; 2008; Egal *et al*., 2009; Bertin *et al*., 2011). The functioning of this microbial community has been proposed to rely on strong trophic interactions (Bertin *et al*., 2011). In particular, *E. mutabilis*, a photosynthetic protist widely found in this AMD has been shown to selectively secrete amino acids and

Received 22 October, 2013; revised 6 December, 2013; accepted 20 December, 2013. *For correspondence. E-mail philippe.bertin@unistra.fr; Tel. +33 (0)3 68 85 18 18; Fax +33 (0)3 68 85 20 28.

sugars that could favour the development and the stability of several bacterial species, some of them being directly involved in bio-remediation processes (Halter *et al*., 2012b). Besides, *E. mutabilis* has been shown to be particularly tolerant to arsenical stress, which could rely on an original response to the presence of this toxic metalloid. Indeed, a previous study has revealed that, when exposed to arsenite, *E. mutabilis* accumulates lower amount of total arsenic in the cell in comparison to *E. gracilis*, a phylogenetically related species not found in AMDs and used as a laboratory model. Moreover, arsenic is mainly found under inorganic forms in *E. mutabilis*, while methylated arsenic species are the major forms in *E. gracilis*, which suggests distinct arsenic metabolisms (Halter *et al*., 2012a). Nevertheless, the underlying genetic determinants conferring such a tolerance as well as those involved in the adaptation to AMDs remain to date largely unknown.

The present study aims at determining more precisely the adaptive mechanisms developed by *E. mutabilis* to cope with arsenic. For that purpose, we used descriptive and functional transcriptomic approaches in the presence or in the absence of arsenite in combination with physiological observations to confirm the most relevant traits. Moreover, in order to precise to what extent these adaptive mechanisms are specific to this protist, the same approaches were applied on *E. gracilis*, a phylogenetically related (Kosmala *et al*., 2009) and arsenic-sensitive species.

Results

The transcriptomes of *E. mutabilis* and *E. gracilis* were sequenced in the presence or in the absence of arsenite using 454 GSFlx instrument with Titanium chemistry and assembled using the ROCHE NEWBLER assembler. The number of nucleotides sequenced for each condition, the size of the datasets after assembly, the mean sizes of the transcripts after assembly and the size of the longest transcript are presented in Table 1. These data allowed us to access to the transcriptomes expressed in each condition and to compare the arsenic response of *E. gracilis* and *E. mutabilis*. These sequence data have been submitted to the European Nucleotide Archive database of the European Bioinformatics Institute (http://www.ebi.ac.uk/ena/data/view/PRJEB4713) and are publicly available.

Previous analyses have revealed that the arsenic response of *E. mutabilis* differs markedly from that is observed in *E. gracilis*, a species more sensitive to this element (Halter *et al*., 2012a). In particular, the accumulation of lower amounts of total arsenic in *E. mutabilis* intracellular compartment may explain its very high tolerance to this toxic metalloid. Such a phenotype results either from a reduced entry of arsenic into the cell, a higher efflux or both. In order to get insights into the underlying genetic determinants, we used a comparative transcriptomic approach and focused in a first step on arsenic metabolism.

Transcripts presenting homologies with Acr3 and an arsenite-activated ATPase related to ArsA, known as transporters and ATPases, respectively, were found in the transcriptome of *E. mutabilis* but not in that of *E. gracilis*. It is worth noting that these transcripts were also observed in the absence of arsenic suggesting a constitutive expression of these defence mechanisms in *E. mutabilis*. Besides, arsenate presents structural analogies with phosphate, and is therefore often incorporated into the cells via low specificity transporters of this latter element (Ali *et al*., 2009; Wysocki and Tamás, 2010). Interestingly, the transcription of genes presenting homologies with specific phosphate transporters (Pst) preventing this phenomenon was only found in *E. mutabilis* and not in *E. gracilis*. This observation suggests that *E. mutabilis* is able to prevent the accumulation of arsenic through low-specific phosphate transporters more efficiently than *E. gracilis*. In addition, functions related to phosphonate transport (GO:0015716) were specifically found in *E. mutabilis* in the presence of arsenic but not in *E. gracilis*. Such organo-phosphoric compounds could correspond to alternative phosphorus sources than Pi in the presence of arsenic, incidentally preventing the accumulation of this toxic metalloid through low-specific phosphate transporters.

Taken together, these observations suggest that *E. mutabilis* has acquired specific mechanisms that allow

Table 1. Summary of the reads obtained after the transcriptome sequencing for *E. mutabilis* (EM) and *E. gracilis* (EG) in the presence or in the absence of arsenic. The sequencing was performed using 454 GSFlx instrument with Titanium chemistry, and the assembly was performed using Roche Newbler assembler.

	EM without AsIII	EM with AsIII	EG without AsIII	EG with AsIII
Number of nucleotide sequenced	66 572 920	80 401 536	51 086 259	71 709 822
Size after assembly (nt)	6 600 095	8 615 937	2 778 693	3 396 167
Number of contigs	10 062	12 287	5 167	4 837
Mean size of the contigs (nt)	655 ± 378	701 ± 413	537 ± 378	702 ± 478
Longest contig size (nt)	6 199	4 808	4 525	5 452

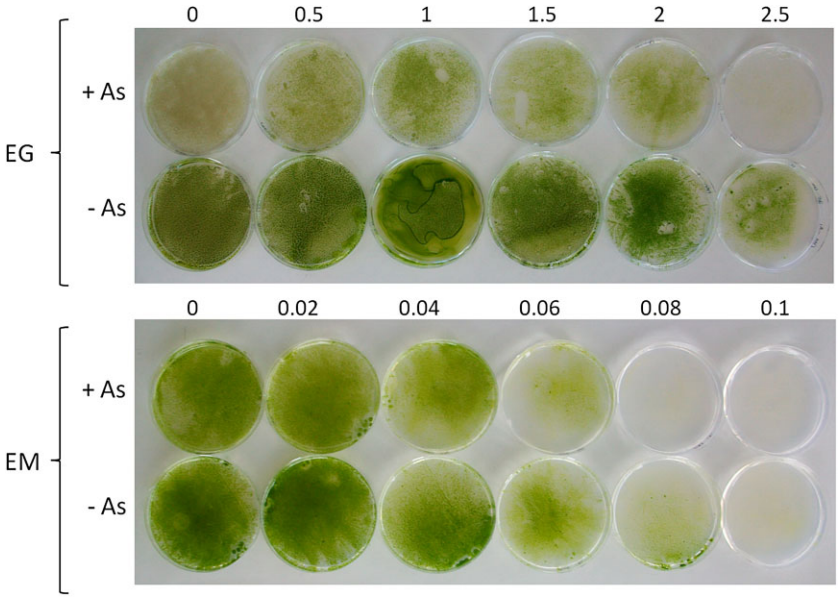


Fig. 1. Growth test of both *Euglena* species in the presence of increasing concentrations of cycloheximide (in $\mu g.ml^{-1}$) in the presence or in the absence of arsenic. Note that the range of cycloheximide concentrations is not the same for *E. mutabilis* (EM) and *E. gracilis* (EG).

it to selectively expulse different arsenical intermediates out of the cell and prevent their accumulation more efficiently than *E. gracilis*. Such genetic determinants may lead to an increased tolerance to this toxic metalloid and explain the lower accumulation level of arsenic observed in its intracellular compartment in comparison to *E. gracilis*.

Interestingly, *E. mutabilis* also differs from *E. gracilis* according to its arsenic metabolism. In particular, it has been previously shown that arsenic methylated forms accumulate strongly in *E. gracilis* intracellular compartment but were not detected in that of *E. mutabilis* under the same conditions (Halter *et al.*, 2012a). In order to identify the genetic determinants that could explain these differences, we searched in our transcriptomic data sets for sequences showing homologies with the arsenite methyltransferases present in databases. Surprisingly, both protists expressed transcripts with homologies with an arsenite methyltransferase, suggesting that *E. mutabilis* as well as *E. gracilis* are able to methylate this toxic element. Nevertheless, no arsenic methylated compound could be detected in an *E. mutabilis* culture, neither in the cell nor in the supernatant (data not shown), as previously observed (Halter *et al.*, 2012a).

Our transcriptomic data sets highlighted some genetic determinants that may explain the differences at the quantitative levels of arsenical species found into the cells. Nevertheless, even if *E. mutabilis* accumulates less arsenic than *E. gracilis* in its intracellular compartment (Halter *et al.*, 2012a), the amount of arsenic found in the cell is still high enough to affect the homeostasis of the cell, as previously shown in other species (Zhou *et al.*, 2005). In particular, the oxidative stress associated with arsenic is known to affect the genomic and proteomic functional integrity of the cell, which requires an adapted response. Interestingly, the over-representation analysis of GO entries revealed that nitrogen metabolism seems to play an important role in the *E. mutabilis* adaptive response. Indeed, *E. mutabilis* transcriptome was significantly enriched in transcripts pertaining to the nitrate transport process (GO:0015706) in cultures with As comparatively to cultures without As and to *E. gracilis* (Fisher exact test, *P*-value = 0). In order to get more insight into this point, we focused our attention on protein and nucleic acid metabolisms, both of them being dependant on nitrogen metabolism.

The over-representation analysis of GO entries revealed that functions related to translation initiation, translation termination and posttranslational modifications were specifically overrepresented in *E. mutabilis*, suggesting a constitutive and higher protein metabolism than in *E. gracilis*. In order to confirm this phenotype, we cultivated both species in the presence or in the absence of cycloheximide, an inhibitor of protein synthesis (Fig. 1). Our results revealed that *E. mutabilis* is dramatically more sensitive to this toxic compound than *E. gracilis*, which supports the hypothesis of a very active protein synthesis metabolism. Moreover, as the generation time of *E. mutabilis* is about twice that of *E. gracilis* (data not shown), it is tempting to speculate that its stronger protein synthesis metabolism is rather linked to a higher protein turnover than to a growth phenotype leading to biomass production. This assertion was further supported by the fact that, in the presence of arsenic, *E. mutabilis* transcriptome was significantly enriched in proteasomal catabolic process (GO:0010498) transcripts relatively to *E. mutabilis* in the absence of arsenic and *E. gracilis* (Fisher exact test, *P*-value = 3.54×10^{-30}). Since this

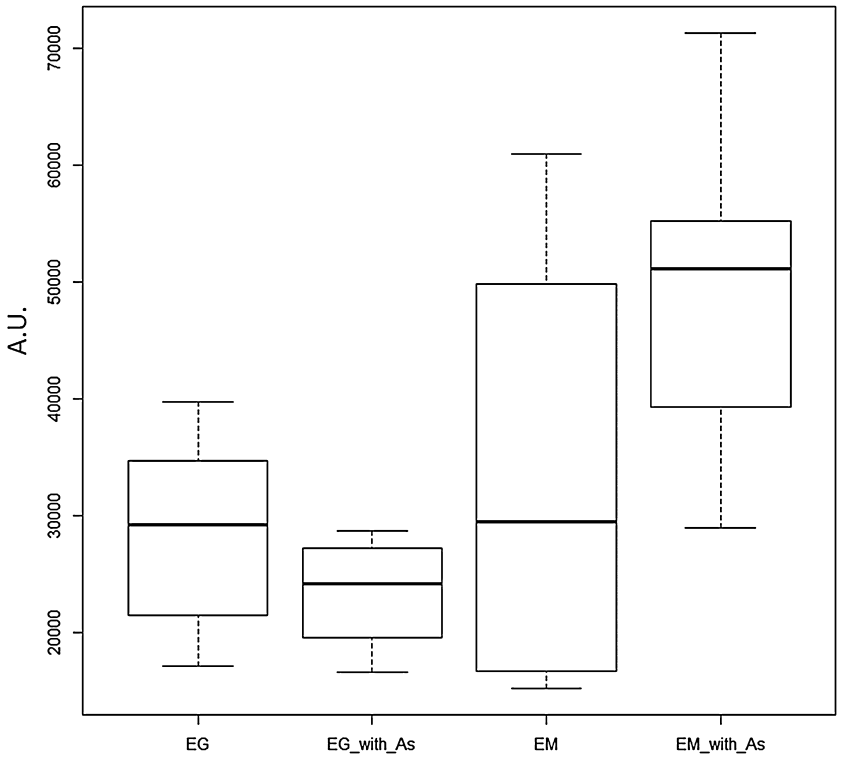


Fig. 2. Proteasome activity test performed in three biological replicates and four technical replicates represented with boxplots. The activity (Arbitrary Units = A.U) of the proteasome was measured in *E. mutabilis* (EM) and *E. gracilis* (EG) in the absence or in the presence of arsenic. A significant induction of the proteasome activity was observed only in *E. mutabilis* in the presence of the metalloid.

process could be linked to protein turnover, we measured the proteasome activity in both species in the presence or in the absence of arsenic (Fig. 2) in order to confirm that its activity is stronger in *E. mutabilis*. A significantly stronger activity (*t*-test, *P*-value = 0.013) of the proteasome was measured in the presence of arsenic in *E. mutabilis*, revealing a more active protein degradation in this protist in comparison to the other conditions. These observations suggest that the higher nitrogen metabolism observed in *E. mutabilis* in comparison to *E. gracilis* may be linked, at least in part, to a higher protein turnover implying proteasome mediated degradation. As arsenic is known to lead to an alteration of the functionality of the proteins, such behaviours could help this protist to maintain the functional integrity of its proteome.

Besides, the over-representation analysis of GO entries revealed that functions related to transcription (GO:0006350) were significantly overrepresented in *E. mutabilis* in comparison to *E. gracilis*, both in the presence and in the absence of arsenic (Fisher exact test, *P*-value = 1.45×10^{-32} and 4.63×10^{-17} respectively). Functions related to DNA repair (GO:0006281) were largely over-represented in *E. mutabilis* transcriptomes, both in the presence and in the absence of arsenic (Fisher exact test, *P*-value = 0). More particularly, nucleotide excision repair mechanisms (GO:0006289) were over-represented in *E. mutabilis* transcriptomes both in the presence and in the absence of arsenic (Fisher exact test, *P*-value = 1.90×10^{-66} and 0 respectively) as well as double-strand break repair via non-homologous end joining reparation process (GO:0006303, Fisher exact test, *P*-value = 3.76×10^{-5} and 7.71×10^{-46} respectively). Interestingly, functions related to DNA mismatch repair (GO:0006298) were specifically over-represented in *E. mutabilis* only in the presence of arsenic (Fisher exact test, *P*-value = 1.05×10^{-56}).

Such observations suggest that *E. mutabilis* prevents DNA damages more efficiently than *E. gracilis* in the presence of this toxic metalloid. To test this hypothesis, a single cell gel electrophoresis approach was used in order to measure DNA integrity in both *Euglena* species in the presence or the absence of this metalloid (Fig. 3). Our results revealed that in *E. gracilis*, the mean size of the tail and its DNA content is more important in the presence of arsenic than in its absence (*t*-test, *P*-value = 1.89×10^{-5}), while the tail size decreased in *E. mutabilis* (*t*-test, *P*-value = 2.02×10^{-8}). These observations suggest that the genomic integrity of *E. mutabilis* is maintained under arsenic exposure, while that of *E. gracilis* is more challenged and accumulates DNA single and/or double strand breaks.

Discussion

Euglena mutabilis is found in multiple AMDs around the world, strongly suggesting that this protist has acquired efficient adaptive mechanisms to cope with the stresses associated to these ecosystems. In particular, this protist

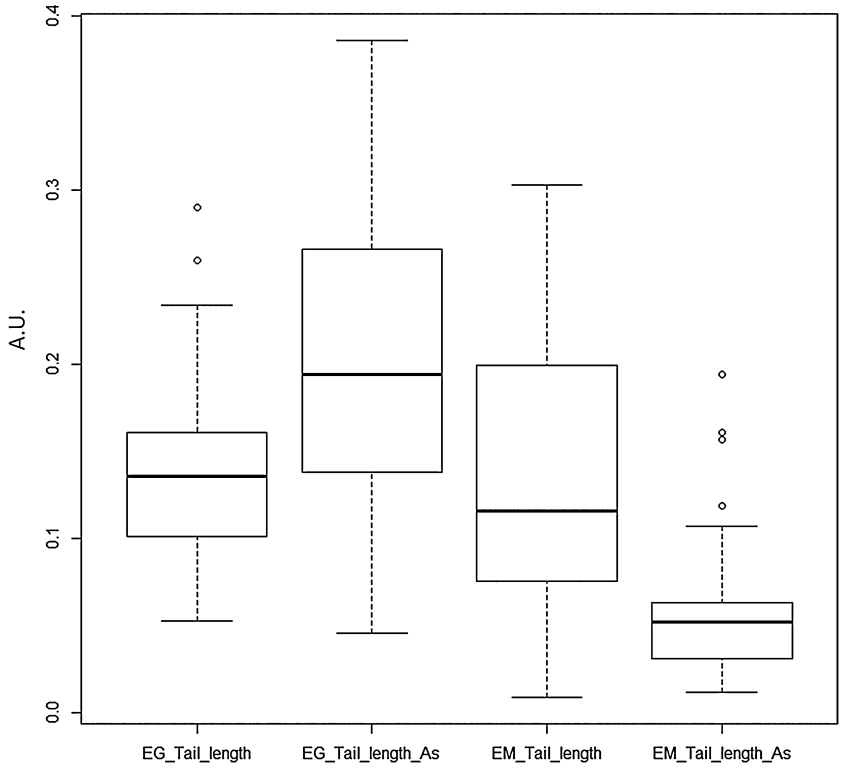


Fig. 3. Boxplots of the tail lengths measured in arbitrary units (A.U.) according to a single cell gel electrophoresis experiment. Measures were performed on at least 41 cells for each condition, i.e. *E. mutabilis* (EM) and *E. gracilis* (EG) in the absence (–As) or in the presence of 300 ppm AsIII (+As).The size of the tail increased significantly in the presence of arsenic in *E. gracilis* while it decreased in *E. mutabilis* under the same conditions.

has been previously shown to be more tolerant to multiple metals and metalloids when compared with *E. gracilis* (Olaveson and Nalewajko, 2000). This resistance has been proposed to rely in part on its hydrophobic cell surface properties (Halter *et al*., 2012a). Nevertheless, this does not rule out the existence of other adaptive mechanisms, which were explored here. Our comparative transcriptomic approach aimed to identify candidate processes that may be specifically active in *Euglena* species in the presence of arsenic allowing the identification of alternative response mechanisms. According to these data, *E. mutabilis* adaptation to AMDs seems to rely on a wide diversity of biological functions which are less or not expressed in *E. gracilis*, a protist absent of such extreme ecosystems.

Indeed, *E. mutabilis* is able to limit the accumulation of arsenic species in the intracellular compartment through different mechanisms. First, the expression of a constitutive arsenic expulsion system coupled to an ATPase could limit the accumulation of this toxic metalloid in its intracellular compartment. As *E. mutabilis* is usually found in AMD impacted ecosystems, this suggests that this protist constitutively expresses an arsenite transporter and increases its efficiency with an associated ATPase. Second, the use of highly specific phosphate transporters, i.e. Pst, or alternative phosphorus transporter, i.e. phosphonate uptake transporters, may prevent the uptake of structural analogues such as arsenate. Third, *E. mutabilis* does not accumulate arsenic methylated forms in its intracellular compartment despite the expression of transcripts presenting 50.3% and 49.5% amino acid identity with arsenite methyltransferase of *Chlamydomonas reinhartdii* and *Coccomyxa subellipsoidea* respectively. This could be of prime importance to explain the arsenite hypertolerance of *E. mutabilis*, some of these methylated compounds being known to present a higher reactivity and genotoxicity than inorganic arsenic compounds (Aposhian, 1997). The reason why no arsenic methylated compounds were detected in *E. mutabilis* despite the expression of transcripts potentially coding for an arsenite methyltransferase still remains unclear. The expression of non-functional enzymes or presenting a low affinity regarding their substrates could be an explanation but will need further investigation to be clarified. These active mechanisms of arsenic expulsion may work in parallel with a more passive system such as highly hydrophobic cell surface properties, and explain the limited accumulation of arsenic in the *E. mutabilis* intracellular compartment.

Such a reduced arsenic accumulation in *E. mutabilis* cells would consequently less challenge its homeostasis. Nevertheless, the intracellular arsenic concentrations previously measured in *E. mutabilis* are still high enough to affect the functional integrity of the cell (Zhou *et al*., 2005; Halter *et al*., 2012a). Indeed, it is known that arsenic generates an oxidative stress leading to protein damage

(Flora, 2011). For instance, in yeast *Saccharomyces cerevisiae*, arsenite induces protein misfolding leading to protein aggregation, a process targeting mainly newly synthesized proteins (Jacobson *et al.*, 2012). Our transcriptomic data combined with physiological observations revealed that *E. mutabilis* has a stronger protein synthesis rate than *E. gracilis* but also activates protein degradation in a proteasome-dependant manner under arsenic exposure. The activation of the proteasome activity under the same conditions has also been described in other species and seems to be an important element of the cellular response to this metalloid. Such results can be analysed in the light of a higher protein turnover in *E. mutabilis* that could prevent the accumulation of misfolded or non-functional proteins, and therefore playing a crucial role in the adaptation of this protist to AMDs.

Arsenic can also generate DNA damages such as DNA single and double-strand breaks, base damage and alkali labile sites (Ding *et al.*, 2009; Flora, 2011). Our transcriptomic data revealed that *E. mutabilis* expressed DNA reparation mechanisms based on non-homologous recombination and mismatch repair more actively than *E. gracilis*. The expression of such biological functions may explain the shorter DNA tail observed in the presence of arsenic. Nevertheless, as this metalloid is known to lead to DNA hypomethylation, other mechanisms such as DNA compaction could also be involved. Such active DNA reparations imply a high consumption of dNTP, which could explain the higher rate of recycling and *de novo* dNTP synthesis observed in *E. mutabilis* in comparison to *E. gracilis*. The constitutive high expression of such DNA reparation mechanisms in *E. mutabilis* coupled to the induction of DNA mismatch reparation processes in the presence of arsenic may play an important role in maintaining the genomic functional integrity observed during our analysis.

Our comparative transcriptomic approach combined with physiological observations highlighted, in addition to traditional defence mechanisms such as arsenic expulsion, some unexpected processes involved in the maintenance of cellular integrity likely implicated in the arsenic hypertolerance of *E. mutabilis* (Fig. 4). The adaptation of this protist to AMDs indeed relies on both a specific set of genetic determinants and a specific organization of its metabolism. Nevertheless, the presence of other specific mechanisms linked to the protection of the cell and presenting no homologies with sequences in databases cannot be excluded. In particular, it would be interesting to address the question of how such extreme ecosystems lead to the acquisition of arsenic-specific defence mechanisms in protists and, more generally, the impact it may have on the organization of their genome.

Experimental procedures

Euglena strains and culture conditions

Euglena mutabilis was isolated from the AMD of Carnoulès as previously described (Halter *et al.*, 2012a), and *E. gracilis* was provided by Sordalab (Etampes, France). All cultures were incubated at 25°C, with a 16 h/8 h light dark photoperiod and a 45 $\mu mol.m^{-2}.s^{-1}$ photon flux density and cultivated in Modified Acid Medium (MAM, Olaveson and Stokes, 1989). When precised, arsenic was added at the desired concentration from sterile stocks solution of 50 000 ppm from

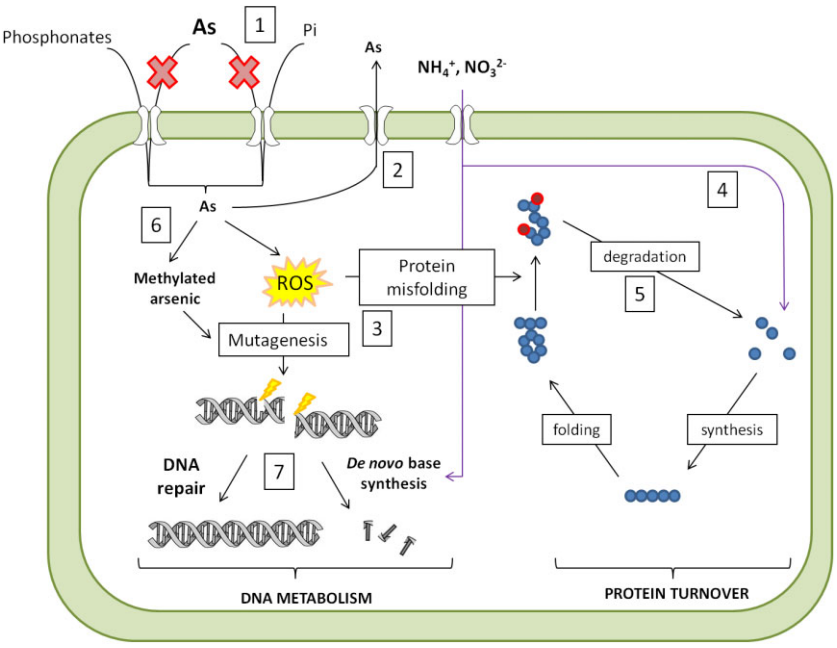


Fig. 4. Model of the main differences between *E. mutabilis* and *E. gracilis* regarding arsenic defence mechanisms. Arsenic accumulation is efficiently prevented in *E. mutabilis* through the expression of specific phosphate transporters (1) and arsenate efflux systems (2). Once in the intracellular compartment, inorganic arsenic leads to an oxidative stress that affects the functional integrity of the cell (3). *E. mutabilis* prevent the accumulation of non-functional proteins through a higher protein turnover, i.e. combination of higher protein synthesis (4) and proteasomal degradation (5). Arsenic-related DNA damages are prevented in *E. mutabilis* by the absence of genotoxic arsenic methylated forms (6) and more active reparation systems combined with a higher DNA metabolism (7).

$NaAsO_2$ salts (Prolabo, Fontenay-sous-bois, France), and the pH was adjusted to 3.2. For protein synthesis inhibition experiment, 300 ppm AsIII were added when required and cycloheximide was provided by Sigma (Saint-Quentin Fallavier, France).

RNA extraction and sequencing

Total RNA were extracted according to González-Mendoza and colleagues (2009) from 10 independent cultures of *E. mutabilis* and *E. gracilis* cultivated either in the absence or in the presence of arsenite (AsIII, 300 ppm). The 10 total RNA extracts of each of the four conditions (*E. gracilis* with AsIII, *E. gracilis* without AsIII, *E. mutabilis* with AsIII, *E. mutabilis* without AsIII) were pooled in order to obtain four different total RNA extracts. mRNA were purified from these total extracts using the Quick Prep mRNA purification kit (Invitrogen, Cergy Pontoise, France) and cDNA were synthesized using the SuperScript Double-Stranded cDNA Synthesis Kit (Invitrogen). Four hundred fifty-four libraries were constructed on cDNA according to the Roche standard procedure using Rapid Library adaptators (GS FLX Titanium Rapid Library Preparation Kit, Roche Diagnostic, USA). Libraries were sequenced using 1/4 Pico Titer Plate on 454 GSFlx instrument with Titanium chemistry (Roche Diagnostic).

Sequence annotation

Transcripts were assembled using ROCHE NEWBLER assembler (454 Life Science, Roche Diagnostic) and were matched against TIGRFAM (Selengut *et al*., 2007) and Pfam (Punta *et al*., 2012) family and domain models with HMMER (Eddy, 1998) as implemented in the RAMMCAP workflow (Li, 2009) available on the CAMERA portal (Sun *et al*., 2011). Hits were considered significant if they had an E-value < 0.001. Gene ontology (GO) annotations (Ashburner *et al*., 2000) derived from TIGRFAM and Pfam hits were assigned to matching Open Reading Frames. The mapping between GO entries and transcripts was propagated from low-level GO entries to their parents in order to take into account the hierarchical nature of the GO graph and homogenize GO annotation levels between all TIGRFAM and Pfam models. Thus, we could ensure that higher level GO entries would be counted even if they were not explicitly mapped to a TIGRFAM or Pfam entry.

Over-representation analysis of GO entries

For every GO entry from the biological process hierarchy, we counted the number of reads that contributed to the assembly of corresponding transcripts in each condition. Those counts were normalized by the length of the corresponding matching hidden Markov model (HMM). Gene ontology categories having less than 10 occurrences in all conditions were discarded to avoid low occurrence classes. The significance of the observed number of occurrences for each GO entry for each condition was assessed using the Fisher exact test (Fisher, 1925), and the *P*-values were corrected for multiple testing using the Benjamini–Hochberg False Discovery Rate method implemented in the R package fdrtool (Strimmer, 2008). Categories having a False Discovery Rate lesser or equal to 1% were selected for further examination.

Target analysis in transcriptome

Sets of reference sequences were extracted from the Refseq database for Acr3, ArsA and ArsM respectively. Sequences in each set were aligned with Clustal Omega multiple sequence alignment program. Each resulting alignment was manually checked and curated before generating an HM model which was subsequently used to search the transcripts translated in all six frames with HMMER using default parameters. Matching translated transcripts were included in the multiple alignments and manually checked against the reference sequences in order to identify false positives and correct frameshift errors. Transcripts that could not be aligned over their whole length were discarded.

Proteasome activity measurements

Euglena mutabilis and *E. gracilis* were grown on solid MAM 0.7% agar plates supplemented or not with 300 ppm As(III). After 1 week of culture, three plates of each condition were harvested to get three samples of approximately 150 µl of cells by condition. Harvesting has been randomly done regarding to the condition in order to avoid a time effect on the final proteasomal activity measurement. Each of the following steps was performed on ice to prevent proteasomal activity or degradation. Cells were suspended in 200 µl of cold lysis buffer containing 10% [v/v] glycerol, 25 mM Tris HCl pH 7.5, 10 mM $MgCl_2$, 1 mM ATP, 1 mM DTT and 1 mM EDTA. Suspensions were sonicated on ice with four pulses of 30 s at 28% of amplitude and 30 s intervals using a VC 750 sonicator (Bioblock Scientific, Illkirch, France). Cellular debris were then removed by two centrifugation at 16 000 $\times$ g for 10 min at 4°C. Protein concentration was determined by the Bradford method (Bradford, 1976) with three replicates per sample. For each sample, 20 µg of proteins from the whole cell extract were deposited in a black µclear 96-well plate (Greiner Bio-One, Frickenhausen, Germany). After addition of 0.02% [w/v] SDS and 200 µM of the fluorogenic proteasome substrate suc-LLVY-AMC (Enzo Life Sciences, Villeurbanne, France), the volume was adjusted to 50 µl with the lysis buffer. Each sample was deposited four times to obtain 12 replicates (three biological and four technical) for each condition. Negative controls containing everything but whole cell extracts or everything but the substrate were used to assess the fluorescent signal that is not due to the proteasomal activity. Wells content was homogenized, and plate was incubated for 30 min at 30°C in the dark. Fluorescence emission of the free AMC formed was then measured with a 360/40 nm excitation filter and a 460/40 nm emission filter using a microplate reader (Synergy HT).

Single-cell gel electrophoresis

Microscopic slides were coated with 0.8% agarose (D5-E, Euromedex, Strasbourg, France) in 10 mM PBS buffer and air dried for one night. Equal volumes of cell suspension and 2% [w/v] low melting point agarose (Sigma, Saint-Quentin

Fallavier, France) in PBS 10 mM were mixed to a final volume of 75 µl and deposited on a coated slide. Then 75 µl of 1% [w/v] LMP was added. Slides were then incubated in the lysis solution (2% [w/v] SDS, 30 mM EDTA, pH 7, Euromedex, Strasbourg, France) for 1.5 min. After 30 min incubation in TAE buffer pH7 for DNA unwinding, slides were subjected to electrophoresis for 15 min at 25 V and 300 mA in the same buffer. Slides were finally air dried at room temperature for 10 min and stained with 18 µM 4',6-Diamidino-2-phenylindole (DAPI, Sigma). Pictures of comets were taken with a fluorescence microscope Leica DM 4000B (Leica Microsystems, Nanterre, France) equipped with a digital camera DFC300 FX at a × 200 magnification and the A4 filter cube. For all observations, exposition, gain, colour saturation and gamma were respectively set at 410 ms, 1X, 0.35 and 1. At least 10 pictures were taken for each slide representing approximately 50 comets for each condition. NIH ImageJ analysis software (Schneider *et al*., 2012) has been used with a comet assay plug-in (http://www.med.unc.edu/microscopy/resources/imagej-plugins-and-macros/comet-assay) to score comets parameters: tail length, tail moment and percentage of DNA in the tail and data were processed with the MINITAB 15 Statistical software (http://it.minitab.com/fr-fr/products/minitab/free-trial.aspx).

Acknowledgements

The study was financed by the ANR 07-BLANC-0118 project (Agence Nationale de la Recherche). David Halter and Jérémy Andres were supported by a grant from the French Ministry of Education and Research. This work was performed within the framework of the research network 'Arsenic Metabolism in Micro-organisms' (GDR2909-CNRS).

References

Ali, W., Isayenkov, S.V., Zhao, F.J., and Maathuis, F.J. (2009) Arsenite transport in plants. *Cell Mol Life Sci* **66:** 2329–2339.

Aposhian, H.V. (1997) Enzymatic methylation of arsenic species and other new approaches to arsenic toxicity. *Annu Rev Pharmacol Toxicol* **37:** 397–419.

Arsène-Ploetze, F., Koechler, S., Marchal, M., Coppée, J.Y., Chandler, M., Bonnefoy, V., *et al*. (2010) Structure, function, and evolution of the *Thiomonas* spp. genome. *PLoS Genet* **6:** e1000859.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., *et al*. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25:** 25–29.

Baker, B.J., Tyson, G.W., Goosherst, L., and Banfield, J.F. (2009) Insights into the diversity of eukaryotes in acid mine drainage biofilm communities. *Appl Environ Microbiol* **75:** 2192–2199.

Bertin, P.N., Heinrich-Salmeron, A., Pelletier, E., Goulhen-Chollet, F., Arsène-Ploetze, F., Gallien, S., *et al*. (2011) Diversity of metabolic interactions between main partners inside an arsenic-rich microbial ecosystem revealed by meta- and proteo-genomics. *ISME J* **5:** 1735–1747.

Bradford, M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* **72:** 248.

Bruneel, O., Duran, R., Casiot, C., Elbaz-Poulichet, F., and Personné, J.-C. (2006) Diversity of microorganisms in Fe-As-rich acid mine drainage waters of Carnoulès, France. *Appl Environ Microbiol* **72:** 551–556.

Bruneel, O., Pascault, N., Egal, M., Bancon-Montigny, C., Goñi-Urriza, M.S., Elbaz-Poulichet, F., *et al*. (2008) Archaeal diversity in a Fe-As rich acid mine drainage at Carnoulès (France). *Extremophiles* **12:** 563–571.

Casiot, C., Morin, G., Juillot, F., Bruneel, O., Personné, J.C., Leblanc, M., *et al*. (2003) Bacterial immobilization and oxidation of arsenic in acid mine drainage (Carnoulès creek, France). *Water Res* **37:** 2929–2936.

Dasgupta, S., Fang, J., Brake, S.S., Hasiotis, S.T., and Zhang, L. (2012) Biosynthesis of sterols and wax esters by *Euglena* of acid mine drainage biofilms: implications for eukaryotic evolution and the early Earth. *Chem Geol* **306:** 139–145.

Denef, V.J., Mueller, R.S., and Banfield, J.F. (2010) AMD biofilms: using model communities to study microbial evolution and ecological complexity in nature. *ISME J* **4:** 599–610.

Ding, W., Liu, W., Cooper, K.L., Qin, X.J., De Souza Bergo, P.L., Hudson, L.G., and Liu, H.S.. (2009) Inhibition of poly(ADP-ribose) polymerase-1 by arsenite interferes with repair of oxidative DNA damage. *J Bio Chem* **284:** 6809–6817.

Duquesne, K., Lebrun, S., Casiot, C., Bruneel, O., Personné, J.C., Leblanc, M., *et al*. (2003) Immobilization of arsenite and ferric iron by *Acidithiobacillus ferrooxidans* and its relevance to acid mine drainage. *Appl Environ Microbiol* **69:** 6165–6173.

Duquesne, K., Lieutaud, A., Ratouchniak, J., Muller, D., Lett, M.C., and Bonnefoy, V. (2008) Arsenite oxidation by a chemoautotrophic moderately acidophilic *Thiomonas* sp.: from the strain isolation to the gene study. *Environ Microbiol* **10:** 228–237.

Eddy, S.R. (1998) Profile hidden Markov models. *Bioinformatics* **14:** 755–763.

Egal, M., Casiot, C., Morin, G., Parmentier, M., Bruneel, O., Lebrun, S., and Elbaz-Poulichet, F. (2009) Kinetic control on the formation of tooeleite, schwertmannite and jarosite by *Acidithiobacillus ferrooxidans* strains in an As(III)-rich acid mine water. *Chem Geol* **265:** 432–441.

Fisher, R.A. (1925) *Statistical Methods for Research Workers*. Edinburgh, Scotland: Oliver and Boyd.

Flora, S.J. (2011) Arsenic-induced oxidative stress and its reversibility. *Free Radic Biol Med* **51:** 257–281.

González-Mendoza, D., Morales-Trejo, A., and Brito-Vera, H. (2009) A rapid and efficient method for isolation of total RNA from *Euglena gracilis* (Euglenoidea). *Genet Mol Res* **8:** 482–486.

Halter, D., Casiot, C., Simon, S., Heipieper, H.J., Marchal, M., Lièvremont, D., *et al*. (2012a) Surface properties and intracellular speciation revealed an original adaptive mechanism to arsenic in the acid mine drainage bio-indicator *Euglena mutabilis*. *Environ Poll* **93:** 1735–1744.

Halter, D., Goulhen-Chollet, F., Gallien, S., Casiot, C., Hamelin, J., Gilard, F., *et al*. (2012b) *In situ* proteo-metabolomics reveals metabolite secretion by the acid mine drainage bio-indicator, *Euglena mutabilis*. *ISME J* **6:** 1391–1402.

Jacobson, T., Navarrete, C., Sharma, S.K., Sideri, T.C., Ibstedt, S., Priya, S., *et al*. (2012) Arsenite interferes with protein folding and triggers formation of protein aggregates in yeast. *J Cell Sci* **125:** 5073–5083.

Johnson, D.B., and Hallberg, K.B. (2003) The microbiology of acidic mine waters. *Res Microbiol* **154:** 466–473.

Kosmala, S., Karnkowska-Ishikawa, A., Milanowski, R., Kwiatowski, J., and Zakrys, B. (2009) Phylogeny and systematic of *Euglena* (euglenaceae) species with axial, stellate chloroplast based on morphological and molecular date – new taxa, emended diagnoses, and epitypifications. *J Phycol* **45:** 464–481.

Li, W. (2009) Analysis and comparison of very large metagenomes with fast clustering and functional annotation. *BMC Bioinformatics* **10:** 359.

Olaveson, M.M., and Nalewajko, C. (2000) Effects of acidity on the growth of two *Euglena* species. *Hydrobiologia* **433:** 39–56.

Olaveson, M.M., and Stokes, P.M. (1989) Responses of the acidophilic alga *Euglena mutabilis* (Euglenophyceae) to carbon enrichment at pH 3. *J Phycol* **25:** 529–539.

Punta, M., Coggill, P.C., Eberhardt, R.Y., Mistry, J., Tate, J., Boursnell, C., *et al*. (2012) The Pfam protein families database. *Nucleic Acids Res* **40:** D290–D301.

Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012) NIH image to ImageJ: 25 years of image analysis. *Nat Methods* **9:** 671–675.

Selengut, J.D., Haft, D.H., Davidsen, T., Ganapathy, A., Gwinn-Giglio, M., Nelson, W.C., *et al*. (2007) TIGRFAMs and Genome Properties: tools for the assignment of molecular function and biological process in prokaryotic genomes. *Nucleic Acids Res* **35:** 260–264.

Strimmer, K. (2008) fdrtool: a versatile R package for estimating local and tail area-based false discovery rates. *Bioinformatics* **24:** 1461–1462.

Sun, S., Chen, J., Li, W., Altintas, I., Lin, A., Peltier, S., *et al*. (2011) Community cyberinfrastructure for Advanced Microbial Ecology Research and Analysis: the CAMERA resource. *Nucleic Acids Res* **39:** D546–D551.

Valente, T.M., and Gomes, C.L. (2007) The role of two acidophilic algae as ecological indicators of acid mine drainage sites. *J Iberian Geol* **33:** 283–294.

Wysocki, R., and Tamás, M.J. (2010) How *Saccharomyces cerevisiae* copes with toxic metals and metalloids. *FEMS Microbiol Rev* **34:** 925–951.

Zhou, J., Meng, R., Sui, X., Meng, L., Jia, J., and Yang, B. (2005) Effects of administration styles of arsenic trioxide on intracellular arsenic concentration, cell differentiation and apoptosis. *Haematologica* **90:** 1277–1279.

DISCUSSION

Les résultats présentés dans ce travail soulignent que la stratégie d'adaptation à l'arsenic mise en place par *E. mutabilis* repose sur une diversité de fonctions biologiques. En effet, la comparaison des transcriptomes des deux espèces d'euglènes obtenus après culture en présence ou en absence d'As(III) à permis d'identifier des fonctions spécifiquement plus exprimées, voire induites par l'arsenic, chez *E. mutabilis*. Ces différences physiologiques soulignent potentiellement l'évolution de cette souche liée à son adaptation à sa niche écologique particulière, le DMA de Carnoulès.

Ce travail confirme tout d'abord qu'*E. mutabilis* prévient plus efficacement l'accumulation intracellulaire d'arsenic qu'*E. gracilis*. Tandis qu'une étude a mis en évidence le rôle de la composition de l'enveloppe cellulaire dans ce processus (Halter et al., 2012a), les résultats présentés ici identifient des mécanismes actifs. En effet, des transcrits liés à l'extrusion de l'As(III), ou au transport spécifique du phosphate sont uniquement présents ou induits chez *E. mutabilis* en présence d'arsenic. Cependant, l'analyse des profils transcriptomiques ne permet pas d'expliquer l'accumulation des formes méthylées d'arsenic observée uniquement chez *E. gracilis*, des transcrits présentant des homologies avec des As(III)-méthyltransferases étant identifiés dans les deux organismes.

Cette étude montre ensuite qu'en plus de limiter la concentration intracellulaire d'arsenic, *E. mutabilis* induit des mécanismes pour pallier aux effets de l'arsenic résiduel et maintenir l'intégrité fonctionnelle des pools protéique et génomique. En effet, l'analyse des transcrits et les cultures en présence d'un inhibiteur de la traduction montrent qu'*E. mutabilis* présente constitutivement un taux de synthèse protéique plus important qu'*E. gracilis*, celui-ci pouvant assurer le remplacement de protéines endommagées. De plus, l'induction de la traduction par l'arsenic dans la première souche est liée à un renouvellement du pool protéique au vu de l'induction de l'activité du protéasome et de l'expression des gènes liés. Ensuite, des fonctions impliquées dans la réparation de l'ADN sont constitutivement plus exprimées, voire induites en présence d'arsenic, chez *E. mutabilis*, et les tests de comète confirment que cette souche accumule moins de cassure dans l'ADN qu'*E. gracilis*. L'ensemble de ces fonctions peuvent finalement être mises en relation avec le métabolisme de l'azote, lié à la synthèse des précurseurs que sont les acides aminés et nucléiques, qui semble plus important chez *E. mutabilis* en présence d'arsenic qu'en son absence ainsi qu'en comparaison à *E. gracilis*.

Enfin, si ces différents processus cellulaires peuvent expliquer en partie l'hypertolérance à l'arsenic d'*E. mutabilis*, sa réponse à l'arsenic repose surement sur une plus grande variété de mécanismes. En effet, si cette étude s'est concentrée sur les différences entre les deux souches d'euglènes afin de mettre en évidence les processus adaptatifs développés par *E. mutabilis*, elle ne dresse pas un inventaire exhaustif des fonctions impliquées dans la réponse au stress arsénié. Par ailleurs, le maintien de l'euglène aussi bien que d'autres organismes dans des environnements toxiques tels que les DMA peut également dépendre d'interactions avec différents partenaires. Cette problématique a ainsi été abordée dans la suite de ce travail.

CHAPITRE III

INTERACTION MÉTABOLIQUE ENTRE *EUGLENA MUTABILIS* ET *THIOMONAS* SP. CB2

INTRODUCTION

La troisième partie de ce travail repose sur l'étude du rôle potentiel des interactions procaryote-eucaryote au sein de la communauté microbienne de Carnoulès et plus précisément celles impliquant *E. mutabilis*. En effet, à l'heure actuelle, peu d'études se concentrent sur les interactions impliquant des protistes photosynthétiques et des bactéries, malgré leur rôle potentiellement important dans le fonctionnement des écosystèmes.

Comme cela a été mentionné précédemment, l'approche de génomique descriptive et fonctionnelle appliquée à l'étude de la communauté bactérienne de Carnoulès ainsi que le métabolisme photosynthétique d'*E. mutabilis* suggèrent l'existence d'interactions trophiques entre ces deux composantes biologiques de l'écosystème de Carnoulès (Johnson and Hallberg, 2003; Das et al., 2009; Bertin et al., 2011). Afin de mieux comprendre ces relations trophiques et notamment l'apport de l'euglène dans la production primaire du site, une stratégie employant des outils métabolomique et protéomique, a permis de montrer qu'*E. mutabilis* sécrète sélectivement différents composés organiques et que certains sont retrouvés *in situ*, à Carnoulès (Halter et al., 2012b). En effet, sur les 26 métabolites identifiés dans l'eau du site, 15 sont activement excrétés par *E. mutabilis* en milieu synthétique. Ces derniers correspondent notamment à des sucres tels que le mannitol, le glucose et le fructose, à des acides aminés comme la glycine, la thréonine et la tyrosine ou à d'autres composés, le glycérol et l'urée en étant des exemples.

Parmi ces derniers métabolites sécrétés par l'euglène et identifiés *in situ* à Carnoulès, l'urée est l'un des plus abondants (Halter et al., 2012b). De manière intéressante, le séquençage du métagénome ainsi qu'une approche comparative par puces CGH, révèlent la présence de gènes codant pour les voies de dégradation de l'urée dans des souches appartenant au genre *Thiomonas* et présentes à Carnoulès, suggérant l'existence d'une interaction entre le protiste et ces bactéries (Arsène-Ploetze et al., 2010; Bertin et al., 2011). De plus, *E. mutabilis* et *Thiomonas* spp. sont des acteurs importants des DMA, participant à la production primaire et l'atténuation de l'arsenic, respectivement, et fréquemment retrouvés dans ces environnements (Johnson and Hallberg, 2003; Valente and Gomes, 2007). La suite de ce travail a ainsi conduit à étudier cette interaction trophique potentielle entre *E. mutabilis* et *Thiomonas* sp. CB2, isolée de ce site.

Dans ce but, différentes expériences ont été menées mêlant des approches de culture en milieu synthétique ou dans l'eau du Reigous filtrée, axéniques ou de cross- et co-cultures, à de la génomique, de la chimie et de la microscopie.

L'ensemble de ces résultats a conduit à la rédaction du manuscrit suivant qui a été soumis à *the ISME Journal*. Néanmoins, les commentaires récemment reçus suggèrent de scinder cet article en deux parties distinctes, ce qui est en cours. Des points soulevés par les rapporteurs ont toutefois été abordés dans la discussion. J'ai contribué à la rédaction de cet article et effectué les différents tests physiologiques concernant l'euglène, des tests biochimiques et métabolomiques ainsi que les observations et analyses liées à la microscopie.

MANUSCRIT

Eukaryote-prokaryote symbiotic interactions promote natural attenuation of toxic metals in acid mine drainage

Florence Arsène-Ploetze[a,1,3], Jérémy Andres[a,1], Corinne Casiot[b], Valérie Barbe[c], Pierre-Etienne Cholley[a], Jacques Faerber[d], Kelle C. Freel[a], David Halter[a,2], Dimitri Heintz[e], Sandrine Koechler[a], Raphael Lugan[e], Ghislaine Magdelenat[c], Marie Marchal[a,3], Frédéric Plewniak[a], Fabienne Seby[f], Philippe N. Bertin[a]

[a]Laboratoire Génétique Moléculaire, Génomique et Microbiologie, UMR7156, CNRS-Université de Strasbourg, Département Microorganismes, Génomes, Environnement, Equipe Ecophysiologie Moléculaire des Microorganismes, Institut de Botanique, 28 rue Goethe, 67083 Strasbourg, France; [b]Laboratoire Hydrosciences Montpellier, UMR 5569, CNRS-UMI, UM II, IRD, Université Montpellier 2, CCMSE, Place E. Bataillon, 34095 Montpellier Cedex 05, France; [c] Laboratoire de Finition, CEA-IG-Genoscope, Evry, France; [d]Institut de Physique et Chimie des Matériaux de Strasbourg, IPCMS-DSI, Département Surfaces et Interfaces, 23, rue de Loess, BP43, 67034 Strasbourg Cedex 2, France ; [e]Plateforme Métabolomique, UPR2357, CNRS, Institut de Biologie Moléculaire des Plantes, Institut de Botanique, 28, rue Goethe, 67083 Strasbourg, France;; and [f]Hélioparc Pau Pyrénées, 2, avenue du Président Angot, 64053 PAU Cedex 9, France.

[1]F.A.-P. and J.A. contributed equally to this work.

[2]Present address: Santé de la Vigne et Qualité du Vin, INRA UMR1131, 28 rue de Herrlisheim, 68021 Colmar, France

[3]To whom correspondence should be addressed. Telephone number: +33 (0)3 68 85 18 12. E-mail: ploetze@unistra.fr

Running title: Metal attenuation promoted by microbial symbiosis

Keywords: Aluminum/Arsenic/*Euglena mutabilis*/*Thiomonas*/Urea degradation

Subject Category: Microbe-microbe and microbe-host interactions

Abstract

Symbiotic relationships between protists and bacteria are likely to be important for the survival of microorganisms in highly toxic environments, but such interactions are poorly understood. In this study, we characterized the metabolic interactions between two microorganisms, the protist *E. mutabilis* and a bacterium of the *Thiomonas* genus, both found in environments heavily polluted with toxic metals, including arsenic. We demonstrated that the protist produces urea in polluted waters. This compound can be degraded by the bacterium, and this degradation activity, in turn, leads to the precipitation of toxic metals and promotes photosynthesis in *Euglena*. Thus, these microbial activities speed up the natural attenuation mechanisms observed in this type of toxic environment. Our data illustrate fundamental syntrophic processes that may result from the co-evolution of two organisms in a hostile environment.

Introduction

Symbiotic interactions between microorganisms are crucial for the functioning of ecosystems and the recycling of chemical elements, particularly in toxic environments such as Acid Mine Drainages (AMD) where growth conditions are harsh (Wilmes and Bond 2009). Generally, mutualistic relationships play an essential role in the development of microbial communities in toxic environments because they enable the survival and growth of microorganisms that would otherwise not be able to thrive (Jones et al. 2012; Johnson and Amarasekare 2013). Due to their ecological relevance, many mutually beneficial relationships have been well studied, e.g. between plants and mycorrhizal fungi (Kiers et al. 2011; Fellbaum et al. 2012), invertebrates and bacteria (Woyke et al. 2006; Engel et al. 2012), bacteria and bacteria (Byrne et al. 2013), and bacteria and fungi (Hughes and Sperandio 2008; McGenity et al. 2012). While symbiotic interactions between bacteria and protists have received relatively little attention, recent studies suggest that they are crucial in various aquatic ecosystems (Baker et al. 2003; Amin et al. 2009; Gast et al. 2009; Nowack and Melkonian 2010; Nancucheo and Johnson 2012; Martinez-Garcia et al. 2012). For instance, metabolic interactions between bacterial species and *Euglena mutabilis*, a photosynthesizing bio-indicator of AMD (Brake et al. 2001a; Brake et al. 2001b; Brake and Hasiotis 2010), have been suggested to be essential for the functioning and development of microbial communities in the AMD-impacted creek of Carnoulès, France (Bertin et al. 2011; Halter et al. 2012). In this AMD, the microbial community composed of bacteria, archaea, and protists has been stable for several years (Bruneel et al. 2003; Duquesne et al. 2003; Bruneel et al. 2006; Duquesne et al. 2008; Bryan et al. 2009; Bruneel et al. 2011; Slyemi et al. 2011; Halter et al. 2011; Volant et al. 2012). Various *Thiomonas* bacteria isolated from this river were previously

characterized and it has been proposed that, through oxidizing arsenite, *Thiomonas* strains promote the precipitation of arsenic, leading to a natural attenuation process (Bruneel et al. 2003; Morin et al. 2003; Duquesne et al. 2008; Bryan et al. 2009; Arsène-Ploetze et al. 2010; Egal et al. 2010; Slyemi et al. 2011; Slyemi et al. 2013). In addition, environmental genomic analyses indicated that *E. mutabilis* is a primary producer, excreting organic compounds which may be consumed by bacterial species (Bertin et al. 2011; Halter et al. 2012). Among the organic substances produced by this protist in synthetic media and found in the interstitial water from sediments of the AMD-impacted creek, urea was one of the most abundant (Halter et al. 2012). Interestingly, genomic sequences revealed that *Thiomonas* strains harbor genes likely involved in urea degradation (Arsène-Ploetze et al. 2010; Bertin et al. 2012). According to these data, we hypothesized that this activity may be a benefit to *E. mutabilis*, and might be crucial for mutualistic interactions (Bertin et al. 2011). In this study we analyzed the role of urea and urea degradation activity on metal oxidation and precipitation, using genomic, chemical, and microscopy based approaches.

Material and methods

Microbial strains, growth conditions and media. *Tm.* sp. CB2 was isolated from AMD-impacted water in the Carnoulès AMD (France), at the source of the Reigous creek (Figure S1) (Arsène-Ploetze et al. 2010) and *Tm. intermedia* K12 was isolated from a corroded concrete wall of the Hamburg sewer system (Milde et al. 1983). Both strains are resistant to arsenic (Table S1) (Arsène-Ploetze et al. 2010) . *E. mutabilis* is present in the upper layer of sediments of the Reigous creek, between the source of the Reigous creek and the confluence with the Amous river (Halter et al. 2011). The strain used in this study was isolated at the sampling point "COWG", 30 m downstream from the creek source where these cells are abundant (Figure S1). The growth capacities of *Thiomonas* and *E. mutabilis* cells were tested in Reigous creek water (AMD-impacted water) collected the 24th of January 2012 (Table S2 and Figure S1) that was filtered twice using 0.22 µm filters. In previous studies, urea was found in the interstitial water from sediment of this AMD-impacted creek (Halter et al. 2012). However, in the AMD-impacted water used for this study and sampled from the creek, using a different metabolomics approach (supplementary data), urea was not detected (Figure S2b). *Thiomonas* strains were routinely cultured on m126 solid or liquid medium (Bryan et al. 2009). After growth, cells were centrifuged for 15 minutes at 4 500 g, washed in 9 g/l NaCl and resuspended in filtered AMD-impacted water at the initial Optical Density (OD_{600nm}) noted in the figure legend. Urea was added to m126 medium or AMD-impacted water when mentioned in the text at a final concentration of 1 g/l. *E. mutabilis* was routinely cultured on minimal solid agar medium (MAM, pH 3) (Olaveson and Stokes 1989) at 25°C with a 16 h/8 h light-dark photoperiod and 45 µmol.m^{-2}.s^{-1} photon flux density (Halter et al. 2011). Cells were harvested and resuspended in 9 g/l NaCl or AMD- impacted water for inoculation. For all experiments, *E. mutabilis* cells were

grown in Erlenmeyer flasks with a 7 g/l agar layer on the bottom and, unless otherwise stated, were inoculated at an initial concentration of 1 mg fresh weight of cells per ml.

Physiological tests. *E. mutabilis* aluminum minimum inhibitory concentration (MIC) determination and growth experiments with aluminium only or a mixture of aluminium, iron and arsenic were performed by inoculating the strain on MAM solid medium consistently from the same *Euglena* suspension, or in liquid media amended with these metals. Aluminum was added from a stock solution of 5 g/l aluminum obtained from aluminum sulfate and the pH value of the media was adjusted to 3. Aluminum concentrations ranging from 0 to 100 mg/l were used to assess the effect of aluminum on *E. mutabilis* growth in liquid medium. These experiments were performed in triplicate for up to 34 days. The synergic effect of iron, arsenic and aluminum was tested in MAM liquid medium supplemented with these metals with the final concentrations corresponding to those measured at the COWG point (Table S2, 891.5 mg/l iron, 115. 2 mg/l arsenite and 36.95 mg/l aluminum). Stock solutions of arsenite [As(III)] and iron [Fe(II)] were prepared at 50 g/l and 8.915 g/l with sodium arsenite and iron sulfate, respectively, as follows: first, hydrochloric acid was added to the arsenite stock solution before adding it to the MAM medium in preparation. Second, a small amount of sulfuric acid was added to the iron before dissolving it in water, and then the pH was slowly adjusted to 3.0 to obtain the stock solution. Third, the pH of the MAM medium was adjusted to 3.0 before adding the metals. Finally the MAM and MAM with metals media were sterilized by filtration with 0.22 µm filters.

Growth of *E. mutabilis* in the different experiments was evaluated by direct cell counts, or by measuring OD_{665nm}. Determination of cell concentration by counting was done in two replicates with a hemocytometer. The OD was read at 665 nm, which corresponds to the maximum of absorption for chlorophyll a, with a Biorad spectrophotometer after pelleting the

cells for 3 min at 3 750 g and lysis of the cell pellets, extraction of chlorophyll with 100 % ethanol, and removal of cell debris by centrifugation for 3 min at 6 000 g.

E. mutabilis was inoculated into filtered AMD-impacted water to assess the ability of this organism to excrete urea in the AMD-impacted water. Samples were taken at multiple time points, and cell concentration and OD were measured. The supernatant of centrifuged samples was filtered and stored at -20°C for subsequent urea detection (supplementary information).

The ability of each *Thiomonas* strain to degrade urea was tested with cells grown in liquid m126 media amended with or without urea. Cells were centrifuged for 15 minutes at 4 500 g, washed and resuspended in PBS buffer (8 g/l NaCl, 0.2 g/l KCl, 1.44g/l Na_2HPO_4 and 0.24 g/l KH_2PO_4) with a pH adjusted to 5.1. One volume of cell suspension was mixed with 1 volume of BCP reagent (1 g/l of Bromo-Cresol Purple with or without 50 g/l urea). A violet color appearing after 5 minutes indicated urea degradation activity whereas an orange color indicated no activity. This urea degradation activity was confirmed by metabolomics: supernatants of centrifuged samples were filtered and stored at -20°C for urea detection (supplementary information). Mortality rate of *Thiomonas* strains was measured with early stationary phase cells grown in m126, centrifuged, washed once with 9 g/l NaCl and resuspended in AMD-impacted water. To determine cell survival, cultures were diluted in 9 g/l NaCl solution and plated on m126. After 10 days of incubation at 30°C, colonies were enumerated. The ferrous iron concentration was measured using the methods described previously (Fadrus and Maly 1975).

The effect of urea degradation activity on the AMD-impacted water was assessed by adding purified urease (Merck) in the sterile AMD-impacted water, in the presence or the absence of 1 g/l urea. 10 U of purified urease were used in 5 ml of water samples. As a negative control, urease was inactivated before addition to samples by heating for 5 min at 95 °C. Samples were then incubated at 30°C for 24 h and pictures of the pellets were taken after a centrifugation step at

3 750 x g for 10 min. Precipitate formation in AMD-impacted water in the presence of urea and active urease was tested in three independent replicates. The elementary compositions of precipitates were then analyzed as described in supplementary information.

Cross-culture experiments. To assess the effect of *Thiomonas* strains on *E. mutabilis* fitness and growth, cross-culture experiments were performed by cultivating *E. mutabilis* in *Thiomonas* growth culture supernatant. To generate the supernatant, *Tm. intermedia* K12 and *Tm*. sp. CB2 were pre-cultivated in m126 medium, then inoculated at OD_{600nm} of 0.2 (from cells in stationary phase) as well as at OD_{600nm} of 0.02 in AMD-impacted water, respectively (similar results were obtained in both cases). Both conditions were tested with and without the addition of 1 g/l urea. After 4 days of *Thiomonas* cultivation, both cultures have reached an OD_{600nm} of 0.2. Samples were centrifuged for 15 min at 4 500 g and filtered through 0.22 µm filters to remove bacterial cells. Filtrates were then inoculated with *E. mutabilis* to a final concentration of 0.5 mg fresh weight per ml. Samples were taken at multiple time points to measure OD_{665nm}. Cell concentration was determined at 16 days of growth.

Co-culture experiments. *Tm. intermedia* K12 or sp. CB2 were pre-cultured in liquid m126 medium supplemented with or without 1 g/l urea, at 30°C for 72 h. Cells were then harvested by centrifugation for 5 min at 3 750 g, washed with 9 g/l NaCl, collected by centrifugation and re-suspended in filtered AMD-impacted water supplemented with or without 1 g/l urea to an initial OD_{600nm} of approximately 0.03. *E. mutabilis* cells were harvested from MAM plates, washed and re-suspended in the AMD-impacted water samples supplemented or not with *Thiomonas* cells and urea. Supernatant samples were taken at different time points to measure the cell concentration, the pH of the culture, and the OD_{665nm}.

Microscopy. To assess physical association of the microorganisms, observations were conducted using transmission electron microscopy (TEM), scanning electron microscopy (FE-SEM) and

environmental SEM (ESEM) (supplementary information).

Chemical analysis. The AMD-impacted water was stored at 4°C during the experiments that were completed during the course of one year. Element concentrations were measured by inductively coupled plasma-mass spectrometry (ICP-MS), inductively coupled plasma-atomic emission spectroscopy (ICP-AES) or by Gas-Chromatography-mass spectrometry (GC-MS) (supplementary information).

Chromatography & mass spectrometry for metabolomic analysis. Urea identification was completed as previously described by (Halter et al. 2012) (supplementary information).

Genome sequence. The genome of CB2 was sequenced by the Eurofins/MWG Company using 454/Roche FLX technology. Two libraries were used for sequencing: one single read shotgun library and one mate-pair library with an 8kb insert size. The assembly with Newbler complemented by optical map data yielded 92 contigs that could be ordered into 9 scaffolds spanning a total of 3 739 437 bases.

Total RNA extraction from sediment of the Reigous creek and reverse transcription. The total microbial community was recovered after Nycodenz gradient density separation as previously described (Bertin et al. 2011), from sediments sampled in May 2007, and stored at -80°C. RNA extraction and RT-PCR were performed as was done previously (Bertin et al. 2011), with some modifications as described in supplementary information.

Results and Discussion

The bacterial urea degradation activity promotes toxic metal precipitation in AMD-impacted water. Two distinct enzymes, urease and urea amidolyase, are known to degrade urea to ammonia and carbon dioxide. In bacteria, the urea amidolyase activity is the result of two separate enzymes, the urea carboxylase and the allophanate hydrolase (UCA-allophanate hydrolase pathway) (Kanamori et al. 2004). The genome sequence of *Tm.* sp. CB2 obtained in this study revealed the presence of genes involved in urea degradation, *i. e. urt* genes implicated in urea import, *ure* genes encoding urease, and genes essential to the Urea Carboxylase (UCA)-allophanate hydrolase pathway in a predicted region of genomic plasticity (RGP) that corresponds to the signature of genomic islands (Vallenet et al. 2009) (Figure S3a). In addition, these genes are absent from the genome of another *Thiomonas* strain, *Tm. intermedia* K12, whose genome has been previously sequenced (http://www.jgi.doe.gov, Table S1). In agreement with genomic data, *Tm*. sp. CB2 was able to degrade urea in m126 synthetic media, whereas *Tm. intermedia* K12 was not (Figure S3b and c). This urea degradation activity was observed in the presence of 1 g/l urea in the culture medium or more, but not at concentrations of 0.1 g/l or lower, suggesting that expression of these genes is inducible. In synthetic media, we observed that urea could not be used by *Thiomonas* strains as the sole carbon or nitrogen source and did not increase growth rate, survival or arsenic resistance (data not shown). This result suggests that these genes are functional but do not allow bacteria to acquire new trophic capacities in the conditions tested, *i. e.* in synthetic media.

We then determined if the urea degradation activity is required for *Tm.* sp. CB2 to grow in the water from where this bacterium was isolated, which is the AMD-impacted creek of

Carnoulès (France), contaminated with several toxic metals (Figure S1, Table S2). When *Tm.* sp. CB2 was incubated in AMD-impacted water, a higher pH was observed in the presence of urea as compared to the absence of urea and as compared to AMD-impacted water inoculated with *Tm. intermedia* K12 in the presence or in the absence of urea (Figure 1a, analysis of covariance performed with MATLAB R2014a *aoctool and multcompare* commands, p-value < 0.01 for intercept difference). This pH increase was explained by the production of ammonia due to urea degradation activity. *Tm.* sp. CB2 and *Tm. intermedia* K12 were able to grow in AMD-impacted water in the absence and presence of urea (Figure S4 and data not shown). These observations and the mortality rates of both *Thiomonas* strains obtained in the same conditions (Figure S4), suggest that the urea degradation activity did not afford the *Tm.* sp. CB2 strain a fitness advantage over *Tm. intermedia* K12 in AMD-impacted water. Remarkably, in these experiments, an orange color was observed in the presence of *Tm*. sp. CB2 and urea (Figure 1b and c). This color was due to an orange precipitate that was more apparent in this condition as compared to conditions where no urea degradation activity was possible (Figure 1b and c). As a negative control, when *Tm. intermedia* K12, which is unable to degrade urea, was inoculated in this water, no orange color was observed, even in the presence of urea (Figure 1b). An orange precipitate was also observed when purified urease and urea were added to sterile AMD-impacted water, whereas less precipitation was observed in samples when only urea was added or when urease was inactivated (Figure 1e). The presence of this orange precipitate suggested that oxidation and subsequent precipitation of iron had occurred when urease or bacteria with urea degradation activity were added. An analysis of covariance performed with MATLAB R2014a (*aoctool* and *multcompare* commands) on the quantity of soluble Fe(II) removed from the medium related to time, showed that the slope of soluble iron removal was significantly higher (p-value < 0.01) for *Tm*. sp. CB2 cultivated with urea than for other cultures (*i. e. Tm*. sp. CB2 without urea and *Tm.*

intermedia K12 with and without urea). A higher speed of soluble Fe(II) attenuation was therefore achieved in the cultures where *Tm*. sp. CB2 could degrade urea as compared to the samples where urea degradation was not possible (Figure 1d). In addition, using inductively coupled plasma-mass spectrometry (ICP-MS), we observed that the soluble iron, arsenic and aluminum concentrations after 11 days were lower in these samples than in the others (Table S3). Consequently, high amounts of these elements were found in the precipitates obtained from samples where urea degradation occurred (samples inoculated with *Tm*. sp. CB2, in the presence of urea) as compared to the other conditions tested (Table S3). No such important precipitation was observed in the case of other metals (Table S3). Finally, iron, arsenic and aluminum precipitation was also observed using X-ray microanalysis in the precipitates obtained from samples where urea degradation occurred (samples where urease was added or inoculated with *Tm*. sp. CB2, in the presence of urea, Figure S5). Altogether, these results revealed that the *Tm.* sp. CB2 urea degradation activity promoted the precipitation of iron, arsenic and aluminum.

Fe(II) present in the AMD-impacted water (Table S2, Figure 1d) was probably first oxidized and then precipitated in these conditions. *Tm.* sp. CB2 does not harbor the genes required for ferrous iron oxidation (Table S1) and was not able to perform ferrous iron oxidation in synthetic medium (data not shown). The iron precipitation observed in our experiments in the presence of urea was therefore probably related to abiotic oxidation. Considering the dependence of the Fe(II) oxidation rate on pH in natural waters (Sigg et al. 2006), the most appropriate explanation for the iron oxidation observed in our experiments is that it is due, at least in part, to the pH increase (as observed in our experiments, Figure 1), resulting from urea degradation and ammonia production. This explanation is consistent with the partial precipitation observed when the pH of the sterile AMD-impacted water was increased with the addition of ammonia (data not

shown). Fe(III) precipitation causes the co-precipitation of arsenic as previously demonstrated in AMD-impacted water (Morin et al. 2003; Casiot et al. 2003; Duquesne et al. 2003; Casiot et al. 2005). Alternatively, *Tm*. sp. CB2 was able to oxidize arsenite and produce arsenate in such conditions (data not shown), which was in agreement with previous studies demonstrating that *Thiomonas* strains expressed arsenite oxidase in AMD-impacted creek (Bertin et al. 2011). Arsenate is less soluble and probably precipitates with Fe(III) more efficiently than arsenite in these experiments, as previously proposed (Morin et al. 2003; Casiot et al. 2003; Casiot et al. 2005).

Remarkably, our study revealed that bacterial urea degradation activity affects aluminum precipitation. This is consistent with the aluminum precipitation observed when purified urease and urea were added to AMD-impacted water (Figure 1e and Figure S5). The X-ray microanalysis data (Figure S5) suggest that aluminum precipitated as aluminum oxyhydroxydes. Previous reports revealed that aluminum hydroxides or microcrystalline gibbsite flocs are formed at a pH ranging from 4.2 to 4.9 in AMD and form aggregates when the pH is higher than 5 (Nordstrom 1982; Nordstrom and Ball 1986; Furrer et al. 2002). The average pH in our samples was less than 4 (Figure 1a) but it is possible that the pH may reach 4.2 or more, in a microscale surrounding environment around the bacterial cells producing ammonia. Alternatively, it has been previously demonstrated that ammonia produced by urease reacts with water producing hydroxyl ions, which react with aluminum sulfate to produce particles of aluminum basic sulfate (Simpson et al. 1998; Kara and Sahin 2000). The AMD-impacted water characteristics, which include a high concentration of sulfate (Table S2), are therefore compatible with conditions required to form such particles. Interestingly, in the AMD-impacted creek, from which *E. mutabilis* and *Tm.* sp. CB2 were isolated, the concentration of aluminum in the water decreased

along the creek at the time of water sample collection, with the highest concentration at the source (Figure S1 and Table S2). In order to demonstrate that aluminum did indeed precipitate in this creek, the decrease of the soluble aluminum concentration (attenuation) between the source and our sampling point (COWG) was compared to the manganese attenuation that was supposedly due mainly to the dilution process. From January 2005 to September 2012, in most samples, the aluminum attenuation was consistently higher over time than that of manganese as shown by a Wilcoxon signed rank test (Figure 2a, MATLAB R2014a *signrank* command, p-value < 0.01,). These data revealed for the first time that aluminum attenuation occurred in this Reigous creek and was due to precipitation of aluminum, as observed in other AMDs or in the circumneutral pH Amous river downstream of the Reigous creek (Furrer et al. 2002; Adra et al. 2013). To test if such a precipitation observed *in situ* could be correlated with bacterial urea degradation activity, the expression of *Thiomonas* genes encoding the urease (*ureC*), the allophanate hydrolase, and as a control, the arsenite oxidase (*aioA*) was determined by RT-PCR with RNA extracted from the community of this AMD-impacted creek, as previously (Bertin et al. 2011). Indeed, we observed that *Thiomonas* cells previously detected in these sediments (Bertin et al. 2011) expressed *aioA* and *ureC in situ* (Figure 2b). Altogether, our *ex situ* experiments and *in situ* observations, suggest that arsenic, iron and aluminum attenuation observed in AMD (Morin et al. 2003; Casiot et al. 2005; Egal et al. 2010) is promoted, at least in part, by ureolytic bacteria, such as *Thiomonas* spp.

The *Tm* sp. CB2 urea degradation activity enhances *E. mutabilis* fitness in AMD-impacted water. We then investigated if the bacterial urea degradation may be beneficial for the urea-producer, *E. mutabilis*. We first observed that *E. mutabilis* was able to grow in the AMD-impacted creek contaminated with several toxic metals (Figure S2a). However, while this protist

grew in these conditions, the total chlorophyll amount remained stable over time, meaning that the average chlorophyll content per cell decreased (Figure S2a). This revealed that *E. mutabilis* photosynthetic activity is inhibited in toxic AMD water. In these harsh conditions, *E. mutabilis* was probably able to grow through organoheterotrophy as previously described for *Euglena gracilis* and *Chlamydomonas acidophila* (Danilov and Ekelund 2002; Perreault et al. 2010).

It has been previously shown that *E. mutabilis* excretes urea when grown in synthetic media (Halter et al. 2012). Consistent with these previous studies, we observed that *E. mutabilis* produced urea when cultivated in AMD-impacted water (Figure S2b). To test whether the protist would benefit from bacterial urea degradation, cross-culture experiments were performed (Figure 3). *Tm.* sp. CB2 was incubated in AMD-impacted water, in the presence or absence of urea. After 4 days, this water was filtered and the cell-free supernatant was inoculated with *E. mutabilis* (Figure 3). A growth augmentation of *E. mutabilis*, revealed by a higher cell density and an increase in concentration of photosynthetic pigments (OD_{665nm}), was observed in water previously incubated with *Tm.* sp. CB2 grown in the presence of urea, but not with *Tm.* sp. CB2 incubated in the absence of urea, nor with water with urea previously incubated with *Tm. intermedia* K12 (Figure 3). To further test if *Tm* sp. CB2 urea degradation activity enhances *E. mutabilis* fitness in AMD-impacted water, we incubated both microorganisms together in sterile creek water (Figure S2c). Again, we observed a beneficial effect of *Tm.* sp. CB2 on *E. mutabilis* growth only in conditions where urea degradation was possible. Altogether, these data demonstrate that *Tm.* sp. CB2 has a beneficial effect on the protist due to the bacterial urea degradation activity. Additionally, TEM, SEM and ESEM observations performed on co-culture experiments in synthetic media or in AMD-impacted water suggested that *Tm*. sp. CB2 cells were

attached to the surface of the protist cells (Figure S6). The fact that both organisms physically interact may favor the exchange of urea produced by the protist and consumed by the bacterium.

We then hypothesized that the precipitation of metals promoted by the ureolytic *Thiomonas* sp. explained the enhanced *E. mutabilis* fitness observed in the presence of *Tm*. sp. CB2 and urea (Figure 3 and Figure S2c). When incubated in AMD-impacted water, the synthesis of *E. mutabilis* photosynthetic pigments was inhibited (Figure 3a and Figure S2a). Since aluminum has been shown to inhibit *Euglena* photosynthesis (Danilov and Ekelund 2002; Perreault et al. 2010), we tested if aluminum that was detected in the AMD-impacted water (Table S2) may cause inhibition of *E. mutabilis* photosynthesis. We evaluated the effect of aluminum on *E. mutabilis* fitness in synthetic liquid or on solid media (Figure 4, Figure S2d). We observed that growth and photosynthetic pigment synthesis are indeed inhibited in *E. mutabilis* in the presence of aluminum, at concentrations higher than 30 mg/l (Figure 4a), which correspond to the concentration measured in the AMD-impacted water (Table S2). It has been previously demonstrated that *E. mutabilis* grows on synthetic media in the presence of 2 500 mg/l arsenite, but a growth delay was observed on solid media in the presence of arsenite, starting with concentrations of 400 mg/l (Halter et al. 2011). Therefore, *E. mutabilis* was grown in synthetic media in the absence or in the presence of arsenic, iron and aluminum at the concentration measured in the AMD-impacted water (Table S2). In the presence of the three metals, the synthesis of photosynthetic pigment was inhibited (Figure 4b). Since bacterial urea degradation led to the precipitation of toxic metals in AMD-impacted water and reduced their bioavailability locally, we proposed that this urea degradation activity indirectly favors the photosynthetic activity of *E. mutabilis* and consequently its colonization of the creek. Moreover, bacterial urea degradation activity produces carbon dioxide, which could be used by the protist as a substrate in

the photosynthesis process. Altogether, these data demonstrate that *Tm.* sp. CB2 has a beneficial effect on the *E. mutabilis* due to its urea degradation activity, which promote iron, arsenic and aluminum precipitation and consequently prevent the inhibitory effect of these toxic metals on the protist photosynthesis.

Conclusions

Our data demonstrate that the bacterial degradation of urea accelerates iron, arsenic and aluminum precipitation, and favors *E. mutabilis* photosynthesis. *E. mutabilis* is able to have a beneficial relationship with *Thiomonas* and to produce urea in the AMD-impacted creek under study. Our data reveal that the bacteria could be attached to the surface of the protist, which would promote the exchange of urea and other metabolites, probably leading to the formation of a microscale environment with specific conditions of pH, toxic metal concentration or oxygen tension, conferring both micro-organisms better growth conditions and increased fitness. The protist photosynthetic activity allows for the production of organic compounds that are likely used by facultative organotrophs. These interactions might be therefore favorable for the bacteria of the microbial community, as previously proposed in the case of other contaminated sites (Nancucheo and Johnson 2012), and in particular for *Thiomonas* strains, which grow optimally in mixotrophic conditions in the presence of organic supplements (Bryan et al. 2009). Remarkably, in the presence of iron, arsenic and aluminum, *Euglena* cells may consume organic compounds via heterotrophic metabolism when they cannot photosynthesize. Thus, without bacterial urea degradation activity, photosynthesis would be inhibited and the beneficial interaction would be lost in favor of a competitive relationship for organic compound uptake. The cooperation we observed between distantly related organisms plays therefore an essential role in enhancing the

survival of both the bacteria and protists, allowing them to inhabit unique environmental niches in an otherwise toxic environment. *Thiomonas* strains are stable over time at the Carnoulès site (Bruneel et al. 2003; Casiot et al. 2003; Bruneel et al. 2006; Duquesne et al. 2008; Bruneel et al. 2011), and the persistence of *Thiomonas in situ* may be due, at least in part, to metabolic interactions with *E. mutabilis*. Interestingly, it is possible that, of the two partners, *Tm.* sp. CB2 may have acquired the capacity for metabolic interactions through horizontal gene transfer, since the genes encoding urea transporters, the urease, the UCA- and allophanate hydrolase were found in the same genomic island. Bacterial urea degradation activity may have allowed for bacterial survival within the unique ecological niche of the AMD and was perhaps selected for and maintained during bacterial and protist co-evolution since it benefits directly or indirectly both microorganisms. More work is needed to fully understand the evolutionary history of such symbiotic interactions and if both organisms are able to exchange other compounds. Nevertheless, our results represent essential progress towards clarifying the role of two important members of AMD microbial communities which were found to be ubiquitous in such toxic environments.

Acknowledgments

This study was financed by the « Observatoire de Recherche Méditerranéen de l'Environnement-Observatoire des Sciences de l'Univers » (OSU OREME) for the Physiochemical monitoring of the AMD, the EC2CO program (INEE, CNRS) and the RARE (ANR 07-BLANC-0118) (Agence Nationale de la Recherche). The authors thank Odile Bruneel for providing the *Tm.* sp. CB2 strain, Didier Lièvremont for ICP-AES experiments, Mathieu Erhardt for the TEM observations and Nicole Liewig for giving us access to the ESEM. Florence Arsène-Ploetze was supported by the CNRS (« mise en délégation pour activité de recherche au CNRS ») and Jérémy Andres and David Halter were supported by a grant from the French Ministry of Education and Research. Marie Marchal and Kelle C. Freel were supported by a studentship and a post-doc fellow from the Agence Nationale de la Recherche, ANR COBIAS, project PRECODD 2007, and ANR THIOFILM (ANR-12-ADAP-0013), respectively.

Conflict of interest statement:

The authors declare no conflict of interest.

Author contributions. F.A.-P., J.A., D.H., M.M., D.H., R.L., C.C., and P.N.B. designed research; F.A.-P., J.A., D.H., C.C., K.F., D.H., S.K., R.L., M.M. and F.S. performed research; V.B., J.F., P.-E.C., and F.P. contributed new reagents/analytic tools; F.A.-P., J.A., D.H., C.C., R.L., P.-E.C., D.H., M.M., F.P. and P.N.B. analyzed data; and F.A.-P., J.A., K.F. and P.N.B. wrote the paper.

Supplementary information is available at the ISME Journal website

References

Adra A, Morin Guillaume, Ona-Nguema G, Menguy N, Maillot F, Casiot Corinne, *et al.* (2013). Arsenic scavenging by aluminum-substituted ferrihydrites in a circumneutral pH river impacted by acid mine drainage. *Environ Sci Technol* **47**:12784–12792.

Amin SA, Green DH, Hart MC, Küpper FC, Sunda WG, Carrano CJ. (2009). Photolysis of iron-siderophore chelates promotes bacterial-algal mutualism. *Proc Natl Acad Sci USA* **106**:17071–17076.

Arsène-Ploetze F, Koechler S, Marchal M, Coppée J-Y, Chandler M, Bonnefoy V, *et al.* (2010). Structure, function, and evolution of the Thiomonas spp. genome. *PLoS Genet* **6**:e1000859.

Baker BJ, Hugenholtz P, Dawson SC, Banfield JF. (2003). Extremely acidophilic protists from acid mine drainage host Rickettsiales-lineage endosymbionts that have intervening sequences in their 16S rRNA genes. *Appl Environ Microbiol* **69**:5512–5518.

Bertin, Heinrich-Salmeron A, Pelletier E, Goulhen-Chollet F, Arsène-Ploetze Florence, Gallien Sébastien, *et al.* (2011). Metabolic diversity among main microorganisms inside an arsenic-rich ecosystem revealed by meta- and proteo-genomics. *ISME J*. **5**:1735-47.

Bertin P, Geist L, Halter D, Koechler S, Marchal M, Arsène-Ploetze F. (2012). Microbial arsenic response and metabolism in the genomics era. In:*The metabolism of arsenite*, Arsenic in the environment, Joanna M. Santini and Seamus A. Ward, pp. 145–159.

Brake S. S., Dannelly HK, Conners KA. (2001). Controls on the nature and distribution of an alga in coal mine-waste environments and its potential impact on water quality. *Environ Geol* **40**:458–469.

Brake S. S., Dannelly HK, Connors KA, Hasiotis S. T. (2001). Influence of water chemistry on the distribution of an acidophilic protozoan in an acid mine drainage system at the abandoned Green Valley coal mine, Indiana, USA. *Appl Geochem* **16**:1641–1652.

Brake Sandra S., Hasiotis Stephen T. (2010). Eukaryote-Dominated Biofilms and Their Significance in Acidic Environments. *Geomicrobiol J* **27**:534–558.

Bruneel, Duran R, Casiot C, Elbaz-Poulichet F, Personné J-C. (2006). Diversity of microorganisms in Fe-As-rich acid mine drainage waters of Carnoulès, France. *Appl Environ Microbiol* **72**:551–556.

Bruneel, Personné J-C, Casiot C, Leblanc M, Elbaz-Poulichet F, Mahler BJ, *et al.* (2003). Mediation of arsenic oxidation by Thiomonas sp. in acid-mine drainage (Carnoulès, France). *J Appl Microbiol* **95**:492–499.

Bruneel, Volant Aurélie, Gallien Sébastien, Chaumande B, Casiot Corinne, Carapito C, *et al.* (2011). Characterization of the Active Bacterial Community Involved in Natural Attenuation Processes in Arsenic-Rich Creek Sediments. *Microb Ecol* **61**:793–810.

Bryan CG, Marchal Marie, Battaglia-Brunet F, Kugler V, Lemaitre-Guillier C, Lièvremont D, *et al.* (2009). Carbon and arsenic metabolism in Thiomonas strains: differences revealed diverse adaptation processes. *BMC Microbiol* **9**:127.

Byrne DP, Potempa J, Olczak T, Smalley JW. (2013). Evidence of mutualism between two periodontal pathogens: co-operative haem acquisition by the HmuY haemophore of Porphyromonas gingivalis and the cysteine protease interpain A (InpA) of Prevotella intermedia. *Mol Oral Microbiol* **28**:219–229.

Casiot C, Lebrun S, Morin G, Bruneel O, Personné J C, Elbaz-Poulichet F. (2005). Sorption and redox processes controlling arsenic fate and transport in a stream impacted by acid mine drainage. *Sci Total Environ* **347**:122–130.

Casiot Corinne, Morin Guillaume, Juillot F, Bruneel Odile, Personné Jean Christian, Leblanc Marc, *et al.* (2003). Bacterial immobilization and oxidation of arsenic in acid mine drainage (Carnoulès creek, France). *Water Res* **37**:2929–2936.

Danilov RA, Ekelund NGA. (2002). Effects of short-term and long-term aluminium stress on photosynthesis, respiration, and reproductive capacity in a unicellular green flagellate (Euglena gracilis). *Acta Hydrochim Hydrobiol* **30**:190–196.

Duquesne K, Lebrun S, Casiot C, Bruneel O, Personné J-C, Leblanc M, *et al.* (2003). Immobilization of arsenite and ferric iron by Acidithiobacillus ferrooxidans and its relevance to acid mine drainage. *Appl Environ Microbiol* **69**:6165–6173.

Duquesne K, Lieutaud A, Ratouchniak J, Muller D, Lett M-C, Bonnefoy Violaine. (2008). Arsenite oxidation by a chemoautotrophic moderately acidophilic Thiomonas sp.: from the strain isolation to the gene study. *Environ Microbiol* **10**:228–237.

Egal M, Casiot C., Morin G., Elbaz-Poulichet F., Cordier MA, Bruneel O. (2010). An updated insight into the natural attenuation of As concentrations in Reigous Creek (southern France). *Applied Geochemistry* **25**:1949–1957.

Engel P, Martinson VG, Moran NA. (2012). Functional diversity within the simple gut microbiota of the honey bee. *Proc Natl Acad Sci U S A* **109**:11002–11007.

Fadrus H, Maly J. (1975). Suppression of iron(III) interference in the determination of iron(II) in water by the 1,10-phenanthroline method. Analyst **100**:549-54

Fellbaum CR, Gachomo EW, Beesetty Y, Choudhari S, Strahan GD, Pfeffer PE, *et al.* (2012). Carbon availability triggers fungal nitrogen uptake and transport in arbuscular mycorrhizal symbiosis. *Proc Natl Acad Sci USA* **109**:2666–2671.

Furrer G, Phillips BL, Ulrich K-U, Pöthig R, Casey WH. (2002). The Origin of Aluminum Flocs in Polluted Streams. *Science* **297**:2245–2247.

Gast RJ, Sanders RW, Caron DA. (2009). Ecological strategies of protists and their symbiotic relationships with prokaryotic microbes. *Trends Microbiol* **17**:563–569.

Halter David, Casiot Corinne, Heipieper HJ, Plewniak F, Marchal Marie, Simon S, *et al.* (2011). Surface properties and intracellular speciation revealed an original adaptive mechanism to arsenic in the acid mine drainage bio-indicator Euglena mutabilis. *Appl Microbiol Biotechnol*. **93**:1735-44.

Halter, Goulhen-Chollet F, Gallien Sebastien, Casiot Corinne, Hamelin J, Gilard F, *et al.* (2012). In situ proteo-metabolomics reveals metabolite secretion by the acid mine drainage bio-indicator, Euglena mutabilis. *ISME J* **6**:1391–1402.

Hughes DT, Sperandio V. (2008). Inter-kingdom signalling: communication between bacteria and their hosts. *Nat Rev Microbiol* **6**:111–120.

Johnson CA, Amarasekare P. (2013). Competition for benefits can promote the persistence of mutualistic interactions. *Journal of Theoretical Biology* **328**:54–64.

Jones EI, Bronstein JL, Ferrière R. (2012). The fundamental role of competition in the ecology and evolution of mutualisms. *Ann N Y Acad Sci* **1256**:66–88.

Kanamori T, Kanou N, Atomi H, Imanaka T. (2004). Enzymatic Characterization of a Prokaryotic Urea Carboxylase. *J Bacteriol* **186**:2532–2539.

Kara F, Sahin G. (2000). Hydrated aluminium sulfate precipitation by enzyme-catalysed urea decomposition. *J European Ceram Soc* **20**:689–694.

Kiers ET, Duhamel M, Beesetty Y, Mensah JA, Franken O, Verbruggen E, *et al.* (2011). Reciprocal rewards stabilize cooperation in the mycorrhizal symbiosis. *Science* **333**:880–882.

Martinez-Garcia M, Brazel D, Poulton NJ, Swan BK, Gomez ML, Masland D, *et al.* (2012). Unveiling in situ interactions between marine protists and bacteria through single cell sequencing. *The ISME Journal* **6**:703–707.

McGenity TJ, Folwell BD, McKew BA, Sanni GO. (2012). Marine crude-oil biodegradation: a central role for interspecies interactions. *Aquat Biosyst* **8**:10.

Milde K, Sand W, Wolff W, Bock E. (1983). Thiobacilli of the corroded concrete walls of the Hamburg sewer system. *Journal of General Microbiology* **129**:1327–1333.

Morin, Juillot F, Casiot Corinne, Bruneel Odile, Personné Jean-Christian, Elbaz-Poulichet Françoise, *et al.* (2003). Bacterial formation of tooeleite and mixed arsenic(III) or arsenic(V)-iron(III) gels in the Carnoulès acid mine drainage, France. A XANES, XRD, and SEM study. *Environ Sci Technol* **37**:1705–1712.

Nancucheo I, Johnson DB. (2012). Acidophilic algae isolated from mine-impacted environments and their roles in sustaining heterotrophic acidophiles. *Front Microbiol* **3**:325.

Nordstrom DK. (1982). The effect of sulfate on aluminum concentrations in natural waters: some stability relations in the system Al2O3-SO3-H2O at 298 K. *Geochimica et Cosmochimica Acta* **46**:681–692.

Nordstrom DK, Ball JW. (1986). The Geochemical Behavior of Aluminum in Acidified Surface Waters. *Science* **232**:54–56.

Nowack ECM, Melkonian M. (2010). Endosymbiotic associations within protists. *Philos Trans R Soc Lond, B, Biol Sci* **365**:699–712.

Olaveson M, Stokes P. (1989). Responses of the acidophilic alga Euglena mutabilis (euglenophyceae) to carbon enrichment at Ph 3. *J Phycol* **25**:529–539.

Perreault F, Dewez D, Fortin C, Juneau P, Diallo A, Popovic R. (2010). Effect of aluminum on cellular division and photosynthetic electron transport in Euglena gracilis and Chlamydomonas acidophila. *Environ Toxicol Chem* **29**:887–892.

Sigg L, Behra P, Stumm W. (2006). Chimie des milieux aquatiques; chimie des eaux naturelles et des interfaces dans l'environnement. Dunod. Paris.

Simpson RE, Habeger C, Rabinovich A, Adair JH. (1998). Enzyme-catalyzed inorganic precipitation of aluminum basic sulfate. *J Am Ceram Soc* **81**:1377–1379.

Slyemi D, Moinier D, Brochier-Armanet C, Bonnefoy Violaine, Johnson DB. (2011). Characteristics of a phylogenetically ambiguous, arsenic-oxidizing Thiomonas sp., Thiomonas arsenitoxydans strain 3As(T) sp. nov. *Arch Microbiol*. **193**:439-49.

Slyemi D, Moinier D, Talla E, Bonnefoy Violaine. (2013). Organization and regulation of the arsenite oxidase operon of the moderately acidophilic and facultative chemoautotrophic Thiomonas arsenitoxydans. *Extremophiles* **17**:911–920.

Vallenet D, Engelen S, Mornico D, Cruveiller S, Fleury L, Lajus A, *et al.* (2009). MicroScope: a platform for microbial genome annotation and comparative genomics. *Database (Oxford)* **2009**:bap021.

Volant A, Desoeuvre A, Casiot C, Lauga B, Delpoux S, Morin G, *et al.* (2012). Archaeal diversity: temporal variation in the arsenic-rich creek sediments of Carnoulès Mine, France. *Extremophiles* **16**:645–657.

Wilmes P, Bond PL. (2009). Microbial community proteomics: elucidating the catalysts and metabolic mechanisms that drive the Earth's biogeochemical cycles. *Curr Opin Microbiol* **12**:310–317.

Woyke T, Teeling H, Ivanova NN, Huntemann M, Richter M, Gloeckner FO, *et al.* (2006). Symbiosis insights through metagenomic analysis of a microbial consortium. *Nature* **443**:950–955.

530

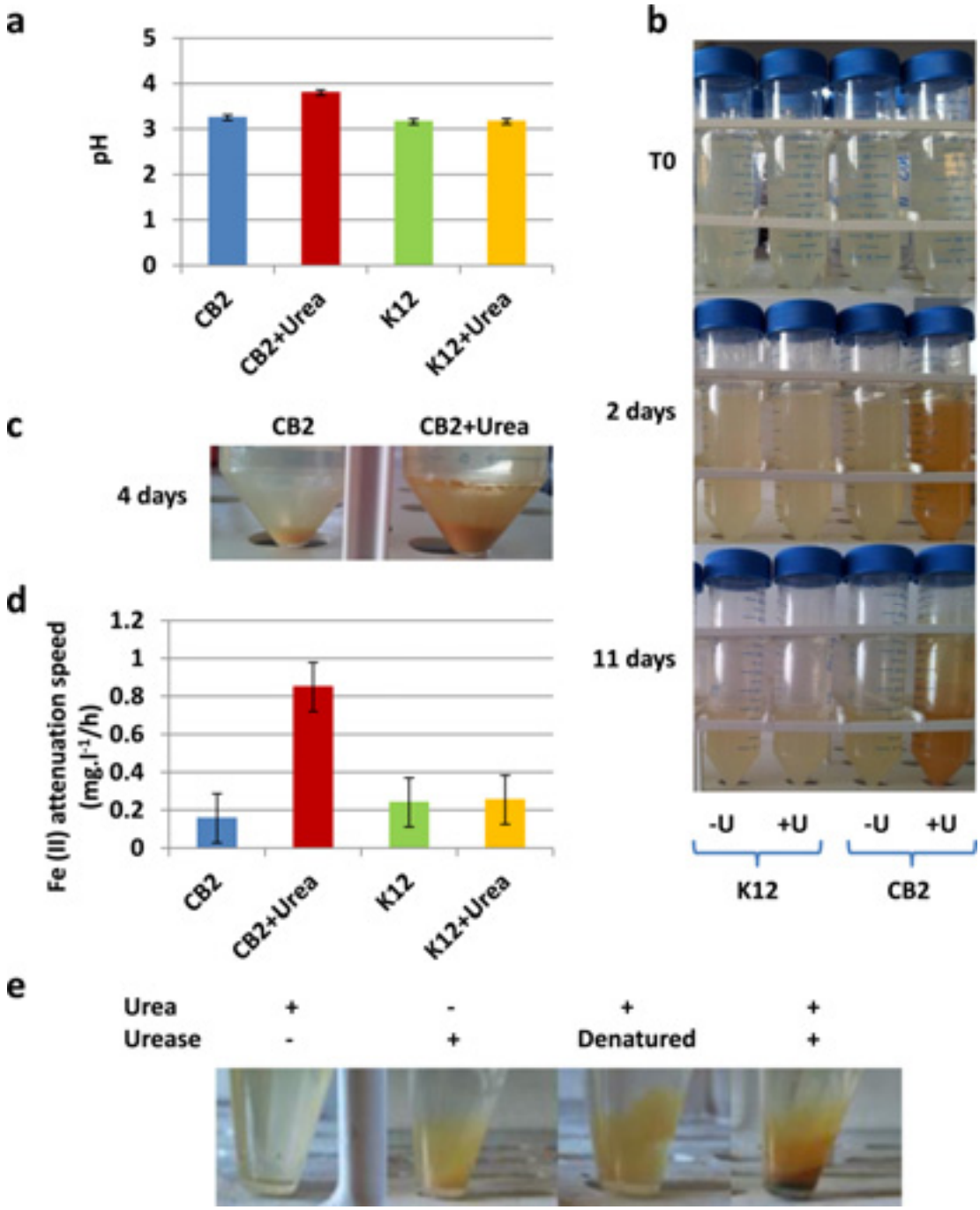


Figure 1. *Thiomonas* urea degradation activity promotes precipitation in AMD-impacted water. *Tm. intermedia* K12 and *Tm.* sp. CB2 were incubated at initial OD600nm of 0.2 in the AMD-impacted water in the absence or presence of 1 g/l urea. (**a**) Comparison of Ph Population marginal mean (PMM) values computed from pH measurements along 11 days of incubation with 4 measures per samples. Error bars represent standard error on mean (SEM) computed from the ANCOVA analysis of the data. (**b**) The color of the medium changed to a shade of orange over time. (**c**) Precipitates observed after centrifugation of 4 days old cultures of *Tm.* sp. CB2 in AMD-impacted water supplemented or not by 1 g/l urea. (**d**) Fe(II) attenuation speed in the soluble fraction is expressed as the difference between the concentration of Fe(II) measured in supernatant of the non-inoculated sample and the concentration of Fe(II) measured in supernatant of an inoculated sample by hour. Error bars represent SEM computed from the ANCOVA analysis of the data. Due to the characteristics of the waters from the creek (Casiot *et al.* 2005; Morin *et al.* 2003), Fe(II) precipitates slowly even in non-inoculated water, because of its abiotic oxidation and subsequent precipitation. (**e**) Effect of urea degradation activity on precipitate formation in the AMD-impacted water. AMD-impacted water was supplemented with 10 U of purified urease in the presence or the absence of 1 g/l urea. As an additional control, urease was inactivated by heating for 5 min at 95 °C.

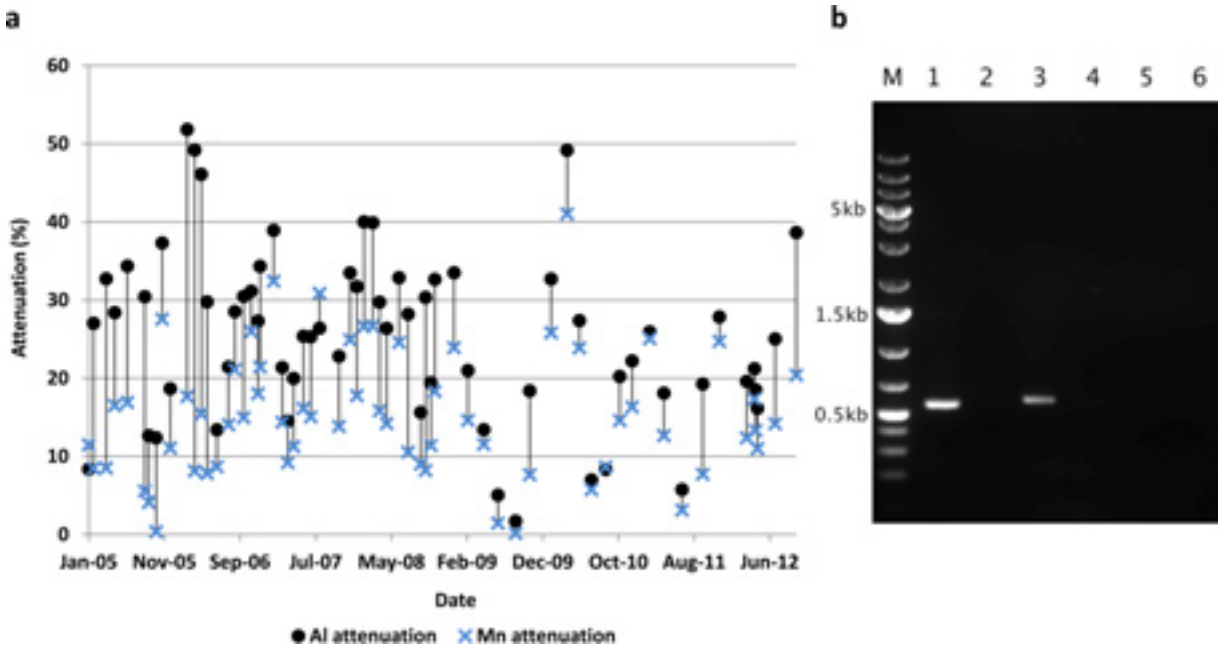


Figure 2. Attenuation of aluminum compared to manganese in the AMD-impacted creek from January 2005 to September 2012 and *in situ* expression of the *Thiomonas* ureaseencoding genes. (**a**) Metal concentrations in water samples were routinely measured by ICP-AES, at two sample points: "Source" and "COWG". Attenuation is expressed as the percentage of metal lost from the "COWG" sample as compared to the "Source" sample. The manganese concentration is of the same order of magnitude as aluminum, and manganese concentration decrease is mainly due to dilution. (•) Al attenuation and (×) Mn attenuation. A Wilcoxon signed ranks test showed that Al attenuation is significantly higher than Mn attenuation (p-value < 0.01). (**b**) Agarose gel analysis of *aioA*, *ureC* and gene coding for allophanate hydrolase transcripts by RT-PCR. Lane M: GeneRulerTM 1kb DNA Ladder Plus (Fermentas). Lane 1: RTPCR product for *aioA* (555bp). Lane 2: negative control (without reverse transcriptase). Lane 3: RT-PCR product for *ureC* (570bp). Lane 4: negative control. Lane 5: RT-PCR product for allophanate hydrolase gene (515bp). Lane6: negative control. The amplification products were sequenced and the expected sequences were obtained (*i. e. aioA* and *ureC*).

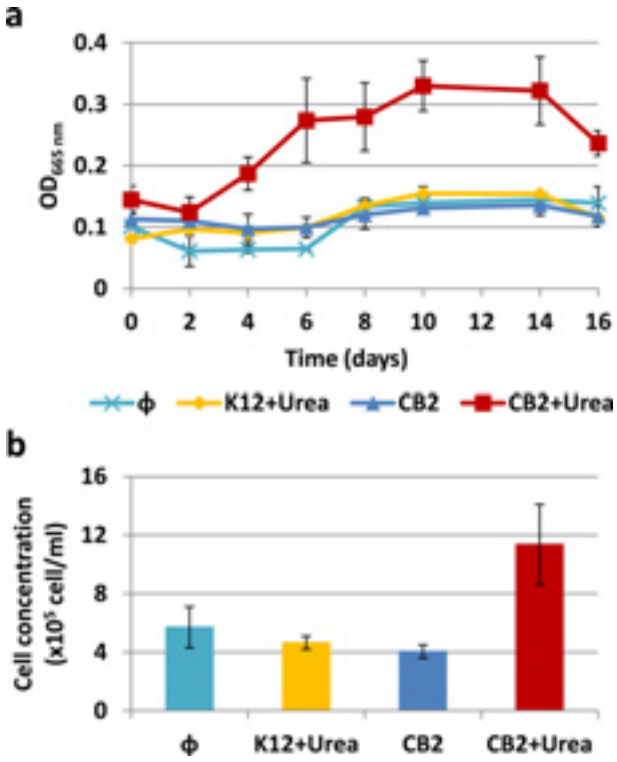


Figure 3. Growth monitoring of *E. mutabilis* in filtrates of *Thiomonas* culture grown in AMD-impacted water. These results highlight the beneficial effect of *Tm.* sp. CB2 urea degradation activity on *E. mutabilis* growth. (**a**) OD665nm as measured over time. (✕) *E. mutabilis,* (◆) *E. mutabilis* + *Tm*. *intermedia* K12 supernatant in the presence of 1 g/l urea, (▲) *and* (□) *E. mutabilis* + *Tm*. sp. CB2 supernatant in the absence and in the presence of 1 g/l urea, respectively. (**b**) Cell concentration determined after 16 days of incubation. Data are representative of two independent experiments and error bars represent SEM.

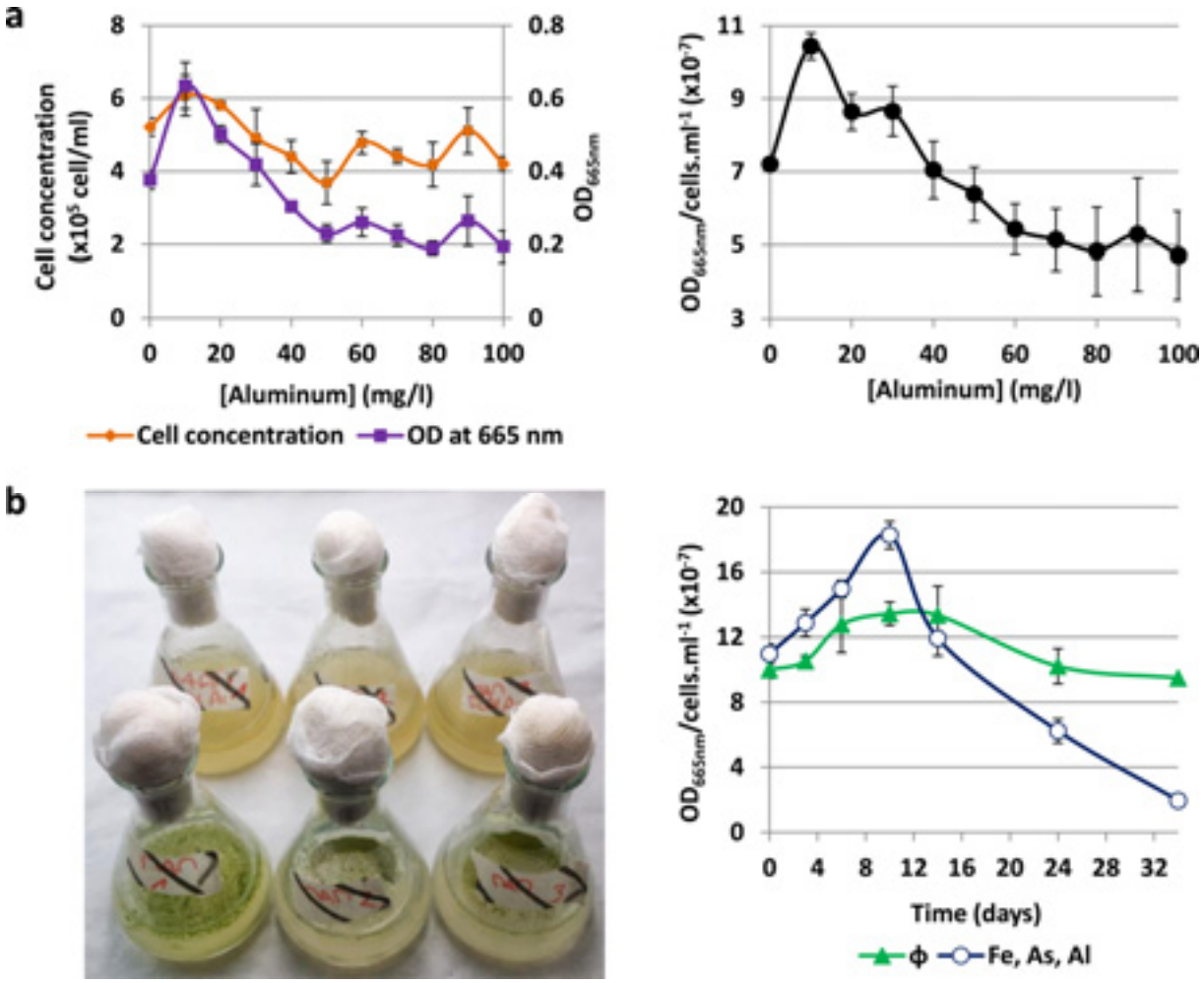


Figure 4. Effect of aluminum, arsenic and iron on *E. mutabilis* growth and photosynthesis. (**a**) Cells were cultivated in liquid MAM media supplemented by different aluminum concentrations. The ratio OD665nm/cells.ml-1 (x10-7) was calculated to evaluate the amount of photosynthetic pigments per cells. In the presence of more than 30 mg/l of aluminum, the protist cells photosynthetic pigments were less abundant. Cell concentration (◆), OD665nm (■) and ratio OD665nm/cells.ml-1 (x10-7) (•) were measured at 24 days of culture. (**b**) Cells were cultivated in liquid MAM media supplemented (o) or not (▲) with 891.5 mg/l iron; 115.2 mg/l arsenite and 36.95 mg/l aluminum. Cell concentration and OD665nm were measured at 24 days of culture**.** In the presence of aluminum, arsenic and iron, the protist cells photosynthetic pigments were less abundant. Data are representative of three independent experiments and error bars represent SEM.

Supplementary Information

Supplementary materials and methods

Microscopy. TEM observations were conducted on co-cultures grown in synthetic media, as described in (Andres *et al.*, 2013), while SEM and ESEM were conducted on co-cultures performed in AMD-impacted water. FE-SEM in high vacuum mode yields well resolved images of *E. mutabilis*, but it was difficult to control the dehydration step, during which *E. mutabilis* cells could not be kept intact, even with a glutaraldehyde fixation. For these reasons, the wet mode was performed and allowed for the localization of *E. mutabilis* and *Tm. sp.* CB2 together. For this purpose, two types of samples were prepared. First, for FE-SEM, cells were taken at day 14 from the co-cultures of *E. mutabilis* and *Tm.* sp. CB2 grown in AMD-impacted water supplemented with 1 g/l urea. Second, for ESEM, *Tm.* sp. CB2 and *E. mutabilis* cells were inoculated at a final OD_{600nm} of 0.3 and concentration of 1 mg/ml, respectively, in AMD-impacted water and cells samples were observed after 1 day of co-culture. FE-SEM (JEOL 6700F) observations were performed in high vacuum mode. The primary beam energy was set to 1 or 3 keV, to obtain pictures with a maximum of surface information and to avoid any charging, since the samples were not metallized. A droplet of suspension was put on a piece of silicon wafer, previously plasma treated to produce a hydrophilic surface, and samples were air-dried. ESEM (FEI XL30-ESEM) observations were performed in wet mode with undried samples. The primary beam energy was set to 10 keV, and the pressure to approximately 5 mb water vapor at room temperature.

FE-SEM and Energy-Dispersive X-ray Spectroscopy (EDXS, Thermo-Noran Vantage) were used to assess the elementary composition of precipitates. For each condition, one pinhead of precipitate was treated with 1 ml hydrogen peroxide to eliminate organic matter, in plastic tubes to avoid metal adsorption. Samples were then centrifuged at 3 000 x g for 15 min and rinsed with Milli-Q water before an additional centrifugation. Precipitates

were then suspended in 100 % ethanol. A droplet of the ultra-sonicated suspension was placed on a plasma treated piece of silicon or carbon wafer, providing a smooth and conductive support. After drying at room temperature the samples were carbon coated. Standard conditions were chosen at 12 keV primary beam energy. Since the arsenic L lines at 1.282 keV overlap with the magnesium K lines at 1.254 keV, some control spectra measurements were collected at 15 keV to verify the presence/absence of the arsenic K lines at 10.543 keV. Depending on the wafer used, the spectra consistently revealed a silicon K line at 1.735 keV or a carbon K line at 0.26 keV, from the silicon or carbon wafer substrate, respectively, the intensity of which depends on the thickness of the analyzed precipitate. Acquisitions were performed either in fixed spot mode or by scanning a selected area and at least 3 analyses were performed per sample. The topography and the chemically inhomogeneous character of the sample did not allow for the calculation of absolute concentrations for all the samples.

Chemical and metabolomics analysis. As(III) and As(V) speciation was assessed by HPLC-ICP-MS. To control the water characteristics during experiments, element concentrations were measured on the 10 month-old 0.22 µm filtered AMD-impacted water and compared to the time of sampling. In this case, concentrations of aluminum, calcium, manganese, iron, nickel, arsenic, strontium, cadmium, thallium, cobalt, barium and lead were measured by inductively coupled plasma-atomic emission spectroscopy (ICP-AES). As(III) and As(V) speciation was also assessed by HPLC-ICP-AES, and iron and Fe(II) was measured by ICP-AES and by the phenanthroline method, respectively (Fadrus & Maly, 1975). No significant decrease in the concentration of elements was apparent except for lead, for which a 2-fold decrease was observed, and for cobalt and barium, which both dropped to below the detection limit. As(III) and As(V) speciation was also assessed by GC-MS profiling analysis performed essentially as reported in Lugan *et al* (Lugan *et al.*, 2010). Briefly, samples were dried under vacuum prior to derivatization by methoxyamine hydrochloride (20 mg/ml in pyridine) and N,O-

Bis(trimethylsilyl)trifluoroacetamide. A mixture of n-alkanes was introduced as retention index standards. The GC-MS system consisted of a 7 683 injector, a 6 890 gas chromatograph and a 5 973 mass spectrometer (Agilent Technologies, Santa clara, USA). Samples were injected twice, in splitless and split mode (split ratio of 50), on an Agilent J&W HP5-MS capillary column, with helium as carrier gas. Identification of arsenite and arsenate was performed by comparison of mass spectra against databases (Wiley7, (Kopka *et al.*, 2005)).

Urea was identified by ultra-performance liquid Zic-PiHILIC chromatography coupled to tandem mass spectrometry (UPLC-MS/MS) on an Acquity UPLC system (Waters corp, Milford, USA) coupled to a triple Quadrupole (Quattro Premier XE Waters). Multiple Reaction Monitoring (MRM) was used for urea identification at the MS/MS level with the transition: 61>44 in the positive Electrospray mode (ESI+). Standard urea was used for chromatography retention time determination and mass spectrometry ionization parameters. Unfortunately, absolute quantification was not possible because of a too high background.

Total RNA extraction from sediment of the Reigous creek and reverse transcription. Cells recovered from sediment were centrifuged at 10 000 rpm 15min at 4°C. Pellets were resuspended in 400µl of suspension solution (25mM Tris-HCl (pH7.6)-10mM EDTA + 20% glucose (vol/vol) and transferred into microtubes containing 0.5g glass beads (0.1 mm diameter) and 0.5µl acidic phenol (pH4.5). Cells were mechanically broken with a Retsch's Mixer Mill apparatus (three 30-s cycles of homogenization at maximum speed with 1-min intervals on ice). Microtubes were centrifuged at 13 000g for 5 min at 4°C. Aqueous phase was placed in a new tube with 1ml the TRIzol reagent (Life Technologies). Homogenization was done by pipetting up and down several times. Microtubes were incubated 5 min at room temperature. 100µl chloroform were added and homogenized by pipetting up and down. Tubes were centrifuged at 13 000g 5 min at 4°C. Aqueous phase was recovered and treated with 200µl chloroform before centrifugation at 13 000g 5min at 4°C. Aqueous phase was

recovered and purified using purified using the RNEasy Plus Mini kit (Qiagen). RNA integrity was checked by electrophoregram using a BioAnalyser (Agilent) and quantified by measuring A_{260} and A_{280} with a Nanodrop. Reverse transcription was performed using the SuperScript® III One-Step RT-PCR System with Platinum® Taq DNA Polymerase (Life Technologies). Each reaction mixture (total volume of 12.5µl) contained 55µg total RNA. Thermocycling conditions were as follow: 30min at 50°C followed by 3min30 at 94°C for RT inactivation; then 40 cycles of 15 seconds at 94°C, 30 seconds at 55°C and 1min at 68°C. A step at 68°c for 5 min was added for final elongation. Primers were designed on three genes: CARN2_aioA_for 5'-CCTGCCATTTCTGCATCG-3' and CARN2_aioA_rev 5'-GCATTCGGAGTTGTACGC-3' for arsenite oxidase large subunit *aioA* gene (CARN2_0821) (Bertin *et al.*, 2011); THICB2v2_ureC_for2 5'-CGAAGGCATGATCCTCAC-3' and THICB2v2_ureC_rev2 5'-CTCGTCGATGGTGTTGAC-3' on urease alpha subunit *ureC* gene (THICB2V2_370058); THICB2v2_360024_for2 5'-GGGCTGTACACCAACTTC-3' and THICB2v2_360024_rev2 5'-GCCATCTTCAAGCTGCAG-3' on putative allophanate hydrolase gene (THICB2V2_360024). Negative controls to check absence of DNA were performed by elimination of the reverse transcription step for each couple of primers. Amplification products were sequenced at Eurofins Genomics and the expected sequences were obtained.

References

Andres J, Arsène-Ploetze F, Barbe V, Brochier-Armanet C, Cleiss-Arnold J, Coppée J-Y, *et al.* (2013). Life in an arsenic-containing gold mine: genome and physiology of the autotrophic arsenite-oxidizing bacterium rhizobium sp. NT-26. *Genome Biol Evol* **5**:934–953.

Bertin, Heinrich-Salmeron A, Pelletier E, Goulhen-Chollet F, Arsène-Ploetze F, Gallien S, *et al.* (2011). Metabolic diversity among main microorganisms inside an arsenic-rich ecosystem revealed by meta- and proteo-genomics. *ISME J*. http://www.ncbi.nlm.nih.gov/pubmed/21562598 (Accessed May 16, 2011).

Casiot C, Lebrun S, Morin G, Bruneel O, Personné J C, Elbaz-Poulichet F. (2005). Sorption and redox processes controlling arsenic fate and transport in a stream impacted by acid mine drainage. *Sci Total Environ* **347**:122–130.

Casiot Corinne, Morin Guillaume, Juillot F, Bruneel Odile, Personné Jean Christian, Leblanc M, *et al.* (2003). Bacterial immobilization and oxidation of arsenic in acid mine drainage (Carnoulès creek, France). *Water Res* **37**:2929–2936.

Egal M, Casiot C., Morin G., Elbaz-Poulichet F., Cordier MA, Bruneel O. (2010). An updated insight into the natural attenuation of As concentrations in Reigous Creek (southern France). *Applied Geochemistry* **25**:1949–1957.

Fadrus H, Maly J. (1975). Suppression of iron(III) interference in the determination of iron(II) in water by the 1,10-phenanthroline method. Analyst.

Kopka J, Schauer N, Krueger S, Birkemeyer C, Usadel B, Bergmüller E, *et al.* (2005). GMD@CSB.DB: the Golm Metabolome Database. *Bioinformatics* **21**:1635–1638.

Lugan R, Niogret M-F, Leport L, Guégan J-P, Larher FR, Savouré A, *et al.* (2010). Metabolome and water homeostasis analysis of Thellungiella salsuginea suggests that dehydration tolerance is a key response to osmotic stress in this halophyte. *Plant J* **64**:215–229.

Table S1. Genomic and phenotypic characteristics of *Tm.* sp. CB2 and *Tm. intermedia* K12

		CB2	K12[1]
Genome characteristics	Accession number	Under submission[4]	NC_014153
	Sequence length[2]	3 970 896 bases	3 396 378 bases
	GC %	63.64	63.88
	Scaffolds	7	1
	Contigs	34	1
	Nosferatu Repeated Regions[2]	12.61 %	4.86 %
	Protein coding density	90.32 %	90.11 %
	CDS (total)	4 141	3 477
	ure/urt genes	Present	Absent
	UCA / allophanate hydrolase genes	Present	Absent
	Genes encoding rusticyanin-like proteins[3]	Absent	Absent
Phenotypic characteristics	MIC for As(III)	10 mM	10 mM
	Urea degradation	+	-
	Iron oxidation activity	-	-

[1]This genome was sequenced by the US DOE Joint Genome Institute (http://www.jgi.doe.gov); [2]Sequence length and Nosferatu Repeated Regions for genomes were calculated without undetermined bases; [3] Genes involved in iron oxidation encode rusticyanin or rusticyanin-like genes in *Acithiobacillus* (45, 46). However, other proteins may be involved in Fe(II) oxidation, for example, Cyc2 and a Rusticyanin-like protein in *Thiobacillus prosperus* (45). Likewise, in *Acidithiobacillus ferrivorans* and some *A. ferrooxidans* strains, iron oxidation involves the Iro protein and a cytochrome c552, cyc1, but the requirement of a rusticyanin like protein was not essential (45, 47). In some bacteria such as *Leptospirillum* spp., other mechanisms were shown to be involved in iron oxidation requiring Cyc_{572} and Cyc_{579}, cbb_3 oxidase and bc_1 complex. In the genome of *Tm*. sp. CB2, genes sharing 45.9 %, 34% and 30% identity with Iro and Cyc1 from *A. ferroxidans* and Cyc2 from *T. prosperus*, respectively, were present, but no iron oxidation activity was detected with *Tm.* sp. CB2; [4] submitted in EMBL, project number PRJEB3960.

Table S2: Chemical parameters of the creek water at the time of sampling.

Sampling point[1]		S	COWG	Gal	Reigous UC
pH		3.79	3.68	3.56	5.14
SO_4^{2-}	mg/l	3 479	2 991	2 639	806
Fe	mg/l	1 060	891.50	647.60	3.23
As	mg/l	154.90	115.20	56.98	0.01
Al	mg/l	44.07	36.95	25.34	6.55
Ca	mg/l	261.60	266.80	300.60	188.40
Mn	mg/l	8.32	7.41	8.28	11.94
Zn	mg/l	21.53	18.53	13.95	10.26
Co	µg/l	291	249	230	287
Ni	µg/l	516	418	361	255
Cu	µg/l	76.4	60.9	57.8	103
Sr	µg/l	207	197	220	631
Cd	µg/l	102	85.4	59.0	49.0
Sb	µg/l	33.2	<QL	<DL	<QL
Ba	µg/l	6.9	6.4	6.9	17.5
Tl	µg/l	354	287	161	9.3
Pb	µg/l	1 394	890	289	606
U	µg/l	26.8	20.5	11.0	6.3

[1]Measurements taken at site source (S), or 30 m, 120m and 1500 m downstream of the sampling point ("COWG", "GAL" and "Reigous UC", respectively) (37). DL: detection limit, QL: quantification limit

Table S3. Metal composition of the Precipitates and Supernatant obtained from cultures of *Tm.* sp. CB2 and *T. intermedia* K12 in AMD-impacted water.

Metal[1]	Supernatants				Precipitates			
	CB2	Urea +CB2	K12	Urea +K12	CB2	Urea +CB2	K12	Urea +K12
Al	45.7	**35.6**	43.6	43.6	69.6	**1 106**	87.5	103
As	80.9	**17.1**	59.4	58.5	2 355	**6 899**	4 797	6 134
Cd	0.066	0.067	0.067	0.067	0.025	0.07	0.039	0.051
Co	0.24	0.23	0.23	0.23	0.27	0.25	0.24	0.29
Cu	0.072	0.07	0.074	0.073	0.22	0.75	0.66	0.58
Fe	882	**714**	847	851	5 961	**20 144**	8 189	10 224
Ni	0.47	0.46	0.47	0.45	0.64	0.73	0.62	0.67
Pb	0.2	0.075	0.065	0.066	10.3	23.9	28.4	35.3
Zn	20.5	20.2	19.9	20.1	21.2	25	21.8	26.9

[1]Analysis of precipitates and supernatants were performed by ICP-AES, at 11 days of culture of *Tm.* sp. CB2 and *T. intermedia* K12 in AMD-impacted water in the presence and the absence of 1 g/l urea. Metal concentrations are given in mg/kg.

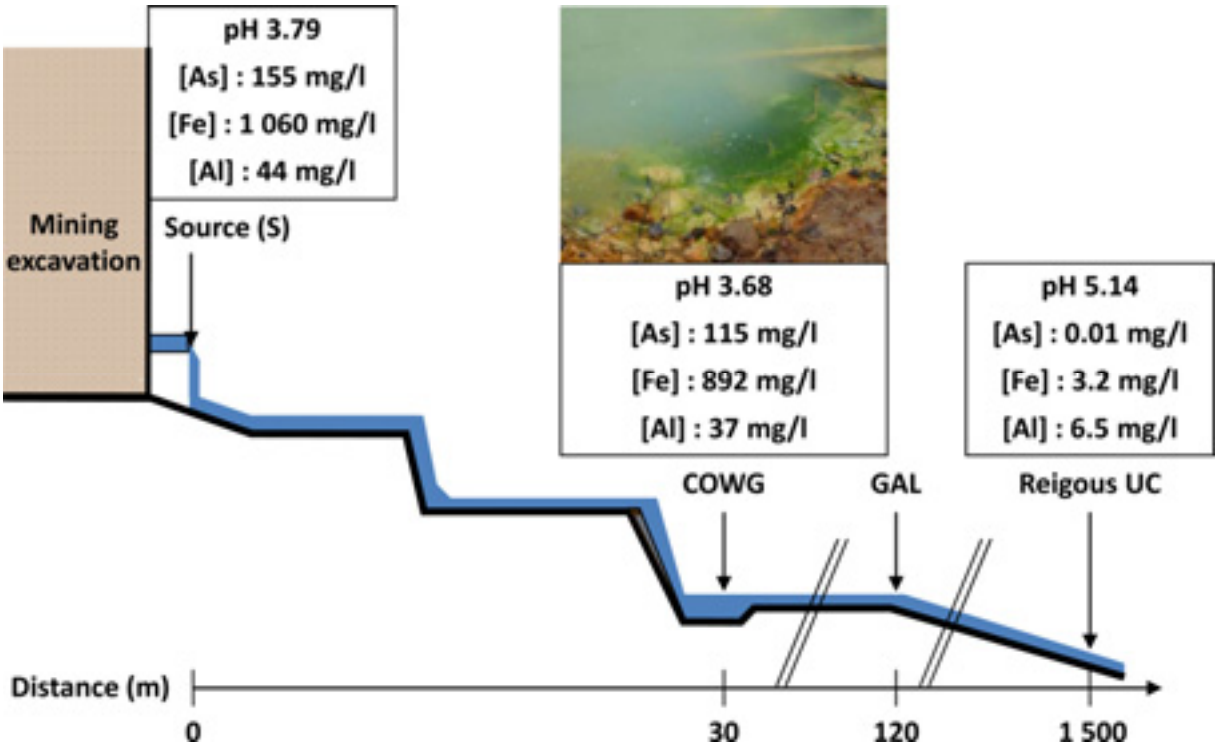


Figure S1. Characteristics of the Carnoulès AMD site. In the AMD of Carnoulès, France, runoff flows into the Reigous creek, which is considered to be an extreme environment because of its high concentrations of toxic metals (iron, 12-25 mmol/l; arsenic, 100-350 mg/l) and its acidic pH (around 3) (Egal *et al.*, 2010). The concentrations of arsenic, aluminum and iron decrease along the creek due to metal oxidation and precipitation (Corinne Casiot *et al.*, 2003; C Casiot *et al.*, 2005). The main sites that were sampled regularly are indicated.

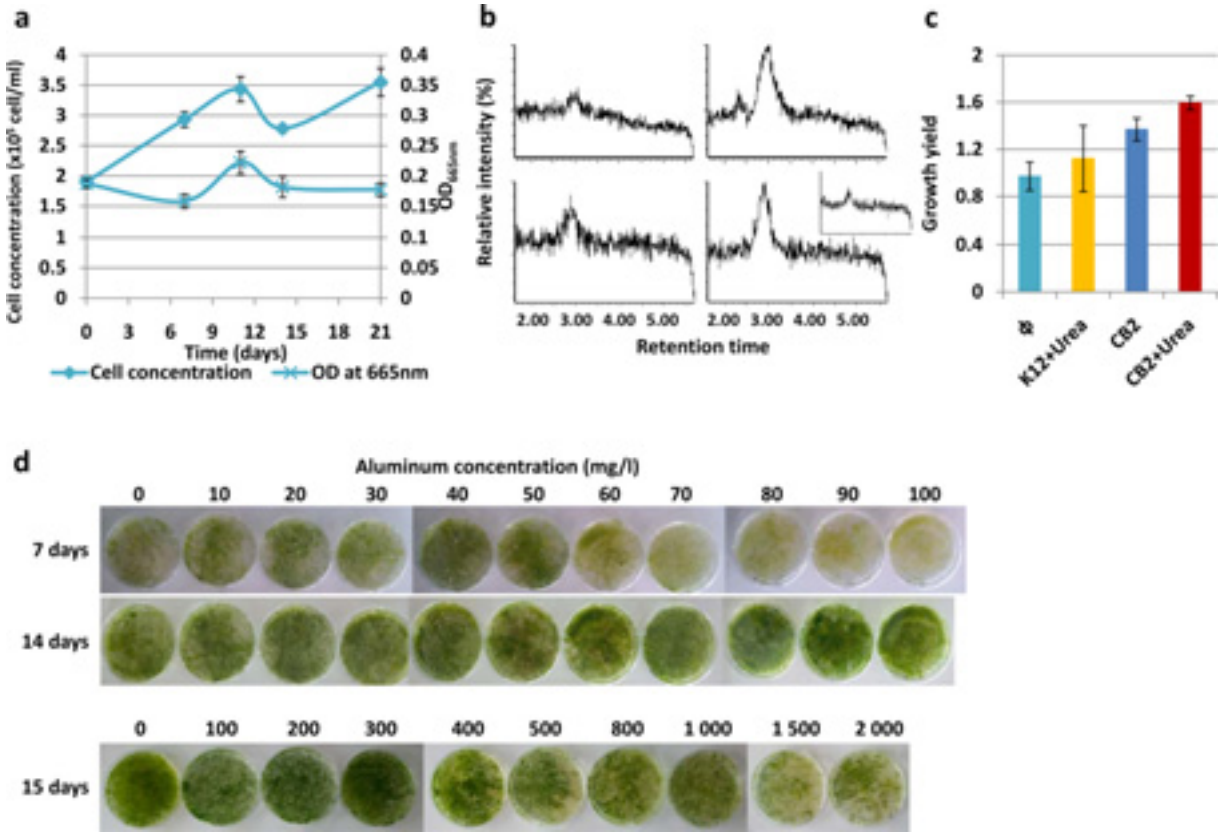


Figure S2. *E. mutabilis* growth in AMD-impacted water and in minimal medium containing aluminum. (**a**) Cell concentration and measure of the OD665nm representing chlorophyll content, as measured over time. Cell concentration increased when *E. mutabilis* cells were incubated in the AMD-impacted water, whereas the chlorophyll content did not. Data are representative of three independent experiments and error bars represent SEM. (**b**) during the growth in AMD-impacted water, urea production by *E. mutabilis* was monitored in several experiments using UPLC-MS/MS approaches. Upper and lower graphs shows urea signal (Relative intensity) in two independent experiments, at day 0 and 6 and day 0 and 21 of incubation, respectively. The small graph corresponds to a control (non-inoculated) at day 21. (**c**) Co-cultures of *Thiomonas* strains with *E. mutabilis*. *E. mutabilis* was co-cultivated with *Tm.* sp. CB2 or *T. intermedia* K12 in AMD-impacted water in the presence or the absence of 1 g/l urea. Growth yield of *E. mutabilis* is expressed as the cell concentration at 14 days of culture divided by the cell concentration at the time of inoculation. (**d**) *E. mutabilis* growth on solid minimal medium containing increasing concentrations of aluminum. *E. mutabilis* growth is delayed beginning with concentrations of 60 mg/l of aluminum at day 7 of culture as compared to the same culture at 14 days. The protist is able to grow in the presence of 2 000 mg/l of aluminum on solid media at day 15 of culture. Pictures are representative of three independent experiments.

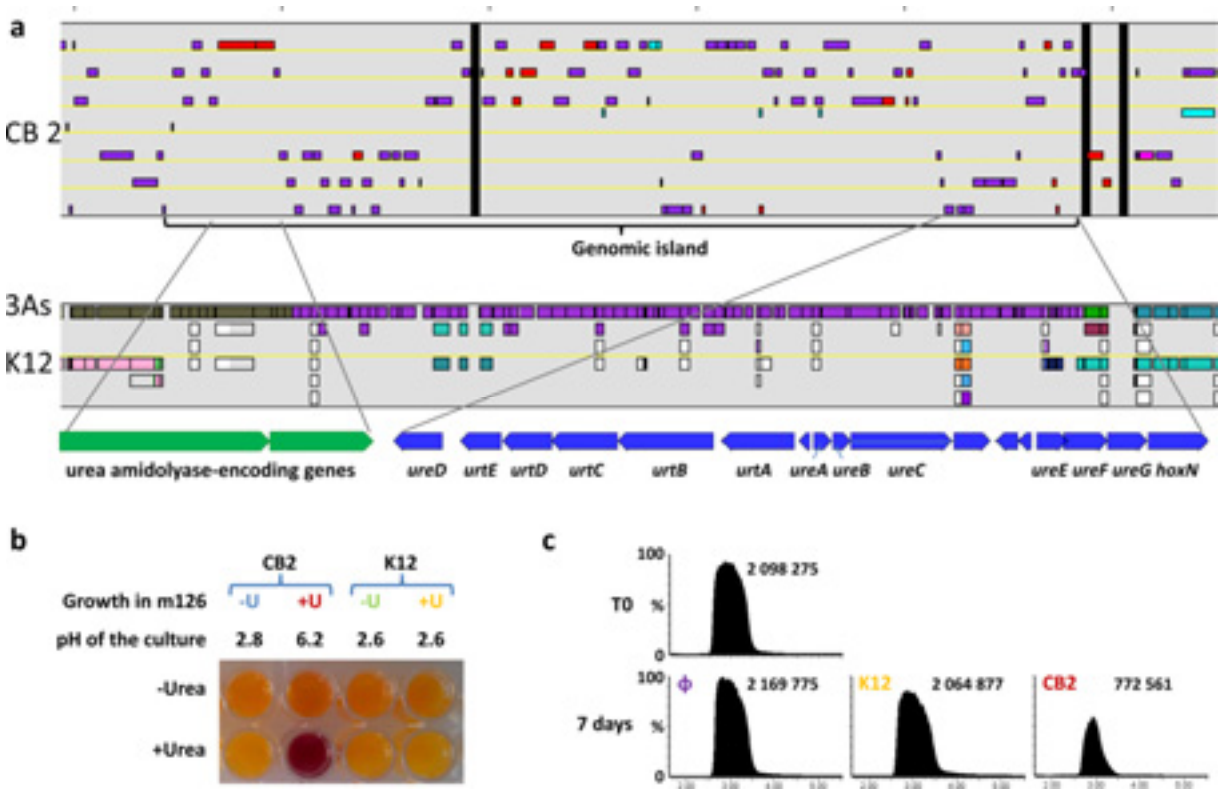


Figure S3. Gene loci involved in urea degradation in *Tm.* sp. CB2 and urea degradation activity in *Thiomonas* strains (**a**) Genes involved in urea import (*urt*) and degradation (*ure* or genes encoding the urea amidolyase involved in the UCA-allophanate hydrolase pathway) are found in one genomic island in the *Tm.* sp. CB2 genome but are absent from the *Tm. intermedia* K12 genome. (**b**) Urea degradation was detected using Bromo-cresol purple as described in Material and methods. The purple color indicates urea degradation, and urea degradation has an impact on media pH when cells are incubated in the synthetic medium m126. These experiments are representative of at least three independent experiments. (**c**) Urea was detected at T0 or after 7 days, by mass spectrometry (UPLC- MS/MS) from the supernatant of *Tm.* sp. CB2 or *T. intermedia* K12 cultures in synthetic medium m126 supplemented or not with 1 g/l urea.

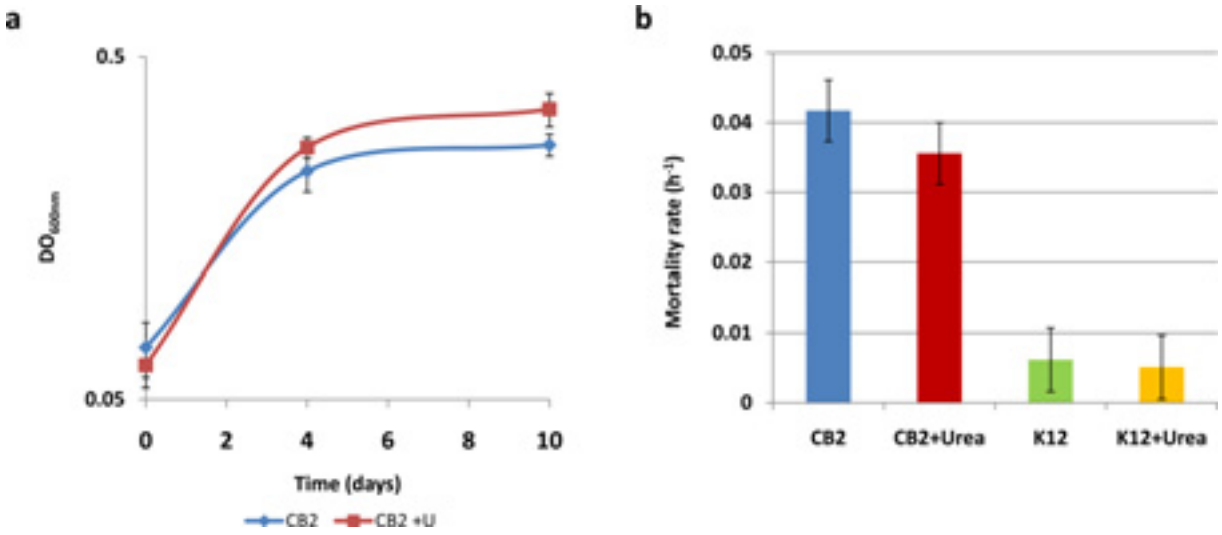


Figure S4. Growth and survival capacities of *Thiomonas* in AMD-impacted water. (**a**) growth curves of *Tm*. sp. CB2 incubation in AMD-impacted water in the presence and in the absence of 1 g/l urea (**b**) Mortality rate of *Tm.* sp. CB2 and *T. intermedia* K12 during incubation in AMD-impacted water in the presence and in the absence of 1 g/l urea. To determine the mortality rate, cells were inoculated in AMD-impacted water at initial OD600nm=0.2 and survival (colony forming unit per ml) was estimated by plating cultures on m126 at different times. Mortality rates were computed from the estimates of the slopes obtained by analysis of covariance (MATLAB R2014a commands *aoctool* and *multcompare*) on the decimal log of cell concentration vs time. The error bars represent the standard error means on the mortality rates computed from standard error means on slope given by *multcompare.*

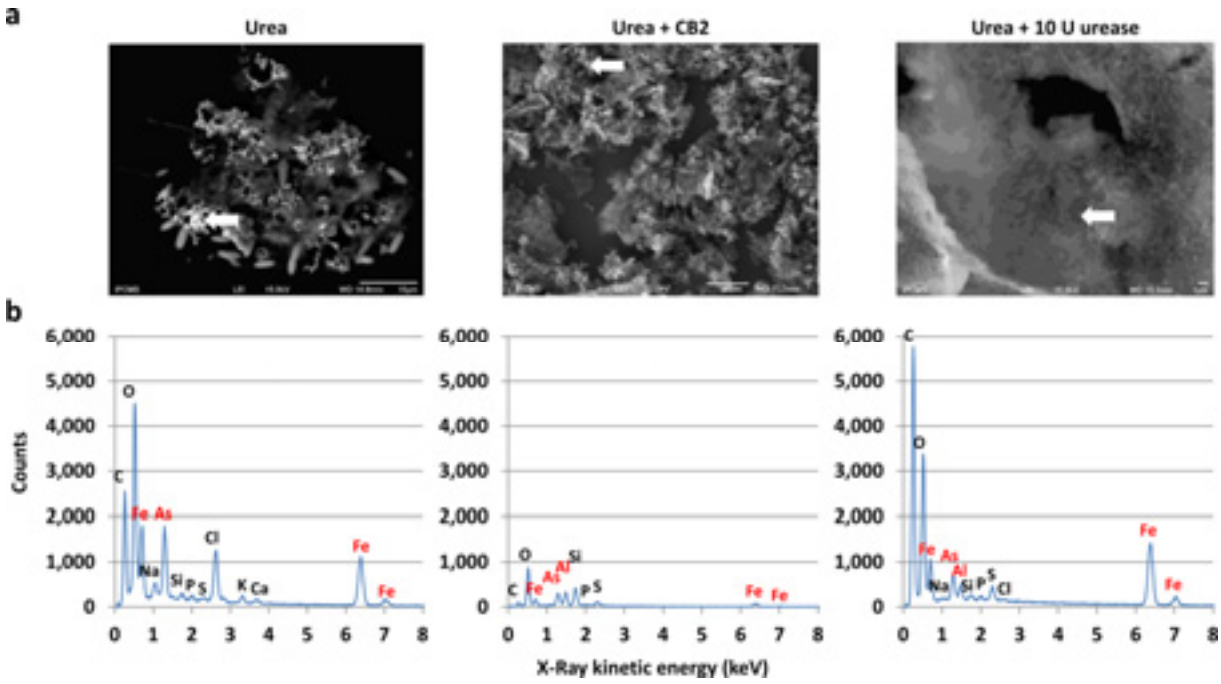


Figure S5. Analysis of precipitates elementary composition. (**a**) SEM photographs and (**b**) X-ray microanalysis graphs of precipitates. These were obtained from AMD-impacted water supplemented with 1 g/l urea, from a 4 days incubation of *Tm.* sp. CB2 and a 1 day incubation of 10 U of urease in AMD-impacted water supplemented with 1 g/l urea, respectively. White arrows on the photographs show spots on which the X-ray microanalyses were conducted to obtain the graphs bellow. Since the magnesium x-rays K lines interfere with the arsenic Llines around 1250-1300eV, the presence of As was confirmed by checking its K-lines at higher energy. Silicon in the second graph and carbon in the first and third ones originate from the supports used in these studies. Pictures and graphs are representative of at least 3 analyzed points. These graphs show that aluminum is present in the precipitate only in the conditions where urea degradation activity was possible. Controls with inactivated urease or with urease but no urea were also performed and supported this conclusion (data not shown).

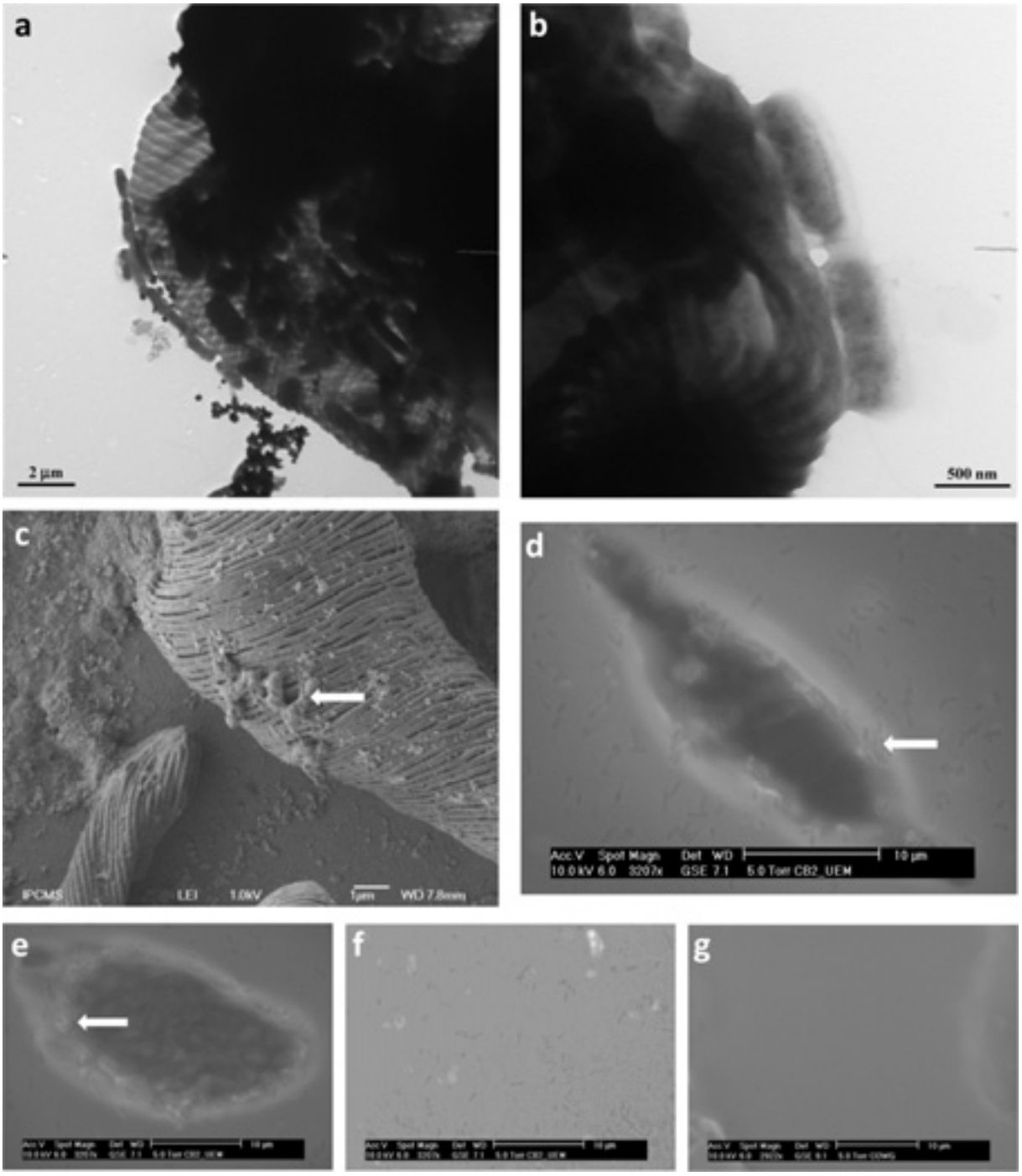


Figure S6. Interactions between *E. mutabilis* and *Tm*. sp. CB2. Co-cultures of *Thiomonas* strains with *E. mutabilis* showing the physical association of the cells *E. mutabilis* was cocultivated with *Tm.* sp. CB2 in synthetic media (A-B) or in AMD-impacted water (C-G) (**a**) and (**b**) TEM pictures of co-cultures of *E. mutabilis* and *Tm*. sp. CB2 in m126 synthetic medium supplemented with iron (600 mg/L). (**c**) SEM of *E. mutabilis* cultivated in AMDimpacted water with *Tm.* sp. CB2 and 1 g/l urea. Pictures are representative of several photographs. (**d**), (**e**), (**f**) and (**g**) ESEM pictures of *E. mutabilis* co-cultivated with *Tm.* sp. CB2 and 1 g/l urea for one day. (**d**) and (**e**) Images of the same co-culture showing bacteria associated with *E. mutabilis*; (**f**) Bacteria alone. (**g**) Non-inoculated AMD-impacted water. Examples of bacteria attached on the protist cell surface are indicated with an arrow.

DISCUSSION

Les résultats présentés dans cet article mettent en évidence une interaction procaryote-eucaryote entre deux organismes majeurs présents à Carnoulès, *Thiomonas* sp. CB2 et *E. mutabilis*. En effet, l'activité de dégradation de l'urée de *Thiomonas* sp. CB2 entraîne la précipitation de certains métaux favorisant ainsi l'activité photosynthétique de l'euglène.

Tout d'abord, le séquençage du génome de *Thiomonas* sp. CB2 et des tests physiologiques montrent sa capacité à dégrader l'urée. L'approche métabolomique a de plus permis de visualiser directement la dégradation de l'urée en milieu synthétique, mais cela n'a pas été possible pour les cultures dans l'eau du Reigous. Néanmoins, l'utilisation de différents contrôles pour les expériences réalisées dans l'eau de la rivière, confirment que les résultats observés sont dûs à cette activité. Tout d'abord, *Thiomonas intermedia* K12, une souche phylogénétiquement proche de *Thiomonas* sp. CB2 mais incapable de dégrader l'urée, a été utilisée en tant que témoin négatif. Un mutant de *Thiomonas* sp. CB2 de phénotype négatif pour l'activité de dégradation de l'urée aurait fourni un contrôle plus direct mais à l'heure actuelle il n'existe pas d'outils génétiques permettant une telle approche sur cette souche. Ensuite de l'uréase purifiée a été utilisée en tant que témoin positif.

De manière surprenante, alors que *Thiomonas* sp. CB2 dégrade l'urée, elle ne retire pas de bénéfice de cette activité. Elle est en effet incapable d'utiliser ce composé comme unique source de carbone ou d'azote en milieu synthétique et tandis que lors de cultures dans l'eau du Reigous l'activité de dégradation de l'urée s'accompagne d'une production d'ammoniaque entraînant une augmentation du pH, *Thiomonas* sp. CB2 n'en tire pas profit. Ces résultats remettent ainsi en cause l'hypothèse émise avant ce travail d'une interaction trophique directe entre *Thiomonas* sp. CB2 et *E. mutabilis*. Cependant, l'élévation du pH coïncide avec la formation d'un précipité orange suggérant l'oxydation et la précipitation du fer présent dans l'eau. Les analyses physico-chimiques, effectuées par spectrométrie de masse et microanalyse X, confirment cette hypothèse et mettent en évidence la précipitation supplémentaire de l'arsenic et de l'aluminium. Si la précipitation du fer s'explique par son oxydation abiotique liée à l'augmentation du pH et celle de l'arsenic à l'oxydation par *Thiomonas* sp. CB2 de l'As(III) en As(V) qui co-précipite plus facilement avec le Fe(III), celle de l'aluminium était plus inattendue. L'utilisation de *T. intermedia* K12 ainsi que l'ajout d'uréase purifiée et d'ammoniaque dans l'eau du Reigous confirment bien cette relation entre l'activité de dégradation de l'urée de *Thiomonas* sp. CB2, l'augmentation du pH de l'eau du Reigous et la précipitation de ces

trois métaux. La figure 3.1 présentée ici apporte quelques précisions sur cette relation, en illustrant des données mentionnées dans l'article. Par ailleurs, ces résultats obtenus *in vitro* sont mis en relation avec l'observation et la détection *in situ* du processus d'atténuation de ces trois métaux le long de la rivière et de l'expression de l'arsénite oxydase et de l'uréase de *Thiomonas* sp. CB2, respectivement.

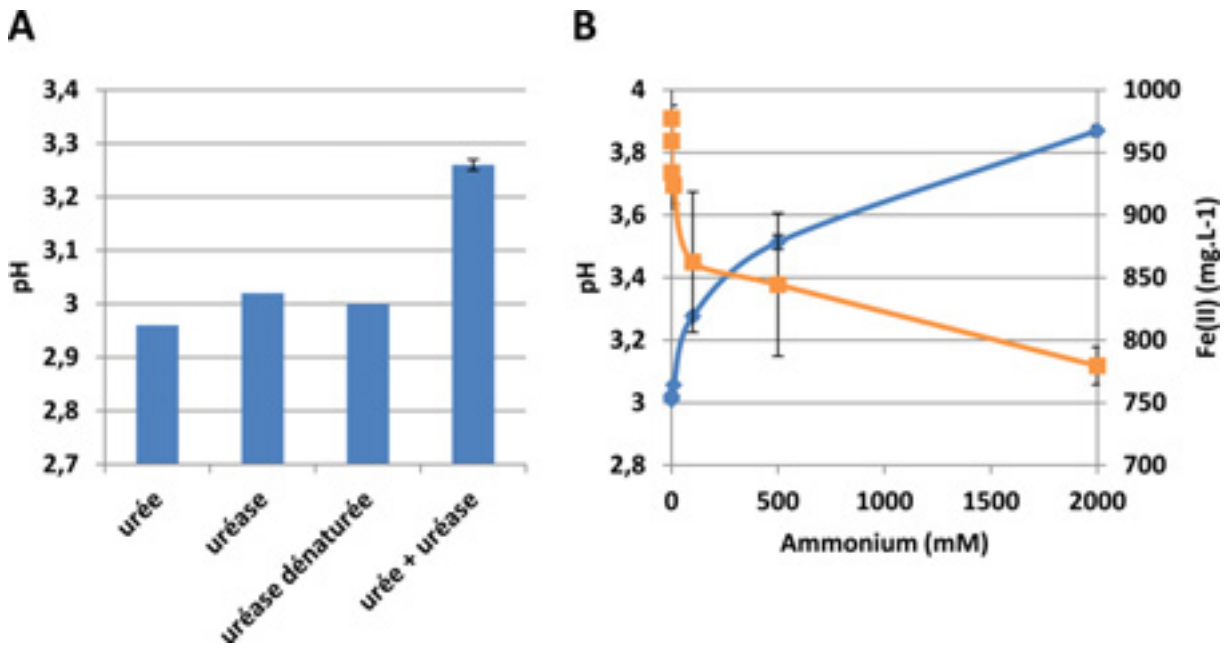


Figure 3.1. Effet de l'ajout d'urée et d'ammonium dans l'eau Du Reigous. **(A)** Effet de la dégradation de l'urée sur le pH de l'eau du Reigous. Ces mesures sont liées à la Figure 1 de l'article. 10 U d'uréase purifiée ont été ajoutés à l'eau du Reigous en présence ou en absence de 1 $g.L^{-1}$ d'urée. Comme contrôle additionnel, l'uréase a été dénaturée par chauffage 5 min à 95 °C. Le pH a été mesuré après 24 h d'incubation à 30°C. **(B)** Effet de l'ajout d'ammonium sur le pH (en bleu) et sur la concentration de Fe(II) (en orange) de l'eau du Reigous. L'ammonium a été ajouté à partir de poudre ou de solutions de sulfate d'ammonium à 1 M et 5 M. Les mesures ont été effectuées après 24 h d'incubation à 30°C. Le Fe(II) a été dosé par la méthode de la phénanthroline (Fadrus and Malý, 1975). La formation d'un précipité orange est également observée en parallèle à l'augmentation du pH et la précipitation du fer. Ces résultats sont issus d'expériences réalisées en triplicats.

L'activité de dégradation de l'urée ne présentant apparemment aucun bénéfice pour *Thiomonas* sp. CB2, nous nous sommes ensuite intéressés à son partenaire potentiel, *E. mutabilis*. Sa capacité à croître dans l'eau de la rivière et à sécréter l'urée dans celle-ci a ainsi été vérifiée. D'une part, ces résultats mettent en évidence la diminution de son activité photosynthétique lorsqu'elle est cultivée dans cette eau et cette baisse d'activité est liée, par des tests effectués en milieu synthétique, à la présence d'arsenic et d'aluminium décrits comme des inhibiteurs de la photosynthèse. D'autre part, si une production d'urée par l'euglène est observée en milieu

synthétique (Halter et al., 2012b), elle semble cependant moindre dans l'eau de la rivière filtrée d'après les données obtenues en métabolomique. La faible production observée ici peut s'expliquer par le fait que trop peu de cellules d'euglènes ont été utilisées. En effet, sur le site elles sont présentes abondamment sous forme d'un tapis (Figure 2.1), et leur contact direct avec les sédiments peut expliquer la relative abondance mesurée dans l'eau interstitielle (Halter et al., 2012b). Une source alternative d'urée n'est cependant pas à exclure.

De manière intéressante, les différentes expériences de cross- et co-culture montrent finalement que la précipitation du fer, de l'arsenic et de l'aluminium, due à l'activité de dégradation de l'urée de *Thiomonas* sp. CB2, favorise la croissance et l'activité photosynthétique d'*E. mutabilis*, mettant bien en évidence une interaction, bien qu'indirecte, entre ces deux organismes. Ainsi, si l'interaction impliquant l'euglène, *Thiomonas* sp. CB2 et l'urée semble moins directe que les données préliminaires ne le suggérait, la dégradation de l'urée par *Thiomonas* sp. CB2 montre un réel effet bénéfique pour *E. mutabilis via* l'augmentation de la précipitation des métaux, suggérant au minimum une relation de commensalisme.

D'autres expériences seront nécessaires pour définir si le terme de mutualisme est applicable ici, cette hypothèse n'étant pas à exclure totalement. En effet, *Thiomonas* sp. CB2, en favorisant l'activité d'un producteur primaire, peut potentiellement en tirer bénéfice, justifiant l'expression de l'activité de dégradation de l'urée. Parmi les composés synthétisés par l'euglène et retrouvés *in situ* et d'après les données génomiques, certains sont potentiellement utilisables par *Thiomonas* sp. CB2 tels que la putrescine ou le glycérol. De plus, si les observations microscopiques ne démontrent pas la spécificité ou la nécessité d'une association physique de ces deux organismes, elles permettent de souligner sa possiblité, celle-ci pouvant favoriser l'échange de métabolites en limitant leur diffusion.

Des approches complémentaires devront être développées afin de préciser cette interaction, notamment par des techniques de marquage isotopique de l'urée ou du CO_2, couramment utilisées dans l'étude des flux métaboliques. En effet, un marquage de l'urée, à l'^{15}N, au ^{13}C ou au ^{14}C couplé à des techniques de spectrométrie de masse à ions secondaires (SIMS) ou de l'autoradiographie permettrait tout d'abord de s'assurer que *Thiomonas* sp. CB2 n'incorpore pas l'urée comme source d'azote ou de carbone. Un marquage de l'urée aux isotopes stables fournirait ensuite une alternative à la chromatographie en phase liquide couplée à la spectrométrie de masse en tandem pour quantifier la dégradation de l'urée par la détection d'ammoniaque marqué. De la même manière, l'utilisation de $^{13}CO_2$ pourrait permettre de suivre l'activité métabolique de l'euglène et sa production d'urée ou d'autres composés. De plus, un tel marquage autoriserait la visualisation du passage de composés, entre *E. mutabilis* et *Thiomonas* sp. CB2.

Pour conclure, cette étude de la relation entre deux membres importants de la communauté microbienne de Carnoulès conduit à une meilleure compréhension de son fonctionnement. De plus, en augmentant l'efficacité de l'atténuation naturelle du site et l'activité d'un producteur primaire, l'activité de dégradation de l'urée est potentiellement favorable à l'ensemble de la communauté microbienne.

CONCLUSION GÉNÉRALE

PERSPECTIVES

Les travaux présentés dans ce manuscrit, mêlant approches de génomique descriptive (séquençage et analyse du génome) et fonctionnelle (transcriptomique, protéomique, métabolomique), mettent en évidence divers processus impliqués dans la réponse adaptative des micro-organismes à l'arsenic. De plus, leur utilisation combinée à des outils plus conventionnels tels que des tests physiologiques, biochimiques ou de biologie moléculaire à permis de dévoiler différents mécanismes impliqués dans la réponse globale des micro-organismes procaryotes et eucaryotes à l'arsenic.

L'analyse du génome de la bactérie *Rhizobium* sp. NT-26 suggère que, lors de son évolution, cette dernière a perdu des fonctions liées à la nodulation et parallèlement gagné certaines capacités métaboliques propices à son adaptation à un environnement riche en arsenic. De la même manière, la comparaison du protiste photosynthétique hypertolérant à l'arsenic *E. mutabilis* à une souche de laboratoire plus sensible souligne des différences génotypiques pouvant refléter son adaptation aux conditions extrêmes du DMA de Carnoulès.

De manière intéressante, les travaux réalisés sur *Rhizobium* sp. NT-26 et *E. mutabilis* montrent des points communs dans la réponse à l'arsenic de ces deux organismes appartenant à des domaines différents du vivant. La résistance à l'arsenic, le transport du phosphate, la réponse au stress oxydatif et la mobilité semblent ainsi être des processus conservés entre ces organismes résistants. Au contraire d'autres mécanismes sont plus spécifiquement mis en évidence chez l'un ou l'autre. Si ceci peut refléter un biais dû aux différentes approches utilisées qui ne permettent pas d'identifier de manière exhaustive les processus impliqués dans la réponse à l'arsenic, cela peut également témoigner d'une réalité biologique où les mécanismes sont réellement spécifiques.

Enfin, ce travail souligne l'importance des interactions entre organismes dans leur maintien dans des environnements contaminés par l'arsenic. En effet, l'exemple développé ici montre que l'activité de *Thiomonas* sp. CB2 participe à l'atténuation naturelle du fer, de l'arsenic et de l'aluminium dans le DMA de Carnoulès, favorisant ainsi celle d'*E. mutabilis* et participant au moins en partie à son maintien. Certaines précisions restent cependant à être apportées quant à cette dernière interaction et pourront être abordées dans des travaux ultérieurs.

De plus, cette interaction bénéfique entre 2 acteurs majeurs de l'écosystème peut également profiter aux autres membres de la communauté bactérienne en favorisant à la fois les processus d'atténuation et la production primaire *in situ*. A moyen et plus long terme, différentes approches pourront ainsi être développées dans le but de comprendre plus en détails les relations trophiques existantes au sein de la communauté microbienne de Carnoulès, expliquant probablement la stabilité de cette dernière dans cet environnement extrême.

La suite de ce travail pourra aller plus avant dans l'étude des interactions trophiques entre *E. mutabilis* et la communauté bactérienne de Carnoulès. En effet, différents éléments témoignent de l'existence potentielle d'autres liens entre ce protiste photosynthétique et les bactéries de la communauté. La comparaison de la liste des composés sécrétés par *E. mutabilis* en milieu synthétique et des données métagénomiques de Carnoulès suggère en ce sens que l'ensemble des composés excrétés par l'euglène en milieu synthétique sont métabolisables par les bactéries de la communauté (Bertin et al., 2011; Halter et al., 2012b).

Plus particulièrement, tandis que l'euglène sécrète plusieurs acides aminés, la reconstruction des génomes de CARN1 et CARN4, 2 des 7 souches majoritaires de Carnoulès appartenant à un nouveau phylum nommé *Candidatus* Fodinabacter communificans suggère que leur métabolisme est résolument orienté vers le transport et le recyclage de ces derniers (Bertin et al., 2011). D'autre part, ces deux organismes portent plusieurs gènes dont les produits sont impliqués dans les voies de biosynthèse de la cobalamine, vitamine pour laquelle *E. mutabilis* est auxotrophe, et des expériences de RT-PCR ciblées ont montré leur expression *in situ* (Bertin et al., 2011). Ces données indiquent ainsi l'existence d'une potentielle interaction trophique entre les CARN1-CARN4 et *E. mutabilis*, permettant d'assurer leur maintien réciproque.

Comme le suggère le chapitre précédent, la confirmation de cette relation repose néanmoins sur l'expérimentation en laboratoire. Dans le but d'étudier cette interaction potentielle, différentes approches sont d'ores et déjà explorées dans le but d'isoler CARN1 et CARN4 qui ne sont pas cultivés à l'heure actuelle. De plus, certaines de ces méthodes permettront également de mieux comprendre le réseau de relations trophiques au sein de la communauté microbienne de Carnoulès dans sa globalité.

1. APPROCHES CULTURALES, MARQUAGE FLUORESCENT ET TRI CELLULAIRE

Tout d'abord, différentes approches culturales ont été entreprises afin d'isoler CARN1 et CARN4. Une première stratégie visait à obtenir un milieu propre à leur isolement grâce à l'analyse de leur génome, à l'instar de l'approche de métagénomique inverse ayant rendu possible la culture de *Leptospirillum ferrodiazotrophum* (Tyson et al., 2005; Delavat, 2012). Des tests ont également été réalisés en anaérobiose ou en essayant de tirer profit de l'hypothèse d'une interaction bénéfique avec *E. mutabilis*. Cependant, aucune de ces approches n'a permis l'isolement de ces deux bactéries à l'heure actuelle, notamment car de telles stratégies reposent sur le choix empirique de conditions

de culture qui ne sont pas forcément adaptées à ces organismes. Des approches alternatives sont à l'étude où pourront l'être à plus long terme.

De manière complémentaire, différentes sondes nucléotidiques fluorescentes ont été développées dans le but de réaliser un marquage de CARN1 et CARN4. Si celles-ci devront tout d'abord être testées afin de vérifier leur spécificité, un tel marquage associé à de la cytométrie en flux et du tri cellulaire permettra d'identifier, de compter et d'obtenir des enrichissements de ces bactéries. Ces derniers pourront permettre de réaliser des tests de culture à plus grande échelle ou l'utilisation d'approches de biologie moléculaire ou génomique.

2. MODÉLISATION BIOINFORMATIQUE DES RÉSEAUX MÉTABOLIQUES D'*E. MUTABILIS* ET DE LA COMMUNAUTÉ BACTÉRIENNE

Le développement des techniques de génomique permet aujourd'hui de nouvelles approches comme la reconstruction des réseaux métaboliques. Ces stratégies visent à modéliser les voies métaboliques présentent dans un organisme *via* l'intégration de données génomiques, physiologiques et de la littérature ainsi que l'utilisation de différentes base de données et outils d'analyse informatique (Feist et al., 2009). Si de telles approches sont utilisées pour mieux comprendre le réseau métabolique de certains organismes et pouvoir ainsi optimiser des fonctions d'intérêt biotechnologique, elles peuvent être transposées en microbiologie environnementale.

Une approche *in silico* de modélisation informatique du réseau métabolique d'*E. mutabilis* et de la communauté est actuellement en cours de développement et vise à identifier et mieux comprendre les interactions trophiques entre ces organismes. Elle pourra également permettre de mettre en évidence des composés nécessaires à la croissance de certains organismes en visualisant des lacunes dans leur réseau métabolique, autorisant l'élaboration de stratégies d'isolement. Ceci est particulièrement intéressant dans l'objectif de l'obtention d'isolats de CARN1 et CARN4.

Par ailleurs, l'utilisation de cette méthode sur un organisme dont le génome n'est pas séquencé tel qu'*E. mutabilis*, et *a fortiori* sur une communauté microbienne, est novatrice : les études de ce type se concentrant d'ordinaire sur des organismes pour lesquels un génome est disponible. Néanmoins, les différentes études génomiques réalisées sur *E. mutabilis* et la communauté microbienne de Carnoulès ont permis l'obtention de différentes données permettant d'envisager la reconstruction des réseaux métaboliques respectifs. En effet, d'une part le séquençage

des ARN messagers et la métabolomique réalisée en laboratoire fournissent des informations sur l'euglène (Halter et al., 2012b; Halter et al., 2014), d'autre part le métagénome et la caractérisation du métaprotéome de Carnoulès renseignent sur la communauté bactérienne (Bertin et al., 2011).

La première étape actuellement développée consiste à reconstruire le réseau métabolique de l'euglène. En effet, séparer les réseaux des différents organismes de la communauté microbienne permettra de mieux visualiser les interactions entre individus. De plus, *E. mutabilis* étant cultivée en laboratoire, il est et sera possible de réaliser différents tests permettant la complétion de son réseau ou bien de vérifier des hypothèses soulevées par ce dernier. L'application de cette approche à l'euglène permet enfin la mise au point d'une méthodologie qui sera plus tard applicable aux bactéries du site.

Brièvement, l'analyse des séquences des transcrits d'*E. mutabilis* à tout d'abord permis d'identifier les enzymes présentes et donc les réactions ou voies métaboliques associées, conduisant à l'ébauche d'un premier réseau. Des réactions non identifiées grâce aux ARN messagers mais liées à diverses observations physiologiques et faits établis ou mentionnées dans la littérature ont ensuite été ajoutées au réseau. L'approche métabolomique réalisée en milieu synthétique, en identifiant des composés que l'euglène est capable de sécréter a conduit à ajouter les réactions conduisant à leur synthèse (Halter et al., 2012b).

Dans le même but, des analyses métabolomiques non ciblées ont été réalisées dans l'eau du Reigous lors de ce travail de thèse afin d'identifier des composés qu'elle est capable de sécréter *in situ*. Cette expérience à permis de distinguer 44 composés, toutes conditions confondues. Une analyse statistique a ensuite été réalisée afin de mettre en évidence ceux présentant une variation significative au cours du temps, indiquant leur production ou consommation et réduisant la liste à 18 molécules. L'identification de ces composés a ainsi permis de compléter le modèle du réseau métabolique d'*E. mutabilis* et de dresser une carte préliminaire de celui-ci.

Si une telle approche reste basée sur la prédiction de fonctions et n'a pas pour prétention d'établir un modèle exact, elle vise à la construction d'un réseau fonctionnel tentant de s'approcher au plus près de la réalité et permettant de débloquer des pistes de réfléxion. L'application de cette méthodologie aux bactéries de Carnoulès permettra potentiellement d'identifier des composés présents en sorties ou entrées de réseaux et donc produits ou nécessaires à la croissance, respectivement, reflétant ainsi des liens trophiques potentiels et permettant de dégager des stratégies d'isolement. Des expériences complémentaires en laboratoire seront alors nécessaires

pour confirmer et caractériser ces relations. L'ensemble de ces éléments apporteront ainsi une meilleure compréhension du fonctionnement de cette communauté soumise au stress arsénié.

BIBLIOGRAPHIE

Aguilera, A., Souza-Egipsy, V., San Martín-Úriz, P., and Amils, R. (2008) Extracellular matrix assembly in extreme acidic eukaryotic biofilms and their possible implications in heavy metal adsorption. *Aquat. Toxicol.* **88**: 257–266.

Aguilera, A., Zettler, E., Gómez, F., Amaral-Zettler, L., Rodríguez, N., and Amils, R. (2007) Distribution and seasonal variability in the benthic eukaryotic community of Río Tinto (SW, Spain), an acidic, high metal extreme environment. *Syst. Appl. Microbiol.* **30**: 531–546.

Andres, J., Arsène-Ploetze, F., Barbe, V., Brochier-Armanet, C., Cleiss-Arnold, J., Coppée, J.-Y., et al. (2013) Life in an arsenic-containing gold mine: genome and physiology of the autotrophic arsenite-oxidizing bacterium *Rhizobium* sp. NT-26. *Genome Biol. Evol.* **5**: 934–953.

Arsène-Ploetze, F., Koechler, S., Marchal, M., Coppée, J.-Y., Chandler, M., Bonnefoy, V., et al. (2010) Structure, function, and evolution of the *Thiomonas* spp. genome. *PLoS Genet.* **6**: e1000859.

Bertin, P.N., Heinrich-Salmeron, A., Pelletier, E., Goulhen-Chollet, F., Arsène-Ploetze, F., Gallien, S., et al. (2011) Metabolic diversity among main microorganisms inside an arsenic-rich ecosystem revealed by meta- and proteo-genomics. *ISME J.* **5**: 1735–1747.

Brake, S.S., Dannelly, H.K., and Connors, K.A. (2001) Controls on the nature and distribution of an alga in coal mine-waste environments and its potential impact on water quality. *Environ. Geol.* **40**: 458–469.

Bromfield, E.S.P., Tambong, J.T., Cloutier, S., Prévost, D., Laguerre, G., Berkum, P. van, et al. (2010) *Ensifer*, *Phyllobacterium* and *Rhizobium* species occupy nodules of *Medicago sativa* (alfalfa) and *Melilotus alba* (sweet clover) grown at a Canadian site without a history of cultivation. *Microbiology* **156**: 505–520.

Bruneel, O., Duran, R., Casiot, C., Elbaz-Poulichet, F., and Personne, J.-C. (2006) Diversity of Microorganisms in Fe-As-Rich Acid Mine Drainage Waters of Carnoulès, France. *Appl. Environ. Microbiol.* **72**: 551–556.

Bruneel, O., Pascault, N., Egal, M., Bancon-Montigny, C., Goñi-Urriza, M.S., Elbaz-Poulichet, F., et al. (2008) Archaeal diversity in a Fe–As rich acid mine drainage at Carnoulès (France). *Extremophiles* **12**: 563–571.

Bruneel, O., Personné, J.-C., Casiot, C., Leblanc, M., Elbaz-Poulichet, F., Mahler, B.J., et al. (2003) Mediation of arsenic oxidation by *Thiomonas* sp. in acid-mine drainage (Carnoulès, France). *J. Appl. Microbiol.* **95**: 492–499.

Bruneel, O., Volant, A., Gallien, S., Chaumande, B., Casiot, C., Carapito, C., et al. (2011) Characterization of the active bacterial community involved in natural attenuation processes in arsenic-rich creek sediments. *Microb. Ecol.* **61**: 793–810.

Das, B.K., Roy, A., Koschorreck, M., Mandal, S.M., Wendt-Potthoff, K., and Bhattacharya, J. (2009) Occurrence and role of algae and fungi in acid mine drainage environment with special reference to metals and sulfate immobilization. *Water Res.* **43**: 883–894.

Delavat, F. (2012) Composition et fonctionnement d'une communauté microbienne au sein d'un drainage minier acide : approches culturales et fonctionnelles.

Delavat, F., Lett, M.-C., and Lièvremont, D. (2013) Yeast and bacterial diversity along a transect in an acidic, As-Fe rich environment revealed by cultural approaches. *Sci. Total Environ.* **463-464**: 823–828.

Duquesne, K., Lebrun, S., Casiot, C., Bruneel, O., Personné, J.-C., Leblanc, M., et al. (2003) Immobilization of arsenite and ferric iron by *Acidithiobacillus ferrooxidans* and its relevance to acid mine drainage. *Appl. Environ. Microbiol.* **69**: 6165–6173.

Egal, M., Casiot, C., Morin, G., Elbaz-Poulichet, F., Cordier, M.-A., and Bruneel, O. (2010) An updated insight into the natural attenuation of As concentrations in Reigous Creek (southern France). *Appl. Geochem.* **25**: 1949–1957.

Fadrus, H. and Malý, J. (1975) Suppression of iron(III) interference in the determination of iron(II) in water by the 1,10-phenanthroline method. *Analyst* **100**: 549–554.

Feist, A.M., Herrgard, M.J., Thiele, I., Reed, J.L., and Palsson, B. (2009) Reconstruction of Biochemical Networks in Microbial Organisms. *Nat. Rev. Microbiol.* **7**: 129–143.

Gast, R.J., Sanders, R.W., and Caron, D.A. (2009) Ecological strategies of protists and their symbiotic relationships with prokaryotic microbes. *Trends Microbiol.* **17**: 563–569.

Giloteaux, L., Duran, R., Casiot, C., Bruneel, O., Elbaz-Poulichet, F., and Goñi-Urriza, M. (2013) Three-year survey of sulfate-reducing bacteria community structure in Carnoulès acid mine drainage (France), highly contaminated by arsenic. *FEMS Microbiol. Ecol.* **83**: 724–737.

Halter, D. (2011) Génomique fonctionnelle de micro-eucaryotes bio-indicateurs d'environnements exposés aux drainages miniers acides.

Halter, D., Andres, J., Plewniak, F., Poulain, J., Da Silva, C., Arsène-Ploetze, F., and Bertin, P.N. (2014) Arsenic hypertolerance in the protist *Euglena mutabilis* is mediated by specific transporters and functional integrity maintenance mechanisms. *Environ. Microbiol.*

Halter, D., Casiot, C., Heipieper, H.J., Plewniak, F., Marchal, M., Simon, S., et al. (2012a) Surface properties and intracellular speciation revealed an original adaptive mechanism to arsenic in the acid mine drainage bio-indicator *Euglena mutabilis*. *Appl. Microbiol. Biotechnol.* **93**: 1735–1744.

Halter, D., Goulhen-Chollet, F., Gallien, S., Casiot, C., Hamelin, J., Gilard, F., et al. (2012b) In situ proteo-metabolomics reveals metabolite secretion by the acid mine drainage bio-indicator, *Euglena mutabilis*. *ISME J.* **6**: 1391–1402.

Johnson, D.B. and Hallberg, K.B. (2005) Acid mine drainage remediation options: a review. *Sci. Total Environ.* **338**: 3–14.

Johnson, D.B. and Hallberg, K.B. (2003) The microbiology of acidic mine waters. *Res. Microbiol.* **154**: 466–473.

Leblanc, M., Achard, B., Ben Othman, D., Luck, J.M., Bertrand-Sarfati, J., and Personné, J.C. (1996) Accumulation of arsenic from acidic mine waters by ferruginous bacterial accretions (stromatolites). *Appl. Geochem.* **11**: 541–554.

Marchal, M., Briandet, R., Halter, D., Koechler, S., DuBow, M.S., Lett, M.-C., and Bertin, P.N. (2011) Subinhibitory arsenite concentrations lead to population dispersal in *Thiomonas* sp. *PloS One* **6**: e23181.

Marchal, M., Briandet, R., Koechler, S., Kammerer, B., and Bertin, P.N. (2010) Effect of arsenite on swimming motility delays surface colonization in *Herminiimonas arsenicoxydans*. *Microbiol. Read. Engl.* **156**: 2336–2342.

Morin, G., Juillot, F., Casiot, C., Bruneel, O., Personné, J.-C., Elbaz-Poulichet, F., et al. (2003) Bacterial formation of tooeleite and mixed arsenic(III) or arsenic(V)-iron(III) gels in the Carnoulès acid mine drainage, France. A XANES, XRD, and SEM study. *Environ. Sci. Technol.* **37**: 1705–1712.

Muller, D., Médigue, C., Koechler, S., Barbe, V., Barakat, M., Talla, E., et al. (2007) A tale of two oxidation states: bacterial colonization of arsenic-rich environments. *PLoS Genet.* **3**: e53.

Santini, J.M., Sly, L.I., Schnagl, R.D., and Macy, J.M. (2000) A new chemolithoautotrophic arsenite-oxidizing bacterium isolated from a gold mine: phylogenetic, physiological, and preliminary biochemical studies. *Appl. Environ. Microbiol.* **66**: 92–97.

Sumida, S., Lyman, H., Kiyohara, N., and Osafune, T. (2007) Mechanism of conversion from heterotrophy to autotrophy in *Euglena gracilis*. *Cytologia (Tokyo)* **72**: 447–457.

Tyson, G.W., Lo, I., Baker, B.J., Allen, E.E., Hugenholtz, P., and Banfield, J.F. (2005) Genome-directed isolation of the key nitrogen fixer *Leptospirillum ferrodiazotrophum* sp. nov. from an acidophilic microbial community. *Appl. Environ. Microbiol.* **71**: 6319–6324.

Valente, T.M. and Gomes, C.L. (2007) The role of two acidophilic algae as ecological indicators of acid mine drainage sites. *J. Iber. Geol.* **33**: 283–294.

Volant, A., Bruneel, O., Desoeuvre, A., Héry, M., Casiot, C., Bru, N., et al. (2014) Diversity and spatiotemporal dynamics of bacterial communities: physicochemical and others drivers along an acid mine drainage. *FEMS Microbiol. Ecol.*

Warren, M.J., Raux, E., Schubert, H.L., and Escalante-Semerena, J.C. (2002) The biosynthesis of adenosylcobalamin (vitamin B12). *Nat. Prod. Rep.* **19**: 390–412.

Jérémy ANDRES

Mécanismes adaptatifs et interactions métaboliques au sein de communautés microbiennes soumises au stress arsénié

L'arsenic est naturellement présent dans la croûte terrestre et de manière abondante dans certains environnements. Si cet élément est toxique pour la plupart des formes de vies, des micro-organismes développent différents mécanismes pour y faire face. Ce travail porte sur l'étude de ces processus impliquant à la fois des réponses individuelles et des interactions entre organismes appartenant à des domaines différents du vivant. Des approches de génomique descriptive et fonctionnelle mettent ainsi en évidence différents mécanismes adaptatifs et fonctions cellulaires impliqués dans la réponse à l'arsenic d'une bactérie et d'un protiste photosynthétique, *Rhizobium* sp. NT-26 et *Euglena mutabilis*, tous deux particulièrement résistants à cet élément. Par ailleurs, tandis que *Rhizobium* sp. NT-26 semble avoir perdu sa capacité à interagir avec les plantes, *E. mutabilis* fait au contraire partie intégrante d'une communauté microbienne incluant différentes bactéries et bénéficie de leur activité.

Mots clés : arsenic, *Rhizobium* sp. NT-26, *Euglena mutabilis*, adaptation, interactions, génomique

Arsenic naturally occurs in earth crust and is particularly abundant in some environments. While this element is toxic for most forms of life, micro-organisms have evolved different mechanisms to cope with it. This work deals with these different processes involving individual responses as well as interactions between organisms belonging to different domains of life. Descriptive and functional genomics approaches highlight several adaptative mechanisms and cellular functions involved in the arsenic response of a bacterium and a photosynthetic protist, *Rhizobium* sp. NT-26 and *Euglena mutabilis*, respectively, both being particularly resistant to arsenic. Also, while *Rhizobium* sp. NT-26 seems to have lost its ability to interact with plants, *E. mutabilis* is on the contrary an integral part of a microbial community including different bacteria and benefits from their activity.

Keywords : arsenic, *Rhizobium* sp. NT-26, *Euglena mutabilis*, adaptation, interactions, genomics