

Thèse présentée pour obtenir le grade de

Docteur de l'Université de Strasbourg



Discipline : Science du Vivant

Mention : Aspects Moléculaires et Cellulaires de la biologie

Par David HALTER

Génomique fonctionnelle de micro-eucaryotes bio-indicateurs d'environnements exposés aux drainages miniers acides

Soutenue publiquement le 3 octobre 2011

Composition du jury :

M. Joseph Schacherer (Université de Strasbourg) – Examineur interne

Mme Céline Brochier (Université de Provence Centre St Charles) – Examineur externe

M. Pierre Peyret (Université de Clermont 1) – Rapporteur externe

M. Philippe Normand (Université Claude Bernard Lyon) – Rapporteur externe

Mme Florence Arsène-Ploetze (Université de Strasbourg) – Co-directrice de thèse

M. Philippe Bertin (Université de Strasbourg) – Directeur de thèse

REMERCIEMENTS

Cette thèse a été financée par le ministère de l'Éducation 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 par Serge Potier. Je le remercie de m'avoir accueilli au sein de son unité.

Je tiens à remercier particulièrement Philippe Bertin et Florence Arsène-Ploetze pour avoir dirigé ce travail de thèse, ainsi que Frédéric Plewniac pour ses remarques constructives dans le domaine de la bio-informatique. Je remercie également les autres membres de l'équipe, Sandrine, Marie, Lucie, François, Jérémy pour leur disponibilité et leur soutien et toutes les personnes du laboratoire, qui d'une manière ou d'une autre, m'ont aidé dans ce travail.

Je remercie vivement Hermann Heipieper et Guillaume Tcherkez pour leurs nombreux conseils et leur implication lors de collaborations scientifiques enrichissantes à tous les niveaux ainsi que le personnel du Génoscope pour son aide précieuse dans les étapes de séquençage et de traitement des données.

Un grand merci à mes parents qui ont su me soutenir durant ces longues années universitaires.

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

SOMMAIRE

LISTE DES TABLES ET DES FIGURES	10
LISTE DES ABREVIATIONS	11

INTRODUCTION

1. LES DRAINAGES MINIERS ACIDES (DMA) : DES ECOSYSTEMES TOXIQUES	14
1.1. Définition des DMA	14
1.2. L'arsenic dans les DMA	15
1.3. Les communautés microbiennes associées aux DMA	17
1.3.1. Impact des communautés bactériennes	17
1.3.1.1. Bio-lixiviation	17
1.3.1.2. Bio-remédiation	18
1.3.2. Diversité micro-eucaryote : quelles espèces ?	19
1.3.3. Diversité micro-eucaryote : comment l'étudier ?	24
1.3.4. Communauté micro-eucaryote : quels impacts sur les DMA ?	25
1.3.4.1. Impact des micro-eucaryotes photosynthétiques	25
1.3.4.2. Impact des micro-eucaryotes hétérotrophes	27
2. METABOLISME DE L'ARSENIC CHEZ LES EUCARYOTES	29
2.1. Mécanismes d'entrée de l'arsenic dans les cellules	29
2.1.1. Chez la <i>Saccharomyces cerevisiae</i>	29
2.1.2. Chez les plantes et les algues	30
2.1.2.1. Plantes non accumulatrices d'arsenic	30
2.1.2.2. Plantes hyper-accumulatrices	31
2.1.3. Chez les mammifères	32
2.2. Impact de l'arsenic au sein des cellules	33
2.2.1. Stress oxydatif	33
2.2.2. Effets mutagènes	35
2.2.3. Impact sur le cycle cellulaire	35
2.3. Mécanismes de défense face à l'arsenic chez les eucaryotes	36
2.3.1. Limitation de la bio-disponibilité	36
2.3.1.1. Synthèse d'exopolysaccharides	36
2.3.1.2. Modification du transport membranaire	37
2.3.1.3. Précipitation extracellulaire de l'arsenic	38

2.3.2. Expulsion de l'arsenic	38
2.3.2.1. Exemple du système <i>arsRDABC</i> procaryotique	39
2.3.2.2. Le système ACR de <i>S. cerevisiae</i>	39
2.3.3 Séquestration intracellulaire de l'arsenic	41
2.3.3.1. Cystéine et glutathion	41
2.3.3.2. Phytochélatines	41
2.3.3.3. Les métallothionéines	43
2.3.4 La méthylation de l'arsenic, un processus de détoxication ?	44
2.3.5 Conclusion	46
..	
3. LES METHODES - OMICS	47
3.1. Intérêts des approches –OMICS	47
3.2. La (méta-)génomique	48
3.2.1. Techniques de séquençage	48
3.2.1.1. Le pyro-séquençage	48
3.2.1.2. La technologie Illumina	50
3.2.2.3. Les technologies SOLiD et de troisième génération	52
3.2.2. Enjeux et limites de la méta-génomique	52
3.3. La (méta-)transcriptomique	54
3.4. La (méta-)protéomique	54
3.4.1. Définition	54
3.4.2. Analyses par spectrométrie de masse	55
3.4.3. Méta-protéome : un contenu non exhaustif	55
3.5. La (méta-)métabolomique	57
3.6. Exemple de l'organisme modèle <i>S. cerevisiae</i>	58
4. OBJECTIFS DE LA THESE	61

RESULTATS

Chapitre I : Génomique descriptive et fonctionnelle de la communauté bactérienne du DMA de Carnoulès	65
A. Introduction	66
B. Résultats	68
C. Discussion et conclusion	70
D. Manuscrit	71
Chapitre II : <i>Euglena mutabilis</i> : quel rôle dans les interactions trophiques ?	73
A. Introduction	74
B. Résultats	75
C. Discussion et conclusion	77
D. Manuscrit	78

Chapitre III : La réponse adaptative d'*E. mutabilis* face au stress arsénié par des approches de physiologie comparative 80

A. Introduction	81
B. Résultats	83
C. Discussion et conclusion	83
D. Manuscrit	84

Chapitre IV : La réponse à l'arsenic chez *E. gracilis* et *E. mutabilis* au stress arsénié par des approches de génomique fonctionnelle 86

A. Introduction	87
B. Résultats	88
1. Approche de protéomique différentielle	88
2. Approche de protéomique différentielle	91
a. Préparation des échantillons	91
b. Mécanismes induits chez <i>E. gracilis</i> en présence d'arsenic	95
c. Mécanismes induits chez <i>E. mutabilis</i> en présence d'arsenic	96
d. Etude transcriptomique comparative entre <i>E. gracilis</i> et <i>E. mutabilis</i>	98
C. Discussion et conclusion	101
1. Approche de protéomique différentielle	101
2. Approche de protéomique différentielle	102
D. Tables supplémentaires	106

CONCLUSION GENERALE ET PERSPECTIVES

ANNEXES

LISTE DES TABLES ET DES FIGURES

Figure 1.1	Spéciation de l'arsenic en fonction du pH et du potentiel d'oxydo-réduction.	15
Figure 1.2	Principales formes arséniées retrouvées dans l'environnement.	18
Figure 1.3	Analyse phylogénétique basée sur les ADNr 18S de la communauté micro-eucaryote du DMA du Rio Tinto.	20
Figure 1.4	Principaux flux de composés arséniés chez les plantes non-accumulatrices et hyper-accumulatrices d'arsenic.	32
Figure 1.5	Comparaison des systèmes de résistance à l'arsenic couplant réduction de l'arséniate et expulsion de l'arsénite chez les procaryotes et chez les eucaryotes.	40
Figure 1.6	Principales étapes des voies de biosynthèse du glutathion et des phytochélatines.	43
Figure 1.7	Mécanismes de défense face aux stress métalliques chez les eucaryotes.	46
Figure 1.8	Principales étapes relatives au pyro-séquençage.	50
Figure 1.9	Principales étapes relatives au séquençage Solexa	51
Figure 1.10	Carte de chaleur des transcrits induits ou réprimés en présence de différents métaux chez la levure.	59
Table 1.1	Diversité des principaux micro-organismes photosynthétiques identifiés dans les milieux aquatiques exposés aux DMA.	21-23
Table 1.2	Définition des principales caractéristiques mesurées dans les biofilms photosynthétiques d'algues présents dans les DMA (Souza-Egipsy et al., 2011).	25
Figure 2.1	Photo du site de Carnoulès au niveau de la confluence entre le Reigous et l'Amous.	66
Table 2.1	Principales caractéristiques physico-chimiques du Reigous.	67
Figure 3.1	Ultra-structure des euglènes et organisation de la pellicule	74
Figure 3.2	Test d'auxotrophie chez <i>E. mutabilis</i>	76
Figure 3.3	Test de résistance au Ph chez <i>E. mutabilis</i> .	76
Figure 4.1	Gel d'acrylamide bi-dimensionnel obtenu avec les protéines extraites d' <i>E. mutabilis</i> .	88-89
Figure 4.2	Profil des ARN extraits des euglènes avant et après enrichissement des ARNm.	92
Figure 4.3	Principales étapes aboutissant à la répartition des séquences soumises à MG-RAST en grands groupes fonctionnels.	94
Figure 4.4	Répartition en 28 grands groupes fonctionnels des séquences annotées par MG-RAST correspondant aux séquences d' <i>E. gracilis</i> et d' <i>E. mutabilis</i> en présence d'arsenic	98
Figure 4.5	Caractéristiques structurales des alginates et des monomères qui les constituent.	99
Figure 4.6	Image de microscopie optique et électronique montrant la synthèse d'une matrice extracellulaire par <i>E. mutabilis</i> .	100

Figure 4.7	Observations en microscopie optique d' <i>E. mutabilis</i> et d' <i>E. gracilis</i> montrant la présence de granules de fer dans le compartiment intracellulaire.	102
Figure 4.8	Observations en microscopie électronique d' <i>E. mutabilis</i> montrant la présence de granules de fer dans le compartiment intracellulaire.	103
Figure 4.9	Principe du « test des comètes » permettant de visualiser l'effet de l'arsenic sur l'intégrité du génome d' <i>E. gracilis</i> et d' <i>E. mutabilis</i> .	104
Figure 4.10	Principales étapes d'un protocole visant à étudier les caractéristiques physico-chimiques des biofilms produit par les euglènes	105
Table 4.1	Identification des protéines présentant une accumulation différentielle chez <i>E. mutabilis</i> en présence d'arsenic.	89-90
Table 4.2	Principales caractéristiques des contigs obtenus pour les transcriptomes d' <i>E. mutabilis</i> et d' <i>E. gracilis</i> en présence ou en absence d'arsenic.	93
Table 4.3	Principales fonctions biologiques induites ou réprimées par <i>E. gracilis</i> en présence d'arsenic.	106
Table 4.4	Principales fonctions biologiques induites ou réprimées par <i>E. mutabilis</i> en présence d'arsenic.	107
Table 4.5	Principales caractéristiques des contigs obtenus pour les transcriptomes d' <i>E. mutabilis</i> et d' <i>E. gracilis</i> en présence ou en absence d'arsenic.	107-108
Figure 5.1	Perspective d'étude des flux métaboliques entre <i>E. mutabilis</i> et la communauté bactérienne de Carnoulès	113
Figure 5.2	Principales étapes d'un protocole visant à cribler les transcrits conférant une tolérance à l'arsenic chez <i>E. mutabilis</i> .	114
Figure 5.3	Localisation du point de prélèvement ayant conduit à l'isolement de <i>Chlamydomonas acidophila</i> et observation microscopique de l'algue	115

LISTE DES ABREVIATIONS

As(GS) ₃	Arsénite tri-glutathion
As(III)	Arsénite
As(V)	Arséniate
DMA	Drainage Minier Acide
DMAs ^{III}	Diméthylarsine
DMAs ^V	acide diméthylarsinique
EOR	Espèces Oxygénée Réactives
EPS	Exopolysaccharides
GS	Glutathion oxydé
GSH	Glutathion réduit
MMAs ^{III}	monométhylarsine
MMAs ^V	acide monométhylarsonique
MT	Métallothionéine
ppm	Partie Par Million
TMA	Triméthylarsine
TMAO	Oxyde de triméthylarsine

INTRODUCTION

1. LES DRAINAGES MINIERES ACIDES (DMA) : DES ÉCOSYSTÈMES TOXIQUES

1.1. DÉFINITION DES DMA

Les anciennes exploitations minières peuvent aboutir à la formation de drainages miniers acides (DMA) lorsque certaines conditions géologiques, chimiques et biologiques sont réunies. En effet, au cours d'une exploitation minière, des volumes très importants de roches excavées sont accumulés sous forme de stériles miniers ce qui les expose davantage à des processus d'oxydation chimique et biologique. Ainsi, l'eau de pluie, l'oxygène atmosphérique mais également les communautés microbiennes peuvent accéder par le biais d'anfractuosités à une très grande surface de roches, inaccessible avant excavation. Ceci conduit à une oxydation rapide des minéraux présents dans les roches, en particulier des minéraux soufrés tels que la pyrite (FeS_2). Cette oxydation de minéraux soufrés génère de l'acide sulfurique qui lessive les roches et facilite la libération des métaux lourds et des métalloïdes présents dans celles-ci. Ainsi, des processus naturels d'oxydation de roches qui, en l'absence d'intervention humaine se seraient déroulés sur une période de temps très longue sur une échelle de temps géologique, se réalisent brutalement. Les effluents qui découlent des stériles miniers, souvent acides et chargés en éléments métalliques, sont connus sous le nom de DMA. Ce processus est sans conteste à l'origine de pollutions d'origines anthropiques parmi les plus étendues et les plus persistantes au monde et peut avoir des effets dévastateurs sur des environnements aquatiques sur le long terme. Cependant, il est à noter que ces processus ne trouvent pas toujours une origine anthropique. En effet, certaines régions dont le fond géochimique est particulièrement riche en minéraux soufrés peuvent conduire à un processus similaire même en l'absence d'exploitation minière (Rio Tinto, Espagne). Par analogie aux DMA, ce processus est alors appelé drainage acide de roches (DRA) pour souligner son origine naturelle.

Le processus à l'origine de la formation des DMA se poursuit jusqu'à ce que les roches constituant le stérile soient complètement oxydées ce qui peut prendre plusieurs siècles. La composition des DMA en éléments métalliques reflète en général la composition des roches desquelles ils sont issus. Cependant, le métabolisme bactérien, en favorisant préférentiellement l'oxydation de certains minerais par rapport à d'autres ou en modifiant la solubilité des éléments en solution, peut affecter les caractéristiques physico-chimiques des effluents.

1.2. L'ARSENIC DANS LES DMA

L'arsenic est un composé relativement abondant dans la croûte terrestre et se présente sous différentes formes plus ou moins toxiques. Ce composé est retrouvé sous 4 états d'oxydation selon le potentiel d'oxydo-réduction de son milieu : l'arsine (As^{3+}), l'arsenic élémentaire (As), l'arsénite (As^{III}) et l'arséniate (As^V). D'une manière générale, l'arsenic est complexé à d'autres métaux ou composés inorganiques sous formes insolubles. Ainsi, ce métalloïde est-il souvent retrouvé sous forme d'arsénopyrite ($FeAsS$) dans les roches excavées, dont l'oxydation conduit à la production d'acide sulfurique et à la mobilisation de fer et d'arsenic.

Lorsqu'il est en solution, la spéciation de l'arsenic est affectée par des facteurs physico-chimiques et biologiques. Le diagramme ci-dessous (Figure 1.1) présente, en l'absence d'activité biologique, les espèces arsénées majoritaires dans un environnement donné en fonction de son pH et de son potentiel d'oxydo-réduction (Eh). On notera que dans les conditions qui prévalent habituellement dans les DMA, l'arsenic est principalement présent sous forme d'AsIII [$As(OH)_3$] et d'AsV [H_2AsO_4 / H_3AsO_4]

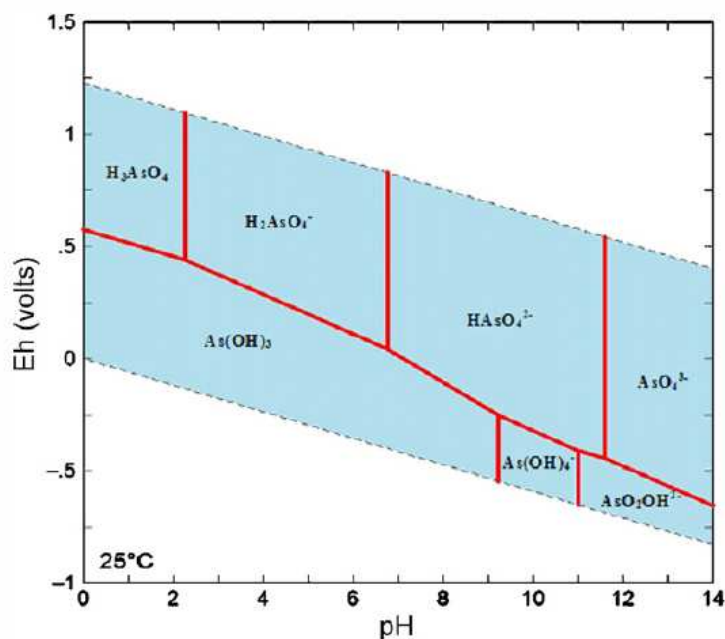


Figure 1.1 : Spéciation de l'arsenic en fonction du pH et du potentiel d'oxydo-réduction du milieu et en l'absence de toute modification d'origine biologique (Cheng et al., 2009).

Outre ces formes inorganiques, l'arsenic est également retrouvé sous forme méthylées. Ces composés résultent principalement d'activités biologiques et conduisent à de multiples intermédiaires présentés ci-dessous (Figure 1.2). En effet, l'arsenic peut être associé à un, deux ou trois groupements méthyles et être sous un état d'oxydation variable (As^{III} ou As^V) ce qui conduit à une diversité assez importante de composés. Il est à noter que la forme arsénée oxydée la plus

méthylée, l'oxyde de triméthylarsine, est une forme volatile, ce qui peut avoir des conséquences importantes dans les cycles bio-géochimiques de ce métalloïde (Cullen & Bentley, 2005). D'autre part, certains micro-organismes ont été décrits pour leur capacité à utiliser ces formes méthylées de l'arsenic comme source de carbone, et ainsi conduire à la production de formes inorganiques (Yoshinaga et al., 2011). Enfin, par des réactions d'oxydo-réduction catalysées biologiquement, le passage d'une forme méthylée oxydée à la même forme méthylée réduite peut également être observé (Lehr et al., 2003; Yoshinaga et al., 2011). Les micro-organismes jouent par conséquent un rôle considérable dans les cycles bio-géochimiques de l'arsenic en affectant ses propriétés physico-chimiques (Figure 1.2). Ces modifications peuvent avoir un impact sur leur toxicité, certaines formes arséniées étant plus facilement assimilables que d'autres. Ainsi, les formes méthylées réduites (MMAs^{III}, DMAs^{III}) sont-elles considérées comme étant 100 à 1000 fois plus toxiques que les formes inorganiques (As^{III}, As^V), elles-mêmes 100 à 1000 fois plus toxiques que les formes arséniées méthylées oxydées (MMAs^V, DMAs^V) (Akter *et al.*, 2006; Mandal & Suzuki, 2002). Cependant la toxicité de ces composés peut varier selon les organismes et les environnements dans lesquels les analyses sont effectuées, leur biodisponibilité n'étant pas toujours extrapolable d'une espèce et d'un environnement à l'autre.

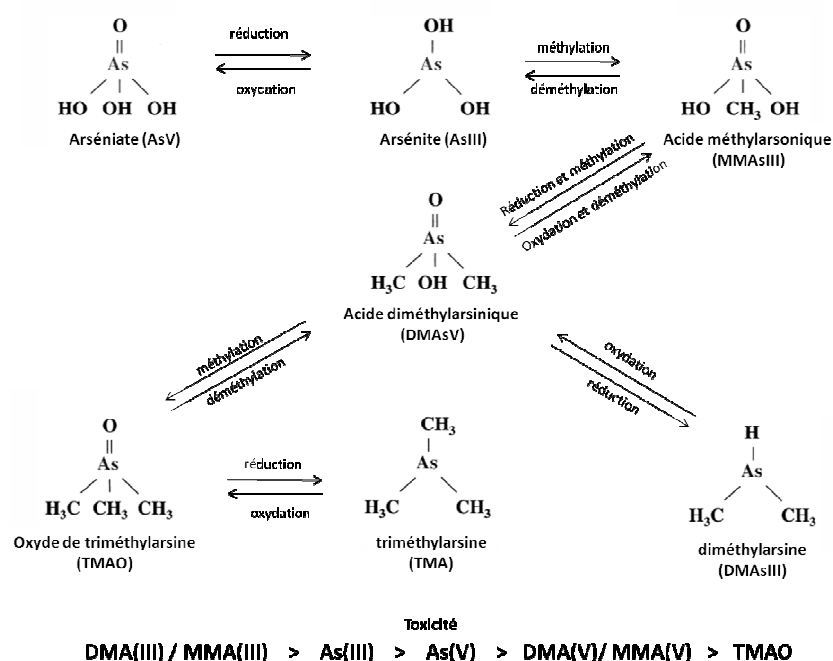


Figure 1.2 : Principales formes arséniées retrouvées dans l'environnement. Le passage d'une forme arséniée à une autre peut avoir lieu lors de réactions biologiques, principalement grâce au métabolisme d'organismes procaryotiques. La toxicité relative des différents composés arséniés est présentée en bas de la figure.

1.3. LES COMMUNAUTÉS MICROBIENNES ASSOCIÉES AUX DMA

Les DMA, de par leurs fortes concentrations en éléments métalliques et leurs fortes acidités, affectent sévèrement les écosystèmes aquatiques situés en aval. En effet, les propriétés physico-chimiques de ces eaux de DMA entraînent un stress acide, oxydatif et osmotique qui est incompatible avec la survie de nombreux organismes vivants. Cependant, certains micro-organismes, eucaryotes et procaryotes, sont couramment retrouvés dans ces niches écologiques particulières et peuvent conduire à travers leur métabolisme à des processus d'atténuation ou d'amplification des DMA. Les impacts des métabolismes microbiens sur ces DMA sont discutés ci-dessous, en distinguant les communautés microbiennes pro- et eucaryotes.

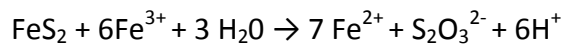
1.3.1. IMPACT DES COMMUNAUTÉS BACTÉRIENNES

Historiquement, l'impact des communautés procaryotes sur les DMA semble avoir été étudié bien avant celui de la communauté micro-eucaryote. En effet, nombre de bactéries sont capables de développer un métabolisme qui affecte significativement les cycles géo-chimiques de métaux retrouvés dans les DMA rendant leur impact sur ces écosystèmes plus visibles. Cette modification des cycles bio-géochimiques des métaux peut conduire à une atténuation des processus de formation des DMA ou au contraire à leur amplification. Ces processus d'atténuation et d'amplification peuvent être tous les deux valorisés soit pour limiter la pollution liée à d'anciennes exploitations minières (bio-remédiation) soit pour faciliter l'extraction de métaux dans des mines en cours d'exploitation (bio-lixiviation).

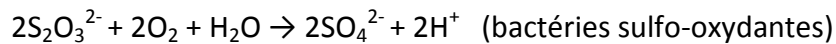
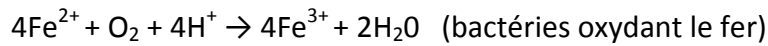
1.3.1.1. Bio-lixiviation

Les métabolismes bactériens à l'origine des processus de bio-lixiviation sont multiples et font généralement intervenir des bactéries utilisant le fer et les composés soufrés comme donneurs d'électrons dans la chaîne respiratoire. Dans le cas de minéraux soufrés tel que la pyrite (FeS_2), le minéral est d'abord oxydé d'une manière abiotique par des ions ferriques Fe^{3+} entraînant la formation de thiosulfate ($\text{S}_2\text{O}_3^{2-}$) et de fer ferreux Fe^{2+} . Dans un second temps, le thiosulfate et le fer ferreux Fe^{2+} sont oxydés respectivement en fer ferrique Fe^{3+} et en sulfate sous l'action du métabolisme bactérien ce qui permet de régénérer le système en ions Fe^{3+} et de produire une acidité favorisant les réactions d'oxydo-réduction entre le fer ferrique et le minéral (Johnson & Hallberg, 2005).

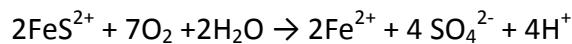
1ère étape (réaction abiotique) :



2^{ème} étape (réactions biotiques)



Bilan:



Un processus similaire peut conduire à la libération d'ions Cu^{2+} dans le cas où les roches renferment des minéraux tels que la chalcopryrite (CuFeS_2), d'ions uranyles (UO_2^{2+}) dans le cas de l'uranite (UO_2) ou d'arsénite (AsO_2^-) dans le cas de l'arsénopyrite (FeAsS) (Baker & Banfield, 2003; Mandl *et al.*, 1992). La bio-lixiviation, bien que très polluante et nécessitant des étapes ultérieures pour la récupération des métaux, permet néanmoins une exploitation à moindre coût de roches renfermant de faibles proportions d'éléments à hautes valeurs ajoutées (Cu, U, Au,...).

1.3.1.2. Bio-remédiation

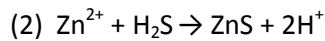
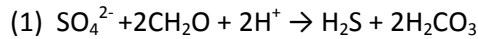
Si le métabolisme bactérien peut être à l'origine de la mobilisation d'éléments métalliques dans les processus de bio-lixiviation, il peut également conduire dans certains cas à des processus d'immobilisation de ces mêmes éléments. Par conséquent, ce processus aboutit à une moindre concentration en éléments métalliques dans les DMA, ce qui peut être apparenté à de la bio-remédiation. D'une manière générale, ces processus d'atténuation reposent sur la modification de la spéciation des ions métalliques présents en solution ou leur complexation avec d'autres composés sous des formes moins solubles (Johnson & Hallberg, 2005).

Ainsi, le métabolisme photosynthétique, l'ammonification, la nitrification ou la méthanogénèse sont des processus conduisant à la production d'alcalinité entraînant la précipitation des ions métalliques sous forme d'hydroxydes (Johnson & Hallberg, 2005). La réaction ci-dessous illustre la production d'alcalinité lors de la photosynthèse :



D'autre part, les bactéries sulfato-réductrices peuvent convertir un acide fort, l'acide sulfurique, en acide faible, le sulfure d'hydrogène, qui forme par ailleurs des complexes insolubles

avec les ions métalliques tels que le cadmium, le zinc ou le cuivre (Johnson & Hallberg, 2005). Les équations bilan conduisant à la production de sulfure d'hydrogène (1) et à l'immobilisation de métaux (2) sont présentées ci-dessous, en prenant l'exemple du zinc :



Enfin, le métabolisme bactérien peut également modifier la spéciation de certains éléments métalliques, conduisant dans certains cas à une diminution de leur solubilité. Ainsi les bactéries oxydant le fer ou l'arsenic peuvent conduire à la formation de minerais insolubles tels que la tooéélite ($\text{Fe}_6^{3+}(\text{As}^{3+}\text{O}_3)_4(\text{SO}_4)(\text{OH})_4 \cdot 4\text{H}_2\text{O}$) (Morin *et al.*, 2003)

Si l'impact de la communauté bactérienne procaryote sur l'évolution des DMA est relativement bien compris, le rôle de la communauté micro-eucaryote est par opposition souvent moins clair.

1.3.2. DIVERSITÉ MICRO-EUCARYOTES : QUELLES ESPÈCES ?

La diversité des micro-organismes eucaryotes semble assez variable d'un DMA à un autre (Baker *et al.*, 2009). Ainsi, si le DMA de Richmond aux États-Unis présente une faible diversité de micro-eucaryotes, principalement dominée par des champignons, d'autres environnements exposés aux drainages acides tel celui du Rio Tinto en Espagne présentent une diversité assez importante de micro-eucaryotes d'origines phylogénétiques très variées (Amaral Zettler *et al.*, 2002; Baker & Banfield, 2003). Ainsi, une étude couplant analyse des séquences d'ADNr 18S, observations microscopiques et isolements en condition de laboratoire a permis de révéler une diversité insoupçonnée de micro-eucaryotes sur ce site. Les principales espèces identifiées dans cette étude sont présentées ci-dessous (Figure 1.3) (Zettler *et al.*, 2002).

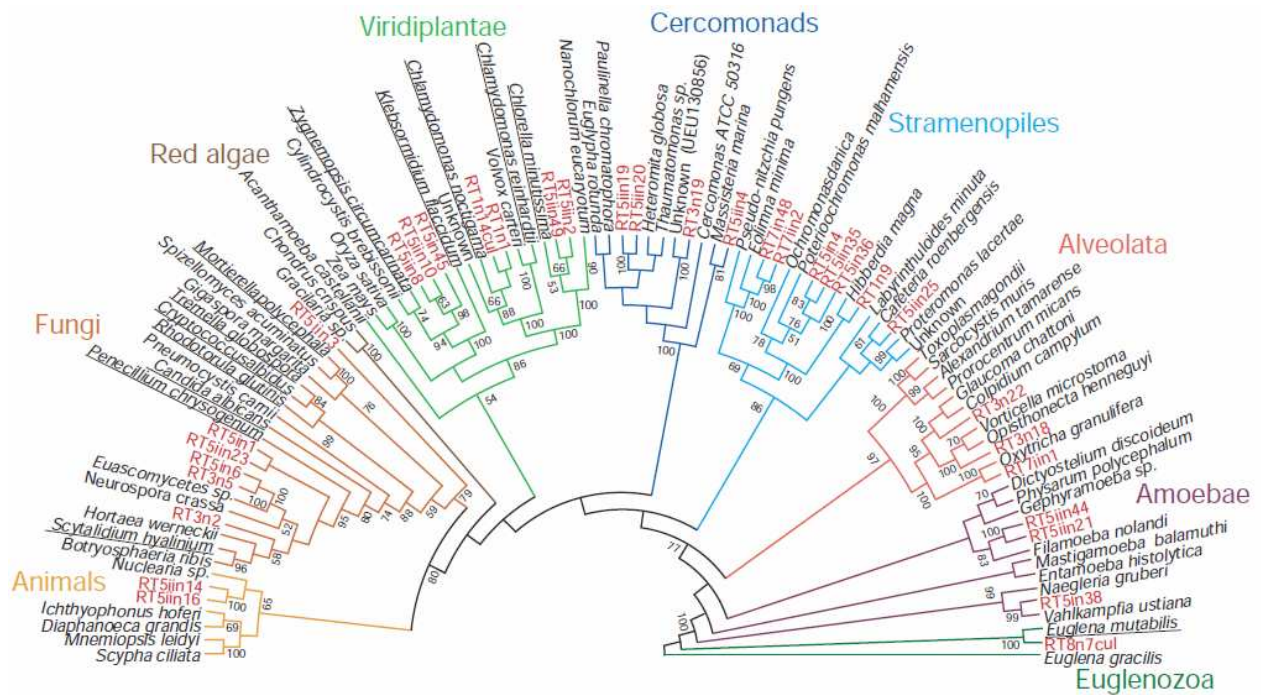

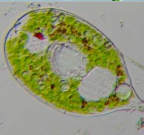








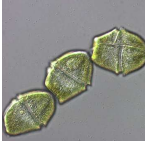


Figure 1.3 : Analyse phylogénétique basée sur les ADNr 18S de la communauté micro-eucaryote du DMA du Rio Tinto. Les espèces en gras représentent des clones environnementaux, RT : les séquences d'espèces identifiées sur le site du Rio Tinto en Espagne. Les taxons soulignés représentent les espèces identifiées par observations microscopiques (Zettler et al., 2002).

Cette communauté micro-eucaryote est en outre représentée par de nombreux biofilms photosynthétiques qui semblent constituer les formes de vie dominantes dans cet écosystème (Souza-Egipsy et al., 2011). Ces organismes photosynthétiques sont particulièrement bien représentés par des chlorophytes (algues vertes) du genre *Chlamydomonas*, *Dunaliella* et *Chlorella* et par des euglénides de l'espèce *Euglena mutabilis*. Des chlorophytes du genre *Zygnemopsis* et *Klebsormidium* et, dans une moindre mesure, *Mesotaenium* et *Stichococcus* sont également retrouvés dans ces environnements (Souza-Egipsy et al., 2011). Enfin, des représentants des Rhodophytes (*Cyanidium*) et des diatomées du genre *Pinnularia* forment occasionnellement de vastes biofilms dans les sites exposés aux DMA. Il est à noter que certaines espèces d'organismes photosynthétiques sont retrouvées d'une manière redondante dans les environnements exposés aux DMA, en particulier *Klebsormidium acidophilum*, *Navicula cincta*, *Chlamydomonas acidophila* et *Euglena mutabilis* et peuvent par conséquent être considérés comme des indicateurs biologiques de tels écosystèmes (Valente & Gomes, 2007).

D'une manière plus générale, d'autres espèces photosynthétiques peuvent être retrouvées dans des environnements exposés aux DMA. Les principales espèces décrites dans la littérature sont représentées ci-dessous (Table 1.1 inspirée de Novis P et Harding J. (2007)).

Micro-organismes photosynthétiques		pH	auteurs
Excavata : euglénides			
<i>Euglena mutabilis</i>		1,4 à 4,7	(Brake <i>et al.</i> , 2001a; Casiot <i>et al.</i> , 2004; Kapfer, 1998; Lessmann <i>et al.</i> , 1999)
<i>Phacus curvicauda</i>		2,5 à 5,8	(Douglas <i>et al.</i> , 1998)
<i>Lepocinclis</i> sp.		2,3 à 2,9	(Lessmann <i>et al.</i> , 2000)
Chromistes : chrysophytes			
<i>Chromulina</i> sp.		2,3 à 2,9	(Lessmann <i>et al.</i> , 2000)
<i>Ochromonas</i> sp.		2,3 à 2,9	(Kamjunke <i>et al.</i> , 2004; Lessmann <i>et al.</i> , 2000; Wollmann <i>et al.</i> , 2000)
Chromistes : cryptomonades			
<i>Cryptomonas ovata</i>		2,5 à 5,7	(Douglas <i>et al.</i> , 1998)
<i>Cyathomonas</i> sp.		2,3 à 2,9	(Lessmann <i>et al.</i> , 2000)
<i>Rhodomonas minuta</i>		2,3 à 2,9	(Lessmann <i>et al.</i> , 2000)
Chromistes : diatomées			
<i>Eunotia</i> sp.		2,3 à 7,7	(Douglas <i>et al.</i> , 1998) (Lessmann <i>et al.</i> , 2000)
<i>Frustulia rhomboides</i>		2,5 à 7,6	(Douglas <i>et al.</i> , 1998)

<i>Navicula mutica</i>		2,5 à 7,3	(Douglas et al., 1998)
<i>Nitzschia</i> sp.		2,3 à 2,9	(Lessmann et al., 2000)
<i>Pinnularia</i> sp.		1,5 à 2,4	(Douglas et al., 1998)
<i>Surirella tenuissima</i>		2,5	(Douglas et al., 1998)
<i>Synedra</i> sp.		2,5	(Douglas et al., 1998)
Chromistes : dinoflagellés			
<i>Gymnodinium</i> sp.		2,6 à 3,6	(Wollmann et al., 2000)
Chromistes : xanthophytes			
<i>Tribonema</i> sp.		3,1 à 3,4	(Winterbourn et al., 2000)
Plantae : chlorophytes			
<i>Chlamydomonas acidophila</i>		2,5 à 3,6	(Kamjunke et al., 2004; Lessmann et al., 2000; Tittel et al., 2005; Woelfl et al., 2000; Wollmann et al., 2000)
<i>Chlorella</i> sp.		3,0	(Souza-Egipsy et al., 2011)
<i>Chlorococcum</i> sp.		8,0	(Garcia-Meza et al., 2005)
<i>Microspora</i> sp.		2,6 à 3,3	(Verb & Vis, 2001)

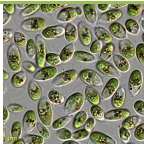
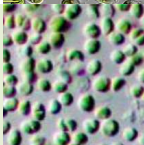
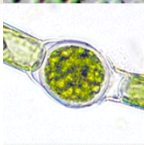
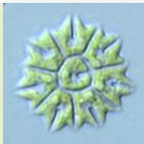
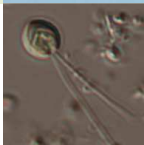
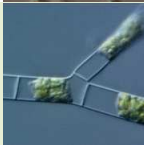

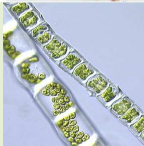



<i>Dunaliella</i> sp.		3,0	(Souza-Egipsy et al., 2011)
<i>Nanochlorum</i> sp.		2,3 à 2,9	(Lessmann et al., 2000)
<i>Oedogonium</i> sp.		2,5 à 7,9	(Douglas et al., 1998)
<i>Pediastrum tetras</i>		2,5 à 5,8	(Douglas et al., 1998)
<i>Scourfieldia cordiformis</i>		2,3 à 2,9	(Lessmann et al., 2000)
<i>Stigeoclonium</i> sp.		8,2	(Pawlik-Skowronska, 2001)
<i>Trentepohlia annulata</i>		2,5	(Douglas et al., 1998)
<i>Ulothrix</i> sp.		3,0 à 11,0	(Rousch & Sommerfeld, 1999)
Plantae : charophytes			
<i>Klebsormidium acidophilum</i>		3,0	(Souza-Egipsy et al., 2011)
<i>Zygnemopsis</i>		3,0	(Souza-Egipsy et al., 2011)
<i>Cyanidium</i>		1,0 à 2,5	(Yin et al., 2008)

Table 1.1 : Diversité des principaux micro-organismes photosynthétiques identifiés dans les milieux aquatiques exposés aux DMA.

1.3.3. DIVERSITÉ MICRO-EUCARYOTES : COMMENT L'ÉtudIER ?

Bien que considérés comme extrêmes pour nombre de micro-organismes, les environnements exposés aux DMA recèlent une diversité assez large d'eucaryotes unicellulaires. Les analyses de la diversité de ces communautés microbiennes reposent habituellement sur l'amplification et le séquençage des gènes codant pour l'ARNr 18S et peuvent par conséquent être sujettes à un certain nombre de biais. En effet, l'efficacité d'extraction de l'ADN à partir des sédiments, le jeu d'amorces dites « universelles » ainsi que la température d'hybridation utilisées pour l'amplification, la compétition entre les séquences de gènes codant pour l'ARNr 18S et les amorces lors de l'hybridation ou encore l'efficacité de clonage dans des vecteurs sont autant de paramètres susceptibles d'affecter la diversité microbienne apparente d'un échantillon.

Des méthodes alternatives de séquençage haut débit telles que les technologies 454 de Roche ou Solexa de Illumina ont été développées (pour une description détaillée de ces techniques de séquençage, se reporter aux chapitres 3.2.1.1 et 3.2.1.2). Ces techniques permettent de s'affranchir des biais liés aux étapes de clonage dans des vecteurs mais sont toujours dépendantes de l'utilisation des amorces universelles. D'autre part, les techniques de séquençage haut débit ne génèrent actuellement au mieux que des séquences d'une taille d'environ 500 nucléotides rendant moins robustes les analyses comparatives sur base de similarité de séquences dans les bases de données.

Plus récemment, des techniques couplant la cytométrie de flux et l'amplification du gène 18S sur une cellule isolée ont permis de contourner la majorité de ces obstacles (Heywood et al., 2010). Le principe de cette technique repose sur la séparation de la communauté microbienne totale d'un échantillon du reste des sédiments puis sur l'isolement de cellules individuelles en plaque de micro-titration par cytométrie de flux. L'identification du gène codant pour l'ARNr 18S est par la suite réalisée sur l'ADN extrait de chaque cellule isolée avec différents couples d'amorces « universelles » ce qui permet de s'assurer que le gène codant pour l'ARNr 18S est bien amplifié. D'autre part, en travaillant sur des cellules individuelles, les phénomènes de compétition pouvant survenir dans des échantillons complexes entre les amorces et les séquences de gènes codant pour les ARNr 18S sont également contournés. Ce type d'analyse a entre autres permis de révéler que les principaux couples d'amorces « universelles » couramment utilisés pour l'étude de la diversité des gènes codant pour les ARNr 18S n'amplifiaient au mieux que 15 à 20% des séquences pour une population donnée, ne donnant par conséquent qu'une vision très partielle de la complexité biologique réelle d'un échantillon (Heywood et al., 2010).

D'autre part, quelle que soit la technique utilisée, l'identification d'un organisme sur base de la séquence de son ARNr 18S sous-entend que la séquence correspondante soit déjà présente et décrite dans les bases de données. Ce dernier point peut être particulièrement limitant lorsqu'on se focalise sur des micro-eucaryotes difficilement cultivables ou évoluant dans des niches écologiques exotiques et peu étudiées. Enfin, l'utilisation de la seule séquence du gène codant pour l'ARNr 18S en tant que marqueur phylogénétique universel peut être sujette à questionnements, la diversité d'une communauté micro-eucaryote n'étant pas nécessairement corrélée à la diversité d'un seul gène. Ces analyses de diversité offrent donc un aperçu certain mais limité de la diversité microbienne dans un échantillon mais restent néanmoins un outil précieux pour identifier rapidement les principaux micro-organismes eucaryotes, cultivés ou non, présents dans un environnement donné.

1.3.4. COMMUNAUTÉ MICRO-EUCARYOTE : QUELS IMPACTS SUR LES DMA ?

Si l'impact des communautés microbiennes procaryotes est relativement bien décrit dans les DMA, celui de la communauté micro-eucaryote est souvent moins clair. D'une manière générale, on distingue les communautés micro-eucaryotes photosynthétiques des communautés micro-eucaryotes hétérotrophes en raison de leur impact parfois antagoniste sur ces écosystèmes.

1.3.4.1. Impact des micro-eucaryotes photosynthétiques

Les communautés micro-eucaryotes photosynthétiques sont considérés comme des producteurs primaires de première importance dans les écosystèmes exposés aux DMA (Das *et al.*, 2009a; Johnson & Hallberg, 2003; Lopez-Archilla *et al.*, 2001). La production primaire des organismes photosynthétiques les plus couramment retrouvés dans les DMA a récemment été étudiée en évaluant différents paramètres liés à leur photosynthèse (Table 1.2) (Souza-Egipsy *et al.*, 2011).

Table 1.2 : Définition des principales caractéristiques mesurées dans les biofilms photosynthétiques d'algues présents dans les DMA (Souza-Egipsy *et al.*, 2011).

Paramètres mesurés	Définitions	Unités
Intensité de lumière saturante	Intensité de lumière au-dessus de laquelle l'efficacité de photosynthèse n'augmente plus.	mmol photons m ⁻² s ⁻¹
Intensité de lumière compensatrice	Intensité de lumière pour laquelle la quantité d'oxygène produit par la photosynthèse compense exactement la quantité d'oxygène consommée par la respiration.	mmol photons m ⁻² s ⁻¹
Intensité de lumière photo-inhibitrice	Intensité de lumière à partir de laquelle l'efficacité de photosynthèse diminue	mmol photons m ⁻² s ⁻¹
Efficacité de photosynthèse	Fraction de l'énergie lumineuse qui est convertie en énergie chimique	mg O ₂ mg Chl a ⁻¹ h ⁻¹
Photosynthèse maximale	Quantité maximale d'oxygène fixée pour une quantité de chlorophylle donnée	mg O ₂ mg Chl a ⁻¹ h ⁻¹
Respiration	Consommation d'oxygène par respiration pendant la phase obscure	mg O ₂ mg Chl a ⁻¹ h ⁻¹

Ces résultats soulignent que les organismes photosynthétiques associés aux DMA sont en général bien adaptés aux conditions de faible luminosité (intensité de lumière saturante et intensité de lumière compensatrice basse) mais que la production primaire (photosynthèse maximale) est faible en comparaison à d'autres écosystèmes où les valeurs peuvent être 10 fois plus élevées. La très forte concentration en composés toxiques présents dans les sites exposés aux DMA est connue pour inhiber la photosynthèse et pourrait en partie expliquer la faible production primaire dans ces sites (Souza-Egipsy et al., 2011).

Les voies anaboliques associées au métabolisme photosynthétique sont supposées jouer un rôle important dans la synthèse et la libération de métabolites qui peuvent être utilisés ensuite par le reste de la communauté microbienne. En effet, ces métabolites pourraient favoriser le développement de bactéries sulfato-réductrices et de bactéries chimio-lithotrophes connues pour affecter significativement ces écosystèmes (Das *et al.*, 2009b). Cependant, si le rôle important des producteurs primaires est souvent souligné dans la littérature scientifique, la nature exacte des composés libérés dans l'environnement ainsi que les flux métaboliques qui peuvent exister entre les différents membres de la communauté restent souvent peu décrits.

L'importante biomasse produite par ces organismes photosynthétiques pourrait également jouer un rôle dans la chélation de nombreux métaux, notamment à travers la synthèse d'exopolysaccharides (Das *et al.*, 2009a). En effet, ces matrices présentent une composition variable en fonction du temps et des espèces qui les synthétisent, mais sont souvent composées de sucres, d'acides nucléiques et de protéines (Beech & Sunner, 2004). Elles renferment par conséquent des groupements chimiques chargés tels que des fonctions carboxylates ($R-COO^-$), phosphates ($R-HPO_4^-$), sulfhydriles ($R-SH$), amines ($R-NH_3^+$), hydroxyles ($R-OH$) ou phénoliques ($R-C_6H_4OH$) susceptibles de chélater les métaux (Diels *et al.*, 2009; Harrison *et al.*, 2007; Pickering *et al.*, 2006). Même si ce processus est saturable et n'affecte pas nécessairement la concentration des métaux de l'écosystème dans son ensemble, cette caractéristique est néanmoins susceptible de conduire à des micro-niches écologiques moins exposées aux métaux et aux métalloïdes (Das *et al.*, 2009a).

D'autre part, le métabolisme photosynthétique associé à certains biofilms conduit localement à une augmentation significative de la teneur en oxygène dissout pouvant affecter la spéciation de certains composés chimiques (Brake et al., 2001a). Ainsi, en présence de fortes concentrations en oxygène dissout, le fer ferreux est oxydé en fer ferrique, moins soluble et qui aura tendance à co-précipiter avec d'autres métaux (Casiot et al., 2003). Cette activité photosynthétique, en diminuant localement la teneur en CO_2 dissout, est également susceptible de générer de l'alcalinité conduisant à la précipitation des métaux lourds présents en solution sous forme

d'hydroxydes (Das *et al.*, 2009a; Rose *et al.*, 1998). En effet, des études menées *in vitro* montrent que le métabolisme photosynthétique du protiste *Euglena mutabilis* conduit à une légère augmentation du pH, uniquement en présence de lumière (Tanaka *et al.*, 2000). D'autre part, certaines algues du genre *Spirulina* sp. ont montré leur capacité à augmenter et à maintenir à pH 8 une solution alimentée quotidiennement par 10% en volume d'effluents de DMA à pH 1,8 (Rose *et al.*, 1998). Si cette génération d'alcalinité n'est probablement pas suffisante pour modifier globalement le pH de rivières impactées par des DMA, elle peut néanmoins conduire localement à des micro-niches moins acides, en particulier dans des structures de type biofilms où les activités photosynthétiques sont très denses. En parallèle, l'oxygène produit durant ce même processus s'accumule sous forme dissoute dans les effluents et peut atteindre 200% des valeurs de saturation à l'équilibre (Brake *et al.*, 2001a). Les modifications physico-chimiques induites par le métabolisme photosynthétique ont donc tendance à favoriser les espèces microbiennes ayant développé un métabolisme aérobie et/ou celles légèrement moins acido-tolérantes.

Par conséquent, la chélation de métaux couplée à leur immobilisation sous forme de précipités dans les matrices extracellulaires produites par les organismes photosynthétiques peut conduire localement à une diminution de la concentration en métaux et maintenir des conditions plus propices à la colonisation par d'autres micro-organismes.

1.3.4.2. Impact des micro-eucaryotes hétérotrophes

L'impact des communautés micro-eucaryotes hétérotrophes sur les sites exposés aux DMA n'est pour l'instant que peu documenté dans la littérature scientifique. A l'instar des communautés photosynthétiques, les organismes hétérotrophes sont susceptibles de chélater un certain nombre de métaux sur les matrices extracellulaires ou sur les parois des cellules (Das *et al.*, 2009a). Cependant, au même titre que chez les organismes autotrophes, ce phénomène est saturable et n'affecte donc probablement que localement la concentration en métaux lourds.

Les métabolismes complémentaires entre les eucaryotes hétérotrophes et autotrophes peuvent également conduire à des interactions de type mutualiste qui favorisent la colonisation de niches écologiques hostiles. De telles observations ont été réalisées entre le protiste *E. mutabilis* et une levure du genre *Cryptococcus* dans un site exposé aux DMA (Nakatsu & Hutchinson, 1988). L'interaction unissant ces deux micro-organismes conduit à la colonisation de niches très acides et très riches en métaux lourds où ces organismes, pris individuellement, ne peuvent évoluer (Nakatsu & Hutchinson, 1988). Par conséquent, même si les mécanismes exacts qui conduisent à ce bénéfice

réciroque ne sont pas clairs, il apparait évident que la communauté eucaryote hétérotrophe peut favoriser le développement d'espèces phototrophes dans les environnements exposés aux DMA.

D'autre part, la communauté microbienne eucaryote est également susceptible de modifier la composition de la communauté bactérienne par des processus de prédation. En effet, certains protozoaires non identifiés et isolés de DMA ont montré leur rôle dans des processus de prédation spécifiquement ciblés sur des espèces bactériennes acidophiles (Johnson & Rang, 1993; McGinness & Johnson, 1992). Ainsi, des observations réalisées *in vitro* ont mis en évidence que les espèces bactériennes *Thiobacillus ferrooxidans*, *Leptospirillum ferrooxidans* et *Acidiphilum cryptum* étaient spécifiquement la proie de certains protozoaires, tandis que des espèces *Thiobacillus thiooxydans* et *Thiobacillus acidophilus* n'étaient pas affectées (McGinness & Johnson, 1992). *A contrario*, une étude réalisée avec un autre protozoaire non identifié a révélé que la croissance de l'espèce *Leptospirillum ferrooxidans* était favorisée par rapport aux autres espèces bactériennes acidophiles lors d'expérimentations de co-culture (Johnson & Rang, 1993). Ces observations soulignent donc que les phénomènes de prédation sélective peuvent affecter la composition de la communauté bactérienne, et par conséquent les processus de bio-remédiation de manière indirecte.

Les micro-organismes eucaryotes semblent donc jouer un rôle important bien qu'indirect dans l'évolution à long terme des écosystèmes exposés aux DMA. La colonisation de ces niches écologiques extrêmes suggère par ailleurs la présence de mécanismes d'adaptation spécifiques, en particulier en regard du stress arsénié très marqué sur ces sites.

2. MÉTABOLISME DE L'ARSENIC CHEZ LES EUCARYOTES

2.1. MÉCANISMES D'ENTRÉE DE L'ARSENIC DANS LES CELLULES

L'arsenic est présent abondamment et sous de nombreuses formes dans l'environnement. Certaines de ces espèces chimiques présentent par ailleurs des analogies structurales avec d'autres composés et sont par conséquent susceptibles d'entrer dans la cellule par des mécanismes divers. Les modes d'incorporation de l'arsenic dans les cellules varient donc d'un organisme à l'autre en fonction de la nature et de la diversité des transporteurs localisés à leur surface.

2.1.1. CHEZ LA *SACCHAROMYCES CEREVISIAE*

Chez *Saccharomyces cerevisiae*, on distingue différents mécanismes d'entrée de l'arsenic selon qu'il s'agisse d'arsénite ou d'arséniate. L'AsV pénètre dans les cellules par le biais de transporteurs habituellement impliqués dans l'incorporation de phosphate en raison d'analogies structurales entre ces deux composés (Berhe et al., 1995). Cet organisme dispose de 2 transporteurs à haute affinité pour le phosphate, Pho84p et Pho89p, et de 2 transporteurs à faible affinité, Pho87p et Pho90p (Persson *et al.*, 1999; Wykoff & O'Shea, 2001). Cependant, seuls Pho84p et Pho87p ont été montrés comme impliqués dans l'incorporation de l'arséniate. De plus, Pho84p pourrait également intervenir dans le transport du manganèse, du cuivre, du zinc et du cobalt (Jensen et al., 2003).

L'AsIII est quant à lui incorporé par deux mécanismes distincts. Le premier fait intervenir une aquaglycéroporine, Fsp1p, présentant des homologues avec les aquaglycéroporines procaryotiques (GlpF) connues pour être impliquées dans l'incorporation de l'arsénite (Gomes *et al.*, 2009; Meng *et al.*, 2004; Zardoya *et al.*, 2002). La protéine Fsp1p est habituellement impliquée dans le transport bidirectionnel du glycérol, molécule présentant des analogies structurales avec l'arsénite (Maciaszczyk-Dziubinska et al., 2010). L'autre principal mécanisme d'entrée de l'arsénite chez la levure fait intervenir des transporteurs d'hexoses (Shah et al., 2010). Ces derniers, nommés Hxt7p et Hxt9p, ont un rôle qui semble loin d'être anecdotique puisque la surexpression de Hxt7p induit une augmentation d'un facteur 4 de l'accumulation d'arsénite dans la cellule (Wysocki et al., 2001). Certains auteurs suggèrent que l'arsénite pourrait être pris en charge par ces transporteurs suite à

l'interaction entre trois molécules d'arsénite qui forment un complexe présentant une analogie structurale avec les hexoses (Liu et al., 2004)

2.1.2. CHEZ LES PLANTES ET ALGUES

2.1.2.1. Plantes non accumulatrices d'arsenic

Les systèmes d'incorporation de l'arsenic sont multiples dans le règne végétal. Ils peuvent être soit communs aux systèmes d'incorporation de l'arsenic observés chez la levure, soit spécifiques aux plantes. L'AsV, comme chez les champignons, les bactéries et les cellules animales, est principalement incorporé par des transporteurs du phosphate en raison de son analogie structurale avec ce composé (Kobayashi *et al.*, 2003; Meharg *et al.*, 1994). Chez les algues unicellulaires, ce mode d'incorporation de l'arsenic semble être dominant alors que dans les cellules végétales des plantes supérieures, l'arséniate aurait tendance à pénétrer moins efficacement que l'arsénite (Abedin *et al.*, 2002; Foster *et al.*, 2008; Meharg & Jardine, 2003). L'AsIII est efficacement incorporé par les plantes supérieures par le biais de plusieurs transporteurs de la famille des NIP (Nodulin26-like Intrinsic Proteins) qui sont une sous-catégorie des MIP (Major Intrinsic Protein) présentant des homologies avec les aquaglycéroporines d'origines bactériennes (Bienert & Jahn, 2010). D'autre part, ces transporteurs ont également été décrits comme étant impliqués dans l'incorporation de formes arséniées méthylées telles que le MMA^V et le DMA^V (Li et al., 2009). Ces transporteurs NIP sont en temps normal responsables de l'incorporation de bore ou de silicium dans les cellules (Jian et al., 2008). Chez le riz, ces transporteurs sont impliqués dans l'entrée du silicium au niveau des racines (transporteur Lsi1) mais également de l'efflux du silicium vers le xylème (transporteur Lsi2). La forte accumulation de silicium par le riz, peut donc expliquer, au moins en partie, les teneurs relativement élevées d'arsenic chez cette espèce par rapport à d'autres plantes (Zhao et al., 2009). L'arsénite présent dans les cellules racinaires peut également être ré-expulsé vers le milieu extérieur mais le transporteur impliqué dans ce processus n'a pour l'heure pas encore été identifié (Xu et al., 2007).

Une fois dans les cellules racinaires, l'arsenic ne présente qu'une mobilité limitée et diffuse de manière variable dans le reste de la plante mais toujours dans des proportions assez faibles. En effet, sur 46 plantes testées, l'arsenic est environ 10 à 100 plus concentré dans les racines que dans le reste de la plante (Raab et al., 2007). Cette différence est probablement liée au fait que l'arsenic est stocké dans les vacuoles des cellules racinaires sous forme de complexes avec le glutathion ou des phytochélatines (Liu et al., 2010). L'arsenic libéré dans le xylème est majoritairement retrouvé sous forme d'arsénite inorganique, pas ou peu de formes méthylées ou complexées avec des fonctions thiols ayant été identifiées dans ce compartiment. Dans le phloème, les teneurs en arsenic sont habituellement basses et l'arsenic est retrouvé sous forme de complexes avec des fonctions

thiols qui ne sont cependant pas très stables en raison du pH légèrement alcalin de ce compartiment (Zhao et al., 2009).

2.1.2.2. Plantes hyper-accumulatrices

Les plantes hyper-accumulatrices d'arsenic contrastent d'une manière marquée avec les autres représentants du règne végétal au niveau de l'absorption et de la mobilité de l'arsenic *in planta* (Figure 1.4). En effet, les plantes hyper-accumulatrices d'arsenic, pour l'instant 12 fougères du genre *Pteris* sur les 400 que compte le genre, se caractérisent, telle *P. vittata*, par une plus grande affinité de leurs transporteurs de phosphate pour l'arséniate qui est la forme la plus absorbée au niveau des racines, et un moindre efflux des formes inorganiques vers le milieu extérieur (Caille et al., 2005).

D'autre part, on ne retrouve qu'une faible proportion d'arsenic complexé à des phytochélatines et/ou du glutathion dans les racines où il est retrouvé principalement sous forme d'AsIII quelle que soit l'espèce arsénisée inorganique présente dans le sol. Contrairement aux observations réalisées sur les plantes non-accumulatrices, l'arsénite est transporté efficacement dans le xylème où il représente entre 93 à 98% de l'arsenic total et forme jusqu'à 2% du poids sec des frondes. Dans les frondes, l'arsénite est majoritairement stocké dans des vacuoles sous forme inorganique mais le transporteur impliqué dans ce processus n'a pas encore été identifié (Gumaelius et al., 2004; Pickering et al., 2006; Su et al., 2008).

Ce transport atypique de l'arsenic chez les plantes hyper-accumulatrices conduit donc, contrairement aux autres plantes, à un rapport de concentration d'arsenic très élevé entre les racines et le reste de la plante (Tu et al., 2002).

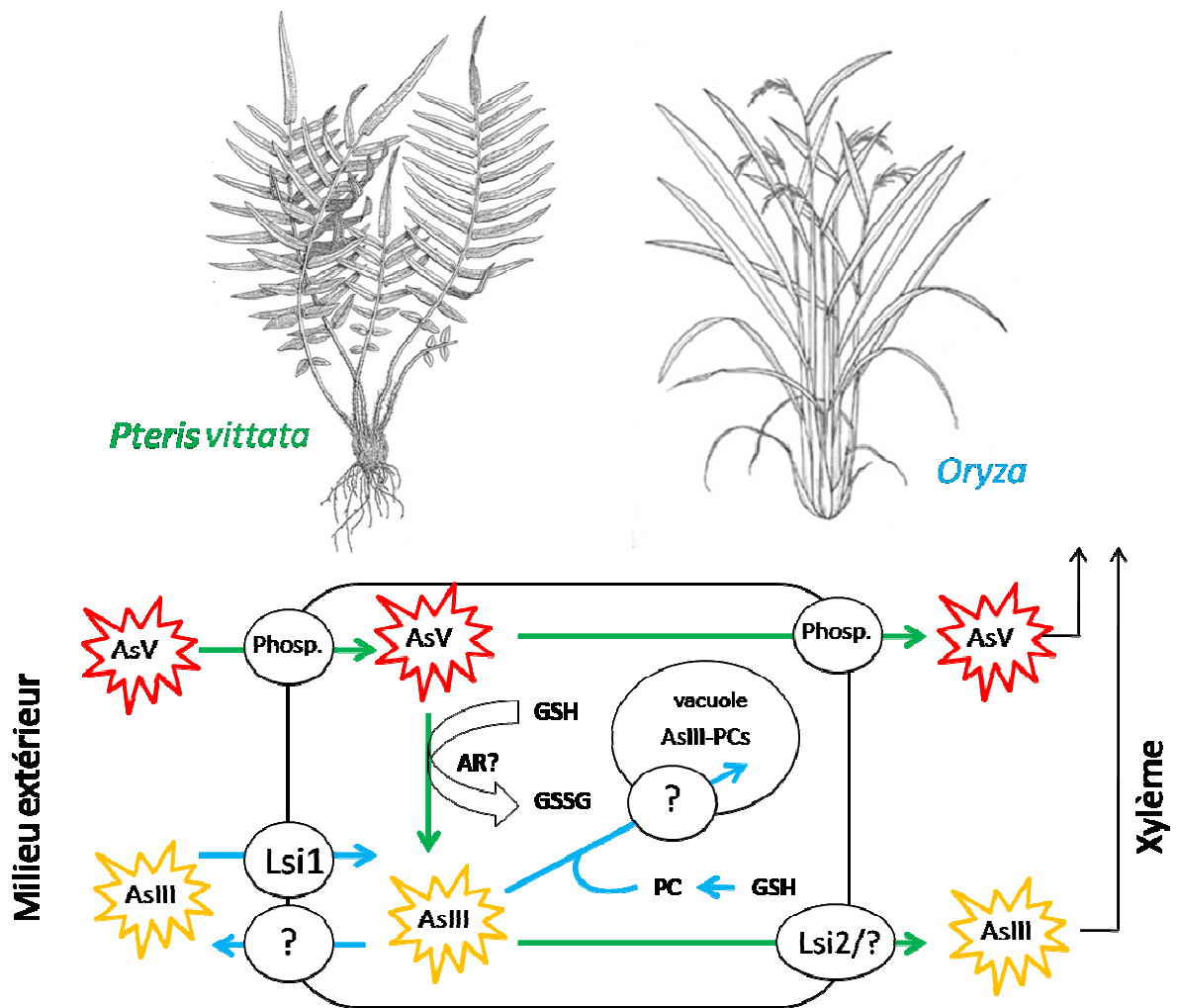


Figure 1.4 : Principaux flux de composés arsénisés chez les plantes non-accumulatrices (en bleu) et hyper-accumulatrices d'arsenic (en rouge). Phosp. : transporteur du phosphate, Lsi1, Lsi2 : transporteurs de silicium, ? : transporteur de composés arsénisés non-identifiés. AsIII : arsénite, AsV : arséniate, PCs : Phytochélatines, GSH : Glutathion oxydé, GSSG, glutathion réduit, AR ? : arséniate réductase non identifiée (Zhao et al., 2009).

2.1.3. CHEZ LES MAMMIFÈRES

Chez les mammifères, les principales sources d'incorporation de l'arsenic sont liées à l'ingestion d'eau ou d'aliments contenant de l'arsenic (surtout AsV, AsIII, DMA^V, MMA^V). Par conséquent, le premier niveau d'entrée de l'arsenic dans l'organisme se situe au niveau des cellules intestinales. En effet, chez l'Homme, le coefficient d'utilisation digestive de l'arsenic total est de l'ordre de 95% (Zheng et al., 2002). Les formes arsénisées méthylées ont été décrites comme faiblement absorbées alors que l'AsV est absorbé plus efficacement, très probablement par le biais de transporteurs du phosphate (Calatayud et al., 2011).

Une fois dans l'organisme, l'arsenic se répartit inégalement dans les différents organes ce qui s'explique notamment par une expression hétérogène des transporteurs de l'arsenic en fonction des types cellulaires. En effet, il existe une variété assez importante d'aquaporines, dont certaines (AQP7 et AQP9) sont chez le rat et l'Homme impliquées dans l'incorporation d'arsénite et de MMAs^{III} (Liu et al., 2002). La protéine AQP9p est préférentiellement exprimée au niveau des poumons, du foie et des leucocytes et les protéines AQP7p au niveau des reins, des testicules et du tissu adipeux (Kageyama et al., 2001; Tsukaguchi et al., 1999). Cette expression hétérogène peut expliquer, au moins en partie, la symptomatologie liée à une intoxication à l'arsenic qui peut conduire à des cancers des poumons, du foie ou des reins (Smith et al., 1992). D'autre part, il a été montré chez le rat qu'il existe une grande variabilité au niveau de l'accumulation de ces protéines sur ces tissus en fonction de l'âge, du régime alimentaire et du sexe, ce qui conduit à des variations interindividuelles marquées en termes de sensibilité à l'arsenic (Nicchia et al., 2001).

Les transporteurs du glucose GLUT1 et GLUT2 ont également été décrits chez les mammifères pour le transport de l'arsénite et de formes arséniées méthylées (Drobna et al., 2010; Liu et al., 2006). GLUT1 est même supposé être le principal mécanisme d'entrée de l'arsénite dans les cellules cardiaques et cérébrales chez le rat, seuls endroits où ces transporteurs sont exprimés (Jiang et al., 2010). Cependant, les mutations ponctuelles au niveau du transporteur GLUT1 connues pour inhiber l'incorporation du glucose n'affectent pas celle des formes méthylées de l'arsenic. Cette observation suggère par conséquent des différences au niveau des mécanismes moléculaires du transport du glucose et des composés arséniés (Jiang et al., 2010).

2.2. IMPACT DE L'ARSENIC AU SEIN DES CELLULES

L'arsenic est un composé très toxique qui peut affecter de manières variées le fonctionnement des cellules. Certains points à l'origine de sa toxicité sont partagés avec d'autres métaux alors que d'autres lui sont spécifiques (Wysocki & Tamas, 2010). D'une manière générale, la toxicité de l'arsenic est liée à l'analogie structurale de l'arséniate avec le phosphate et à la très forte affinité de l'arsénite pour les fonctions sulfhydriles. Ces deux caractéristiques sont à l'origine de la majorité des dysfonctionnements cellulaires décrits ci-dessous :

2.2.1. STRESS OXYDATIF

La capacité de l'arsenic à induire un stress oxydatif au sein de la cellule est largement décrite dans la littérature scientifique même si les mécanismes moléculaires sous-jacents restent relativement flous (Flora, 2011). En effet, une augmentation de la teneur en espèces oxygénées

réactives (EOR) tels que les radicaux hydroxyles OH^\bullet , les oxygènes singulets O^\bullet , les anions superoxydes $\text{O}_2^{\bullet-}$ et les peroxydes d'hydrogène H_2O_2 est mesurée dans les cellules de mammifère en présence d'arsenic (Flora, 2011; Liu *et al.*, 2001). De plus, des manifestations indirectes du stress oxydatif arsénié ont également été décrites chez la levure (Menezes *et al.*, 2008). La capacité de l'arsenic à induire un stress oxydatif est d'ailleurs partagée par d'autres métaux tels que le cadmium, le plomb ou le mercure bien qu'ils soient inactifs dans les réactions d'oxydo-réduction *in vivo* contrairement à l'arsenic (Jin *et al.*, 2008; Wysocki & Tamas, 2010).

Plusieurs hypothèses peuvent être émises au vu des observations réalisées en condition de laboratoire pour expliquer l'origine du stress oxydatif arsénié. Premièrement, l'arsénite est connu pour sa très forte affinité avec les fonctions sulfhydriles et peut par conséquent réagir et limiter le pool de glutathion libre de la cellule, l'un des principaux métabolites antioxydants (Stohs & Bagchi, 1995). De plus, l'arsénite interagit également avec la thiorédoxine réductase, protéine impliquée dans la réduction d'un autre agent réducteur majeur de la cellule, la thiorédoxine (Zhang *et al.*, 2007). Cependant, d'autres mécanismes sont probablement impliqués dans la genèse du stress oxydatif puisque les effets de ce dernier sont observés à des concentrations en arsenic bien moindre que le stock d'anti-oxydants de la cellule (Lafaye *et al.*, 2005).

Une deuxième hypothèse pourrait être liée au fait que l'arsenic induit des modifications du métabolisme conduisant à la production d'EOR. En effet, les formes méthylées de l'arsenic ont été décrites comme pouvant déstabiliser le complexe fer/ferritine conduisant à la libération de fer à l'origine de la production de radicaux hydroxyles OH^\bullet selon la réaction de Fenton-Haber-Weiss (Ahmad *et al.*, 2000; Kitchin & Wallace, 2008).

- | | |
|--|---|
| 1) Réaction de Fenton : | $\text{M}^{n+} + \text{H}_2\text{O}_2 \rightarrow \text{OH}^- + \text{OH}^\bullet + \text{M}^{(n+1)+}$ |
| 2) Réduction par le superoxyde : | $\text{O}_2^{\bullet-} + \text{M}^{(n+1)+} \rightarrow \text{O}_2 + \text{M}^{n+}$ |
| 3) Réaction de Haber-Weiss (Bilan 1+2) : | $\text{O}_2^{\bullet-} + \text{H}_2\text{O}_2 \rightarrow \text{OH}^- + \text{OH}^\bullet + \text{O}_2$ |

D'autre part, l'arsenic entraîne des dysfonctionnements de la chaîne respiratoire mitochondriale, ces organites étant décrits comme les principales sources EOR induites au sein des cellules (Novis & Harding, 2007). Il reste cependant difficile de comprendre si l'augmentation des EOR libérées par les mitochondries sous stress arsénié est la cause ou la conséquence des dysfonctionnements induits par l'arsenic sur ces organites (Flora, 2011). Ces effets se conjugueraient donc à la déplétion du stock de métabolites antioxydants. Les EOR générées par l'arsenic perturbent dans un second temps le fonctionnement global de la cellule en déstabilisant la conformation

tertiaire des protéines, en induisant des peroxydations des lipides et en entraînant des mutations au sein de l'ADN (Halliwell & Gutteridge, 1984).

2.2.2. EFFETS MUTAGÈNES

L'arsenic est également décrit comme un agent mutagène pour l'Homme par l'Agence Internationale de Recherche sur le Cancer (IACR). Aucune étude n'a pour l'instant permis de démontrer sans ambiguïté les mécanismes moléculaires à l'origine des cancers et aucune augmentation des cassures d'ADN double brin n'a été observée chez la levure en présence d'arsenic (Jo et al., 2009). Cependant, des études épidémiologiques menées dans différents pays (Mexique, Amérique du Sud, Bangladesh) ont permis de conclure que l'ingestion d'eau contenant une centaine de microgramme d'arsenic était associée à plusieurs types de cancers (poumon, peau, vessie, rein, foie et colon) (United States National Protection Agency, 2001). Ces effets mutagènes dérivent très probablement à la fois des EOR générées au sein de la cellule en présence d'arsenic, mais également par l'inactivation, directe ou indirecte, des systèmes de réparation de l'ADN (Kligerman & Tennant, 2007; Shi *et al.*, 2004; Yamanaka *et al.*, 1990). Il n'est pas à exclure que l'analogie structurale de l'arséniate avec le phosphate puisse être susceptible d'induire son incorporation dans les acides nucléiques et de bloquer les systèmes de réparation conduisant à des mutations. Cependant, bien que l'incorporation d'arsenic au sein de molécules d'ADN ait été soutenue récemment (Wolfe-Simon et al., 2010), la nature du protocole expérimental ainsi que la stabilité chimique des intermédiaires formés, ont conduit à remettre en question la réalité de cette observation (Oehler, 2011; Redfield, 2011).

2.2.3. IMPACTS SUR LE CYCLE CELLULAIRE

L'activité biologique d'une protéine est étroitement liée à sa structure tri-dimensionnelle qui est en partie assurée par des ponts disulfures intra- ou inter-moléculaires. Ces repliements dans l'espace sont dictés en partie par les fonctions thiols des cystéines qui entrent dans la composition de la protéine. D'autre part, ces groupements fonctionnels peuvent également intervenir directement dans l'activité catalytique de la protéine dans la mesure où ils sont susceptibles d'interagir avec son substrat. La très forte affinité de l'arsénite pour les fonctions thiols peut donc perturber le fonctionnement d'une large gamme de protéines, soit directement en bloquant leurs sites catalytiques, soit indirectement en perturbant leurs repliements tri-dimensionnels (Flora, 2011; Ramadan *et al.*, 2009). Cet effet se conjugue à la déstabilisation des protéines suite au stress oxydatif observé en présence d'arsenic. En effet, certaines protéines, telle la tubuline, semblent être particulièrement affectées par l'arsénite dans leurs repliements tri-dimensionnels conduisant à la

déstabilisation rapide du cytosquelette dans son ensemble et affectant les principales étapes du cycle cellulaire comme la prolifération, la division, la différenciation et l'apoptose (Hoffman & Lane, 1992; Shao *et al.*, 1998; Shen *et al.*, 1999; Zhang *et al.*, 2007). D'autres protéines, telle la pyruvate déshydrogénase, semblent être plutôt affectées par le stress oxydatif arsénié que par la fixation de l'arsenic lui-même (Samikkannu *et al.*, 2003). Enfin, l'arsenic peut également entraîner l'apoptose d'une manière indirecte, notamment par l'activation d'une cascade de régulation induite par le peroxyde d'hydrogène accumulé dans les cellules en présence d'arsenic (Jing *et al.*, 1999).

L'arsenic conduit donc à des perturbations complexes du fonctionnement cellulaire. Ces dysfonctionnements multiples et lourds au sein de la cellule nécessitent la mise en place de mécanismes de détoxification efficaces.

2.3. MÉCANISMES DE DÉFENSE FACE À L'ARSENIC CHEZ LES EUCARYOTES

Chez les eucaryotes, les principaux mécanismes de détoxification de l'arsenic reposent sur l'augmentation de l'efflux de l'arsenic, une diminution de son incorporation dans le compartiment intracellulaire, sa séquestration par des métabolites spécifiques et/ou sa compartimentalisation (Tsai *et al.*, 2009). Ces processus de détoxification nécessitent souvent des modifications de la spéciation de l'arsenic qui peut être oxydé, réduit ou méthylé par la cellule. Les principaux systèmes de détoxification connus à l'heure actuelle chez les eucaryotes sont décrits ci-dessous. Ces mécanismes sont parfois similaires ou homologues à des mécanismes de détoxification procaryotiques qui seront le cas échéant décrits brièvement :

2.3.1. LIMITATION DE LA BIODISPONIBILITÉ

2.3.1.1. Synthèse d'expolysaccharides

Un des mécanismes le plus communément décrit dans la littérature scientifique et connu pour limiter l'exposition à l'arsenic chez les micro-organismes repose sur la synthèse d'une matrice d'exo-polysaccharides (EPS) extracellulaires (Costerton *et al.*, 1995; Van Hille *et al.*, 1999). Ces matrices extracellulaires ont été largement décrites chez les bactéries mais peuvent également être synthétisées par des algues, des champignons ou des protistes dans quels cas elles sont souvent qualifiées de mucilages (Baffico *et al.*, 2004). Les mécanismes de résistance liés à la synthèse d'une matrice extracellulaire sont multifactoriels mais reposent surtout sur une limitation de l'exposition des cellules à de nombreux composés toxiques en raison de la capacité de cette matrice à adsorber des molécules chargées, incluant les ions métalliques, l'arséniate et l'arsénite (Stewart, 1998). Ce

mécanisme de défense, bien que non spécifique à l'arsenic, peut être considéré comme un premier niveau de protection.

D'autre part, les biofilms, outre leur capacité à séquestrer l'arsenic, sont également caractérisés par une très forte densité cellulaire. Le métabolisme très dense associé à ces structures conduit habituellement à l'accumulation locale de carbonates, phosphates, hydroxydes ou de sulfates qui entraînent une précipitation des métaux sous forme de complexes insolubles à la surface des cellules (Douglas & Beveridge, 1998). D'une manière analogue, les conditions localement très oxydantes observées dans le cas de biofilms photosynthétiques peuvent favoriser l'oxydation abiotique du fer ferreux et de l'arsénite respectivement en fer ferrique et en arséniate (Brake et al., 2001a). Ces composés auront tendance à co-précipiter et à induire par la même occasion une diminution de leur biodisponibilité (Dupraz & Visscher, 2005).

2.3.1.2. Modification du transport membranaire

L'AsIII et l'AsV pénètrent dans les cellules par le biais de transporteurs membranaires impliqués en temps normal dans l'incorporation de métabolites essentiels au fonctionnement de la cellule. Un des mécanismes de défense face au stress arsénié consiste donc à modifier soit la nature, soit le nombre de ces transporteurs afin de limiter l'incorporation de l'arsenic. Dans le cas de l'arséniate, il a été observé chez une bactérie du genre *Herminiimonas* que les transporteurs classiques du phosphate (Pit) étaient réprimés en présence d'arsenic, alors que des transporteurs à haute affinité limitant l'entrée de l'arséniate (Pst) étaient induits (Weiss et al., 2009). Ce mécanisme conduit donc à une plus haute spécificité des systèmes de transport et entraîne par conséquent une moindre exposition du compartiment intracellulaire à l'arséniate.

D'autre part, il a été observé chez la levure exposée à l'AsIII une inhibition de la transcription de gènes codant pour les transporteurs des hexoses, une activation de la glycolyse et une inhibition de la biosynthèse du glycogène (Jin et al., 2008). Ces observations suggèrent donc qu'en présence d'un stress arsénié, la levure inhibe l'incorporation des hexoses et puise dans ses réserves afin de compenser ce manque. Il n'est pas à exclure que cette réponse résulte du fait que les transporteurs d'hexoses sont les principales voies d'entrée de l'arsénite chez la levure et que par conséquent, cette adaptation du métabolisme énergétique constituerait également un mécanisme de défense à part entière face au stress arsénié (Jin et al., 2008).

2.3.1.3. Précipitation extracellulaire de l'arsenic

L'exposition de la cellule à l'arsenic peut également être limitée par une diminution de la biodisponibilité des composés toxiques dans le milieu extérieur. La capacité d'oxyder enzymatiquement l'arsénite en arséniate, moins soluble, est pour l'instant décrit principalement chez les bactéries (Koechler *et al.*, 2010; Quéméneur *et al.*, 2008). Le bénéfice lié à cette oxydation de l'arsénite est potentiellement double. En effet, l'arséniate a tendance à interagir avec le fer ferrique présent dans le milieu extérieur pour former des complexes peu solubles qui auront tendance à diminuer localement la concentration extracellulaire en arsenic bio-disponible (Casiot *et al.*, 2003). D'autre part, il a été montré chez la bactérie *Herminiimonas arsenicoxydans* la co-induction du transporteur spécifique du phosphate Pst en parallèle à l'activité d'oxydation de l'arsenic (Weiss *et al.*, 2009). La combinaison de ces deux processus conduirait donc à une incorporation limitée de l'arséniate dans la cellule.

Chez les eucaryotes, seule l'algue rouge *Cyanidioschyzon* sp. semble être capable d'oxyder l'arsénite en arséniate, mais rien n'indique qu'il s'agisse d'une réaction catalysée enzymatiquement (Qin *et al.*, 2009). En effet, la production d'oxygène liée au métabolisme photosynthétique pourrait également, d'une manière indirecte, être à l'origine de ce phénomène et aucun gène eucaryote conférant la capacité d'oxyder l'arsénite n'a été isolé pour l'instant. D'autre part, le protiste photosynthétique *Euglena mutabilis* a également été décrit comme oxydant l'arsénite mais ces observations semblent être remises en question par des travaux présentés dans cette thèse (chapitre III).

2.3.2. EXPULSION DE L'ARSENIC

Un autre moyen habituellement utilisé par les micro-organismes pour limiter l'exposition du compartiment intracellulaire à l'arsenic consiste à expulser activement ce composé hors de la cellule. A l'instar des organismes procaryotes, les eucaryotes ont développé des mécanismes de défense face au stress arsénié couplant réduction de l'arséniate et expulsion de l'arsénite. Chez les procaryotes, le système *arsRDABC* impliqué dans ce processus de résistance a été abondamment décrit dans la littérature scientifique (Tsai *et al.*, 2009). Avant de décrire les mécanismes retrouvés chez les eucaryotes, un rapide aperçu du système procaryotique sera décrit ici afin de souligner les analogies et différences qui existent entre ces deux processus.

2.3.2.1. Exemple du système *arsRDABC* procaryotique

Le système *arsRDABC* procaryotique fait intervenir des gènes regroupés en organisation opéronique et dont *arsR* et *arsB* sont les éléments indispensables. Ces deux gènes codent respectivement pour un régulateur de la transcription et pour une pompe d'efflux de l'arsénite utilisant la force protomotrice (Kuroda *et al.*, 1997; Wu & Rosen, 1991). Les gènes *arsD*, *arsA* et *arsC* codent respectivement pour une protéine chaperon, une arséniate réductase et une ATPase conférant une plus grande efficacité au système mais non indispensables à son fonctionnement (Kuroda *et al.*, 1997; Lin *et al.*, 2006). L'ATPase ArsA interagit avec la pompe d'efflux ArsB ce qui conduit à un efflux de l'arsénite plus efficace (Kuroda *et al.*, 1997). La protéine ArsD, stabilise le complexe ArsA/ArsB impliqué dans l'efflux de l'arsénite et facilite le transfert de l'arsénite vers la sous-unité ArsA du système (Lin *et al.*, 2006). Enfin, ArsC est une arséniate réductase qui permet de réduire l'arséniate en arsénite. La présence de cette enzyme dans le système confère donc également une résistance à l'arséniate, autre forme inorganique de l'arsenic couramment retrouvée dans l'environnement. La réduction de l'arséniate en arsénite par ArsC nécessite la présence d'éléments réducteurs qui peuvent être représentés soit par le complexe glutathion/glutaredoxine, soit par la thiorédoxine (Ji & Silver, 1992; Mukhopadhyay & Rosen, 2002). Les arséniate réductases procaryotiques du type ArsC peuvent donc être divisées en deux sous familles en fonction de l'agent réducteur utilisé et de l'organisation de leur site catalytique. Il est à noter qu'un autre mécanisme de réduction de l'arsénite est retrouvé chez les procaryotes. Ce système, impliquant les gènes *arrA* et *arrB*, est lié au métabolisme énergétique en utilisant l'arséniate comme accepteur final d'électron dans la chaîne respiratoire (Saltikov & Newman, 2003). Ce système, exprimé uniquement en anaérobiose et en présence de très faibles concentrations d'arséniate (de l'ordre de la nanomole), n'intervient pas dans les mécanismes de détoxication et ne sera par conséquent pas décrit ici.

2.3.2.2. Le système ACR de *S. cerevisiae*

Si le système *arsRDABC* ne trouve pas d'homologue chez les levures, des systèmes similaires du point de vue fonctionnel couplant réduction de l'arséniate et expulsion de l'arsénite sont cependant décrits chez ces organismes (Figure 1.5). Ce système est composé chez *S. cerevisiae* de 3 gènes nommés *ACR1*, *ACR2* et *ACR3* pour « Arsenite Resistance Compound » codant respectivement pour un facteur de transcription, une arséniate réductase et une pompe d'efflux (Rosen, 2002). La taille de la pompe d'efflux ACR3p (45,8 kDa) est similaire à celle d'ArsB (45 kDa) mais ces protéines ne présentent pas d'homologies de séquences entre elles et diffèrent au niveau de leur structure et de leur spécificité de substrat (Ghosh *et al.*, 1999). En effet, si ArsB est capable d'expulser à la fois l'antimoine et l'arsénite, ACR3p est spécifique à l'arsénite. Le système de réduction de l'arséniate et

d'expulsion de l'arsénite de la levure est donc fonctionnellement très proche du système de détoxification procaryotique mais semble résulter d'une convergence évolutive soulignant la grande efficacité de ce système de défense face au stress arsénié (Rosen, 2002).

En plus des pompes ACR3 et ArsB, une troisième famille de transporteurs de composés arséniés a été identifiée chez *S. cerevisiae*. Ce système fait intervenir le gène *YCF1* (pour yeast cadmium resistance factor) et code pour un transporteur de l'arsenic au niveau des vacuoles (Ghosh et al., 1999). Ce transporteur, contrairement à ACR3p, est ATP-dépendant et n'est pas spécifique de l'arsénite. En effet, le transporteur YCF1 fait partie de la grande famille des MRP (Multidrug Resistance-associated Protein) et peut aussi prendre en charge le transport du cadmium, de mercure ou d'autres composés toxiques lorsqu'ils sont sous forme de complexes avec le glutathion (Gueldry et al., 2003; Li et al., 1997). L'expression du gène *YCF1* conduit à une accumulation de ces complexes au sein des vacuoles et induit par la même occasion une moindre exposition du compartiment intracellulaire à l'arsenic.

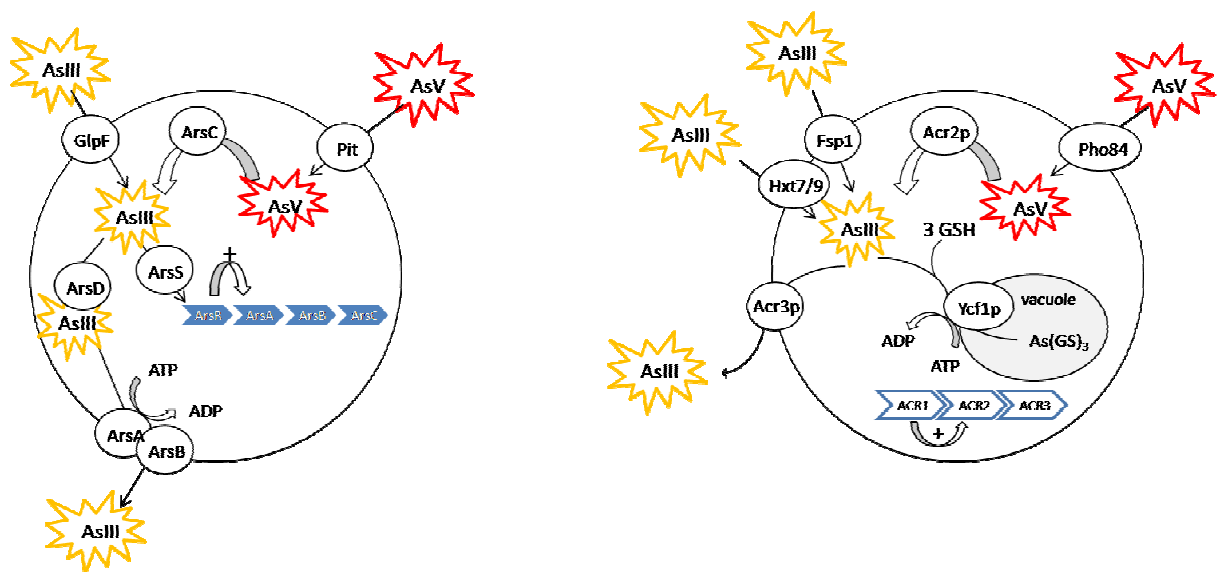


Figure 1.5 : Comparaison des systèmes de résistance à l'arsenic couplant réduction de l'arséniate et expulsion de l'arsénite chez les procaryotes (à gauche) et chez les eucaryotes (à droite). Les mécanismes d'incorporation de l'arsenic sont également représentés. AsIII : arsénite, AsV : arséniate, GS : glutathion oxydé, GSH : glutathion réduit (Rosen, 2002).

2.3.3. SÉQUESTRATION INTRACELLULAIRE DE L'ARSENIC

Un troisième mécanisme de défense face à l'arsenic largement répandu dans le monde vivant consiste à inactiver ce composé sous forme de complexe au sein de la cellule. Ce système est principalement basé sur la très haute affinité de l'arsenic pour les fonctions thiols.

2.3.3.1. Cystéine et glutathion

A l'instar de certains procaryotes, surtout des bactéries à Gram négative, les eucaryotes ont développé des mécanismes de défense face aux stress arséniés faisant intervenir des composés chélateurs tels que le glutathion (GSH) (Fahey *et al.*, 1978; Meister & Tate, 1976). Ce tri-peptide est formé par la condensation d'acide glutamique (E), de cystéine (C) et de glycine (G). La liaison unissant l'acide glutamique et la cystéine est atypique (liaison gamma) et ne peut être synthétisée qu'enzymatiquement et non pas par les ribosomes (Perales-Vela *et al.*, 2006). Ce composé est impliqué au niveau intracellulaire dans la chélation de métaux lourds et d'arsenic grâce à la fonction thiol de la cystéine. En général, 3 molécules de glutathion sont impliquées dans la chélation de chaque molécule d'arsénite et forment un complexe du type $As(GS)_3$. Ces complexes peuvent par la suite être transférés dans des vacuoles où les formes arséniées sont accumulées sous forme inactive. Ce transport fait intervenir chez l'algue verte *Chlamydomonas reinhardtii* une protéine qui présente des homologues de séquences avec le transporteur Ycf1 de *S. cerevisiae* impliqué dans un processus similaire (Wang & Wu, 2006). Bien que considéré comme l'un des principaux mécanismes de détoxification chez les plantes vasculaires face aux métaux lourds et à l'arsenic, le transporteur impliqué dans ce processus n'a pas encore été identifié (Pawlik-Skowronska *et al.*, 2004; Schat *et al.*, 2002). Cependant, la synthèse de composés chélateurs tel que le glutathion n'explique pas l'hyper-tolérance à l'arsenic chez les plantes hyper-accumulatrices du genre *Pteris* présentant des teneurs en GSH relativement faibles et où l'arsenic est majoritairement retrouvé sous forme non complexée (Zhao *et al.*, 2009).

2.3.3.2. Phytochélatines

Chez les plantes, les algues et certains champignons, le glutathion peut également servir de substrat à une enzyme, la phytochélatine synthase qui catalyse la condensation de 2 à 11 molécules de glutathion entre elles pour former des complexes de 2000 à 10 000 Da présentant plusieurs fonctions thiol, les phytochélatines (PCs) (Figure 1.6) (Cobbett & Goldsbrough, 2002; Perales-Vela *et al.*, 2006; Steffens, 1990). Les processus de détoxification faisant intervenir les PCs sont également retrouvés chez certains protistes tels qu'*Euglena mutabilis* (Navarro *et al.*, 1997). Au même titre que le glutathion, trois PCs sont généralement sollicitées pour la chélation d'une molécule d'arsénite,

formant des complexes du type $\text{As}(\text{PCs})_3$. Les formes méthylées de l'arsenic peuvent également être complexées par le glutathion et/ou les PCs ce qui conduit à une diversité assez importante de complexes arséniés. Par exemple, chez le tournesol 14 complexes arséniés différents ont été identifiés, les majoritaires étant $\text{As}^{\text{III}}(\text{PC})_2$, $\text{As}^{\text{III}}-(\text{PC})_3$, $\text{GS-As}^{\text{III}}-(\text{PC})_2$, $\text{MMA}-(\text{PC})_2$ (Raab et al., 2005). Les complexes As-PCs de faible poids moléculaire peuvent conduire à la formation de complexes de haut poids moléculaire (CHPM) de plusieurs centaines de kDa chez *E. gracilis* ou de 30 kDa chez les plantes et les champignons (Rodriguez-Zavala et al., 2007). Ces CHPM sont composés à 75% de petites molécules du type cystéine, glutathion et gamma-glutamyl cystéine, à 25% de PCs et sont stabilisés par des ions S^{2-} qui conduisent à la formation de particules de taille micrométrique (Kneer & Zenk, 1997).

Les complexes As-PC sont stockés dans des vacuoles chez les plantes et chez les levures (Mendoza-Cozatl & Moreno-Sanchez, 2005). Chez *E. gracilis*, un protiste photosynthétique ne disposant pas de vacuole, la localisation subcellulaire des complexes arséniés n'a été que peu étudiée à l'heure actuelle. Cependant des études réalisées sur le cadmium, composé qui induit des réponses faisant également intervenir les PCs, ont montré chez *E. gracilis* une accumulation de 60% du cadmium intracellulaire total dans les chloroplastes sous forme de complexes Cd-PCs (Mendoza-Cozatl et al., 2002). D'autre part, les PCs sont également retrouvées dans les mitochondries lorsque cet organisme est cultivé en condition hétérotrophique, ce qui pourrait suggérer que ces organites pourraient également servir de compartiments pour le stockage des métaux lourds et de métalloïdes comme l'arsenic (Devars et al., 2000). Ces protistes ont donc développé des procédés de compartimentalisation originaux pour pallier l'absence de vacuoles, mais la nature des transporteurs impliqués dans ce processus n'a pour l'heure pas été identifiée (Mendoza-Cozatl & Moreno-Sanchez, 2005). Cependant, il est à noter que les mitochondries et les chloroplastes sont le siège de réactions indispensables au fonctionnement de la cellule et que l'accumulation de métaux lourds dans ces compartiments affecte leur métabolisme (Mendoza-Cozatl et al., 2002; Nagel et al., 1996). Par conséquent, l'accumulation de métaux lourds dans les chloroplastes et les mitochondries de certains micro-organismes ne correspond peut être pas à un mécanisme de défense à proprement parler, mais résulterait de la présence de nombreux transporteurs d'ions essentiels tels que le Zn^{2+} , le Mn^{2+} ou le Ca^{2+} à la surface de ces organites (Mendoza-Cozatl & Moreno-Sanchez, 2005).

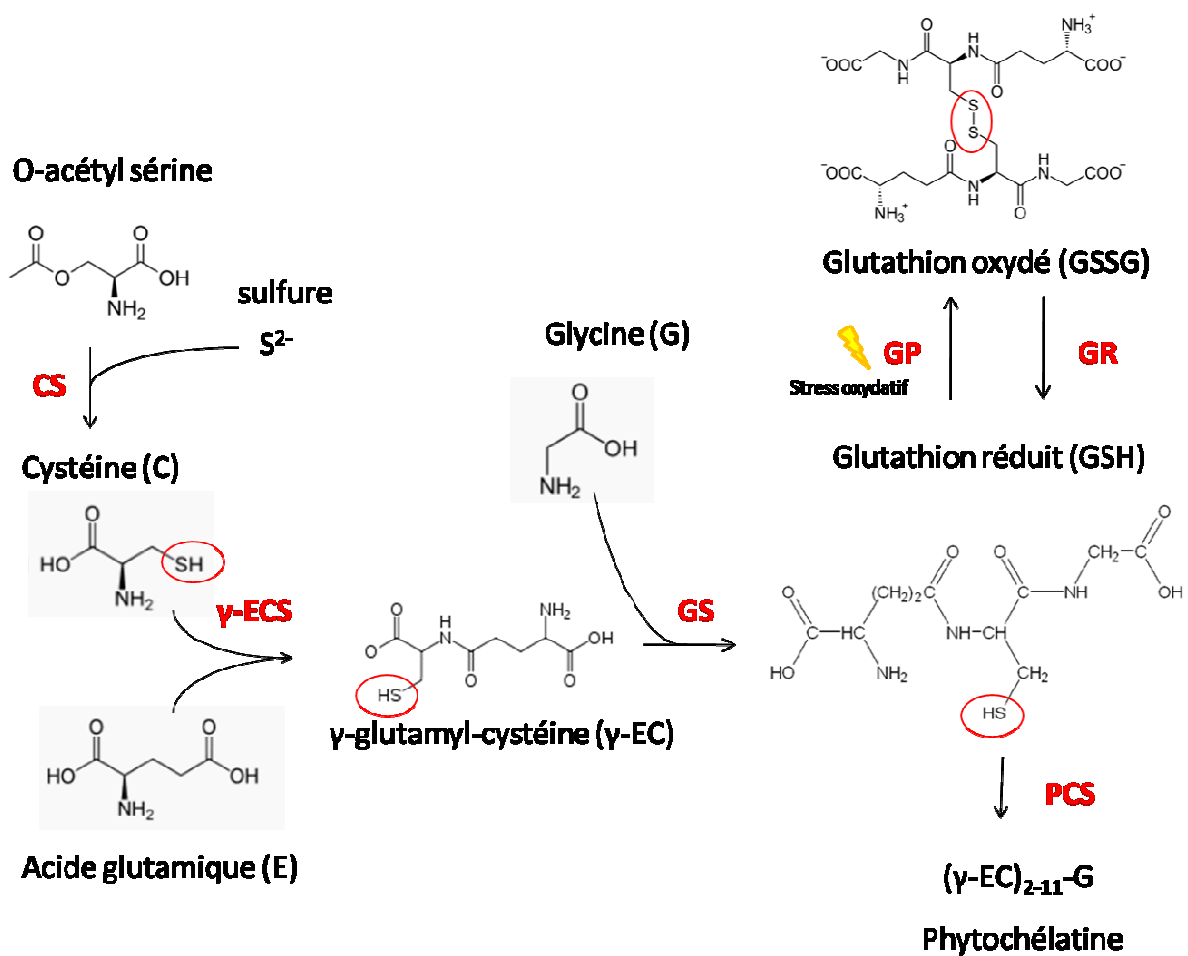


Figure 1.6 : Principales étapes des voies de biosynthèse du glutathion et des phytochélatines. CS : cystéine synthase ; γ -ECS : gamma-glutamylcystéine synthase, GS : glutathion synthase ; PCS : phytochélatine synthase ; GR : glutathion réductase, GP : glutathion peroxydase. La fonction thiol impliquée dans le chélation de l'arsenic est entourée en rouge.

2.3.3.3. Les métallothionéines (MT)

Les métallothionéines (MT) sont de petites protéines qui peuvent également être impliquées dans des processus de séquestration des métaux lourds (Hassinen et al., 2011). Ces protéines sont largement distribuées dans le monde vivant et sont retrouvées chez les cyanobactéries, les champignons, les animaux et les végétaux. Chez les plantes, les MT ont une taille variable (entre 45 et 87 acides aminés) et présentent des domaines riches en cystéine (entre 10 et 17 cystéines)(Hassinen et al., 2011). Ces clusters de cystéine sont séparés par des séquences d'acides aminés relativement longues (jusqu'à 40 acides aminés) dont le rôle dans l'activité de la protéine n'a pas encore été élucidé. Les MT ont la capacité de lier avec une forte affinité mais peu de stabilité le cuivre, le zinc et le cadmium (Maret, 2004). Les complexes MT-métaux sont localisés dans le cytosol

et ne semblent pas être compartimentalisés contrairement aux complexes formés avec le glutathion (Lee et al., 2004; Zhigang et al., 2006). Chacun des complexes formé avec les MT peut contenir jusqu'à 6 atomes de cadmium et 8 atomes de cuivre, ce qui en fait l'un des principaux modes de séquestration des métaux dans la cellule (Hassinen et al., 2011). Les plantes hyper-accumulatrices ou fortement tolérantes aux métaux lourds montrent d'ailleurs une accumulation plus importante que la moyenne de ces composés. Les MT semblent également jouer un rôle dans la séquestration de l'arsenic, dans la mesure où la délétion des gènes codant pour ces protéines chez *A. thaliana* conduit à une moindre accumulation de ce métalloïde dans les feuilles (Grispen et al., 2009). Plus récemment, les MT ont également été décrites comme intervenant dans la séquestration de EOR ce qui pourrait également être une caractéristique importante dans la compréhension des mécanismes de résistance face aux métaux lourds (Hassinen et al., 2011).

Les mécanismes de séquestration de l'arsenic sont donc multiples et représentés dans tous les grands règnes du vivant. La combinaison de mécanismes limitant l'entrée de l'arsenic dans la cellule et/ou conduisant à sa séquestration semble donc jouer un rôle efficace pour lutter contre ce métalloïde toxique. Il est à noter que les formes méthylées de l'arsenic peuvent également être prises en charge lors de processus de séquestration ce qui soulève la question du rôle des processus de méthylation de l'arsenic dans les mécanismes de défense.

2.3.4. LA MÉTHYLATION DE L'ARSENIC, UN PROCESSUS DE DÉTOXICATION ?

La méthylation de l'arsenic est catalysée par une enzyme, l'arsénite méthyltransférase, dont le gène correspondant est retrouvé chez les archaeae, les bactéries et les eucaryotes (Thomas *et al.*, 2007). Cette enzyme catalyse le transfert successif de plusieurs groupements méthyls sur l'arsénite à partir d'un donneur, la S-adénosylméthionine. La large distribution de ce gène dans le monde vivant suggère un avantage évolutif conséquent associé à cette fonction biologique. En effet, cette réaction peut conduire chez les organismes procaryotes à la synthèse d'une forme arséniée volatile, le TMAO, supposée limiter l'exposition à l'arsenic dans le compartiment intracellulaire ou dans l'environnement immédiat des micro-organismes exprimant cette fonction (Cullen & Bentley, 2005). D'autre part, la plupart des eucaryotes supérieurs expulsent dans les urines l'arsenic présent dans l'organisme sous forme d'arsenic méthylé (Draguet et al., 2005). Ces observations ont longtemps amené à considérer la méthylation de l'arsenic comme un mécanisme de défense (Aposhian, 1997). Cette hypothèse est d'ailleurs soutenue par des études montrant que l'expression hétérologue chez *Escherichia coli* d'un gène codant pour une arsénite méthyltransférase isolée de l'algue verte

unicellulaire, *Cyanidioscyzon sp.*, induit une plus grande tolérance à l'arsenic (Qin *et al.*, 2009; Yuan *et al.*, 2008).

Plus récemment, des observations conduisant à des interprétations plus ambiguës ont amené à remettre en question le rôle exact de la méthylation de l'arsenic dans des processus de détoxification (Petrick *et al.*, 2000). Ainsi, chez les plantes vasculaires cultivées en condition de laboratoire, des formes méthylées telles que le MMA^V ou le DMA^V ont également été identifiées (Wu *et al.*, 2002). Cependant, les formes méthylées de l'arsenic sont ultra-minoritaires dans la majorité des plantes et la question de la capacité d'un certain nombre d'entre elles à produire les formes arséniées volatiles reste posée. En effet, une étude menée sur le tournesol cultivé 1 jour en présence d'arsenic puis repiqué dans un nouveau milieu dépourvu d'arsenic révèle que la concentration totale de ce métalloïde *in planta* reste constante durant les 30 jours suivants (Raab *et al.*, 2007). Par conséquent, la capacité à produire des formes arséniées volatiles, si elle existe, est négligeable chez cette espèce et n'est probablement pas impliquée dans des processus de défense. D'autre part, les plantes hyper-accumulatrices d'arsenic comme *P. vittata*, lorsqu'elles sont exposées à des formes inorganiques de l'arsenic, ne renferment des formes arséniées méthylées que sous forme de traces suggérant que le processus de méthylation de l'arsenic n'est pas non plus impliqué dans l'hyper-tolérance à l'arsenic chez cette espèce (Kertulis *et al.*, 2005). Enfin, au cours des réactions de méthylation catalysées par l'arsénite méthyltransférase, de nombreux intermédiaires réactionnels sont produits, dont certains, tels que MMAs^{III}, le DMAs^{III} ou le TMA sont décrits comme étant plus cytotoxiques et génotoxiques que les formes inorganiques habituellement retrouvées dans l'environnement (Petrick *et al.*, 2000; Thomas *et al.*, 2001).

Par conséquent, le rôle de détoxification de cette enzyme face au stress arsénié est remis en cause et sa large distribution dans le vivant pourrait être liée à un avantage évolutif autre que celui initialement supposé. Plusieurs hypothèses peuvent être émises: l'arsénite méthyltransférase pourrait catalyser en parallèle une réaction de méthylation autre que celle de l'arsénite, réaction qui conférerait un avantage évolutif considérable chez les organismes porteurs de ce gène. Selon cette hypothèse, la pression de sélection conduirait au maintien du gène codant pour l'arsénite méthyltransférase malgré, et non en raison, de la production de formes arséniées méthylées. Une autre hypothèse fait suite à une observation inattendue effectuée dans le métabolisme des composés arséniés. En effet, lorsque le régime alimentaire de certaines espèces animales (hamster, rat, chèvre, poulet,...) est totalement dépourvu d'arsenic (moins de 1ng par g d'aliment), des troubles de la reproduction et de la croissance sont décrits (Uthus, 1992). Ces observations suggèrent un rôle physiologique de l'arsenic sous forme de traces qui peut alors être considéré comme un micro-

oligonutrimment mais dont le mécanisme d'action reste pour l'heure inconnu (Uthus, 1992). Il n'est pas à exclure que cet éventuel rôle physiologique pourrait être assuré par l'un des intermédiaires réactionnels produits durant la méthylation de l'arsenic, ce qui expliquerait le maintien de ce gène dans tous les règnes du vivant.

2.3.5. CONCLUSION

Les micro-organismes ont donc développé des procédés multiples de résistance face au stress arsénié. Ces mécanismes de défense sont souvent communs à ceux développés pour d'autres métaux et sont représentés ci-dessous (Figure 1.7). Cependant, la réponse d'un micro-organisme face à un stress arsénié ne se limite pas nécessairement aux processus de détoxification. Ainsi, en raison de son impact sur une multitude de composants cellulaires, l'arsenic est susceptible d'induire une réorganisation complète du métabolisme et la physiologie de la cellule. Afin d'étudier la réponse globale des micro-organismes face à un stress donné, des approches complémentaires à celles employées traditionnellement ont été développées au cours des dernières années et sont décrites dans la partie suivante.

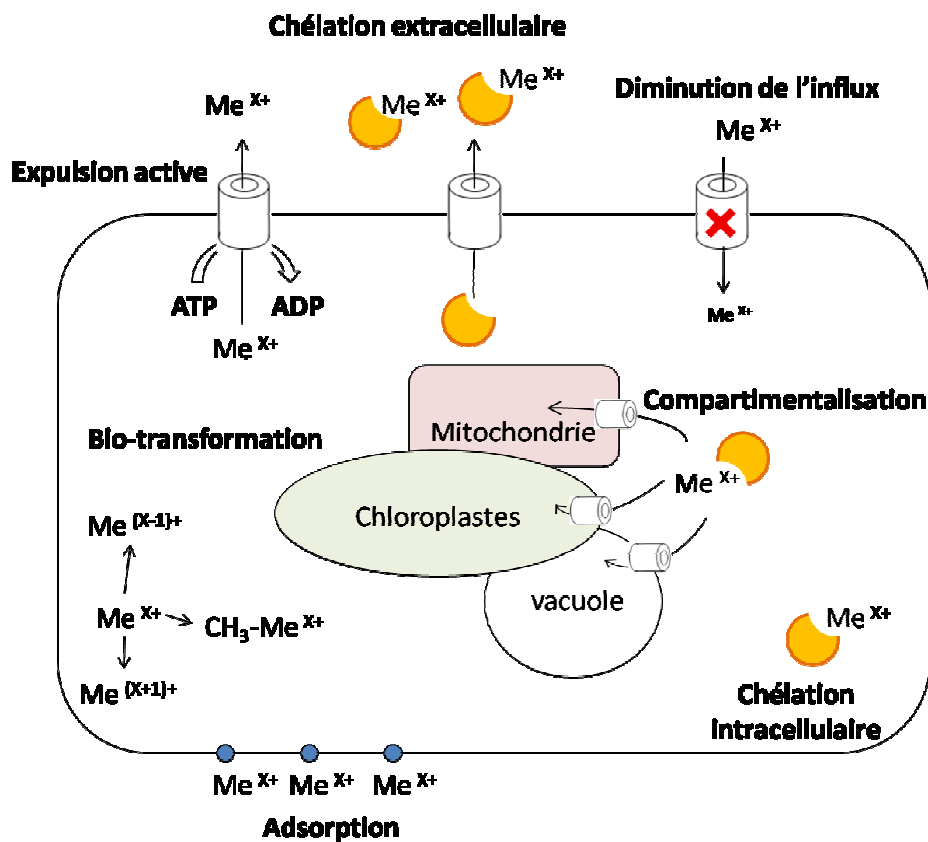


Figure 1.7 : Principaux mécanismes de défense face aux stress métalliques chez les eucaryotes. Me^{x+} : ion métallique.

3. LES MÉTHODES - OMICS

3.1. INTÉRÊTS DES APPROCHES –OMICS

Les exemples précédents illustrent bien que la réponse biologique d'un micro-organisme face à un stress donné est rarement simple, mais qu'elle est au contraire multifactorielle et fait intervenir différents acteurs moléculaires interagissant entre eux d'une manière plus ou moins complexe et hiérarchisée. De la même manière, l'étude d'une communauté microbienne dans son ensemble ne peut être réduite à un seul individu et des interactions, parfois complexes, unissent les différents organismes d'une même communauté. Par conséquent, il peut s'avérer indispensable d'envisager la description et la compréhension d'un écosystème microbien ou de la réponse d'un organisme face à un stress donné d'une manière globale.

Afin de déterminer la réponse globale d'un ou de plusieurs organismes face à un stress, des approches traditionnellement regroupées sous le terme de génomique descriptive et fonctionnelle ont récemment été développées. Ces approches couplent l'analyse de l'intégralité des gènes (génomique), des transcrits (transcriptome), des protéines (protéome) et/ou des métabolites (métabolome) d'un système biologique (organisme ou communauté). Dans certains cas, des outils moléculaires particulièrement bien développés chez les organismes modèles peuvent être utilisés pour déterminer le rôle de chaque gène dans une réponse biologique donnée. Ainsi, la souche BY4743 de *S. cerevisiae* a été utilisée comme outil dans le « systematic deletion project » visant, comme son nom l'indique, à déléter systématiquement tous les cadres de lecture ouverts de plus de 100 nucléotides depuis le codon « start » jusqu'au codon « stop » ce qui permet le criblage de tous les mutants à différents stress. Par analogie aux techniques plus courantes de génomique fonctionnelle citées précédemment, cette approche porte le nom de délétome (Jin et al., 2008). La combinaison de tous ces outils offre la possibilité d'obtenir une large vision de la réponse biologique à différents stress et d'identifier les gènes qui sont indispensables à la résistance à ceux-ci.

La génomique, la transcriptomique, la protéomique et la métabolomique sont autant d'approches globales qui se focalisent chacune sur un niveau hiérarchique biologique, c'est-à-dire, respectivement sur le contenu de l'information génétique, sur les gènes exprimés, sur le produit de la traduction ou sur de petites molécules biologiquement actives. L'analyse des données apportées par chacune de ces approches ainsi que la comparaison des données les unes par rapport aux autres permet d'avoir une image de la réponse biologique d'un organisme ou d'une communauté

microbienne. Cependant, chacune de ces techniques présente des limites qui lui sont propres et dont il faut impérativement tenir compte lors de l'analyse des résultats pour limiter au maximum les erreurs ou abus d'interprétations. D'autre part, la quantité énorme d'informations générées par ces approches ne doit pas conduire à faire l'économie d'analyses physiologiques plus traditionnelles permettant souvent de confirmer ou d'infirmer un certain nombre d'hypothèses. Les variantes, atouts et limites de chacune de ces approches de génomique descriptive et/ou fonctionnelle utilisées durant ma thèse sont décrites dans les paragraphes suivants. Le préfixe « méta » est ajouté ou non au nom de ces techniques pour souligner le fait que l'analyse ait été effectuée respectivement sur une communauté microbienne ou sur une population mono-spécifique.

3.2. LA (MÉTA-) GÉNOMIQUE

La (méta-) génomique repose sur l'étude du contenu génétique d'un seul ou d'un ensemble de micro-organismes. Par conséquent, cette approche est indissociable des techniques de séquençage qui ont connu récemment une évolution particulièrement rapide et qui seront décrites brièvement ci-dessous.

3.2.1. TECHNIQUES DE SÉQUENÇAGE

La technique de Sanger, développée depuis le milieu des années 1970, a connu des améliorations successives, incluant des procédures d'automatisation qui ont permis d'en réduire les coûts et d'augmenter la taille des séquences générées jusqu'à plus de 1kpb. Cependant, cette technique de séquençage repose sur des étapes préliminaires de clonage et de transformation bactérienne relativement pénalisantes en termes de coûts et de temps. Durant cette dernière décennie ont donc été développées des systèmes de séquençage à haut débit alternatifs reposant sur des principes radicalement différents de ceux utilisés précédemment et qualifiés pour cette raison de technologie de seconde génération (Metzker, 2010).

3.2.1.1. Le pyro-séquençage

Parmi les nouvelles techniques de séquençage, on peut citer la technologie de pyro-séquençage 454 développée en 2003 par Life Sciences, première technologie alternative au séquençage Sanger. Cette technique, permet actuellement de générer environ 450 Mpb par run avec des fragments d'une taille moyenne de 330 pb (Metzker, 2010).

Le principe repose sur la fragmentation de l'ADN à séquencer puis sur la ligation d'adaptateurs aux extrémités des fragments. Ces fragments d'ADN sont ensuite dénaturés sous

forme simple brin et liés individuellement à une micro-bille recouverte d'oligo-nucléotides complémentaires aux adaptateurs.

Après amplification PCR, des millions de séquences identiques recouvrent chaque bille (Figure 1.8, étape 1). Ces billes sont isolées individuellement dans des micro-puits contenant également de la polymérase, de la luciférase, de l'apyrase, de la luciférine et de l'adénosine phosphosulfate (APS). Des nucléotides (dATP, dGTP, dCTP, dTTP) sont ensuite ajoutés séquentiellement dans chaque puit. Soit le nucléotide est incorporé lors de la synthèse du brin complémentaire par la polymérase, dans quel cas il y aura une émission lumineuse selon les réactions décrites ci-dessous (Figure 1.8), soit le nucléotide n'est pas complémentaire, dans quel cas il n'y a pas d'émission lumineuse. Entre chaque ajout de nucléotides, le puit est lavé pour supprimer les nucléotides non incorporés puis le nucléotide suivant est injecté (Figure 1.8, étape 2).

La succession d'émission lumineuse (ou non) ainsi que son intensité sont enregistrées pour chaque puit ce qui permettra de déduire la séquence nucléotidique des fragments d'ADN présents sur chaque bille (Figure 1.8, étape 3). Chaque puce étant composée de millions de micro-puits, un run de séquençage permet d'accéder simultanément à la séquence de millions de molécules d'ADN. Ces séquences peuvent par la suite faire l'objet d'un assemblage pour reconstituer le ou les génomes correspondants. D'autre part il est à souligner que cette technique de séquençage est particulièrement sensible à la présence d'homopolymères dans les molécules d'ADN à séquencer, ce qui peut conduire à un décalage de cadre de lecture (Metzker, 2010).

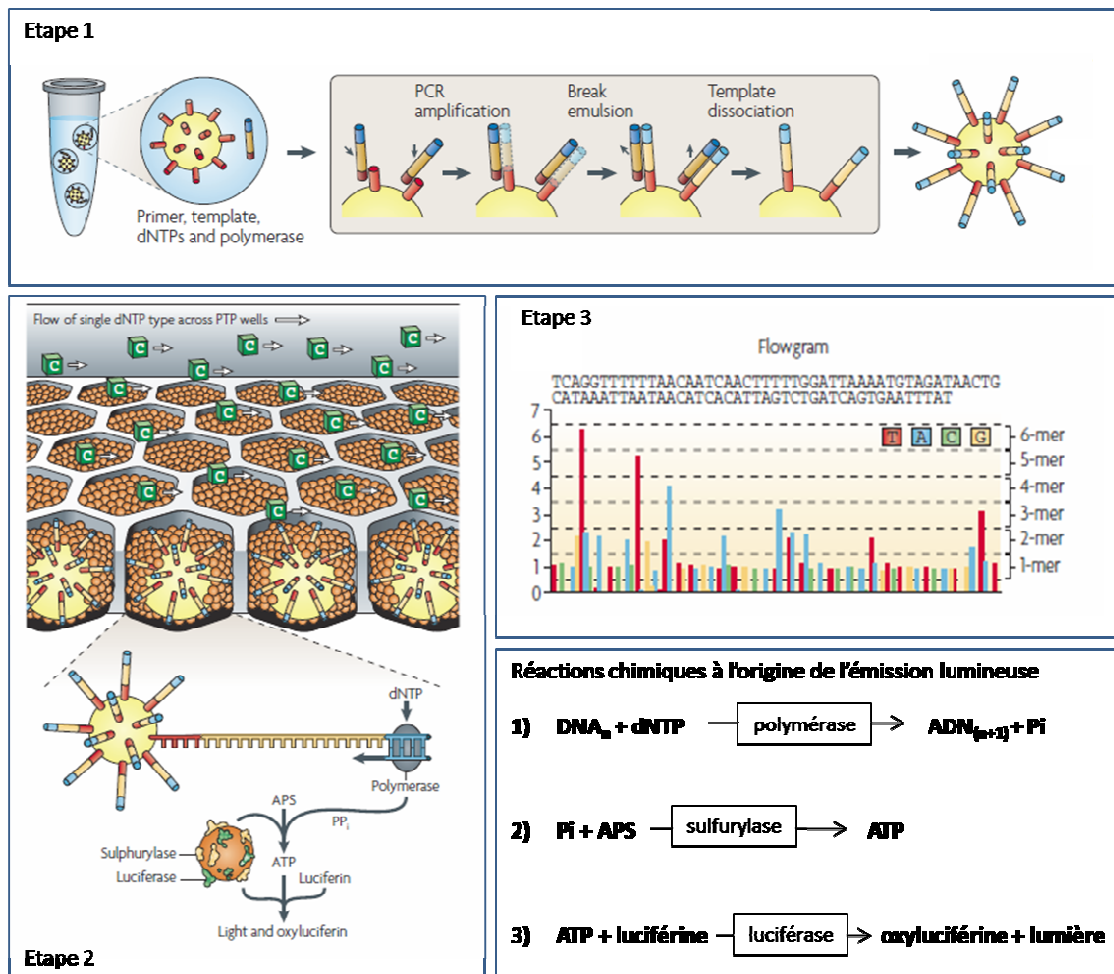


Figure 1.8 : Principales étapes relatives au pyroséquençage d'ADN (Metzker, 2010).

3.2.1.2. La technologie Illumina

La société Illumina (anciennement Solexa) a développé une autre technique de séquençage haut débit peu après l'avènement des techniques de pyro-séquençage. La première étape de cette technique de séquençage consiste en la fragmentation de l'ADN à séquencer puis à la ligation d'adaptateurs à chacune des extrémités. Ces molécules d'ADN sont ensuite dénaturées sous forme simple brin puis fixées sur une lame de verre recouverte d'oligo-nucléotides complémentaires aux adaptateurs et présents en excès par rapport aux fragments d'ADN. Les molécules d'ADN isolées sur la lamelle sont amplifiées par PCR pour produire localement des clusters de fragments d'ADN répartis sur la lamelle. La densité de ces clusters sur la lamelle dépend directement du rapport entre la quantité d'oligo-nucléotides sur le verre et la quantité de fragments à séquencer (Figure 1.9 étape 1).

Dans un second temps, une polymérase synthétise le brin complémentaire de chaque fragment d'ADN en incorporant des nucléotides marqués en partie 3' par un fluorochrome spécifique

bloquant l'élongation. Après lavage, seul le nucléotide complémentaire à la séquence d'ADN à séquencer reste au niveau du cluster (Figure 1.9 étape 2). Une image haute résolution de la lamelle de verre est alors enregistrée, la longueur d'onde émise par chaque cluster renseignant directement sur la nature du nucléotide incorporé (Figure 1.9 étape 3). Après acquisition de l'image, le fluorochrome du dernier nucléotide incorporé est libéré par action enzymatique et un nouveau jeu de nucléotides marqués peut être ajouté pour accéder à la séquence du nucléotide suivant. La succession de ces étapes permet d'accéder à la séquence complète du fragment d'ADN de chaque cluster dans la limite des 100 premiers nucléotides mais générant au total plus de 18 Gpb par lecture.

D'autre part, afin de faciliter l'assemblage des données obtenues, une étape supplémentaire décrite sous le terme de « mate pair » est souvent utilisée. Cette approche repose sur le séquençage Illumina des deux extrémités d'un même fragment d'ADN dont la taille est connue suite à un retournement physique des brins d'ADN sur la lamelle de verre. De cette manière, on obtient à la fois des informations sur la séquences des 100 nucléotides localisés sur les 2 extrémités du fragment d'ADN, mais également sur la distance qui sépare ces deux séquences. Par conséquent, l'utilisation de banques de fragments d'ADN de différentes tailles (par exemple 1, 2, 5 et 10Kpb +/- 5%) permet de les assembler plus efficacement.

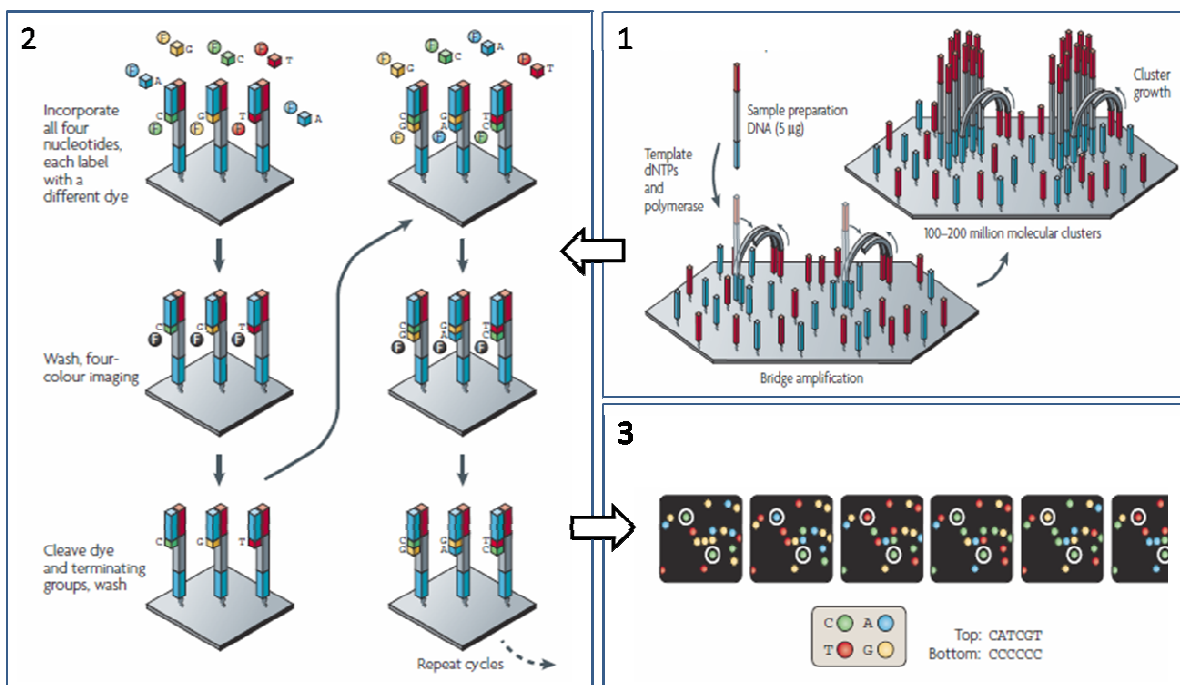


Figure 1.9 : Principales étapes relatives au séquençage Solexa (Metzker, 2010).

3.2.1.3. Les technologies SOLiD et de troisième génération

Applied Biosystems a développé la technologie SOLiD (*Sequencing by Oligonucleotide Ligation and Detection*) basée sur un principe de séquençage par ligation générant 30 GPb par run mais avec des fragments d'une taille moyenne de l'ordre de 50 pb (Metzker, 2010). Ce principe de séquençage, utilisé principalement dans le cadre de la détection de mutations nucléotidiques ponctuelles n'a pas été utilisé durant ma thèse et ne fera par conséquent pas l'objet d'une description détaillée dans ce manuscrit. D'autre part, des techniques de séquençage dites de troisième génération sont en cours de développement, notamment par la société Helicos avec l'Hélioscope et par la société Pacific Biosciences. Cette dernière a développé un procédé de séquençage en temps réel, c'est-à-dire qui permet d'accéder à la séquence nucléotidique d'un fragment d'ADN au fur et à mesure de la synthèse de son brin complémentaire. Ce procédé permet de générer des séquences nucléotidiques d'une taille moyenne de 1 kpb facilitant les étapes d'assemblage ultérieures. Bien que cette technologie souffre encore d'un taux d'erreur très élevé, elle semble être une alternative intéressante pour remplacer à l'avenir les techniques de séquençage hydrides utilisées actuellement pour l'assemblage de données (méta-)génomiques. En effet, lorsque l'objectif est d'assembler le plus efficacement possible des séquences nucléotidiques, l'idée est actuellement de combiner une technique de séquençage générant des séquences relativement longues (type pyro-séquençage) avec d'autres techniques à très haut débit (Illumina) permettant d'obtenir une très forte couverture.

Par conséquent, ces nouvelles méthodes de séquençage, si elles offrent la possibilité de générer d'importantes quantités de données à un faible coût, soulèvent de nouveaux défis en terme d'assemblage et de traitement des données, surtout en raison de la taille limitée des séquences nucléotidiques générées. Selon la question biologique posée (ré-séquençage, séquençage *de novo*, assemblage, single nucleotide polymorphisme) on utilisera donc plutôt l'une ou l'autre technique de séquençage, voire une combinaison de plusieurs d'entre elles.

3.2.2. ENJEUX ET LIMITES DE LA (MÉTA-)GÉNOMIQUE

La (méta-) génomique repose sur l'étude du contenu génétique d'un seul ou d'un ensemble de micro-organismes. Le génome renfermant par définition toute l'information génétique nécessaire au fonctionnement d'un organisme, cette approche apporte en théorie une vision exhaustive des potentialités fonctionnelles d'une population microbienne donnée, à la condition toutefois que l'intégralité du ou des génomes soit restituée.

Ce dernier point dépend à la fois de la couverture utilisée lors du séquençage, c'est-à-dire du nombre de fois en moyenne qu'un nucléotide donné est séquencé, ainsi que la qualité de l'assemblage. Si les contraintes de couverture et d'assemblage sont en général relativement bien maîtrisées pour des populations monoclonales de micro-organismes, le défi est tout autre pour des communautés constituées de différentes populations microbiennes et ceci d'autant plus que la complexité biologique de l'échantillon biologique est grande. Dans ce dernier cas, la couverture et/ou l'assemblage n'est en général pas suffisant pour reconstituer indépendamment le génome des différents micro-organismes présents dans l'échantillon. Les informations générées s'apparentent alors à un pool de séquences non assemblées et on parle alors de « méta-génome » (Handelsman *et al.*, 1998). Il est à noter sur ce dernier point que le terme « méta-génomique » est souvent utilisé avec excès dans la littérature scientifique pour décrire le séquençage haut débit de marqueurs phylogénétiques d'une population microbienne et préalablement amplifiés par PCR (notamment les gènes codant pour l'ARNr 16S et 18S). Cette dernière approche, bien qu'apportant des informations sur la composition phylogénétique d'une population microbienne, n'apporte aucun éclairage direct sur ses potentialités fonctionnelles reposant sur les génomes et ne peut par conséquent pas être qualifiée d'analyse méta-génomique *sensu stricto*.

L'interprétation des données qui en résultent est donc radicalement différente selon qu'il s'agisse de l'étude d'un génome ou d'un méta-génome. En effet, pour les données génomiques, on peut espérer, dans le cas où la couverture et l'assemblage sont suffisants et de qualité, restituer le génome d'un micro-organisme dans son intégralité. Toutes les potentialités fonctionnelles de l'organisme en question sont alors accessibles et l'on peut émettre des hypothèses sur ce que l'organisme est susceptible de pouvoir faire ou de ne pas faire. *A contrario*, pour les données méta-génomiques, l'étude ne repose plus sur des individus microbiens isolés mais sur une partie ou sur l'ensemble de la communauté microbienne qui est définie comme « unité fonctionnelle ». La couverture des génomes étant alors plus faible et l'assemblage souvent impossible ou partiel, la méta-génomique conduit à l'obtention de données non exhaustives et l'analyse doit alors se focaliser sur les potentialités fonctionnelles retrouvées dans l'échantillon sans tirer de conclusions sur celles qui ne le sont pas.

Les analyses (méta-)génomiques offrent également un aperçu taxonomique de l'échantillon dans la mesure où les séquences des gènes pouvant être utilisés comme marqueurs phylogénétiques sont accessibles, mais surtout un aperçu des potentialités fonctionnelles de l'échantillon biologique. Cependant, si les analyses (méta-)génomiques apportent des informations sur les potentialités fonctionnelles de l'organisme ou de la communauté microbienne étudiée, elles ne permettent pas de

prédire si les gènes identifiés sont effectivement fonctionnels et/ou s'ils sont exprimés au moment de l'analyse. Pour pallier ces limites, les analyses (méta-)génomiques descriptives sont souvent combinées à des analyses fonctionnelles telles que la transcriptomique, la protéomique et/ou la métabolomique permettant de confirmer ou d'infirmer l'expression d'un certain nombre de fonctions prédites.

3.3. LA (MÉTA-)TRANSCRIPTOMIQUE

Cette technique repose sur l'analyse des transcrits exprimés par un seul ou un ensemble de micro-organismes à un instant donné. Tous les transcrits exprimés (ARNm, ARNt, ARNr, siARN, miARN, ...) sont en théorie susceptibles d'être étudiés lors d'analyses (méta-)transcriptomiques. Cependant, dans la plupart des cas, ces études se focalisent uniquement sur les ARNm du fait qu'ils codent potentiellement pour des protéines et apportent par conséquent un éclairage sur les fonctions exprimées par l'organisme ou la communauté microbienne étudiée. Comme pour les analyses (méta-)génomiques, les analyses (méta-)transcriptomiques souffrent parfois de quelques abus de langage dans la littérature scientifique, dans la mesure où des études se focalisant uniquement sur les ARNr sont présentées comme faisant appel à la (méta-)transcriptomique. Là aussi, même si ces études permettent d'identifier phylogénétiquement les populations microbiennes physiologiquement actives, elles n'apportent pas d'informations directes sur les fonctions biologiques exprimées par celles-ci, ces informations reposant sur le contenu génomique.

Les analyses (méta-)transcriptomiques permettent d'accéder en théorie au contenu total des transcrits exprimés par une communauté microbienne, mais ne permettent en général pas d'identifier les micro-organismes qui les ont exprimés. Par conséquent, les données méta-transcriptomiques sont idéalement combinées aux données méta-génomiques. En effet, si ces dernières ont conduit à l'assemblage de génomes, les transcrits identifiés peuvent être assignés aux organismes correspondants.

3.4. LA (MÉTA-)PROTÉOMIQUE

3.4.1. DÉFINITION

La (méta-)protéomique est un volet de la génomique fonctionnelle qui se focalise sur l'analyse des protéines exprimées par un seul ou un ensemble de micro-organismes (Wilkins, 1996). Les protéines sont directement effectrices de nombreuses fonctions biologiques constitutives ou induites lors d'une réponse biologique ce qui souligne l'intérêt de cette approche. D'autre part, des

modifications post-traductionnelles ou des épissages alternatifs au niveau des transcrits sont susceptibles de modifier les fonctions biologiques associées aux protéines et ne sont donc accessibles qu'en étudiant les protéines elles-mêmes (Wilmes & Bond, 2006).

3.4.2. ANALYSES PAR SPECTROMÉTRIE DE MASSE

Lors de l'analyse par spectrométrie de masse, le signal relatif aux peptides majoritaires aura tendance à masquer celui des peptides minoritaires et par conséquent à rendre l'analyse moins sensible. Avant d'en effectuer l'analyse, le protéome est donc fragmenté, souvent par gel de polyacrylamide en une ou deux dimensions. Les protéines sont ensuite digérées pour générer des peptides dont la taille est compatible avec les analyses de spectrométrie de masse. Les peptides obtenus sont par la suite séparés les uns des autres, le plus souvent par chromatographie liquide haute performance (HPLC). Cette fragmentation de l'échantillon se superpose à la fragmentation des protéines déjà réalisée sur gel 1D ou 2D et joue un rôle important dans la qualité de l'analyse en limitant la co-élution des peptides analysés (Domon & Aebersold, 2006). D'autres techniques plus résolutive de séparation des peptides ont été développées, telle que la nano-HPLC (système miniaturisé plus résolutif que l'HPLC traditionnelle) ou l'UHPLC (chromatographie liquide à ultra-haute performance) basée sur l'utilisation de billes plus fines dans la phase stationnaire de la colonne chromatographique conduisant à un meilleur pouvoir résolutif.

Les peptides élués sont ensuite ionisés (par MALDI ou ESI) avant d'entrer dans l'analyseur où leur rapport masse sur charge sera analysé. Il existe 4 types d'analyseurs couramment utilisés en protéomique : l'analyseur à temps de vol (TOF), l'analyseur quadripolaire (Q), les trappes d'ions (IT) ou les analyseurs à transformée de Fourier (FT-ICR, Orbitrap). Chacun de ces analyseurs présente une résolution, une sensibilité, une gamme et précision de masse ainsi qu'une capacité à produire des spectres de masses qui lui sont spécifiques. Ces analyseurs peuvent être utilisés isolément ou fonctionner en association dans le cas des analyses en tandem (MS/MS) (Domon & Aebersold, 2006). Dans ce dernier cas, les peptides ionisés initiaux, appelés ions parents, sont fragmentés en sous-produits avant de pénétrer dans le deuxième analyseur où le rapport masse sur charge de chaque sous-fragment est analysé. Cette analyse en tandem permet de déduire la séquence en acides aminés de chaque peptide et ainsi de faciliter l'identification de la protéine correspondante dans les bases de données.

3.4.3. MÉTA-PROTÉOME : UN CONTENU NON EXHAUSTIF

Le méta-protéome est défini comme l'ensemble du contenu protéique d'un organisme ou d'une communauté microbienne. Contrairement aux acides nucléiques, les protéines présentent des

propriétés physico-chimiques différentes les unes des autres (charge, hydrophobicité, solubilité ...) en fonction de la nature des acides aminés qui les composent. Cette caractéristique peut être problématique dans les procédures d'extraction des protéines dans la mesure où cela peut conduire à un enrichissement et/ou à une perte de certaines protéines par rapport à d'autres en fonction de leur solubilité dans le tampon d'extraction utilisé (Hegde *et al.*, 2003).

D'autre part, dans le cas d'analyses de protéomique différentielle, les protéines peuvent être séparées par un gel de polyacrylamide bi-dimensionnel. La séparation des protéines s'effectue alors en fonction de leur point isoélectrique lors de la première dimension et en fonction de leur masse la seconde dimension. La séparation des protéines en fonction de leur point isoélectrique sous-entend que la procédure d'extraction des protéines ne doit pas modifier la charge de celles-ci, et que par conséquent l'usage de détergents anioniques puissants tel que le SDS habituellement utilisé pour la lyse et la solubilisation des protéines est proscrite. Ceci conduit à une faible solubilité des protéines hydrophobes, en particulier les protéines membranaires, qui ne sont donc pas ou peu accessibles par cette technique de séparation bi-dimensionnelle, contrairement aux gels de polyacrylamide uni-dimensionnels (Groen & Lilley, 2010; Krause, 2006). D'autre part, les protéines de très haut poids moléculaire, auront tendance à ne pas pénétrer dans le gel et seront par la même occasion exclues de l'analyse. Enfin, l'identification des protéines extraites d'un échantillon repose habituellement sur des techniques de spectrométrie de masse en tandem. Lors de telles analyses, deux problèmes majeurs sont limitants dans l'exhaustivité de l'analyse :

Le premier problème résulte du fait que seuls les ions parents les plus intenses sont sélectionnés dans un premier balayage de spectre de masse pour être analysés dans un second temps par dans un deuxième analyseur. Par conséquent, ce type de procédure automatisée conduit à une perte des données relatives aux ions minoritaires et donc des protéines correspondantes. Dans certains cas extrêmes et bien définis, par exemple pour le plasma sanguin où les 9 protéines majoritaires représentent 90% des protéines totales, celles-ci peuvent être extraites de l'échantillon par immuno-extraction mais ces étapes de purification conduisent inévitablement à la perte de protéines (Huang *et al.*, 2003).

L'autre problème principal lors de l'analyse par spectrométrie de masse résulte du fait que les ions élués pendant la fragmentation et l'acquisition MS/MS d'un autre peptide ne sont pas analysés non plus. L'utilisation de spectromètres de masse ayant une vitesse de balayage élevée couplé à une durée d'acquisition des données courte et à une bonne fragmentation de l'échantillon permet néanmoins de limiter ce phénomène.

Pour toutes ces raisons techniques, l'analyse d'un (méta-)protéome ne peut donc pas être considérée comme exhaustive et le nombre de protéines identifiées dans un échantillon donné est en général largement sous-estimé par rapport au contenu réel en protéines et ceci d'autant plus qu'il s'agit d'un échantillon complexe. L'analyse qui en résulte doit donc se focaliser sur les fonctions biologiques retrouvées dans l'échantillon en évitant autant que possible d'interpréter l'absence de telle ou telle protéine. D'autre part, il est à noter que la synthèse d'une protéine ne signifie pas nécessairement que l'activité biologique associée soit également présente. En effet, l'accumulation d'une protéine dans une cellule peut résulter d'une induction de sa synthèse mais également de son accumulation sous forme d'agrégats qui ne sont pas nécessairement biologiquement actifs ou résulter d'un dysfonctionnement des voies de dégradation. D'autre part, l'activité biologique d'une protéine peut être régulée par des modifications post-traductionnelles (phosphorylation, acétylation, ...). Par conséquent, l'interprétation biologique des données protéomique doit être réalisée avec prudence et doit être confirmée, dans la mesure du possible, par des approches alternatives et complémentaires (mesure directe de l'activité, dosage des substrats et produits, ...).

3.5. LA (MÉTA-)MÉTABOLOMIQUE

La (méta-)métabolomique est un volet de la génomique fonctionnelle se focalisant sur le métabolome qui peut être défini comme « l'ensemble des petites molécules impliquées dans le réseau bio-chimique » (Daviss, 2005; Oliver *et al.*, 1998). Par conséquent, leur synthèse, transport, transformation et/ou dégradation sont en général régulés et catalysés par des partenaires protéiques, principales molécules effectrices de la réponse biologique. L'étude des métabolites peut donc se révéler fondamentale pour confirmer ou infirmer l'activité biologique des protéines qui leur sont associées.

De par la définition très large donnée aux métabolites, leur nature et leur propriétés physico-chimique sont également très diverses. Par conséquent, il est difficile de définir un mode opératoire permettant de faire une analyse globale du métabolome qui reste plus proche du concept biologique que d'une réalité expérimentale (Kaddurah-Daouk *et al.*, 2008). Les techniques développées pour identifier les métabolites se focalisent donc sur un seul ou un ensemble de métabolites partageant des propriétés communes (solubilité dans un solvant, fonction chimique commune, température de volatilité, affinité pour une enzyme...). On peut par exemple citer la chromatographie en phase gazeuse pour l'identification des acides gras après extraction au chloroforme ou l'identification des métabolites cytoplasmique dans une solution d'eau/méthanol (80/20% v/v). A l'instar des analyses (méta-)protéomiques, des techniques analytiques basées sur la spectrométrie de masse sont également utilisées lors d'analyses (méta-)métabolomiques.

3.6. EXEMPLE DE L'ORGANISME MODÈLE *S. CEREVISIAE*

La levure *S. cerevisiae* est étudiée depuis plusieurs décennies en conditions de laboratoire et est l'un des premiers organismes dont le génome a été séquencé. Ces analyses ont notamment confirmé la présence de 16 chromosomes chez cet organisme présentant un génome d'une taille totale de 12 Mpb et contenant 5 885 cadres de lecture ouverts (Goffeau *et al.*, 1996). Cette levure est considérée comme un organisme modèle en biologie moléculaire et de nombreux outils, tel le délétome de cette espèce, ont été créés (Jin *et al.*, 2008). Le criblage de cette banque de mutants pour leur tolérance à l'arsenic a permis d'identifier les gènes qui sont, directement ou indirectement, impliqués dans les mécanismes de résistance à ce composé (Thorsen *et al.*, 2009). Ces analyses ont souligné que les gènes impliqués dans le métabolisme du soufre, la biosynthèse de glutathion, la perception de l'environnement, la synthèse d'ARNm et la transcription, le transport vacuolaire, la régulation du cycle cellulaire, le métabolisme des lipides, la biogénèse des mitochondries et la synthèse du cytosquelette jouent un rôle dans l'adaptation au stress arsénié (Thorsen *et al.*, 2009). Des études du délétome de la levure portant sur d'autres métaux ont d'ailleurs mis en avant le fait que les gènes impliqués dans le repliement tri-dimensionnel de la tubuline, la transduction du signal et les voies de sécrétion sont spécifiquement indispensables à la tolérance au stress arsénié chez cette espèce (Jin *et al.*, 2008).

D'autre part, des approches de transcriptomique comparative réalisées grâce à l'utilisation de puces Agilent ont révélé l'induction de gènes impliqués dans le transport de métaux, la détoxification des EOR, le métabolisme des sucres et des lipides, le transport des polyamines et la transcription des ARN quel que soit le métal auquel la levure est exposée tandis que les gènes impliqués dans l'assignement et le transport de protéines étaient globalement réprimés (Figure 1.10) (Jin *et al.*, 2008). Concernant le stress arsénié en particulier, le métabolisme du soufre, la régulation du repliement tri-dimensionnel des protéines, la protéolyse, la réponse au stress oxydatif, l'homéostasie du fer et la biosynthèse du tréhalose sont des fonctions biologiques dont la transcription est spécifiquement induite en présence d'arsenic. Il est à noter que ce métalloïde est le composé qui affecte l'expression du plus grand nombre de gènes en comparaison aux autres métaux testés (Ag, Cu, Cd, Hg, Zn, Cr), soit 762 gènes sur les 4 700 que compte la levure. Ces résultats ont été en partie confirmés par des analyses de protéomique et de métabolomique ciblée soulignant respectivement l'induction de protéines impliquées dans le métabolisme du soufre et l'augmentation de la concentration intracellulaire en glutathion et cystéine (Thorsen *et al.*, 2007).

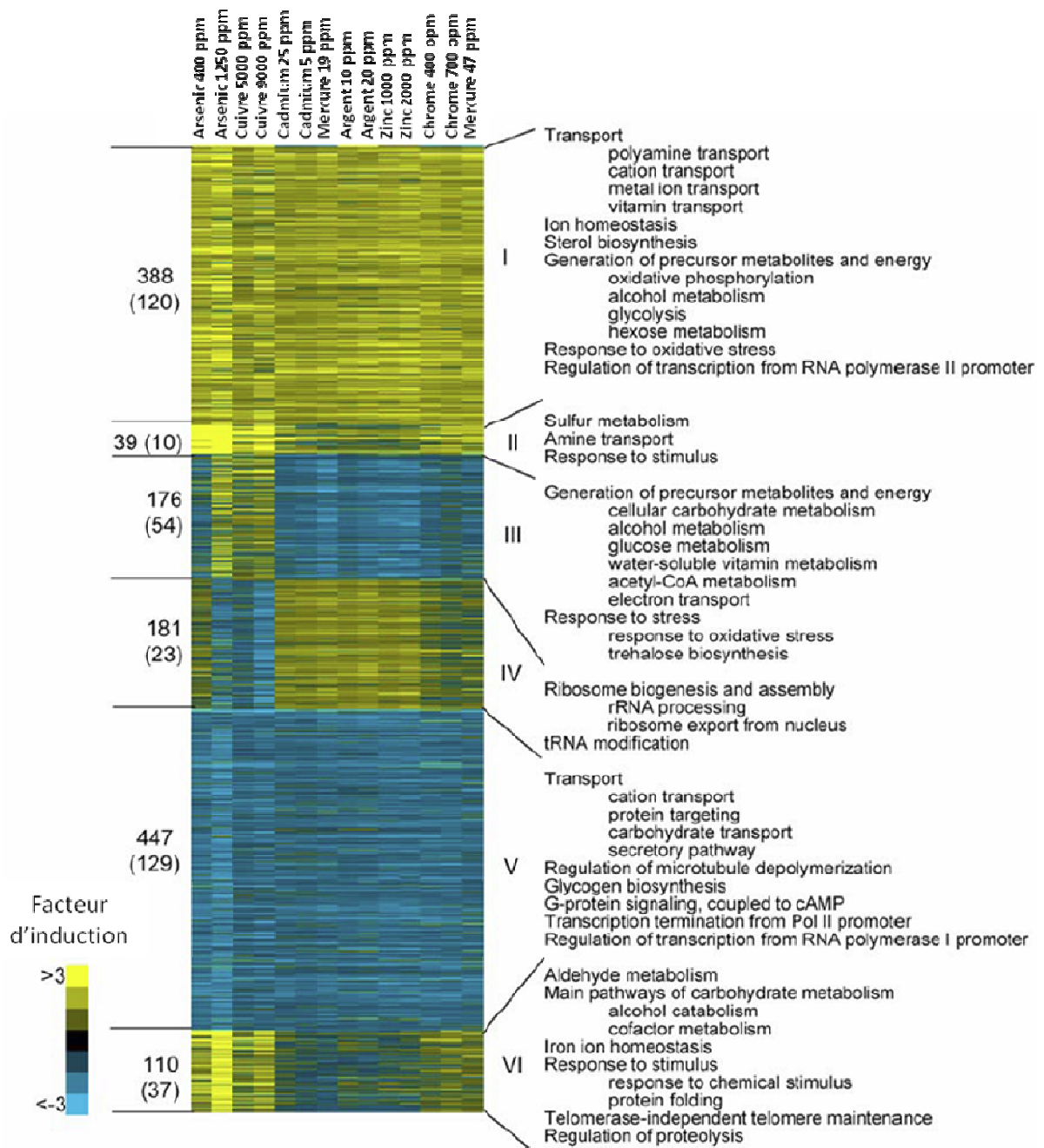


Figure 1.10 : Carte de chaleur des transcrits induits ou réprimés en présence de différents métaux chez la levure. Les transcrits présentant un profil de variation similaire ont été rassemblés en différents groupes (I à VI) (Jin et al., 2008).

Ces résultats soulignent donc le fait que l'usage de différentes techniques d'analyses –OMICS permet d'accéder à des résultats complémentaires les uns des autres et que l'adaptation au stress arsénié est un processus complexe et hautement régulé chez la levure, dont une partie au moins est spécifique à ce composé.

La génomique fonctionnelle est un puissant outil d'observation du vivant au niveau moléculaire qui offre la possibilité d'apporter un éclairage original et sans a priori sur un échantillon biologique donné y compris sous contraintes environnementales. Chacun des volets de la génomique fonctionnelle présente des limites propres qui peuvent néanmoins être contournées, au moins en partie, lorsque les différentes approches sont combinées et en utilisant d'autres techniques alternatives. Les données générées permettent par conséquent d'émettre un certain nombre d'hypothèses sur le fonctionnement d'un écosystème microbien qui peuvent être testées et vérifiées dans un second temps par des approches plus ciblées.

4. OBJECTIFS DE LA THÈSE

Le DMA de Carnoulès dans le Gard (France) fait l'objet d'un suivi de ses caractéristiques physico-chimiques depuis plusieurs décennies. Ces observations ont notamment mis en évidence un processus naturel d'atténuation le long du cours d'eau conduisant à une diminution rapide de la concentration en arsenic, en fer et en sulfate. Ce processus a été proposé comme résultant, au moins en partie, du métabolisme bactérien de deux espèces affiliées au genre *Thiomonas* et *Acidothiobacillus* oxydant respectivement l'arsénite et le fer. Bien qu'apportant un éclairage sur l'impact potentiel de ces bactéries sur les cycles bio-géochimiques de l'arsenic et du fer, ces analyses ont été effectuées sur des cultures pures *in vitro* et n'apportent par conséquent que peu d'information sur le fonctionnement et l'impact en conditions environnementales de la communauté microbienne dans son ensemble.

Les objectifs de mes travaux de thèse étaient donc de déterminer grâce à des approches de génomiques descriptive et fonctionnelle, l'impact de la communauté microbienne dans son ensemble et en particulier celui des euglènes très abondantes sur ce site. Ainsi, l'apport de ces protistes dans les interactions trophiques et leur éventuel rôle dans les processus de bio-remédiation étaient à déterminer. L'étude des mécanismes adaptatifs qui permettent à ces euglènes de coloniser efficacement les DMA, en particulier en regard du stress arsénié, étaient également très peu documentés et ont donc constitué le second volet d'étude de ma thèse.

Afin d'atteindre ces objectifs, les interactions trophiques au sein de la communauté bactérienne ont été étudiées par des approches de protéo-génomique et de transcriptomique ciblée sur les gènes d'intérêts. L'apport des euglènes dans ces interactions trophiques a été abordé à la fois par des approches de protéomique et de métabolomique *in situ* et ont été validées par des observations réalisées en condition de laboratoire sur les euglènes isolées du site. Enfin, le métabolisme de l'arsenic chez les euglènes et son éventuel impact sur le site a été étudié en couplant des observations physiologiques et des analyses de génomique fonctionnelle (transcriptomique et protéomique différentielle).

Ces approches ont conduit à plusieurs publications scientifiques acceptées, soumises ou en cours de rédaction qui ont été utilisées dans ce manuscrit comme fil conducteur pour la présentation de mes travaux de thèse.

RESULTATS

CHAPITRE I :

GENOMIQUE DESCRIPTIVE ET

FONCTIONNELLE DE LA

COMMUNAUTE BACTERIENNE DU

DMA DE CARNOULES

A. INTRODUCTION

Le site de Carnoulès correspond à une ancienne exploitation minière abandonnée depuis les années 1960 qui a conduit à l'accumulation d'environ 1,5 million de tonnes de roches sous forme de stérile minier (Leblanc et al., 1996). Ces roches ainsi exposées à l'eau de pluie, au métabolisme microbien et à l'oxygène atmosphérique, sont soumises à une forte oxydation conduisant à des DMA étudiés depuis plusieurs années (Leblanc et al., 1996). Ces drainages se mélangent aux eaux de ruissellement d'une source localisée sous le stérile et forment ensemble un ruisseau, le Reigous, fortement exposé à ces DMA (Figure 2.1).

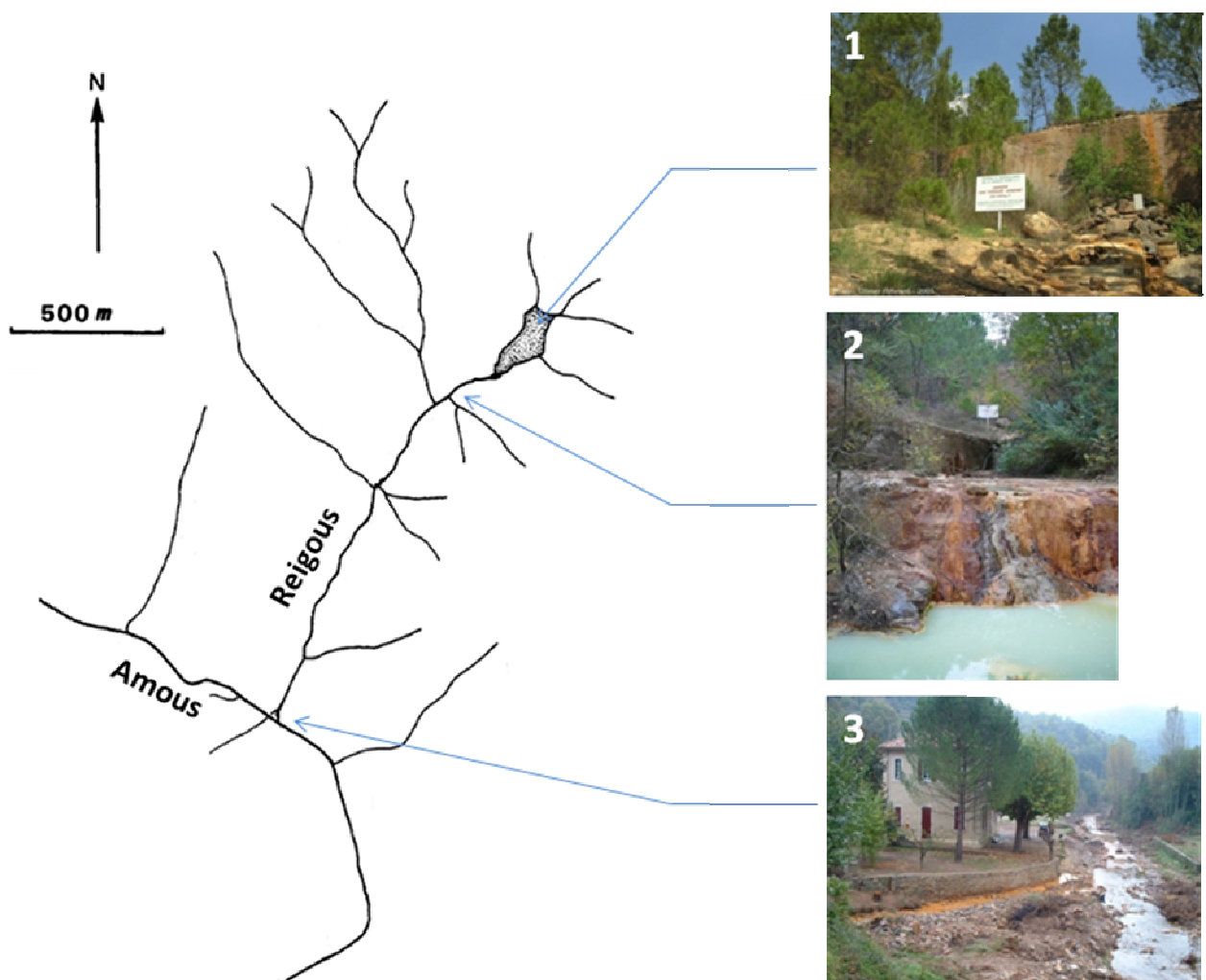


Figure 2.1 : Topologie du DMA de Carnoulès présentant la localisation du stérile minier (1), le point CowG localisé 30m en aval et correspondant à une petite retenue d'eau (2) et la confluence entre le Reigous et l'Amous (3). Photos personnelles, schéma réalisé d'après Leblanc et al. (1996).

Un suivi des caractéristiques physico-chimiques de ce ruisseau a notamment révélé qu'il présentait de manière constante de fortes concentrations en métaux et en sulfates, mais également un faible pH (Table 2.1). De plus, le Reigous renferme d'importantes teneurs en arsenic, notamment en raison du fait que les roches excavées durant l'exploitation minière sont riches en un minéral arsénié, l'arsénopyrite (Fe-As-S) (Leblanc et al., 1996). Au fur et à mesure de l'écoulement du Reigous, les concentrations en métaux, en arsenic et en sulfates diminuent rapidement pour atteindre des valeurs qui, bien que toujours très élevées, témoignent d'un processus naturel d'atténuation (Table 2.1).

Lieu d'échantillonnage	Période d'échantillonnage	pH (± SD)	T°C	DO (± SD)	As(III) (± SD)	As(V) (± SD)	Fe(II) (± SD)	Fe(III) (± SD)	
Stérile minier	S5	Novembre 2004	5.78 (±0.05)	15.5	2	277 (±14)	42 (±2)	1299 (±104)	4195 (±420)
		Avril 2005	4.05 (±0.05)	17.3	0.1-0.2	128 (± 6)	66 (±3)	784 (±62)	3264 (±326)
		Septembre 2005	3.73 (±0.05)	20.6	4-5	78 (± 4)	53 (±3)	778 (±62)	3629 (±363)
Reigous	S1	Novembre 2005	2.5 (±0.05)	14.6	1	30.0 (±0.8)	39 (±2)	879 (±70)	4388 (±441)
	CowG		2.74 (±0.05)	10.6	5-6	22.0 (±0.8)	22.0 (±0.8)	501 (±40)	1785 (±182)
	CowS		3.43 (±0.05)	7.2	5-6	4.5 (±0.2)	1.50 (±0.08)	95 (±8)	902 (±90)
	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. Ces analyses mettent en évidence la forte contamination du cours d'eau en métaux, en arsenic et en sulfate ainsi que le processus d'atténuation observé le long du cours d'eau (S5 : stérile minier, S1 : source du Reigous, COWG : point à 30 m de la source, COWS : point à 150 m de la source, CONF : confluence avec l'Amous à 1,5 km en aval de la source). Les unités sont données en mg/l (Bruneel et al., 2008).

Ces caractéristiques conduisent à la formation d'un écosystème qui, bien que toxique pour la plupart des organismes vivants, est propice au développement d'une communauté microbienne visiblement bien adaptée à ces contraintes environnementales. Ainsi, la présence dans les sédiments d'archées et de bactéries a été mise en évidence par des analyses de séquençage des gènes codant pour les ARNr 16S (Bruneel et al., 2006; Bruneel et al., 2008). L'atténuation naturelle observée sur le site a été proposée comme résultant, au moins en partie, du métabolisme de certains micro-organismes procaryotes dont le métabolisme est supposé conduire à la précipitation de fer, de sulfate et d'arsenic sous forme d'un minéral insoluble, la tooéélite ($\text{Fe}_6^{3+} (\text{As}^{3+}\text{O}_3)_4(\text{SO}_4)(\text{OH})_4 \cdot 4\text{H}_2\text{O}$) (Morin et al., 2003). En effet, des isolats bactériens obtenus à partir de sédiments ont montré que des espèces bactériennes affiliées à *Acidothiobacillus ferroxydans* et *Thiomonas* sp. pouvaient oxyder

in vitro le fer ferreux et l'arsénite en fer ferrique et arséniate respectivement, étapes préliminaires à la formation de ce minéral (Bruneel *et al.*, 2003; Duquesne *et al.*, 2003; LeBlanc *et al.*, 2002).

La communauté microbienne procaryote semble donc jouer, via son métabolisme, un rôle important dans les processus de bio-atténuation observés sur le site de Carnoulès. Cependant, l'expression *in situ* des gènes impliqués dans oxydation de l'arsenic et d'oxydation du fer n'a pas été démontrée et d'autres fonctions biologiques impliquées directement ou indirectement dans les processus de bio-atténuation ne sont pas exclues. Afin de comprendre plus en détail le fonctionnement de la communauté bactérienne en condition environnementale, des approches de génomique descriptive et fonctionnelle ont été entreprises sur les bactéries récupérées directement à partir des sédiments. Outre l'avantage de pouvoir étudier les micro-organismes non cultivables/non cultivés, ces analyses de génomique descriptive et fonctionnelle offrent également la possibilité d'étudier sans *a priori* les fonctions exprimées *in situ* par ceux-ci et d'émettre des hypothèses sur les interactions trophiques qui les unissent.

Des analyses couplant séquençage Sanger et 454 de l'ADN extrait de la fraction bactérienne récupérée des sédiments ainsi que des analyses méta-protéomiques ont été réalisées. De plus, certaines fonctions biologiques jugées essentielles pour le fonctionnement de la communauté microbienne et prédites par les données de méta-génomique ont été spécifiquement ciblées par des approches de RT-PCR sur les ARNm extraits des sédiments. Ce dernier point constitue la part expérimentale de mon travail dans cette analyse, dont les principaux résultats sont présentés dans le manuscrit à la fin de ce chapitre.

B : RESULTATS

Les techniques d'assemblage utilisées ont conduit à l'assemblage des génomes des 7 espèces microbiennes majoritaires du site, 2 d'entre elles, CARN 1 et 4, étant jusqu'ici totalement passées inaperçues par des approches cultivables malgré leur caractère abondant et leur métabolisme visiblement très actif sur ce site (environ 2/3 des protéines identifiées sur le site sont affiliées à ces deux espèces).

Les analyses méta-génomiques ont conduit à l'identification des potentialités fonctionnelles des différents génomes reconstitués. La synthèse des protéines correspondantes a été validée, dans la mesure du possible, par des approches de méta-protéomique permettant l'identification de plus de 500 protéines qui ont pu être associées dans la plupart des cas aux génomes reconstitués correspondants. L'expression *in situ* des protéines impliquées dans la production de sulfate,

d'arsénite et de fer ferrique a ainsi été démontrée lors de cette analyse, soulignant le fait que la formation de tooéléite, minéral résultant de la co-précipitation de ces trois composés, était bien liée, au moins en partie, au métabolisme bactérien. Si certaines de ces fonctions telle l'oxydation de l'arsénite ou l'oxydation du fer ferreux étaient déjà décrites précédemment par des approches *in vitro*, d'autres fonctions comme l'oxydation du soufre et la méthylation de l'arsenic n'étaient pas encore connues bien que jouant directement un rôle dans les processus de bio-remédiation sur ce site.

Ainsi, la présence du gène codant pour l'arsénite méthyltransférase a été révélée par l'approche méta-génomique. L'expression du gène codant pour cette enzyme a été vérifiée par des approches de transcriptomique ciblée, basée sur la rétro-transcription et l'amplification PCR de transcrits grâce à l'utilisation d'amorces spécifiques (RT-PCR). Dans ce cas, la transcription du gène codant pour l'arsénite méthyltransférase a pu être validée, ce qui était en accord avec des analyses physico-chimiques *in situ* mettant en évidence la présence de formes arséniées méthylées dans les sédiments. La transcription du gène ainsi que l'indentification *in situ* des produits de l'enzyme permettent de suggérer fortement l'expression de cette fonction biologique.

D'autre part, une analyse MFA (Multiple Factorial Analysis) réalisée à partir des données méta-génomique et méta-protéomique a également fait ressortir le fait que les génomes reconstitués CARN 1 et 4 présentaient un métabolisme très actif résolument orienté vers le recyclage et le métabolisme des acides aminés et des nucléotides. Par ailleurs, ces deux espèces disposent, contrairement aux autres micro-organismes, de gènes impliqués dans la synthèse de la cobalamine. Ce dernier point pourrait jouer un rôle indirect très important dans les processus de bio-remédiation dans la mesure où des observations réalisées *in vitro* ont montré que ce composé favorisait fortement l'oxydation du fer ferreux en fer ferrique par *Acidothiobacillus ferroxydans*. Les approches de génomique fonctionnelle ont par conséquent fait ressortir la présence d'interactions trophiques potentiellement très fortes au sein de la communauté bactérienne, en particulier concernant CARN 1 et 4.

Par ailleurs, d'autres interactions trophiques potentielles au sein de cette communauté bactérienne ont également été mises en évidence. Ainsi, les approches de génomique descriptive ont souligné le fait que certains génomes reconstitués regroupaient des gènes impliqués dans des processus de fixation du carbone et/ou de l'azote (CARN 2, 5 et 7), tandis que d'autre renfermaient des fonctions impliquées dans le recyclage de la matière organique (CARN 1, 4 et 6). L'expression de gènes jouant un rôle dans la production ou le recyclage de la matière organique a été validée par une approche de RT-PCR développée sur certains gènes cibles dont les protéines correspondantes

n'ont pas été identifiées. Ainsi, la transcription de gènes jouant un rôle dans la dégradation de purines (*punA*) et de la fixation de l'azote (*nif Q* et *nif A*) ont pu être mis en évidence tandis que l'expression de gènes impliqués dans la fixation du carbone (*rubisco*) ou le transport des acides aminés a été validée par l'analyse méta-protéomique. L'intégralité de ces analyses a permis de proposer un modèle du fonctionnement *in situ* de la communauté bactérienne (voir manuscrit en fin de chapitre).

C : DISCUSSION ET CONCLUSION

Les génomes reconstitués de CARN1 et CARN4, bien que très proches au niveau phylogénétique et de leur organisation génétique, ont pu être assemblés indépendamment. De la même manière, le génome reconstitué de CARN5 est quasiment identique aux 2 génomes de la population clonale d'*Acidothiobacillus ferrooxydans* présent dans les bases de données (ATCC 23270 et ATCC 53993), à l'exception notable d'un fragment d'environ 300 Kpb qui n'est pas retrouvé dans le génome reconstitué de CARN5 ni dans la souche ATCC 53993. L'absence d'un tel fragment, confirmé par deux techniques de séquençage indépendantes (Sanger et 454), souligne le fait qu'il ne s'agit pas d'un biais de séquençage. Ces observations soulignent donc que les techniques d'assemblage utilisées sont suffisamment discriminantes pour souligner des différences d'organisation génétique relativement fines.

Par opposition, les espèces du genre *Thiomonas* ont récemment été démontrées comme présentant une grande plasticité au niveau de leur organisation génomique ce qui suggère la présence de différentes sous populations au sein du site de Carnoulès (Arsène-Ploetze et al., 2010). L'analyse méta-génomique n'ayant conduit à la reconstruction que d'un seul génome phylogénétiquement apparenté au genre *Thiomonas* (CARN 2), cet assemblage peut éventuellement être considéré comme un génome « moyen » des différents écotypes de *Thiomonas* présents sur le site de Carnoulès.

Par conséquent, bien que l'approche méta-génomique ne permette pas de restituer complètement les génomes de populations clonales de bactéries, elle offre la possibilité d'accéder aux potentialités fonctionnelles de populations bactériennes visiblement assez homogènes, y compris pour les organismes non-cultivables. L'interprétation des différents méta-génomes a donc été effectuée comme s'il s'agissait de génomes de 7 bactéries différentes, en prenant soin de ne discuter que les fonctions retrouvées, sans tirer de conclusions sur celles qui ne l'étaient pas, les données méta-génomiques ne présentant pas un caractère exhaustif.

Ces résultats, bien qu'apportant des éclaircissements sur la fonctionnalité de la communauté bactérienne, ont également soulevé des questions sur les interactions trophiques au sein de la communauté microbienne dans son ensemble, en particulier par rapport à *Euglena mutabilis*. En effet, ce protiste, abondamment retrouvé sur le site, est décrit comme l'un des principaux producteurs primaires dans les environnements exposés aux DMA. Ces observations ont donc tout naturellement conduit à étudier le métabolisme de l'euglène en condition environnementale, en particulier sur son rôle dans les échanges trophiques.

D : MANUSCRIT

L'intégralité de ces résultats ont conduit à une publication scientifique dans « the ISME Journal » dont le manuscrit correspondant est présenté ici.

ORIGINAL ARTICLE

Metabolic diversity among main microorganisms inside an arsenic-rich ecosystem revealed by meta- and proteo-genomics

Philippe N Bertin¹, Audrey Heinrich-Salmeron¹, Eric Pelletier^{2,3,4}, Florence Goulhen-Chollet¹, Florence Arsène-Ploetze¹, Sébastien Gallien⁵, Béatrice Lauga⁶, Corinne Casiot⁷, Alexandra Calteau^{3,4,8}, David Vallenet^{3,4,8}, Violaine Bonnefoy⁹, Odile Bruneel⁷, Béatrice Chane-Woon-Ming^{3,4,8}, Jessica Cleiss-Arnold¹, Robert Duran⁶, Françoise Elbaz-Poulichet⁷, Nuria Fonknechten^{2,3,4}, Ludovic Giloteaux⁶, David Halter¹, Sandrine Koechler¹, Marie Marchal¹, Damien Mornico^{3,4,8}, Christine Schaeffer⁵, Adam Alexander Thil Smith^{3,4,8}, Alain Van Dorsseleer⁵, Jean Weissenbach^{2,3,4}, Claudine Médigue^{3,4,8} and Denis Le Paslier^{2,3,4}

¹Génétique Moléculaire, Génomique et Microbiologie, UMR7156 CNRS and Uds, Strasbourg, France;

²CEA, DSV, IG, Genoscope, Laboratoire de Métagénomique des Procaryotes, France; ³CNRS, UMR8030, Evry, France; ⁴UEVE, Université d'Evry, France; ⁵Laboratoire de Spectrométrie de Masse Bio-Organique, Institut Pluridisciplinaire Hubert Curien, UMR7178 CNRS and Uds, Strasbourg, France; ⁶Equipe Environnement et Microbiologie, UMR5254 CNRS and UPPA, Institut Pluridisciplinaire de Recherche sur l'Environnement et les Matériaux, Pau, France; ⁷Laboratoire HydroSciences Montpellier, UMR 5569, Montpellier, France; ⁸CEA, DSV, IG, Genoscope, Laboratoire d'Analyses Bioinformatiques pour la Génomique et le Métabolisme, France and ⁹Laboratoire de Chimie Bactérienne, UPR9043 CNRS et Université de la Méditerranée, Marseille, France

By their metabolic activities, microorganisms have a crucial role in the biogeochemical cycles of elements. The complete understanding of these processes requires, however, the deciphering of both the structure and the function, including synecologic interactions, of microbial communities. Using a metagenomic approach, we demonstrated here that an acid mine drainage highly contaminated with arsenic is dominated by seven bacterial strains whose genomes were reconstructed. Five of them represent yet uncultivated bacteria and include two strains belonging to a novel bacterial phylum present in some similar ecosystems, and which was named 'Candidatus Fodinabacter communificans.' Metaproteomic data unravelled several microbial capabilities expressed *in situ*, such as iron, sulfur and arsenic oxidation that are key mechanisms in biomineralization, or organic nutrient, amino acid and vitamin metabolism involved in syntrophic associations. A statistical analysis of genomic and proteomic data and reverse transcriptase-PCR experiments allowed us to build an integrated model of the metabolic interactions that may be of prime importance in the natural attenuation of such anthropized ecosystems.

The ISME Journal advance online publication, 12 May 2011; doi:10.1038/ismej.2011.51

Subject Category: integrated genomics and post-genomics approaches in microbial ecology

Keywords: microbial genomics; extreme environment; uncultured microorganism; bacterial phylum; synecology; functional ecology

Introduction

Microorganisms are involved in biogeochemical nutrient cycles and have, therefore, a crucial role in the biosphere (Haferburg and Kothe, 2007; Falkowski *et al.*, 2008; Konopka, 2009). Although

microbial ecosystems constitute a major reservoir of our planet's genetic biodiversity, it is generally recognized that most of the microorganisms present in the environment are not accessible by the current culture-dependent techniques. The recent developments in genomics have given an unprecedented opportunity to gain insight into the structure and the functioning of microbial communities (Allen and Banfield, 2005; Bertin *et al.*, 2008; Wilmes *et al.*, 2009). Indeed, environmental genomics has extended their analysis far beyond the sole taxonomic studies, allowing the characterization of the whole community gene pool and its expression. Such

Correspondence: PN Bertin, Génétique Moléculaire, Génomique et Microbiologie, Institut de Botanique, Université de Strasbourg, 28 rue Goethe, Strasbourg 67000, France.

E-mail: philippe.bertin@unistra.fr

Received 19 November 2010; revised 17 February 2011; accepted 29 March 2011

approaches give rise to an integrated picture of ecosystems and are, therefore, of great interest to interpret the metabolisms present in ecological niches considered as extreme, including those impacted by human activities. Among the most toxic anthropogenic contaminants, arsenic is at the origin of serious forms of water pollution in industrial and post-industrial areas all over the world. In particular, high arsenic contents are encountered in mine drainage waters, where the metalloid is usually associated with sulfur, iron and other metals (Vaughan, 2006). Few forms of life are known to thrive in these often acid waters, as compared with neutral waters (Johnson and Hallberg, 2003; Deneff *et al.*, 2010). In this respect, the As(III) form (arsenite) can provide chemolithotrophic organisms with energy (Silver and Phung, 2005; Stolz *et al.*, 2006).

The former mine of Carnoulès, Gard (France) provides an outstanding example of such an extreme environment. The sulfurous wastes contain As-rich pyrite and the leached waters are the source of a small stream called the Reigous that contains between 50 and 350 mg l⁻¹ of soluble arsenic, mainly in the form of As(III) (Casiot *et al.*, 2003; Egal *et al.*, 2010). However, although the arsenic levels remain still high, this concentration decreases by 95% between the source of the Reigous and its confluence with the river Amous, 1.5 km downstream. This natural process of attenuation seems to be mainly due to microbial metabolism, leading to the oxidation of iron and arsenic into Fe(III) and As(V), and their co-precipitation with sulfur. In addition, laboratory experiments suggest that bacteria belonging to *Thiomonas* or *Acidithiobacillus* genera are involved in this process *in situ* (Bruneel *et al.*, 2003; Casiot *et al.*, 2003; Duquesne *et al.*, 2003; Morin *et al.*, 2003; Egal *et al.*, 2009). However, 16S-based community analyses have revealed that other genera are present in this ecosystem (Duquesne *et al.*, 2003, 2008; Bruneel *et al.*, 2003, 2006, 2008, 2011). The biological activity of these uncultured bacteria may have a significant impact on the functioning of this ecosystem. In the present study, we used a multidisciplinary approach that took advantage of the 'omics' methods, to decipher the role of microorganisms, including uncultured bacteria, in the complex metabolic processes at work in the Carnoulès acid mine drainage (AMD), an arsenic-rich ecosystem.

Materials and methods

Sampling and chemical analysis

Samples were collected in May 2007 at the station called COWG (Carnoulès Oxydizing Wetland, G) located 30 m downstream of the spring (Bruneel *et al.*, 2003). In all, 5 cm deep white sediments were collected at the bottom of the creek using a sterile tube and pooled (Global positioning system (GPS) coordinates: 44°07'01.80"N/4°00'06.90"E), while the running water (that is a thin column, <10 cm)

covering these sediments was collected in triplicate and filtered (300 ml) through sterile 0.22 µm nucleopore filters. These filters were transferred into a collection tube, frozen in liquid nitrogen and stored at -80 °C until further analysis. The main physico-chemical parameters (pH, T°, dissolved oxygen) were determined in the field and arsenic speciation, Fe(II) and sulfate analyses were performed as previously described (Bruneel *et al.*, 2008).

DNA isolation and sequencing

DNA was extracted from the cellular fraction either directly using the UltraClean Soil DNA Isolation kit (MoBio Laboratories Inc., Carlsbad, CA, USA), or after separation of microbial cells by Nycodenz gradient, using the Wizard Genomic DNA Extraction kit (Promega, Madison, WI, USA) and stored at -20 °C (Supplementary Information). Nebulized DNA fragments ranging from 3 to 5 kb were used to construct a genomic library and DNA inserts were sequenced, as previously described (Muller *et al.*, 2007), giving rise to 550 920 Sanger reads. In parallel, 281 758 DNA reads were obtained by GS-FLEX pyrosequencing using standard procedures. Both methods produced a total of 430.3 Mbp.

Clone library and phylogenetic analyses

Bacterial diversity was analyzed by cloning and sequencing PCR-amplified 16S rRNA genes (Supplementary Information). Sequences were compared with the RDP database (<http://rdp.cme.msu.edu>) (Wang *et al.*, 2007) and BLAST (Basic Local Alignment Search Tool) online searches (Altschul *et al.*, 1997). Phylogenies were constructed with the molecular evolutionary genetics analysis v4.0 program (Tamura *et al.*, 2007) using maximum composite-likelihood model and neighbor-joining algorithm. The sequences of clones CG determined in this study were submitted to the EMBL database and were given accession numbers FN391809 to FN391849.

Bioinformatics, statistical analysis and phylogenomic approach

Two-third of the metagenomic sequences were successfully organized in seven bins (Supplementary Information) and were then integrated into the MicroScope platform for the prediction of coding sequences, followed by automatic and expert annotation (Vallenet *et al.*, 2009). The mean polymorphism frequency in the population was assessed using SNIPer (Ning *et al.*, 2001). Molecular phylogenies were inferred using 27 universal marker genes chosen from a reference gene set (Ciccarelli *et al.*, 2006) (Supplementary Information). For each marker, the corresponding family of homologous genes from the HOGENOM database (Perrière *et al.*, 2000) was identified. Each family data set was

then aligned with the program MUSCLE (Edgar, 2004) and filtered using the program GBLOCKS (Castresana, 2000; Talavera and Castresana, 2007). Maximum-likelihood phylogenies were reconstructed with PhyML (v2.4.4; Guindon and Gascuel, 2003) using the Jones–Taylor–Thornton model of amino acid substitution (Jones *et al.*, 1992). The metabolic network was predicted by the ‘Pathway Tools’ software (Karp *et al.*, 2002) using MetaCyc (Caspi *et al.*, 2008) as a reference pathway database and then analyzed by a multiple factor analysis from a two-dimensional matrix (Supplementary Information). DNA sequences were submitted to GenBank and were given the following ProjectIDs: 38045—Carnoules metagenome study (top level); 41535—Carnoules metagenome study, bin 1; 41537—Carnoules metagenome study, bin 2; 41539—Carnoules metagenome study, bin 3; 41541—Carnoules metagenome study, bin 4; 41543—Carnoules metagenome study, bin 5; 41545—Carnoules metagenome study, bin 6; 41549—Carnoules metagenome study, bin 7.

Protein extraction, gel electrophoresis and mass spectrometry identification

Proteins were extracted from cells recovered using Nycodenz gradient (Supplementary Information), separated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) using 12% gradient slab gels (PROTEAN II, Bio-Rad Laboratories, Hercules, CA, USA), stained with Coomassie Brilliant Blue R-250 and *in gel* digested, as previously described (Weiss *et al.*, 2009). The resulting peptide extracts were analyzed by liquid chromatography and mass spectrometry (nanoLC-MS/MS) on a nanoACQUITY Ultra-Performance-LC (UPLC, Waters, Milford, MA, USA) coupled to SYNAPT hybrid quadrupole orthogonal acceleration time-of-flight tandem mass spectrometer (Waters). Data were analyzed using the MASCOT 2.2.0. algorithm (Matrix Science, London, UK) to search against a target-decoy protein database (Supplementary Information).

PCR and reverse transcription

Total RNA was extracted as described (González-Mendoza *et al.*, 2009) after Nycodenz gradient density separation of the bacterial community for 15 min at 4 °C from sediments stored at –80 °C. Total RNA was purified using the RNEasy Mini kit (Qiagen, Valencia, CA, USA) and digested with

DNase (Fermentas, Glen Burnie, MD, USA). The reverse transcription of *punA*, *soxC* and *cobS* transcripts was performed using SuperScript III One-Step reverse transcriptase–PCR System (Invitrogen, Carlsbad, CA, USA) and 5'-CTCCATCCGAAAAAG TGCTC-3'/5'-AAAGAGGTTTTGCTGCGGTA-3', 5'-ATCGACGGATTTCTGGATTG-3'/5'-GGTAACGCTGC CATCTAAGC-3' and 5'-TTGCTGATTGTTGCTCTG G-3'/5'-TCCAGCAGACTGGACAACAC-3' primers, respectively. A negative control was carried out without the first reverse transcription step. Amplified fragments were sequenced by Millgen and the resulting sequences were analyzed using BLAST program (Altschul *et al.*, 1997).

Bacterial culture and growth medium

Iron oxidation was tested with *Acidithiobacillus ferrooxidans* B5 strain isolated from Carnoules (Casiot *et al.*, 2003). A total of, 5×10^6 cells ml⁻¹ were grown at 30 °C for 5 days in 100:10 liquid medium (Schrader and Holmes, 1988) supplemented or not with 1 mg l⁻¹ cobalamin (Sigma, St Louis, MO, USA) and shaken at 140 r.p.m. Iron oxidation was followed by the appearance of an orange precipitate and cellular growth was estimated by colony-forming unit (CFU) counting.

Results

Phylogeny and genome reconstruction

The main physico-chemical parameters (pH, T°, dissolved oxygen, iron concentration) were determined at the sampling site: a pH value of roughly 3.5 was measured, and arsenic, iron and sulfate concentrations reached 87, 625 and 3209 mg l⁻¹, respectively, which correspond to 1.16 mM As, 11.16 mM Fe and 33.43 mM SO₄ (Table 1). This further supports a persistent contamination by arsenic and other inorganic elements (Casiot *et al.*, 2003; Egal *et al.*, 2010). To identify the microbial species present in this ecosystem, DNA was extracted from the upper zone of sediments, 16S rRNA gene sequences were amplified, cloned and sequenced as described (Bruneel *et al.*, 2008). Sequence analyses revealed a dominance of bacteria belonging to the Proteobacteria phylum (β- and γ-Proteobacteria with 38% and 23%, respectively) (Figure 1). Remaining clones belonged to Firmicutes (4%) and Acidobacteria (3%), and to a lesser extent to classes such as Spirochaetes (1%) and α-Proteobacteria (1%). This demonstrates that the microbial community at Carnoules includes a few other bacterial

Table 1 Physico-chemical characteristics of the Carnoules water during the sampling at COWG^a

PH	T (°C)	DO ^b (mg l ⁻¹)	Eh ^c (mV)	As(III) (mg l ⁻¹)	As(V) (mg l ⁻¹)	Fe(II) (mg l ⁻¹)	Fe(III) (mg l ⁻¹)	SO ₄ ²⁻ (mg l ⁻¹)
3.83	15.1	8.3	459	71 (±4)	16 (±1)	625(±30)	<DL ^d	3209 (±300)

^aCarnoules Oxydizing Wetland, G. ^bDissolved oxygen. ^cRedox potential. ^dDetection limit.

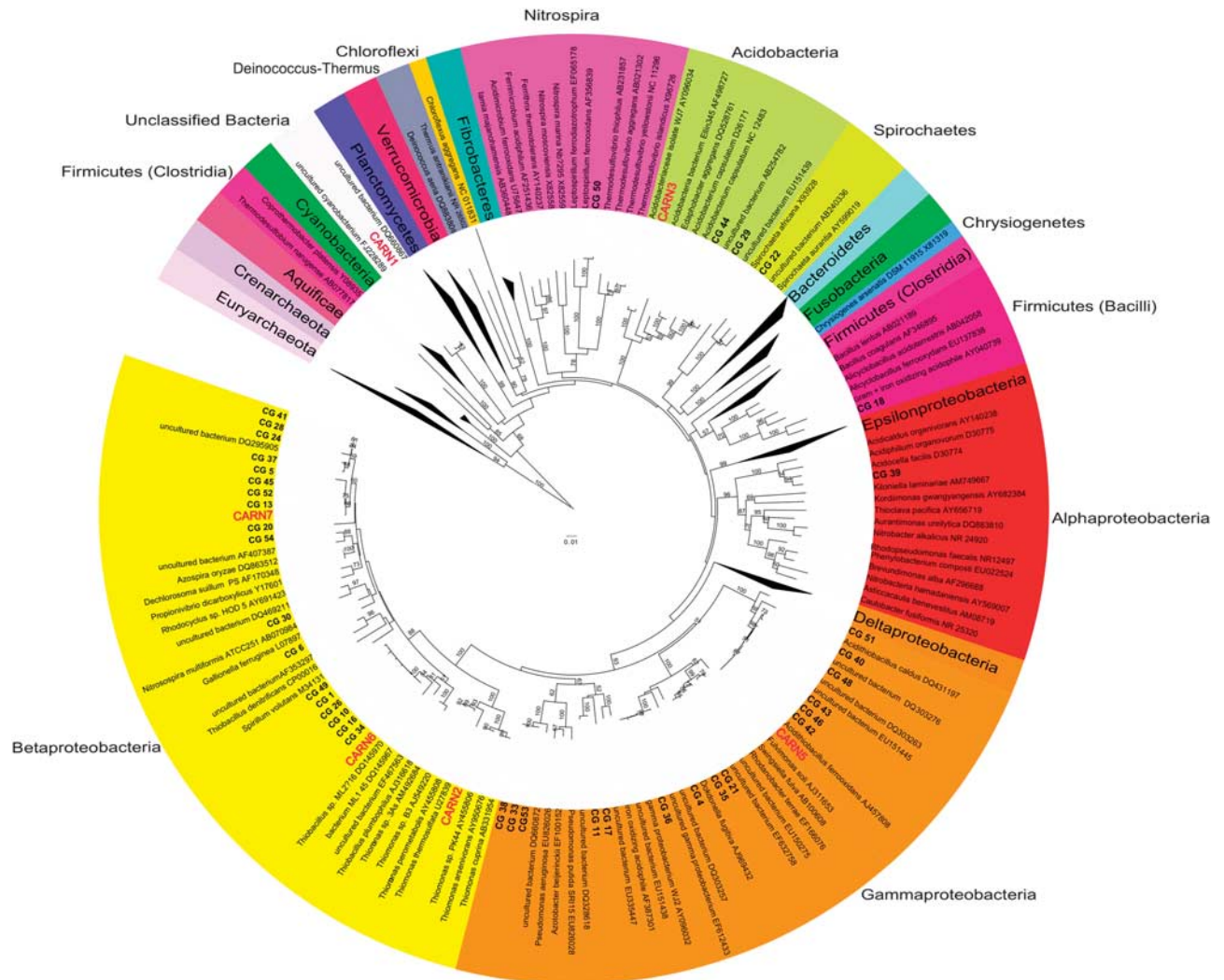


Figure 1 Phylogenetic tree representing the taxonomic affiliation of the Carnoules community microorganisms. The 16S rRNA gene sequences were obtained from DNA sediments after PCR amplification (clones CG-X) or metagenomic sequencing (CARN bins, except for CARN4, see text). A total of 759 positions in the final data set were obtained after all positions containing gaps and missing data were eliminated. The distances were computed using the maximum composite-likelihood method and the tree was inferred using the neighbor-joining method. The scale bar corresponds to 0.05 substitutions per site. Percentages of 1000 bootstrap resamplings that supported the branching orders in each analysis are shown above or near the relevant nodes. Bootstrap values are shown for branches with >50% bootstrap support.

phyla in addition to those previously identified (Bruneel *et al.*, 2006, 2011).

To decipher the possible role of microorganisms in the biotransformation of toxic elements present at the study site, total DNA extracted from sediments was next fully sequenced. These genomic data were assembled in seven major bins called CARN1–CARN7, using a combination of 16S rRNA gene sequences, guanine and cytosine content (GC%), mean coverage of the various contigs, mean polymorphism frequency and similarity to already sequenced genomes, including *At. ferrooxidans* and *Thiomonas* sp. 3As (<http://www.genoscope.cns.fr/agc/microscope/carnoulescope>). Even though we demonstrated the existence of other species (see above), the presence of numerous dominant organisms would probably not have allowed to construct

such large bins. Their size ranged from 1.5 to >4.0 Mbp and their guanine and cytosine content ranged from 52% to 65% (Table 2). DNA sequencing led to a low coverage for some contigs, in particular those of CARN3, which suggests that the size of the corresponding genome is underestimated. Finally, although the sequence assembling from ecotypes cannot be excluded, each supercontig group was considered to correspond to the genome of one single major organism (Supplementary Figure S1).

We evaluated the origin of the seven strains by a phylogenetic analysis of 16S rRNA sequences obtained from both the clone library and metagenomic sequencing (Figure 1), except for CARN4, whose missing rRNA sequence precluded its 16S-based classification. Nevertheless, at the extremity of a contig, this bin contained a 23S rRNA gene

Table 2 General features of the seven Carnoulès bins

	Carnoulès 1	Carnoulès 2	Carnoulès 3	Carnoulès 4	Carnoulès 5	Carnoulès 6	Carnoulès 7
Number of base pairs	255 8671	406 9759	156 7831	276 0983	298 4465	282 4537	270 8310
Guanine and cytosine content (GC%)	62.13	64.66	58.55	61.89	58.4	58.51	51.7
No. of predicted protein-encoding genes	2664	4302	1498	2806	3292	2851	2949
No. of predicted proteins of unknown function (%)	1267 (47.56 %)	1561 (38.29 %)	781 (52.14 %)	1359 (48.43 %)	1500 (45.57 %)	1531 (53.70 %)	1322 (44.83 %)
	<i>Number (%)</i>	<i>Number (%)</i>	<i>Number (%)</i>	<i>Number (%)</i>	<i>Number (%)</i>	<i>Number (%)</i>	<i>Number (%)</i>
<i>Cellular processes and signaling</i>							
Cell cycle control, cell division, chromosome partitioning	33 (1.24)	50 (1.16)	9 (0.06)	49 (1.75)	46 (1.4)	38 (1.33)	51 (1.73)
Cell wall/membrane/envelope biogenesis	139 (5.22)	202 (4.7)	70 (4.67)	177 (6.31)	193 (5.86)	153 (5.37)	137 (4.65)
Cell motility	60 (2.25)	112 (2.6)	29 (1.94)	61 (2.17)	53 (1.61)	63 (2.21)	97 (3.29)
Post-translational modification, protein turnover, chaperones	101 (3.79)	177 (4.1)	46 (3.07)	119 (4.24)	136 (4.13)	88 (3.09)	118 (4)
Signal transduction mechanisms	121 (4.54)	230 (5.35)	55 (3.67)	121 (4.31)	100 (3.04)	105 (3.68)	156 (5.29)
Intracellular trafficking, secretion and vesicular transport	62 (2.33)	120 (2.79)	52 (3.47)	79 (2.82)	131 (3.98)	89 (3.12)	104 (3.56)
Defense mechanisms	56 (2.1)	107 (2.49)	36 (2.4)	66 (2.35)	74 (2.25)	59 (2.07)	62 (2.1)
Extracellular structures	1 (0.04)	1 (0.02)	1 (0.07)	3 (0.11)	2 (0.06)	1 (0.04)	1 (0.03)
RNA processing and modification		1 (0.02)			1 (0.03)		1 (0.03)
<i>Information storage and processing</i>							
Chromatin structure and dynamics	1 (0.04)	1 (0.02)		1 (0.04)	1 (0.03)		1 (0.03)
Translation, ribosomal structure and biogenesis	161 (6.04)	177 (4.1)	73 (4.87)	166 (5.92)	161 (4.89)	75 (2.63)	103 (3.49)
Transcription	120 (4.5)	230 (5.35)	66 (4.41)	148 (5.27)	141 (4.28)	109 (3.82)	143 (4.85)
Replication, recombination and repair	202 (7.58)	372 (8.65)	99 (6.61)	182 (6.49)	272 (8.26)	284 (9.26)	307 (10.41)
<i>Metabolism</i>							
Energy production and conversion	161 (6.04)	335 (7.79)	64 (4.27)	177 (6.31)	199 (6.05)	149 (5.23)	172 (5.83)
Amino acid transport and metabolism	313 (11.75)	488 (11.34)	120 (8.01)	309 (11.01)	277 (8.41)	214 (7.51)	207 (7.02)
Nucleotide transport and metabolism	69 (2.59)	94 (2.19)	25 (1.67)	61 (2.17)	73 (2.22)	63 (2.21)	63 (2.14)
Carbohydrate transport and metabolism	139 (5.22)	232 (5.39)	101 (6.74)	173 (6.17)	171 (5.19)	107 (3.75)	125 (4.24)
Coenzyme transport and metabolism	108 (4.05)	152 (3.53)	42 (2.8)	110 (3.92)	113 (3.43)	72 (2.53)	102 (3.46)
Lipid transport and metabolism	112 (4.20)	164 (3.81)	36 (2.4)	112 (3.99)	75 (2.28)	63 (2.21)	65 (2.2)
Inorganic ion transport and metabolism	117 (6.64)	322 (7.48)	78 (5.21)	183 (6.52)	231 (7.02)	136 (4.77)	158 (5.36)
Secondary metabolite biosynthesis, transport and catabolism	64 (2.40)	132 (3.07)	29 (1.94)	76 (2.71)	68 (2.07)	36 (1.26)	44 (1.49)

sequence, which was similar to that of CARN1. We also performed a phylogenomic analysis of the bins using 27 universal marker genes (individually or combined) sufficiently conserved to build a tree of life (Ciccarelli *et al.*, 2006), except for CARN6 where those markers were not found. Precisely, the candidates CARN2, CARN3 and CARN5 were affiliated to the β -Proteobacterium *Thiomonas* genus, the Acidobacteria clade and to the *At. ferrooxidans* γ -Proteobacterium, respectively (Supplementary Table S1). These phylogenetic results were supported by an analysis of gene order conservation known to be correlated with evolutive distances (Huynen and Bork, 1998). Finally, classification using 16S rRNA gene and RDP classifier indicated with confidence that CARN6 and CARN7 belong to the β -Proteobacteria class. In addition, 16S rRNA gene BLAST searches against NCBI nr database showed that the closest cultivated microorganisms to these two bins belong to the *Thiobacillus* and *Gallionella* genera, respectively (Supplementary Table S1).

Remarkably, the candidates corresponding to the CARN1 and CARN4 bins showed an important phylogenetic relationship and may represent two subpopulations according to the polymorphism distribution along their genome (Supplementary Figure S2). RDP classifier indicated that those bins did not correspond to any known taxonomic phylum and, according to the metabolic properties identified in the present study (see below), this new genus was herein named '*Candidatus Fodinabacter communicans*' (from fodina, mine and communicare, share), in accordance with the recommendations for incompletely characterized microorganisms (Murray and Stackebrandt, 1995).

Metabolic potentialities of and in situ expressed activities by the seven dominant bacteria

To determine the major metabolic potentialities of each bin, an in depth *in silico* analysis of their gene content was performed. In parallel, we investigated the functioning of the bacterial community by analyzing the proteins synthesized *in situ* by all bins. This metaproteomic approach allowed the reliable identification of >500 unique proteins belonging to various functional classes, for example membrane and transport, stress response and energy metabolism (Supplementary Tables S2ab). The experimental protein pattern was representative of the theoretical profile inferred from metagenomic data and the number of proteins identified from each bin was in agreement with the level of sequence coverage, which may reflect species abundance. Indeed, while a few proteins originated from CARN3 and CARN6, CARN1 and CARN4 bins were shown to express 70% of the identified proteins and emerged as abundant and very active in the ecosystem (Figure 2). Due to a low homology ranging from 20% to 30% with proteins present in databases, about 25% of the proteins synthesized by these last two bins were annotated as hypothetical

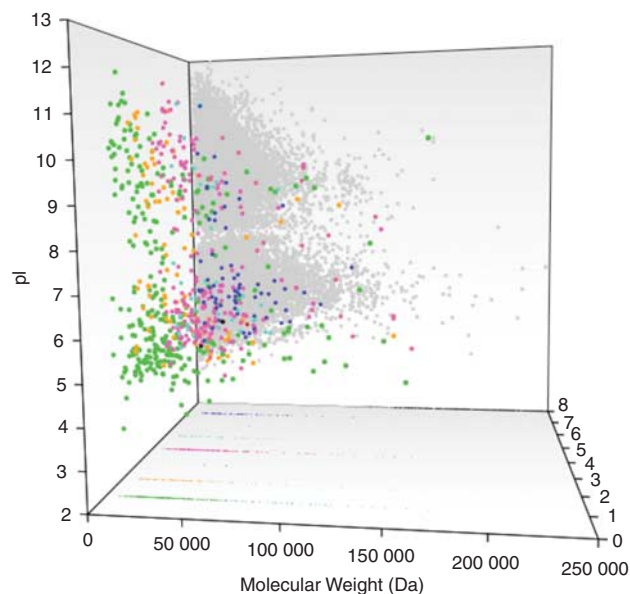


Figure 2 Experimental metaproteomic pattern obtained by MS/MS identification of the proteins expressed *in situ* (CARN1, green; CARN2, orange; CARN3, black; CARN4, pink; CARN5, light blue; CARN6, brown and CARN7, blue). As a background, the theoretical distribution predicted from metagenomic data is represented in gray.

(Supplementary Table S2b). One-third might, however, represent exported hydrolases, membrane transport proteins or sensors and modulators of chemotaxis and motility (data not shown).

The genome of each bin was shown to contain at least one *ars* operon encoding arsenite efflux pumps and arsenate reductases. These genes are involved in arsenic resistance, and the presence of the corresponding proteins was demonstrated in protein extracts (Supplementary Table S2b). An *arsM* gene coding for an arsenite S-adenosylmethyltransferase was also identified in the genome of CARN6, in agreement with the presence of methylated forms of arsenic in Carnoulès sediments, that is monomethylarsonate and dimethylarsinate (7.10^{-4} and 3.10^{-4} mg, respectively, per mg dry weight). A few *arsM* homologous genes were also identified in unassembled sequences, which further supports the existence of arsenic methylation at the study site (data not shown). Finally, because of the structural similarity between As(V) and phosphate, arsenic metabolizing strains may preferentially transport phosphate via the specific Pst phosphate transport system rather than the Pit general transport mechanism, in order to reduce the entry of As(V). Accordingly, no Pit protein was identified in metaproteomic data, while several Pst proteins were detected (Supplementary Table S2b).

The microbial response to arsenic is known to result in various biological effects, including oxidative stress, DNA damage, exopolysaccharide synthesis and biofilm formation (Beyersmann and Hartwig, 2008; Marchal *et al.*, 2010). Metaproteomic data showed that Carnoulès strains indeed express

functions protecting against general and oxidative stress, including superoxide dismutase or thioredoxin, and chaperones such as DnaK or GroEL, which have previously been detected in AMD (Ram *et al.*, 2005; Bruneel *et al.*, 2011). Several proteins such as RecA and DPS involved in DNA recombination and repair were also identified. In addition, several flagellar proteins synthesized by CARN1 and CARN4 were identified in protein extracts (Supplementary Table S2b), as well as a protein involved in the synthesis of type 4 pilus expressed by *Thiomonas* sp. and known to have a role in twitching motility and adhesion (Li *et al.*, 2010). The corresponding operon, found in the genome of this bin, was also present in the genomes of CARN5 and CARN7. These adaptive processes typically depend on multiple regulatory mechanisms that allow bacteria to respond to a wide panel of stimuli. This was supported by the proteomic identification of proteins such as the nucleoid-associated proteins HU and H-NS, which have a major role in both the structure and the function of chromosomal DNA (Tendeng and Bertin, 2003; Grove, 2010). Several regulators, including alternative σ factors such as σ^H and regulatory proteins belonging to the two-component systems or those involved in cell-to-cell signaling, were also expressed by various strains (Supplementary Tables S2ab), in agreement with the presence in their genome of multiple genes coding for proteins involved in stress response.

In a partly oligotrophic environment, such as the Carnoulès AMD, various metabolic reactions may be needed for autotrophic metabolism. In this respect, CARN5 and CARN7 may fix nitrogen by the expression of *nif* genes, which encode a nitrogenase (Dixon and Kahn, 2004), although this oxygen-sensitive enzyme was not detected in protein extracts. In addition, carbon fixation depends on proteins such as ribulose 1,5-biphosphate carboxylase/oxygenase (involved in Calvin cycle) (Badger and Bek, 2008), carboxysome structural proteins and carbon monoxide dehydrogenase (involved in acetyl-coenzyme A synthesis). These enzymes were identified in protein extracts, in agreement with the presence of the corresponding genes in *Thiomonas* sp. (CARN2), *Acidithiobacillus* sp. (CARN5) or *Gallionella*-related (CARN7) bins. These observations suggest that these three strains have a key role as organic compound primary producers. Nevertheless, other microorganisms may preferentially metabolize and recycle some organic compounds released by others. For example, the CARN6 bin carries the cellulase-encoding gene *bczS* and the α -amylase *amyM* gene, suggesting that this strain is able to metabolize complex carbohydrates. Similarly, several enzymes required for amino acid transport and metabolism were identified in the bins lacking the carbon and nitrogen fixation genetic determinants, in agreement with a mixotrophic or organotrophic metabolism. In particular, the CARN1 and CARN4 bins of '*Candidatus* Fodinabacter

communificans' contain genes such as *liv* and *opp* involved in branched amino acid and oligopeptide transport, respectively. In addition, both strains were shown to encode multiple peptidases, such as the metallopeptidase M61 and the serine protease S41 (Supplementary Table S2b), and to express the *punA* gene (Supplementary Figure S3), which suggests that they can use purine as a carbon source (Schuch *et al.*, 1999). The CARN1 and CARN4 bins also expressed Gcv proteins of the glycine cleavage system that catalyzes the degradation of glycine and PdxS proteins involved in active vitamin B6 biosynthesis, which has an important role in amino acid metabolism (Fitzpatrick *et al.*, 2007).

A wide diversity of bioenergetic electron chains may be needed to accommodate the presence in this ecosystem of various electron acceptors such as O₂ or Fe(III) oxides and electron donors such as Fe(II) or sulfides. Accordingly, several terminal oxidases might be operative in all strains of the Carnoulès community, for example the cytochrome oxidase *cox* and *cta* operons and the *cyo* and *cyd* operons encoding quinol oxidases. Some operons involved in anaerobic respiration were also identified, in particular *ntr*, *nar* and *nas* in CARN2 and CARN7 bins, which suggests the possible recourse to anaerobic nitrate and nitrite metabolism. Nevertheless, the identification of several nicotinamide adenine dinucleotide-ubiquinone and cytochrome *c* oxidases in protein extracts supports the existence of an active aerobic metabolism at the sampling site (Supplementary Table S2b). The oxidation of H₂ to protons may be a source of energy due to the presence in CARN2, CARN5 and CARN6 of several hydrogenase-encoding genes, in particular *hox* and *hup* operons (Friedrich and Schwartz, 1993). In addition, possible inorganic electron donors may comprise sulfur compounds. This is suggested by the presence and the expression (Supplementary Figure S3) in *Thiomonas* sp. (CARN2), *Acidithiobacillus* sp. (CARN5) or *Gallionella*-related sp. (CARN7) of genes such as *hdr*, *sor*, *sox*, *tetH* or *sqr* involved in the oxidation of reduced inorganic sulfur compounds, for example sulfide, sulfur, sulfite, thiosulfate or tetrathionate. Finally, the CARN7 bin harbors a *dsrABEFHCMKLJOPN* operon. The DsrA sequence showed only 39% identity with that of the sulfato-reducing *Desulfovibrio vulgaris*, but 72% with the amino acid sequence of the sulfoxidizing *Thiobacillus denitrificans*. This suggests that the *Gallionella*-related bin (CARN7) is able, like *T. denitrificans*, to use sulfur compounds as electron donors in its energy metabolism, as suggested in *Gallionella ferruginea* (Lutters-Czekalla, 1990).

Additionally, arsenite may constitute a possible inorganic electron donor and three arsenite oxidases with a nucleotide and amino acid sequence similarity >95% and 98%, respectively, and possibly involved in the bioenergetic transformation of As(III) to As(V) were shown to be expressed by *Thiomonas* sp. (CARN2). Although we cannot rule

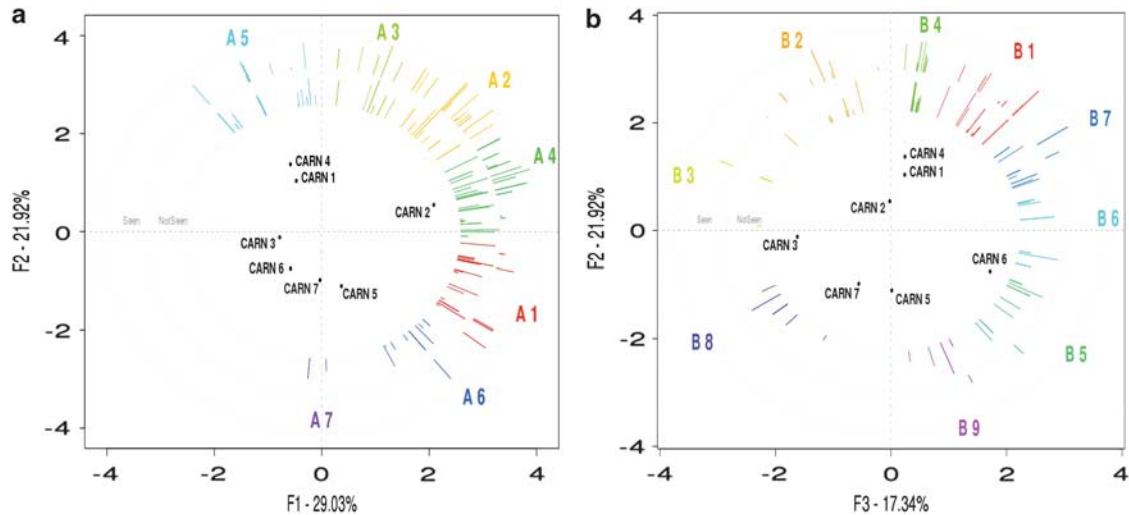


Figure 3 Multiple factorial analysis of the seven Carnoulès bins performed on a two-dimensional matrix combining bins and enzymatic reactions, respectively. To highlight possible metabolic distinctions between bins, three axes (F1 to F3) capturing the highest dispersion in the resulting dot cloud were selected; they represent more than half the total dispersion. Colored lines represent the projection of vectors corresponding to the enzymatic reaction frequencies, the external disk differentiating reactions whose enzymes were identified in the metaproteomic data. The reaction vectors were then hierarchically clustered, which led to 7 (A1 to A7) and 9 (B1 to B9) classes (indicated by the colors) for the first (a) and second (b) factorial planes, respectively. For each cluster, the corresponding reactions are listed in Supplementary Tables S3ab. For example, several reactions of Cluster A4 are concerned with energy and inorganic nutrient metabolism and in arsenic detoxification. As they are in the same quadrant of the plot as CARN2, these reactions are thus mostly specific to this bin. Similarly, Cluster A1 groups contain reactions common to CARN2 and CARN5 (Calvin–Benson–Bascham cycle and urea degradation pathways); Cluster A5 contains CARN6-specific reactions (cellulose metabolism) and Clusters A6, B1 and B4 include reactions in CARN1 and CARN4 (lysine fermentation other amino acid or nucleoside degradation pathways and cobalamin biosynthesis).

out that those sequences may originate from various ecotypes, the presence of multiple copies of *aox* operons has been recently demonstrated in *Thiomonas* strains (Arsène-Ploetze *et al.*, 2010). Arsenate respiratory reductase-encoding gene, that is *arrA*, which allows anaerobic respiration of As(V), was neither identified in any of the seven bins nor in unassembled sequences, in agreement with the aerobic conditions prevailing in the ecosystem under study (Table 1). Importantly, the *Acidithiobacillus* sp. bin (CARN5) expressed the sole RusA protein (Supplementary Table S2b), which is involved in electron transport with iron used as an energy source. This further supports a major role of *Acidithiobacillus* sp. bin (CARN5) in the iron oxidation observed at Carnoulès (Bruneel *et al.*, 2003; Duquesne *et al.*, 2003; Morin *et al.*, 2003). Interestingly, two cytochromes *c* synthesized by CARN7 were also identified in protein extracts (Supplementary Table S2b). They showed 61% and 43% amino acid similarity, respectively, with *Cyc1* and *Cyc2* of *At. ferrooxidans*. *Cyc1* belongs to the *c4* family and *Cyc2* is an outer membrane cytochrome *c* proposed to receive electrons directly from ferrous iron (Yarzabal *et al.*, 2002). Even though the other genes of the *rus* operon known to be involved in the electron transfer between Fe(II) and oxygen in *At. ferrooxidans* (Appia-Ayme *et al.*, 1999) have not been detected in CARN7, it is tempting to hypothesize that this organism may be involved in iron cycling in the AMD under study.

Synecologic interactions within the microbial community

Autotrophic, mixotrophic or heterotrophic metabolisms present at Carnoulès suggest the existence of metabolic and nutrient exchanges that may be of prime importance inside the microbial community, revealing the importance of syntrophic relationships. To support and extend these observations, a multiple factor analysis of both metagenomic and metaproteomic data was performed on a two-dimensional matrix combining bins and enzymatic reactions. The most discriminant axes, that is capturing the highest dispersion in the resulting dot cloud, suggest an important and specific role for each organism inside the Carnoulès community. Indeed, factorial planes segregated CARN2 from the other bins (Figure 3a) and separated CARN1/CARN4 from CARN5/CARN7. In addition, CARN3 was opposed to CARN6, revealing CARN6-specific reactions (Figure 3b). No such correlation was observed between our data and those from the AMD biofilm (Tyson *et al.*, 2004; Deneff *et al.*, 2010), which further supports the marked difference between these two ecosystems (Supplementary Figure S4). From the variable classification results, several clusters of reactions were then associated to each bin group (Figure 3). These include, for example, Cluster A4, which is linked to CARN2 (energy metabolism, inorganic nutrient metabolism and arsenic detoxification) (Supplementary Table S3a), and Cluster A1, which groups reactions

common to CARN2 and CARN5 (Calvin–Benson–Bassham cycle and urea degradation pathways). In addition, Cluster A5 contains CARN6-specific reactions, such as those involved in cellulose metabolism, while Clusters A6, B1 and B4 include reactions related to lysine fermentation and to other amino acid or nucleoside degradation pathways in CARN1 and CARN4, and also gather reactions related to cobalamin biosynthesis (Supplementary Table S3b). Indeed, the CARN1 and CARN4 bins carry the *cobSTU* operon involved in cobalamin (vitamin B12) synthesis (Escalante-Semerena, 2007), which was shown by reverse transcriptase–PCR to be expressed *in situ* (Supplementary Figure S3). Remarkably, iron oxidation catalyzed by *At. ferrooxidans* was strongly increased in the presence of cobalamin (Supplementary Figure S5). These data suggest that this vitamin synthesized by CARN1 and CARN4 may be used by CARN5, leading to an increase in iron oxidation.

Discussion

In the last few years, a huge amount of genomic sequences have been published in databases,

including a complete characterization of several bacteria metabolizing arsenic (Muller *et al.*, 2007; Arsène-Ploetze *et al.*, 2010). In microbial ecology, the major challenge remains, however, to determine more precisely the role of microorganisms and their relationships with other members of the microbial community that result in an efficient functioning of the ecosystem. In this respect, metagenomic approaches based on high-throughput technologies may be of great interest in synecology, since they allow to investigate the role of complex microbe consortia as a whole, including that of uncultured microorganisms in the processes taking place *in situ*. In the present study, we performed an in-depth analysis of descriptive and functional genomic data that allowed us to identify the dominant diversity present in an arsenic-contaminated ecosystem (Figure 4). Some microorganisms belong to bacterial genera already characterized, for example *Thiomonas* (CARN2) and *Acidithiobacillus* (CARN5). Others correspond, however, to uncultured microorganisms such as the one related to *Gallionella* sp. (CARN7) and the first representative of a novel bacterial phylum, here named ‘*Candidatus Fodinabacter* communicans’ (CARN1 and CARN4). Interestingly, only

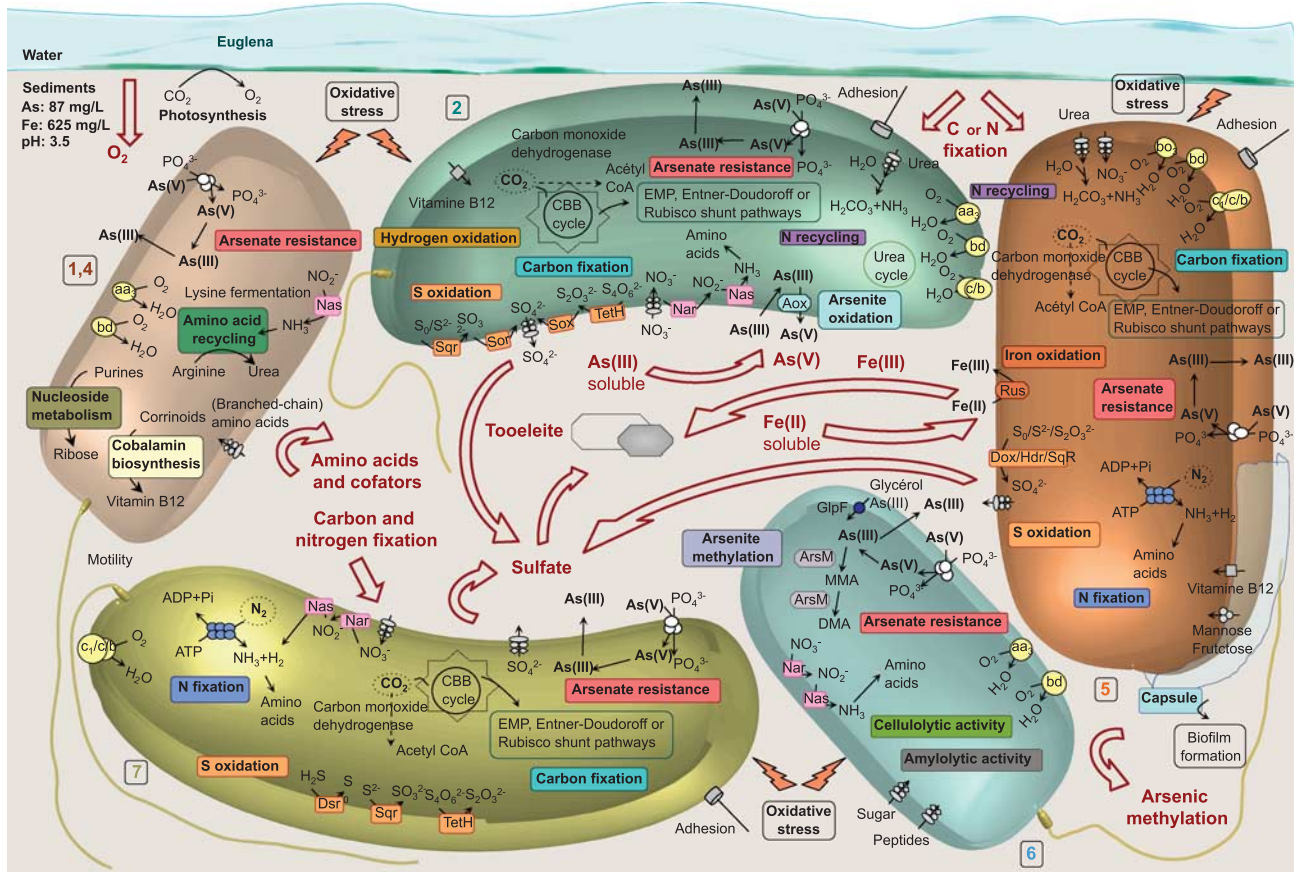


Figure 4 Model of the Carnoulès bacterial community highlighting the major functions identified by metagenome sequencing or metaproteome characterization. These activities include carbon and nitrogen fixation, energy metabolism, flagellum and capsule biosynthesis, amino acid transport and degradation, detoxification and stress response, arsenic and iron metabolism. The possible interactions between these microorganisms or with other chemical or biological compounds present on the study site are indicated by arrows. CARN bins are numbered from 1 to 7.

two 16S rRNA sequences (97% similarity) similar to CARN1 are present in databases and originate from other acid mine environments, suggesting that uncultivated bacteria of this genus are widespread in such ecosystems. Finally, the presence at the sampling site of two subpopulations of 'Candidatus Fodinabacter communicans' (Supplementary Figures S1 and 2) is reminiscent to what has been observed in other environmental contexts (Simmons *et al.*, 2008; Wilmes *et al.*, 2008). Such ecotypes, which may reflect an important genetic variability resulting from gene gain or loss, has been recently described in *Thiomonas* strains (Arsène-Ploetze *et al.*, 2010).

Metagenomic and metaproteomic data emphasized the diversity of mechanisms that may be important *in situ*, allowing the seven dominant bacteria to adapt their metabolic activities to changes in environmental conditions (Figure 4). One example concerns the ability of microorganisms to live in organized surface communities called biofilms, where aggregated cells are embedded in an exopolymeric matrix. Such a lifestyle confers on them a higher resistance to various environmental stresses (Harrison *et al.*, 2007) and favors the physical interactions between the cells (Davey and O'toole, 2000). In addition, flagellum biosynthesis and motility are often known to have a role in the first steps of biofilm formation (Soutourina and Bertin, 2003). Metagenomics revealed that most bins contain multiple flagellar and pili genes and metaproteomics allowed us to identify the corresponding proteins in CARN1/CARN4 and in CARN2, respectively (Supplementary Table S2b). While no flagellar operon was identified in *Acidithiobacillus* sp. (CARN5), as it may be observed in already sequenced related bacteria (<http://www.genoscope.cns.fr/agc/microscope/carnoulescope>), this bin contains several genes, such as *galE* and *pgm* involved in capsular biosynthesis in *At. ferrooxidans* (Barreto *et al.*, 2005), suggesting a role of this strain in the formation of a biofilm.

More importantly, combined with multiple factorial analysis of the data, our observations demonstrated both the metabolic specificity and partnerships that may exist inside an arsenic-rich environment at the benefit of the microbial community as a whole. These processes include the fixation of inorganic carbon and nitrogen by several strains, in particular those belonging to the *Thiomonas* (CARN2), *Acidithiobacillus* (CARN5) and *Gallionella*-related (CARN7) genera. Indeed, these strains were shown to synthesize proteins such as the Rubisco, that is ribulose-1,5-biphosphate carboxylase oxygenase. This enzyme of the Calvin cycle catalyzes the first step of CO₂ fixation. In addition, the genome of CARN5 and CARN7 also contains the *nif* genes, which may encode a nitrogenase. This further supports a key role of these organisms in nitrogen cycle inside the Carnoules community. Their autotrophic metabolisms may be essential to reach an equilibrium between auto- and heterotrophy

providing other partners with organic nutrients. In this respect, the presence of *fucP* and *exuT* genes coding for L-fucose and hexuronate transporters, respectively, in CARN3 may support, at least in part, the carbon requirements of this organism. In addition, CARN6 carries the cellulase-encoding gene *bczS* and the α -amylase *amyM* gene, suggesting its ability to metabolize more complex carbohydrates.

Remarkably, our results highlighted the role of several members of the Carnoules community in the recycling of both mineral and organic resources, such as arsenic, iron, sulfur, urea, vitamins, nucleosides and amino acids (Figure 4). In this respect, the arsenite oxidase activity expressed by *Thiomonas* sp. (CARN2), associated with metabolisms such as iron oxidation and sulfur oxidation by *Acidithiobacillus* sp. (CARN5) and the *Gallionella*-related strain (CARN7), seems to be of prime importance in the co-precipitation of these inorganic elements, leading to the partial but efficient attenuation of this arsenic-contaminated ecosystem (Bruneel *et al.*, 2003; Casiot *et al.*, 2003; Duquesne *et al.*, 2003; Egal *et al.*, 2009, 2010). Other microorganisms, in particular 'Candidatus Fodinabacter communicans,' may recycle organic compounds released by others or provide them with essential cofactors, improving their biomineralization activities. First, CARN1 and CARN4 bins expressed genes involved in branched-chain amino acid and oligopeptide transport. Second, and unlike other Carnoules bins, their genome carries *drm* and *punA* genes, suggesting that they can use purine as a carbon source (Schuch *et al.*, 1999). Third, genes involved in amino acid fermentation were also identified in both bins, for example those of the lysine pathway, which converts lysine into butyrate, acetate and ammonia (Kreimeyer *et al.*, 2007). Fourth, the CARN1 and CARN4 bins of 'Candidatus Fodinabacter communicans' carry the arginase-encoding gene *rocF* involved in urea biosynthesis, while a complete urease-encoding *ure* operon was present in *Thiomonas* sp. bin (CARN2). Fifth, genes involved in cobalamin biosynthesis, including the *cobSTU* operon, were identified in both CARN1 and CARN4 bins, while BtuC cobalamin transporters are possibly encoded by other bins, such as *Thiomonas* (CARN2) and *Acidithiobacillus* sp. (CARN5), as it may be observed in already sequenced related bacteria (<http://www.genoscope.cns.fr/agc/microscope/carnoulescope>). In *Rhodopseudomonas palustris* strain TIE-1, a CobS-like protein has been previously shown to be involved in iron oxidation, but the mechanism remains unclear (Jiao *et al.*, 2005). Interestingly, we showed that cobalamin strongly activates iron oxidation in *At. ferrooxidans*, the first step in the natural remediation observed at the study site. Such a metabolic cooperation may thus be of great importance in the natural biomineralization observed at Carnoules. Finally, cobalamin biosynthesis may also be useful to eukaryotes, in particular *Euglena* sp. present at the study site (Casiot *et al.*, 2004) whose cell cycle is

known to require this cofactor (Olaveson and Stokes, 1989; Bré *et al.*, 1981).

Taken together, our data provide evidence that at least seven bacteria are involved in the functioning of the AMD ecosystem under study. In particular, our observations support the existence of multiple metabolic cooperations between the Carnoulès microorganisms (Figure 4). They also highlight an indirect but crucial role of the first representative of a novel and uncultured bacterial phylum, that is 'Candidatus Fodinabacter communicans,' in the biomineralization processes in this arsenic-rich ecosystem. In the future, descriptive and functional genomic approaches such as those presented here should give an integrated view of the biological objects present in any environment, their relationships, their role in the nutrient biogeochemical cycles and their possible use in the development of novel methods of bioremediation. More routinely used, such strategies will lead to important advances in microbial ecology, revealing what was recently regarded as impossible to explore.

Acknowledgements

AH-S, JC-A and DH were supported by a grant from the French Ministry of Education and Research and MM and FG-C by a grant from the Agence Nationale de la Recherche (ANR). Financial support came from the Université de Strasbourg (UdS), the Consortium National de Recherche en Génomique (CNRG), the Centre National de la Recherche Scientifique (CNRS) and the Agence Nationale de la Recherche (ANR) in the frame of the RARE and MicroScope projects. This work was done in the frame of the 'Groupement de Recherche—Métabolisme de l'Arsenic chez les Microorganismes (GDR2909-CNRS)' (<http://gdr2909.alsace.cnrs.fr/>).

References

- Allen EE, Banfield JF. (2005). Community genomics in microbial ecology and evolution. *Nat Rev Microbiol* **3**: 489–498.
- Altschul S, Madden T, Schaffer A, Zhang J, Zhang Z, Miller W *et al.* (1997). Gapped blast and psi-blast: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389–3402.
- Appia-Ayme C, Guiliani N, Ratouchniak J, Bonnefoy V. (1999). Characterization of an operon encoding two *c*-type cytochromes, an *aa3*-type cytochrome oxidase, and rusticyanin in *Thiobacillus ferrooxidans* ATCC 33020. *Appl Environ Microbiol* **65**: 4781–4787.
- 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.
- Badger MR, Bek EJ. (2008). Multiple Rubisco forms in proteobacteria: their functional significance in relation to CO₂ acquisition by the CBB cycle. *J Exper Bot* **59**: 1525–1541.
- Barreto M, Jedlicki E, Holmes DS. (2005). Identification of a gene cluster for the formation of extracellular polysaccharide precursors in the chemolithoautotroph *Acidithiobacillus ferrooxidans*. *Appl Environ Microbiol* **71**: 2902–2909.
- Bertin PN, Médigue C, Normand P. (2008). Advances in environmental genomics: towards an integrated view of micro-organisms and ecosystems. *Microbiology* **154**: 347–359.
- Beyersmann D, Hartwig A. (2008). Carcinogenic metal compounds: recent insight into molecular and cellular mechanisms. *Arch Toxicol* **82**: 493–512.
- Bré MH, El Ferjani E, Lefort-Tran M. (1981). Sequential protein-dependent steps in the cell cycle initiation and completion of division in vitamin B12 replenished *Euglena gracilis*. *Protoplasma* **108**: 301–318.
- Bruneel O, 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 O, Pascault N, Egal M, Bancon-Montigny C, Goñi-Urriza MS, 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 JC, 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 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*; e-pub ahead of print 12 February 2011.
- Casiot C, Bruneel O, Personné JC, Leblanc M, Elbaz-Poulichet F. (2004). Arsenic oxidation and bioaccumulation by the acidophilic protozoan, *Euglena mutabilis*, in acid mine drainage (Carnoules, France). *Sci Total Environ* **320**: 259–267.
- Casiot C, Morin G, Juillot F, Bruneel O, Personné JC, Leblanc M *et al.* (2003). Bacterial immobilization and oxidation of arsenic in acid mine drainage (Carnoulès creek, France). *Water Res* **37**: 2929–2936.
- Caspi R, Foerster H, Fulcher CA, Kaipa P, Krummenacker M, Latendresse M *et al.* (2008). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res* **36**: D623–D631.
- Castresana J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* **17**: 540–552.
- Ciccarelli FD, Doerks T, von Mering C, Creevey CJ, Snel B, Bork P. (2006). Toward automatic reconstruction of a highly resolved tree of life. *Science* **311**: 1283–1287.
- Davey ME, O'toole GA. (2000). Microbial biofilms: from ecology to molecular genetics. *Microbiol Mol Biol Rev* **64**: 847–867.
- Denef VJ, Mueller RS, Banfield JF. (2010). AMD biofilms: using model communities to study microbial evolution and ecological complexity in nature. *ISME J* **4**: 599–610.
- Dixon R, Kahn D. (2004). Genetic regulation of biological nitrogen fixation. *Na Rev Microbiol* **2**: 621–631.
- Duquesne K, Lebrun S, Casiot C, Bruneel O, Personné JC, Leblanc M *et al.* (2003). Immobilization of arsenite and ferric iron by *Acidithiobacillus ferrooxidans* and its relevance in acid mine drainage. *Appl Env Microbiol* **69**: 6165–6173.
- Duquesne K, Lieutaud A, Ratouchniak J, Muller D, Letto MC, 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.
- Edgar RC. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**: 1792–1797.
- Egal M, Casiot C, Morin G, Elbaz-Poulichet F, Cordier MA, Bruneel O. (2010). An updated insight into the natural attenuation of As concentrations 3 in Reigous Creek (southern France). *Appl Geochem* **25**: 1949–1957.
- Egal M, Casiot C, Morin G, Parmentier M, Bruneel O, Lebrun S et al. (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.
- Escalante-Semerena JC. (2007). Conversion of cobinamide into adenosylcobamide in Bacteria and Archaea. *J Bacteriol* **189**: 4555–4560.
- Falkowski PG, Fenchel T, Delong EF. (2008). The microbial engines that drive Earth's biogeochemical cycles. *Science* **320**: 1034–1039.
- Fitzpatrick TB, Amrhein N, Kappes B, Macheroux P, Tews I, Raschle T. (2007). Two independent routes of de novo vitamin B6 biosynthesis: not that different after all. *Biochem J* **407**: 1–13.
- Friedrich B, Schwartz E. (1993). Molecular biology of hydrogen utilization in aerobic chemolithotrophs. *Ann Rev Microbiol* **47**: 351–383.
- González-Mendoza D, Morales-Trejo A, Brito-Vera H. (2009). A rapid and efficient method for isolation of total RNA from *Euglena gracilis* (Euglenoidea). *Genet Mol Res* **8**: 482–486.
- Grove A. (2010). Functional evolution of bacterial histone-like HU proteins. *Curr Issues Mol Biol* **13**: 1–12.
- Guindon S, Gascuel O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* **52**: 696–704.
- Haferburg G, Kothe E. (2007). Microbes and metals: interactions in the environment. *J Basic Microbiol* **47**: 453–467.
- Harrison JJ, Ceri H, Turner RJ. (2007). Multimetal resistance and tolerance in microbial biofilms. *Nat Rev Microb* **5**: 928–938.
- Huynen MA, Bork P. (1998). Measuring genome evolution. *Proc Natl Acad Sci USA* **95**: 5849–5856.
- Jiao Y, Kappler A, Croal LR, Newman DK. (2005). Isolation and characterization of a genetically tractable photoautotrophic Fe(II)-oxidizing bacterium, *Rhodospseudomonas palustris* strain TIE-1. *Appl Environ Microbiol* **71**: 4487–4496.
- Johnson DB, Hallberg KB. (2003). The microbiology of acidic mine waters. *Res Microbiol* **154**: 466–473.
- Jones DT, Taylor WR, Thornton JM. (1992). The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci* **8**: 275–282.
- Karp P, Paley S, Romero P. (2002). The pathway tools software. *Bioinformatics* **18**: S1–S8.
- Konopka A. (2009). What is microbial community ecology? *ISME J* **3**: 1223–1230.
- Kreimeyer A, Perret A, Lechaplais C, Vallenet D, Médigue C, Salanoubat M et al. (2007). Identification of the last unknown genes in the fermentation pathway of lysine. *J Biol Chem* **282**: 7191–7197.
- Li Y-Q, Wan D-S, Huang S-S, Leng F-F, Yan L, Ni Y-Q et al. (2010). Type IV Pili of *Acidithiobacillus ferrooxidans* are necessary for sliding, twitching motility, and adherence. *Curr Microbiol* **60**: 17–24.
- Lutters-Czekalla S. (1990). Lithoautotrophic growth of the iron bacterium *Gallionella ferruginea* with thio-sulfate or sulfide as energy source. *Arch Microbiol* **154**: 417–421.
- 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.
- Morin G, Juillot F, Casiot C, Bruneel O, Personné JC, 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 environment. *PLoS Genet* **3**: e53.
- Murray RGE, Stackebrandt E. (1995). Taxonomic note: implementation of the provisional status *Candidatus* for incompletely described procaryotes. *Int J Syst Bacteriol* **45**: 186–187.
- Ning Z, Cox AJ, Mullikin JC. (2001). SSAHA: a fast search method for large DNA databases. *Genome Res* **11**: 1725–1729.
- Olaveson MM, Stokes PM. (1989). Responses of the acidophilic alga *Euglena mutabilis* (Euglenophyceae) to carbon enrichment at pH 3. *J Phycol* **25**: 529–539.
- Perrière G, Duret L, Gouy M. (2000). HOBACGEN: database system for comparative genomics in bacteria. *Genome Res* **10**: 379–385.
- Ram RJ, VerBerkmoes NC, Thelen MP, Tyson GW, Baker BJ, Blake RC et al. (2005). Community proteomics of a natural microbial biofilm. *Science* **308**: 1915–1920.
- Schrader JA, Holmes DS. (1988). Phenotypic switching of *Thiobacillus ferrooxidans*. *J Bacteriol* **170**: 3915–3923.
- Schuch R, Garibian A, Saxild HH, Piggot PJ, Nygaard P. (1999). Nucleosides as a carbon source in *Bacillus subtilis*: characterization of the *drm-pupG* operon. *Microbiology* **145**: 2957–2966.
- Silver S, Phung LT. (2005). Genes and enzymes involved in bacterial oxidation and reduction of inorganic arsenic. *Appl Environ Microbiol* **71**: 599–608.
- Simmons SL, DiBartolo G, Denef VJ, Goltsman DSA, Thelen MP, Banfield JF. (2008). Population genomic analysis of strain variation in *Leptospirillum* Group II bacteria involved in acid mine drainage formation. *PLoS Biol* **6**: e177.
- Soutourina OA, Bertin PN. (2003). Regulation cascade of flagellar expression in Gram-negative bacteria. *FEMS Microbiol Rev* **27**: 505–523.
- Stolz JF, Basu P, Santini JM, Oremland RS. (2006). Arsenic and selenium in microbial metabolism. *Annu Rev Microbiol* **60**: 107–130.
- Talavera G, Castresana J. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol* **56**: 564–577.
- Tamura K, Dudley J, Nei M, Kumar S. (2007). MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol Biol Evol* **24**: 1596–1599.
- Tendeng C, Bertin PN. (2003). H-NS in Gram-negative bacteria: a family of multifaceted proteins. *Trends Microbiol* **11**: 511–518.
- Tyson GW, Chapman J, Hugenholtz P, Allen EE, Ram RJ, Richardson PM et al. (2004). Community structure

- and metabolism through reconstruction of microbial genomes from the environment. *Nature* **428**: 37–43.
- 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* **2009**: bap021.
- Vaughan DJ. (2006). Arsenic. *Elements* **2**: 71–75.
- Wang Q, Garrity GM, Tiedje JM, Cole JR. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* **73**: 5261–5267.
- 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 *Hermiimonas arsenicoxydans* by proteomics data. *Biochimie* **91**: 192–203.
- Wilmes P, Andersson AF, Lefsrud MG, Wexler M, Shah M, Zhang B *et al.* (2008). Community proteogenomics highlights microbial strain-variant protein expression within activated sludge performing enhanced biological phosphorus removal. *ISME J* **2**: 853–864.
- Wilmes P, Simmons SL, Deneff VJ, Banfield JF. (2009). The dynamic genetic repertoire of microbial communities. *FEMS Microbiol Rev* **33**: 109–132.
- Yarzabal A, Brasseur G, Ratouchniak J, Lund K, Lemesle-Meunier D, DeMoss JA *et al.* (2002). The high-molecular-weight cytochrome *c* Cyc2 of *Acidithiobacillus ferrooxidans* is an outer membrane protein. *J Bacteriol* **184**: 313–317.

Supplementary Information accompanies the paper on The ISME Journal website (<http://www.nature.com/ismej>)

Supplementary Figures

Diversity of metabolic interactions inside an arsenic-rich microbial ecosystem revealed by meta- and proteo-genomics

Philippe N. Bertin *et al.*

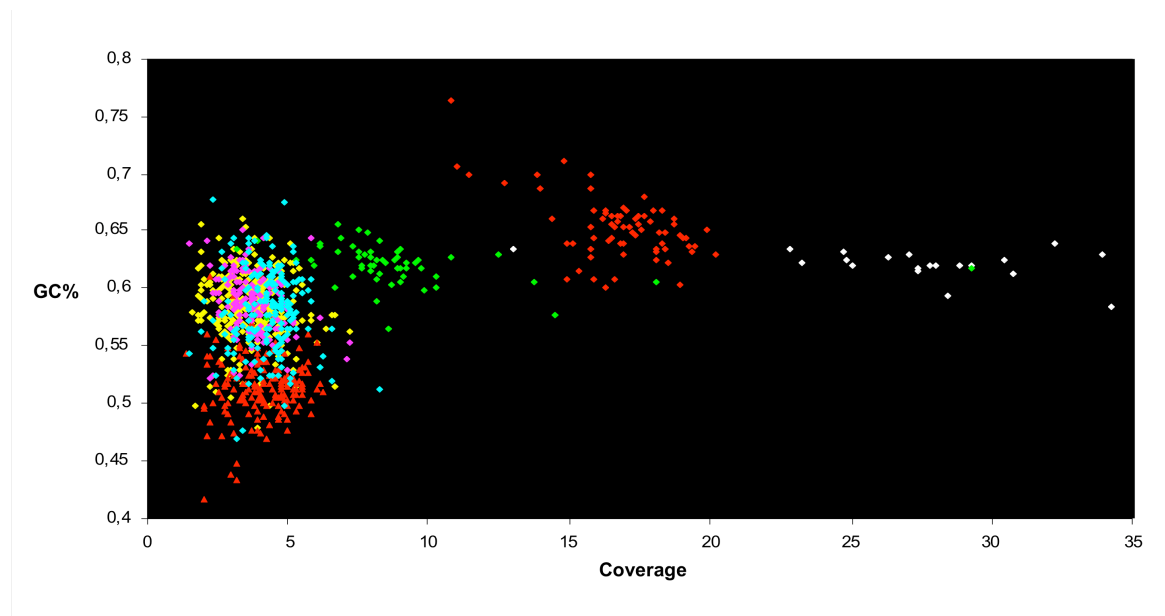


Figure S1 Distribution plot showing the repartition of the seven major contigs based on their respective GC% and mean coverage. Colors indicate the different bin assignation of supercontigs (CARN1, white, CARN2, red; CARN3, purple; CARN4, green, CARN5, blue; CARN6, yellow and CARN7, light red).

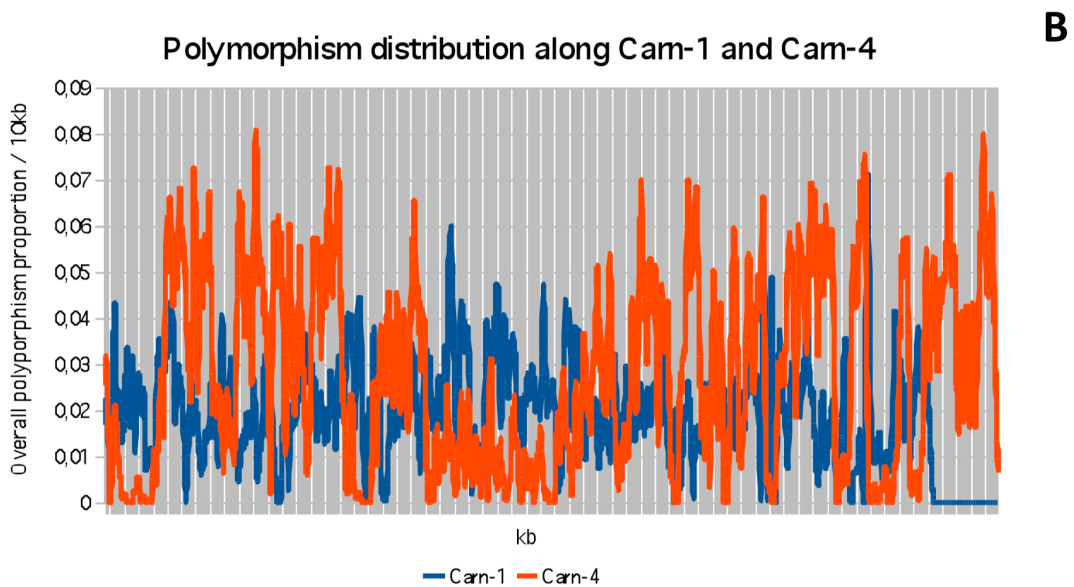
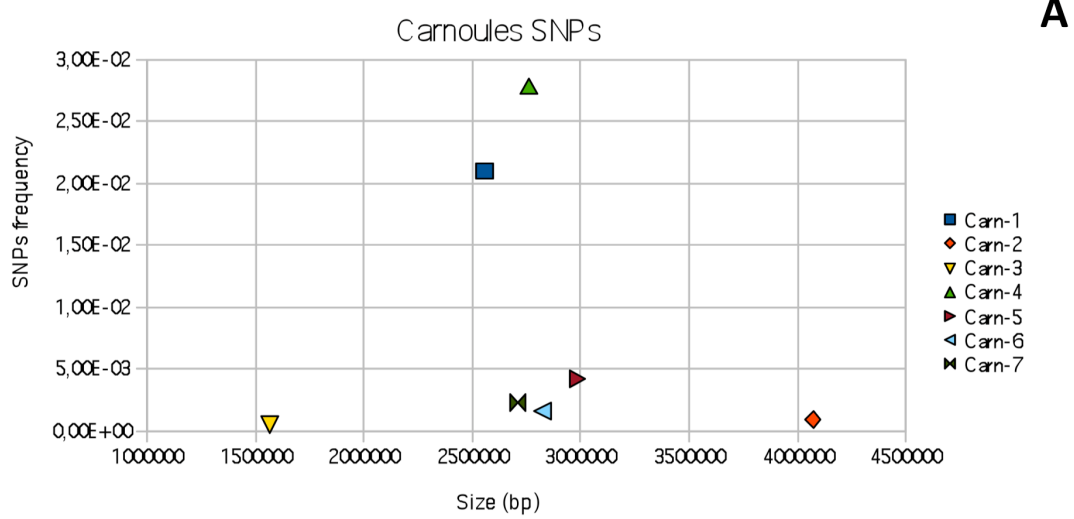


Figure S2 Amount of polymorphism in the sequenced population regarding the CARN bin reference sequences. **A:** SNPs+indel events frequencies relative to each bin sequence length; **B:** distribution of the polymorphism along the respective sequences of CARN1 and CARN4 bins.

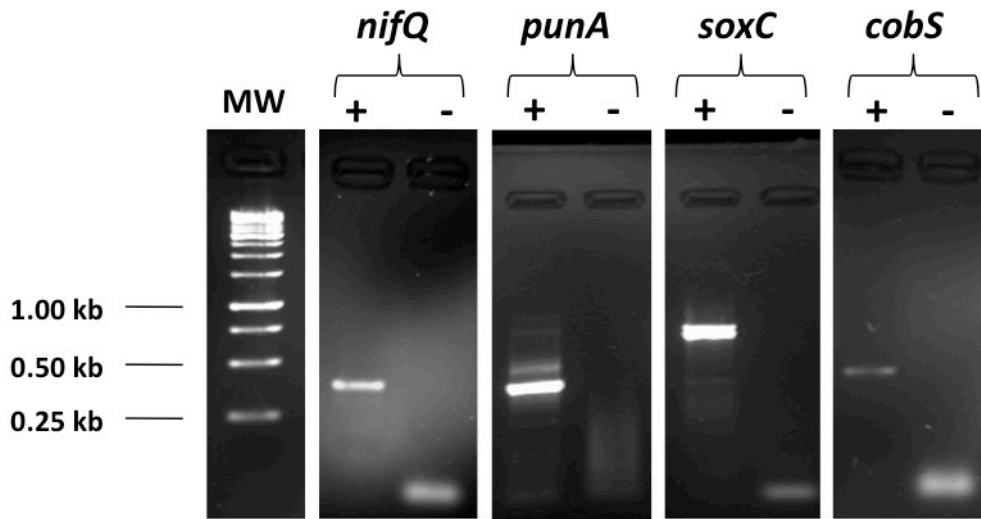


Figure S3 Agarose gel electrophoresis of *nifQ*, *punA*, *soxC* and *cobS* amplified fragments using total microbial community RNA as a template. A reverse transcription step was performed (+) or not (-) before PCR amplification.

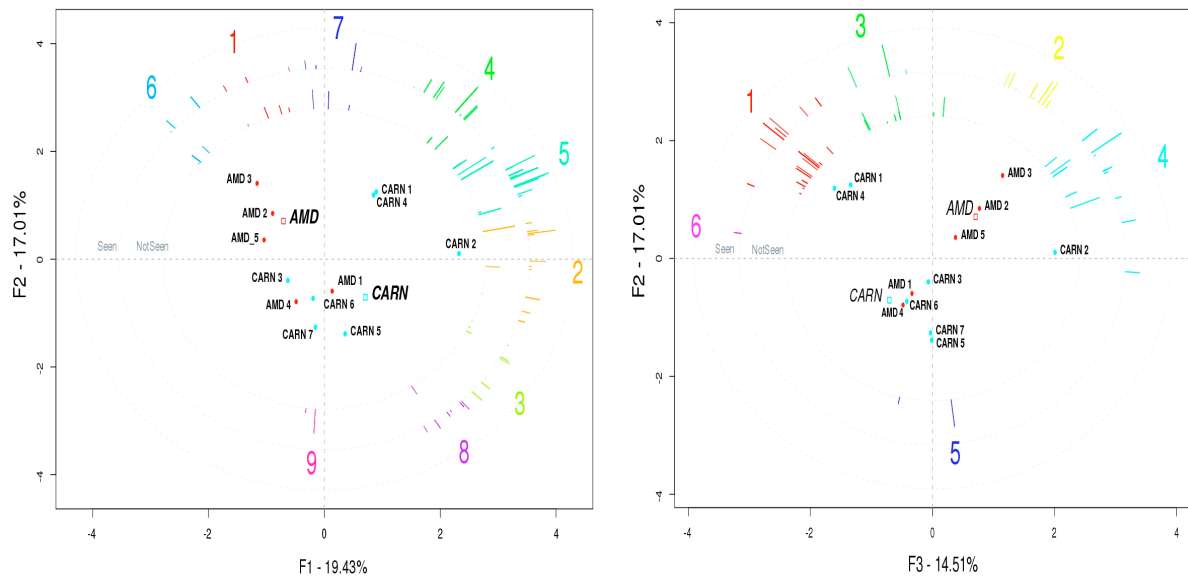


Figure S4 Multiple Factorial Analysis of the Carnoulès and AMD bins, performed on a two dimensional-matrix combining bins and enzymatic reactions, respectively. To highlight possible metabolic distinctions between bins, 3 axes (F1 to F3) capturing the highest dispersion in the resulting dot cloud were selected; they represent more than half the total dispersion. Colored lines represent the variable vectors corresponding to the enzymatic reaction frequencies, the external disk indicating those identified in the metaproteomic data. The reaction variables were then hierarchically clustered, which led to 9 and 6 classes (indicated by the colors) for the first and second factorial planes, respectively. The corresponding functions are listed in Supplementary Tables S3ab. AMD1 = *Leptospirillum* group II bin, AMD2 = *Ferroplasma* type II bin, AMD3 = *G-plasma*-bin, AMD4 = *Leptospirillum* group III bin and AMD5 = *Ferroplasma* type I bin.

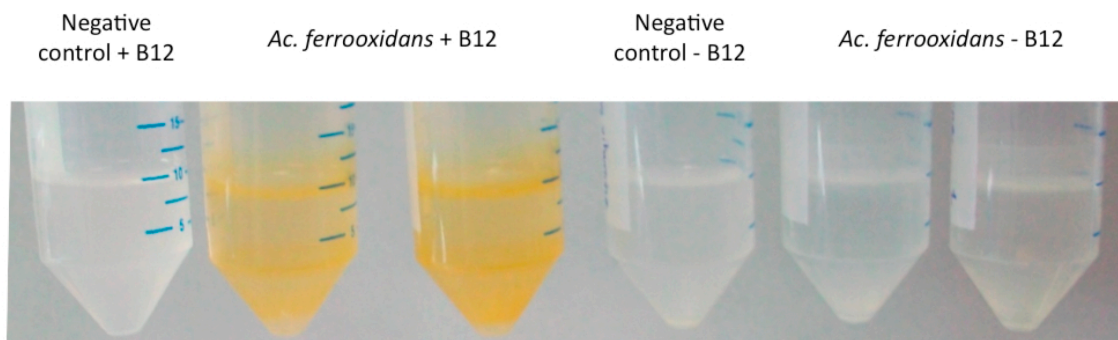


Figure S5 Effect of cobalamin on iron oxidation in *Acidithiobacillus ferrooxidans*. The activity was tested in liquid medium supplemented or not with 1 mg.L⁻¹ cobalamin (Sigma), and compared to non-inoculated samples used as negative controls. Cultures were grown at 30°C under agitation. Iron oxidation was followed by the appearance of an orange precipitate after 4 days incubation. The addition of cobalamin did not modify the pH (pH4.2) and had no effect on cell growth rate (data not shown).

CHAPITRE II

EUGLENA MUTABILIS :

QUEL ROLE DANS LES

INTERACTIONS TROPHIQUES ?

A : INTRODUCTION

Environ une centaine d'espèces d'euglènes sont actuellement répertoriées à travers le monde et ont été regroupées dans un groupe monophylétique par Marin et al. (2003). L'ultra-structure de ces organismes est présentée ci-dessous, en particulier l'organisation de la pellicule de ces organismes (Figure 3.1). Cette pellicule permet aux euglènes de maintenir leur forme originale tout en leur conférant une souplesse qui est impliquée chez *E. mutabilis* dans les mouvements de type gliding (Hader & Melkonian, 1983). Les euglènes disposent par ailleurs de 3 à 4 chloroplastes, d'un noyau, de nombreuses mitochondries, d'une vacuole contractile, d'un photorécepteur impliqué dans le phototactisme et de deux flagelles qui ne sortent pas de l'invagination antérieure chez *E. mutabilis* (Figure 3.1). Cette espèce se distingue par ailleurs des autres euglènes par la colonisation d'environnements acides tels que les tourbières, les lacs acides ou les sites exposés aux DMA au point d'être considérée comme un indicateur biologique de ces derniers (Brake et al., 2001b).

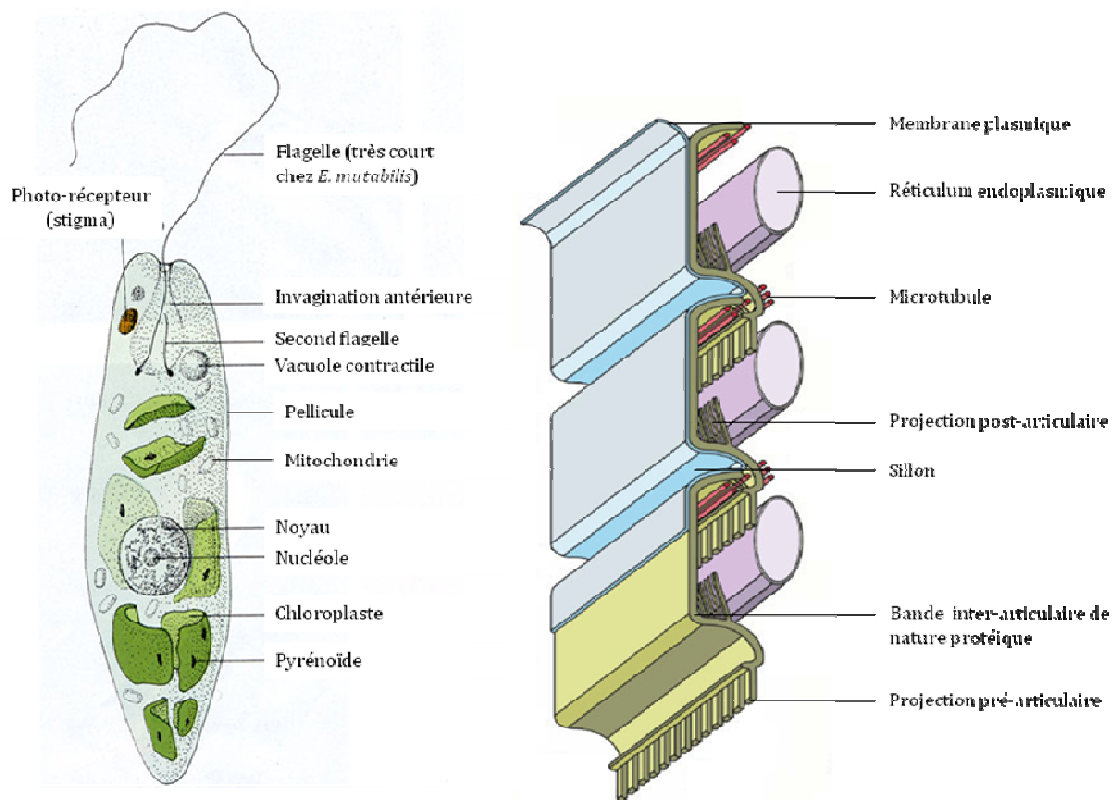


Figure 3.1 : Ultra-structure des euglènes et organisation de la pellicule conférant à ces protistes une membrane à la fois rigide et souple leur permettant de se déplacer par gliding (Leander et al., 2007).

En raison de leur métabolisme photosynthétique, les euglènes ont longtemps été considérées comme des algues vertes unicellulaires (Christensen, 1964). Cependant, une observation ultra-structurale plus approfondie de leurs chloroplastes laisse supposer une acquisition par endosymbiose secondaire de ces organites (Gibbs, 1978). Cette caractéristique souligne par conséquent une histoire évolutive très différente de celles des algues unicellulaires mais aboutissant à un mode de vie similaire. Cette histoire évolutive originale est d'ailleurs trahie par certaines particularités trophiques, les euglènes pouvant alterner entre un mode de vie photo-autotrophe et un mode de vie strictement chimio-hétérotrophe en perdant leur chloroplastes (Sumida *et al.*, 2007). Malgré leur capacité à pouvoir développer un métabolisme chimio-hétérotrophe, ces protistes disposent toujours des voies anaboliques leur conférant la capacité de synthétiser tous les composants nécessaires pour leur croissance en milieu minéral, à l'exception notoire de la thiamine, de la cobalamine et de la biotine (Olaveson & Stokes, 1989).

La présence ubiquitaire d'*E. mutabilis* dans les DMA ainsi que son métabolisme photosynthétique ont donc conduit à considérer cet organisme comme un producteur primaire de première importance dans ces écosystèmes par ailleurs souvent très pauvres en matière organique (Das *et al.*, 2009a; Johnson & Hallberg, 2003). Afin de comprendre plus en détail les interactions trophiques potentielles au sein de la communauté microbienne du site de Carnoulès, et en particulier l'apport de l'euglène dans la production primaire, une étude portant sur la synthèse et la sécrétion de métabolites par *E. mutabilis* a été réalisée. Cette approche repose sur l'usage de techniques de protéo-métabolique effectuées *in situ* et confirmées par des études complémentaires menées en condition de laboratoire sur une culture pure d'*E. mutabilis*. Les principaux résultats relatifs à cette étude sont présentés sous forme de manuscrit soumis à *The ISME Journal* (voir fin de chapitre).

B : RESULTATS

Les analyses de protéomique et métabolomique menées sur *E. mutabilis in situ* et *in vitro* ont permis de souligner que ce protiste sécrète sélectivement certains métabolites, principalement des acides aminés et certains sucres, mais pas d'acides gras ni certains acides organiques. Cette observation pourrait expliquer la présence de la majorité des métabolites identifiés dans les eaux interstitielles du site de Carnoulès et conforte le rôle de producteur primaire de cette euglène.

Afin de déterminer les interactions trophiques dont pourrait profiter en contrepartie *E. mutabilis*, des tests d'auxotrophie ont également été réalisés en conditions de laboratoire. Les

résultats présentés ci-dessous (Figure 3.2), mettent en évidence une auxotrophie pour la vitamine B12 mais pas pour la thiamine et la biotine contrairement aux données de la littérature (Olaveson & Stokes, 1989). Ces résultats suggèrent que si *E. mutabilis* est capable de sécréter de nombreux métabolites au sein du DMA de Carnoulès, ce protiste y trouve également une source exogène de cobalamine lui permettant de croître.

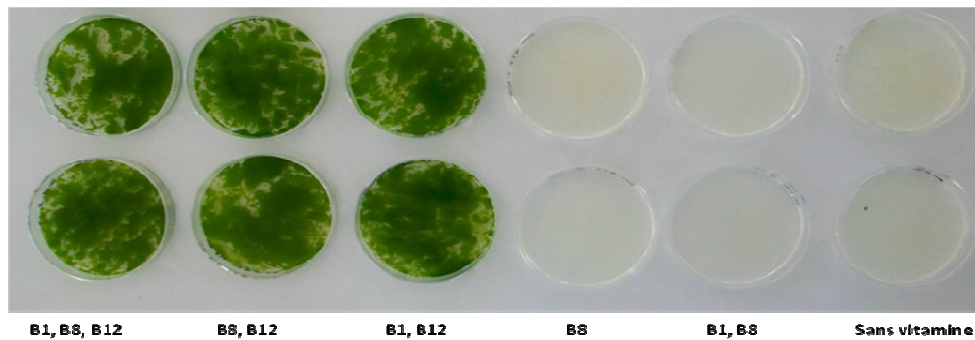


Figure 3.2 : Test d'auxotrophie pour les vitamines B1 (thiamine), B8 (biotine) et B12 (cobalamine). La cobalamine est nécessaire et suffisante pour permettre la croissance d'*E. mutabilis* en milieu minéral.

Par ailleurs, les analyses de protéo-métabolomique réalisées sur *E. mutabilis in situ* ont également conduit à l'identification de multiples mécanismes de défense face aux stress abiotiques du site de Carnoulès. Ainsi, l'expression de multiples pompes d'efflux de protons (membranaires et vacuolaires) ont été mises en évidence et sont probablement impliquées dans la tolérance au stress acide chez ce protiste. Cette hypothèse est confortée par des analyses effectuées en conditions de laboratoire qui ont notamment souligné le caractère acidophile/acidotolérant de cette espèce (Figure 3.3).

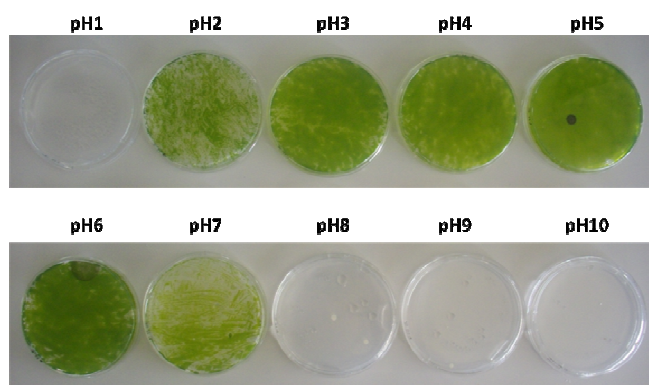


Figure 3.3 : Le caractère acidotolérant (une légère croissance est observée à pH 7) de la souche d'*E. mutabilis* isolée du site a été mis en évidence par un test de croissance à différents pH.

C : DISCUSSION ET CONCLUSION

La sécrétion de métabolites par l'euglène confère très probablement un avantage à la communauté microbienne dans son ensemble en favorisant la croissance de nombreux micro-organismes. Cependant, la bio-synthèse de métabolites essentiels tels que les sucres et les acides aminés est un processus très coûteux en terme d'énergie pour *E. mutabilis* et leur sécrétion ne peut que difficilement se concevoir sans un avantage conséquent en contre partie. L'hypothèse la plus probable est que les métabolites sécrétés par l'euglène favorisent le développement d'autres micro-organismes permettant à *E. mutabilis* de coloniser plus efficacement cet écosystème.

Ainsi, les sucres sécrétés par *E. mutabilis* en conditions de laboratoire (principalement du glucose, du fructose, de l'arabinose et du galactose) sont susceptibles de favoriser le développement d'autres bactéries décrites comme étant impliquées *in situ* dans les processus de bio-atténuation. En effet, *Thiomonas* sp. et *Acidithiobacillus ferrooxidans* disposent toutes deux des gènes impliqués dans l'assimilation de ces sucres et *Thiomonas* sp. a été montrée comme étant capable d'utiliser le fructose comme unique source de carbone (Slyemi *et al.*, 2011). Par conséquent, en accélérant le développement d'espèces bactériennes impliquées directement dans des processus de bio-atténuation, *E. mutabilis* faciliterait indirectement sa propre colonisation des DMA.

D'autre part, certains membres de la communauté procaryotique (CARN 1 et 4) expriment *in situ* des gènes impliqués dans la biosynthèse de cobalamine et sont susceptibles d'être favorisés par ce protiste. En effet, les résultats de l'analyse MFA (chapitre I) ont souligné que le métabolisme de CARN 1 et 4 est clairement orienté vers le transport et le recyclage des acides aminés qui sont abondamment sécrétés par *E. mutabilis*. La synthèse de cobalamine par ces bactéries favoriserait donc en retour le développement de ce protiste auxotrophe pour cette vitamine mais favoriserait également l'activité d'oxydation du fer par *A. ferrooxydans* tel que cela a été décrit précédemment (chapitre I, supplementary Figures).

Ces observations soulignent par conséquent de nombreuses interdépendances trophiques assimilables à du mutualisme au sein de la communauté microbienne, dont l'euglène, via la sécrétion de sucres et d'acides aminés, serait l'un des principaux acteurs. Les analyses de protéométabolomiques réalisées *in situ* ont également conduit à l'identification de protéines accumulées par *E. mutabilis* en conditions environnementales. Si la plupart de ces protéines ont permis de conforter le rôle de producteur primaire de l'euglène, de multiples protéines impliquées dans des réponses au stress oxydatif (superoxyde dismutase, ascorbate peroxydase, hydroperoxyde reductase,

thioredoxine peroxydases, ...) et au stress acide (pompes d'efflux de protons) ont également été identifiées. Ces résultats soulignent qu'*E. mutabilis* disposent de mécanismes adaptatifs pour faire face aux contraintes abiotiques les plus marquées sur ce site. Cependant, aucun mécanisme de défense au stress arsénié n'a pu être mis en évidence, et ceci malgré les très fortes concentrations de ce métalloïde sur le site de Carnoulès. Afin de comprendre plus en détail la physiologie de l'euglène, et en particulier sa réponse au stress arsénié, des études ont été réalisées sur une culture pure en conditions de laboratoire.

D : Manuscrit

L'intégralité de ces résultats ont été soumis à « the ISME Journal » dont le manuscrit correspondant est présenté ici.

1 ***In situ* proteo-metabolomics revealed metabolite secretion by the acid mine**
2 **drainage bioindicator, *Euglena mutabilis***

3 David Halter¹, Florence Goulhen-Chollet¹, Sébastien Gallien², Corinne Casiot³, Jérôme
4 Hamelin⁴, Françoise Gilard⁵, Christine Schaeffer², Christine Carapito², Alain Van Dorsselaer²,
5 Guillaume Tcherkez⁵, Florence Arsène-Ploetze¹, Philippe N. Bertin^{1*}

6

7 ¹ UMR7156 Université de Strasbourg/CNRS, Génétique Moléculaire, Génomique et
8 Microbiologie, Département Micro-organismes, Génomes, Environnement, 28 rue Goethe,
9 67083 Strasbourg Cedex, France.

10 ² Laboratoire de Spectrométrie de Masse Bio-organique, Institut Pluridisciplinaire Hubert
11 Curien, UMR7178 CNRS-Université de Strasbourg, 25 rue Becquerel, 67087 Strasbourg,
12 France.

13 ³ Laboratoire HydroSciences Montpellier, UMR 5569 (CNRS - IRD - Universités Montpellier
14 I et II), Université Montpellier II, CC MSE, Place Eugène Bataillon, 34095 Montpellier
15 Cedex 05, France.

16 ⁴ INRA, UR0050, Laboratoire de Biotechnologie de l'Environnement, Avenue des Etangs,
17 11100 Narbonne, France

18 ⁵ Plateforme Métabolisme-Métabolome, CNRS, UMR 8618, Institut de Biologie des Plantes -
19 Bâtiment 630 - Université Paris Sud 11. 91405 Orsay cedex, France

20

21 *** Corresponding author.**

22 Mailing address: UMR7156 Université de Strasbourg/CNRS, Génétique Moléculaire,
23 Génomique et Microbiologie, Département Micro-organismes, Génomes, Environnement, 28

24 rue Goethe, 67083 Strasbourg Cedex, France. Phone: +33 3 68 85 20 08. Fax: +33 3 68 85 20

25 28. E-mail: philippe.bertin@unistra.fr

26

27 **Running title.** Metabolite secretion in AMD by *Euglena mutabilis*

28 **Keywords.** proteomics / metabolomics / *Euglena mutabilis* / primary production / AMD

29

30 **Abstract**

31 *Euglena mutabilis* is a photosynthetic protist found in acidic aquatic environments
32 such as peat bogs, volcanic lakes and acid mine drainages (AMD). Through its photosynthetic
33 metabolism, this protist is supposed to have an important role in primary production in such
34 oligotrophic ecosystems. Nevertheless, the exact contribution of *E. mutabilis* in organic
35 matter synthesis remains unclear and no evidence of metabolite secretion by this protist has
36 been established to date. Here we combined *in situ* proteo-metabolomic approaches to
37 determine the nature of the metabolites accumulated by this protist and those secreted into an
38 AMD. Our results revealed that the secreted metabolites are represented by a large number of
39 amino acids, polyamine compounds, urea and some sugars but no fatty acids, suggesting a
40 selective organic matter contribution in this ecosystem. Such a production may have a crucial
41 impact on the bacterial community present on the study site, since it has been previously
42 shown that prokaryotes transport and recycle *in situ* most of the metabolites secreted by *E.*
43 *mutabilis*. Consequently, this protist may have an indirect but important role in AMD
44 ecosystems but also in other ecological niches often described as nitrogen-limited.

45 **Introduction**

46 *Euglena mutabilis* is a photosynthetic protist widespread in acidic environments all
47 around the world. It has been described as a part of the eukaryotic microbial community in
48 peat bogs (Pentecost, 1982), volcanic lakes (Sittenfeld et al., 2002), acid mine drainages
49 (AMDs) (Casiot et al., 2004a) and is considered as a bio-indicator of these latter ecosystems
50 (Brake et al., 2001a; Brake et al., 2001b). AMDs exposed environments are usually
51 considered as very toxic to biota as they present, besides low pH values, very high
52 concentrations of metallic compounds (Leblanc et al., 1996). Such drainages are observed in
53 ancient mining exploitations where the oxidation of the extruded rocks leads to the production
54 of sulfuric acid and to the mobilization of elements such as iron, copper, aluminum and
55 arsenic into the percolating waters (Baker et al., 2004; Canovas et al., 2008; Leblanc et al.,
56 1996). This process leads to a persistent contamination of downstream aquatic environments
57 and is a major concern for environmental and public health in many countries.

58 The extreme physico-chemical characteristics of AMDs make these environments
59 hostile for most of the living cells (Baker et al., 2009; Leblanc et al., 1996). Nevertheless, it is
60 now well established that the microbial communities they harbour are well adapted and can
61 have a decisive impact on the evolution of such ecosystems (Das et al., 2009; Tichy et al.,
62 1998). For instance, the bacterial community of the Carnoulès AMD (Gard, France) has been
63 extensively studied and appears to have an active role in bioattenuation processes. Bacteria
64 related to *Acidithiobacillus ferrooxidans* and *Thiomonas* spp. have been shown to oxidize
65 Fe(II) and As(III), respectively, leading to the co-precipitation of less bioavailable As-Fe
66 complexes (Bertin et al., 2011; Bruneel et al., 2003; Duquesne et al., 2003). Besides these
67 two bacteria, the Carnoulès AMD prokaryotic community has also been shown to be
68 dominated by 5 other bacteria, two of which being related to a novel phylum (Bertin et al.,
69 2011). Interestingly, several of these dominant bacteria, including those having a role in

70 bioattenuation processes, express *in situ* biological functions involved in metabolite transport
71 and recycling (Bertin *et al.*, 2011). Such an observation suggests that the organic matter
72 present at the study site may be crucial for the activities of the prokaryotic community as a
73 whole.

74 *E. mutabilis* has been previously identified in the Carnoulès AMD where it forms
75 easily observable green mats at the sediment/water interface (Casiot *et al.*, 2004a). Due to its
76 photosynthetic metabolism, this protist is supposed to have an important role in primary
77 production in such an oligotrophic environment, which could influence the composition and
78 the metabolism of the whole microbial community (Das *et al.*, 2009; Johnson and Hallberg,
79 2003). Nevertheless, no evidence of such a metabolic contribution has been established so far.
80 To address this question, the synthesis of organic compounds by this protist was investigated
81 by identifying proteins and metabolites accumulated by *Euglena* cells *in situ* and compared to
82 the metabolites identified in the water at the sampling site. In addition, both the synthesis and
83 secretion of metabolites were investigated under laboratory conditions after the isolation of
84 this protist. Metabolites accumulated within cells, i.e. endo-metabolome, and those liberated
85 into the extracellular medium, i.e. exo-metabolome, were characterized. Taken together, the
86 results allowed us to draw a model of the metabolic contribution of *E. mutabilis* in the AMD
87 of Carnoulès.

88

89 **Materials and methods**

90 **Sampling and chemical analyses:**

91 The Carnoulès AMD is localized in the south of France. Abandoned in 1960, the
92 mining exploitation led to the accumulation of 1.5 MT of rocks, containing high amounts of
93 heavy metals and metalloids present in sulfide minerals. The oxidation of these rocks led to the
94 production of an acid mine drainage which pours into the Reigous creek presenting high levels

95 of dissolved metals and metalloids (LeBlanc et al., 2002). Samples were collected in this creek
96 in October 2008 at the station called COWG and located 30 m downstream of the spring
97 (Bruneel et al., 2003). Two phases were distinguished at this sampling site: 6 cm deep white
98 sediments covering the bottom of the creek, and a thin column (i.e. less than 10 cm) of running
99 water covering these sediments.

100 The main physico-chemical parameters (pH, dissolved oxygen concentration,
101 conductivity, total dissolved solid) were determined in the field at the sampling site. Oxygen
102 measurements were done at the sediment/water interface using a microsensor (Unisense®).
103 Sediments were collected in triplicates using a sterile bottle, and the running water was filtered
104 (300 mL) through sterile 0.22 µm Nuclepore filters which were then transferred into collection
105 tube, frozen in liquid nitrogen, and stored at -80°C until further analysis. Samples for total iron
106 and arsenic determination were acidified to pH 1 with HNO₃ (14.5 M), and stored at 4°C in
107 polyethylene bottles. Samples for Fe(II) determination were stabilized with 1,10-
108 phenanthroline chloride in acetate buffer (pH 4.5) (Casiot et al., 2003). Samples for arsenic
109 speciation were preserved by addition of 5% (v/v) 0.25 M EDTA solution (Bednar et al.,
110 2002). Samples for arsenic speciation, Fe(II) and sulfate determination were stored in the dark
111 and analyzed within 24 h.

112 The determination of total dissolved arsenic was performed by Inductively Coupled
113 Plasma-Mass Spectrometry (ICP-MS, Thermo X7 Series). Elevated As concentrations were
114 diluted and no interference due to ArCl was detected. ICP-MS was calibrated using peak
115 intensity, acquired in peak jump mode, with standard solutions. Indium-115 was used as an
116 internal standard. Analyses of arsenic species (As(III), As(V)) were carried out using anion-
117 exchange chromatography (25 cm × 4.1 mm i.d. Hamilton PRP-X100 column) with
118 SpectraSystem SCM 1000 solvent delivery pump coupled to ICP-MS. The detection limit was
119 2 nM for As(III) and 4 nM for As(V) with a precision better than 5%. Total dissolved iron
120 was measured by Flame Atomic Absorption Spectrometry. Fe(II) concentration was

121 determined by reading the absorbance at 510 nm after complexation with 1,10-phenanthroline
122 chloride solution in buffered samples (pH 4.5) (detection limit: 0.2 μ M; precision better than
123 5%). Sulfate concentration was determined after precipitation of BaSO₄ with BaCl₂ and by
124 reading the absorbance at 650 nm (Casiot et al., 2003).

125

126 **Eukaryotic microbial community analysis:**

127 The eukaryotic microbial diversity was analyzed using capillary electrophoresis
128 coupled with single strand conformation polymorphism, as previously described (Quéméneur
129 et al., 2010). DNA was extracted from sediments using the UltraClean Soil DNA Isolation Kit
130 according to the recommendations of the manufacturer (MoBio Laboratories Inc., Carlsbad,
131 CA, USA). All extracted genomic DNA samples were stored at -20°C until further processing.
132 18S rRNA genes of the eukaryotic microbial community were amplified using primers Euk1A
133 and Euk516r (Scanlan and Marchesi, 2008). Three independent amplifications were performed
134 as follows: denaturation 94°C 1 min, annealing 64°C 1 min, amplification 72°C 1 min. and
135 repeat 35 times. 16S rRNA genes were amplified as previously described (Bruneel *et al.*,
136 2006)

137

138 ***Euglena mutabilis* cells recovery and 18S/16S rRNA analyses:**

139 To recover *Euglena* cells, 10 g of sediments were homogenized for 30 s in 10 mL of
140 saline buffer (NaCl 0.8%, KCl 0.02%, Na₂HPO₄ 0.15%, KH₂PO₄ 0.02%). After 10 min
141 decantation, 7.5 mL of supernatant were added without mixing to 17.5 mL of Nycodenz[®]
142 solution 65% (w/v) (Axis-Shield, Dundee, Scotland), and then centrifuged 1 h at 14, 000 x g.
143 The Nycodenz[®] gradient separated two distinct phases composed of *E. mutabilis* cells (green
144 upper phase) and other micro-organisms (brownish lower phase). The upper phase
145 corresponding to *E. mutabilis* cells was recovered by pipetting, washed by adding 2 volumes of
146 NaCl 0.9% and centrifuged (30 s, 1, 000 x g at 4°C). Genomic DNA of these cells was

147 extracted using the Wizard Genomic extraction kit (Promega, U.S.A.) and used as a template
148 for 18S rRNA and 16S rRNA amplifications as described above. A part of these *Euglena* cells
149 was plated on minimal solid agar medium (pH 3.2) (Olaveson and Stokes, 1989) and
150 cultivated by successive streaking on this solid media at least 5 times to ensure purity (25°C,
151 16 h / 8 h light/dark photoperiod / 45 $\mu\text{mol m}^{-2} \text{s}^{-1}$ photon flux density) (Halter *et al.*, 2011).

152

153 **Protein extraction, mass spectrometry analyses and protein identification:**

154 Proteins were extracted from *Euglena* cells using cell lysis buffer (Urea 6M, SDS 2%, DTT 2
155 mM, Glycerol 4%, Tris HCl pH 6.8 0.05M and Bromophenol Blue 0.05%) and heated for 1
156 min. at 100°C. After centrifugation (5 min. 4°C, 16, 000 x g), the supernatant was recovered
157 and proteins were separated by 1D SDS-PAGE (Laemmli, 1970) using a 12% gradient slab gel
158 (PROTEAN II, Bio-Rad laboratories). Electrophoresis was carried out at 60 mA/gel. Proteins
159 were stained with Coomassie brilliant blue R-250. Bands were systematically excised from the
160 gels and stored at -20°C before mass spectrometry analysis. *In gel* digestion of gel bands was
161 performed as previously described (Weiss *et al.*, 2009). The peptide extracts were analyzed by
162 nanoLC-MS/MS on a nanoACQUITY Ultra-Performance-LC (UPLC, Waters, Milford, MA)
163 coupled to a SYNAPT hybrid quadrupole orthogonal acceleration time-of-flight tandem mass
164 spectrometer (Waters, Milford, MA). The capillary voltage was set at 3 500V and the cone
165 voltage at 35 V. For tandem MS experiments, the system was operated with automatic
166 switching between MS and MS/MS modes. The 3 most abundant peptides, preferably doubly
167 and triply charged ions, were selected on each MS spectrum for further isolation and CID
168 fragmentation with 2 energies set using collision energy profiles. The complete system was
169 fully controlled by MassLynx 4.1 (SCN 566, Waters, Milford, MA). Raw data collected during
170 nanoLC-MS/MS analyses were processed and converted with ProteinLynx Browser 2.3 (23,
171 Waters, Milford, MA) into .pkl peak list format.

172 The MS/MS data were analyzed using the MASCOT 2.2.0. algorithm and searched
173 against two in-house generated databases. The first database included expressed sequence
174 tags (EST) of *Euglena* downloaded from <http://www.ncbi.nlm.nih.gov>, and the second
175 database was composed of the protein sequences downloaded from <http://www.uniprot.org/>.
176 For the estimation of the false positive rate in protein identification, a target-decoy database
177 search was performed (Elias and Gygi, 2007). In this approach, peptides were matched
178 against a database consisting of the native protein sequences found in the database (target)
179 and of the sequence-reversed entries (decoy). Spectra were searched, allowing a maximum of
180 1 missed cleavage, with a mass tolerance of 30 ppm for MS and 0.1 Da for MS/MS data and
181 with carbamidomethylation of cysteines and oxidation of methionines specified as variable
182 modifications.

183 Protein identifications using the EST database were validated when at least two
184 peptides with high quality MS/MS spectra (less than 12 points below Mascot's threshold score
185 of identity at 95% confidence level) were detected. In the case of one-peptide hits, the score of
186 the unique peptide had to be greater (minimal "difference score" of 10) than the 95%
187 significance Mascot threshold. Protein identifications using the Uniprot database were
188 validated when at least two peptides with a minimal Mascot ion score of 30 were detected. In
189 the case of one-peptide hits, the score of the unique peptide had to be greater (minimal
190 "difference score" of 16) than the 95% significance Mascot threshold.

191 All identifications were included into the "InPact" proteomic database developed in our
192 laboratory (<http://inpact.u-strasbg.fr/~db/>) (Bertin et al., 2008).

193

194 **Metabolite extraction, identification and GC-TOF analyses:**

195 Metabolite extraction was performed using methanol/water solvent (80/20 v/v).
196 Metabolite derivatization, GC-TOF-MS, data processing and profile analyses were carried out
197 as already described (Noctor et al., 2007). Metabolites were extracted either from *E. mutabilis*

198 cell pellets recovered from sediments or directly from Carnoulès water. In this latter case,
199 NaCl 0.9% (2 ml) was added to 10 g of water/sediment sample and gently mixed for 5 min at
200 room temperature. After 2 min decantation, the upper phase was recovered and filtered first
201 through a 10 µm membrane (Millipore) and then through a 0.22 µm pore membrane to remove
202 microbial cells. Lyophilized samples were then stored at -20°C for further analysis. For *in vitro*
203 analysis, metabolites were extracted from *Euglena* cell pellets recovered by centrifugation after
204 1 and 7 days incubation in inorganic liquid media. The supernatant was collected in parallel,
205 filtered through a 0.22 µm pore membrane to remove microbial cells, lyophilized and stored at
206 -20°C.

207 Each metabolite was identified by the comparison of spectra with those obtained with
208 the corresponding pure metabolite and the amount was calculated on the basis of the
209 corresponding peak area related to an internal standard (ribitol) and compared to the amount
210 of starting material. Integrated peak areas were obtained after deconvolution by the LECO
211 PEGASUS III software. The covariance of the metabolites was evaluated by hierarchical
212 classification using Mev algorithm.

213 **Results**

214 **Characterization of the Carnoulès AMD**

215 To characterize the natural environment of *E. mutabilis*, the physical, chemical and
216 biochemical parameters of the Carnoulès AMD were determined. These analyses were
217 performed at the sampling site called CowG, located 30 m downstream of the spring, where *E.*
218 *mutabilis* forms abundant green mats and where the bacterial community has been recently
219 characterized (Bertin *et al.*, 2011). The physico-chemistry of the running water covering the
220 sediments revealed a low pH (3.1), a high dissolved oxygen content at the sediment / water
221 interface (5.72 mg.L⁻¹) and a high concentration of sulfate (3 409 (±341) mg.L⁻¹). High
222 conductivity (4 900 µS.cm⁻¹) and total dissolved solid values (3 630 mg.L⁻¹) were also
223 measured, suggesting the occurrence of osmotic stress. Moreover, our analysis revealed the
224 presence of both oxidized and reduced form of arsenic (20.5 (±0.6) mg.L⁻¹ of As(V) and 133
225 (±2) mg.L⁻¹ of As(III)) and reduced form of iron (1 220 (±61) mg.L⁻¹ of Fe(II)). Finally, the
226 biochemical analysis at the sampling site led to the identification of 26 metabolites in the
227 water. These compounds were mainly glycerol, urea, amino acids or derivatives, unsaturated
228 fatty acids, sugars and organic acids (Table 1, Supplementary Table 1). Taken together, these
229 results confirmed the extreme environmental conditions prevailing in this AMD ecosystem and
230 revealed the presence of organic compounds in the water that may be synthesized by the
231 microbial community.

232

233 ***Euglena* cell recovering**

234 To investigate the role of *E. mutabilis* in primary production, this protist was
235 recovered by a cultivation-independent step using a density gradient separation. This
236 experimental procedure allowed the separation of *Euglena* cells from the remaining part of the
237 microbial community according to their specific density (Figure 1). Optical microscope
238 observations performed on the *Euglena* cell fraction revealed both the absence of any

239 apparent microbial contaminant and the existence of a homogenous population of *E. mutabilis*
240 (data not shown).

241 To determine the efficiency of this fractionation procedure, both 18S rRNA and 16S
242 rRNA genes were amplified from DNA extracted from the green phase corresponding to *E.*
243 *mutabilis*. The profiles of the amplified sequences were determined by CE-SSCP and
244 compared to the CE-SSCP profiles of the 16S rRNA and 18S rRNA sequences amplified from
245 DNA directly extracted from sediments (Figure 1). These fingerprinting profiles revealed the
246 presence of low eukaryotic and prokaryotic diversity in the sediments. A single highly
247 dominant peak for both 18S rRNA and 16S rRNA sequences was observed in the *E. mutabilis*
248 cell fraction, which was shown to correspond to the *E. mutabilis* 18S rRNA gene and to the
249 chloroplast 16 rRNA gene, respectively (data not shown). Although we cannot rule out that
250 the *E. mutabilis* cells represent different ecotypes, they were not distinguished from each
251 other according to morphological observations and 18S rRNA sequencing and were
252 consequently considered as homogenous. These results suggest that the eukaryotic microbial
253 community present in the Carnoulès sediments is dominated by *E. mutabilis* and that the
254 fractionation procedure allowed us to recover efficiently these cells without any cultivation
255 step.

256

257 **Proteins expressed *in situ* by *E. mutabilis***

258 To identify the main biological functions expressed by *E. mutabilis in situ*, total
259 proteins were extracted from *Euglena* cells. These extracted proteins were then separated
260 according to their molecular weight by 1D SDS-PAGE. NanoLC-MS/MS analyses of all gel
261 bands led to the identification of 189 unique proteins using two databases. The first one, an
262 expressed sequence tags (ESTs) database of *E. gracilis* was used to identify proteins already
263 described in *Euglena* species and led to the identification of 162 proteins. The second one, an
264 Uniprot subset database, was used to exhaustively identify proteins, including those not yet

265 described in *Euglena* species. This second analysis led to the identification of 27 additional
266 proteins, including 9 bacterial proteins and 18 proteins identified in species related to the
267 *Euglena* either phylogenetically or functionally, i.e. photosynthetic organisms (Figure 2 and
268 Supplementary Tables 2 and 3). Even though some protein contaminants, i.e. less than 5%,
269 were still present, these results confirmed that the cellular fraction obtained by density gradient
270 separation was highly enriched in *Euglena* cells.

271 The majority of the identified proteins are known to be involved in photosynthesis, i.e.
272 PSI and PSII systems, ATP synthase and cytochrome b₆-f complex, highlighting the active role
273 of *E. mutabilis* in photosynthetic oxygen production. Moreover, proteins associated with cell
274 division, transcription and protein turnover were also identified suggesting that *Euglena* cells
275 are physiologically active and able to grow at the study site. Phototactism related proteins, i.e.
276 photoactivated adenylyl cyclase, and proteins involved in oxidative stress defense mechanisms,
277 i.e. thioredoxin peroxidase, ascorbate peroxidase, superoxide dismutase and hydroperoxide
278 reductase, were also accumulated suggesting the existence of adaptation processes to
279 environmental stresses (Lengfelder and Elstner, 1979; Seaver and Imlay, 2001; Shigeoka *et al.*,
280 1980). Similarly, a membrane-associated ATP-dependent H⁺ transporter was found,
281 presumably reflecting an acclimation to the low pH conditions prevailing in the Carnoulès
282 AMD.

283 Most of the proteins accumulated by *Euglena* cells, i.e. approximately 70%, are known
284 to be involved in energy production and/or anabolic pathways (Figure 2). For instance, a large
285 amount of enzymes involved in carbohydrate metabolism (Calvin cycle, glycolysis and pentose
286 phosphate pathway), amino acid, nucleotide and fatty acid biosynthesis or nitrogen metabolism
287 (urea cycle and ammonium assimilation) were identified, suggesting that *E. mutabilis*
288 expressed functions possibly leading to specific organic matter production.

289

290 **Metabolite identification in *E. mutabilis* cells *in situ***

291 To determine which metabolites were preferentially accumulated in *E. mutabilis* cells
292 *in situ*, a metabolomic approach was performed. 57 metabolites, mainly fatty acids (C:8, C:10,
293 C:12, C:14, C:16, C:18, C:18(2), C:18(3)), sugars (arabinose, fructose, glucose, galactose,
294 ribose, trehalose, xylose), amino acids (excepted leucine and histidine), organic acids
295 (ascorbic, benzoic, lactic, malic, nicotinic, hydroxybenoic, pipercolic, pyroglutamic, succinic
296 and sinapinic acid) and other metabolites such as urea, spermidine, putrescine, ornithine,
297 phytol, myo-inositol, mannitol, glycerol derivatives, GABA and ethanolamine were identified
298 in *E. mutabilis* cells (Supplementary Table 1). These results correlated with the expression of
299 the anabolic pathways identified by the proteomic approach, which may have led to the
300 accumulation of multiple metabolites in *Euglena* cells. Interestingly, among the 26 metabolites
301 identified in the water (Table 1), 22 were also found in the protist, i.e. all metabolites excepted
302 octanoic acid, decanoic acid, levoglucosan and 2-furancarboxylic acid (Supplementary Table
303 1). Although we cannot rule out that the metabolites identified in *E. mutabilis* cells were
304 accumulated from the environment, these data suggest that this protist may be able to
305 synthesize such metabolites and secreted them into the AMD water.

306

307 **Metabolite synthesis and secretion by *E. mutabilis* *in vitro***

308 A metabolic profile was established *in vitro* from a pure culture to determine which
309 metabolites found in the *E. mutabilis* cells are synthesized by the protist itself and may be
310 liberated into the water. *E. mutabilis* cells were grown in mineral medium containing only
311 cobalamine since this protist is auxotroph for this cofactor (Olaveson and Stokes, 1989). After
312 2 weeks growth, *E. mutabilis* cells were transferred into a mineral liquid medium mimicking
313 the acidic conditions of the AMDs and incubated for one week. Metabolites accumulated
314 within the cells, i.e. the endo-metabolome, and those secreted into the culture medium, i.e. the
315 exo-metabolome, were analyzed after one day and one week of incubation. 78 metabolites
316 were identified from *E. mutabilis* cells *in vitro*, which is significantly higher than the 57

317 metabolites identified from *Euglena* cells *in situ*. This observation may be explained, at least in
318 part, by the amount of biological material, which was higher in the experiment performed in
319 laboratory conditions. Only three metabolites were found *Euglena* cells *in situ* but not in the
320 *Euglena* cells cultivated *in vitro*, i.e. 5-oxopyrrolidine-2-carboxamide, pipercolic acid and
321 sinapinic acid. This results indicates that *E. mutabilis* is able to synthesize the majority of the
322 metabolites found within the *Euglena* cells *in situ*.

323 To determine which metabolites may be liberated by *E. mutabilis*, the exo-metabolites
324 were identified and compared to the endo-metabolites after one day and one week of
325 incubation (Supplementary Table 1). The amount of each metabolite was quantified using
326 ribitol as an internal standard. The metabolite variations in both the medium and *E. mutabilis*
327 cells were analyzed by hierarchical classification. Variation patterns (p value <0.05) were
328 represented as a heat map highlighting the existence of three main groups of metabolites
329 (Figure 3). The first one corresponds to metabolites (phytol and fatty acids) accumulated
330 within the cells after 7 days of incubation but not in the medium (I). Surprisingly, fatty acids
331 and some organic acids (fumaric acid, ascorbic acid, maleic acid, malic acid, glyceric acid,
332 oxalic acid, hydroxyglutaric acid) were not detected in the medium despite their accumulation
333 in the cells. The second group is represented by metabolites (glucose-6-phosphate and
334 downstream products such as organic acids and ascorbate) that disappeared upon *in vitro*
335 incubation (II). The third group is composed by metabolites accumulated in the external
336 medium (III). These compounds were mainly amino acids (proline, isoleucine, alanine, valine,
337 leucine, phenylalanine, glycine, asparagines, tyrosine and proline), sugars (glucose, arabinose,
338 mannose, ribose, xylose, fructose and mannitol) and other compounds (putrescine, glycerol,
339 glycolic acid, spermidine, galactosyl glycerol, hydroxybenzoic acid and ornithine).

340 To quantify the observed accumulation or depletion patterns in both the extracellular
341 medium and *Euglena* cells, we calculated the variations in metabolite content (% mg⁻¹ FW)
342 (Supplementary Figure 1). Fatty acids accumulated in *Euglena* cells (by up to 6 fold) and

343 putrescine appeared to be the most specifically excreted metabolite, with a 16-fold increase
344 upon cell incubation. Sugars such as glucose and arabinose (10-fold increase) and amino acids
345 (from 2 to 8-fold increase) were also shown to be abundantly excreted. These quantification
346 patterns suggest that the accumulation of some metabolites in the medium is not simply due to
347 cell lysis but rather relies on a selective secretion by *Euglena* cells.

348 The exo-metabolites identified *in vitro* were compared to those identified at the
349 sampling site (Figure 4). Out of the 26 metabolites identified in AMD water, 15 were shown to
350 be secreted under laboratory conditions by *E. mutabilis*. These metabolites were mainly sugars
351 (mannitol, glucose, fructose), amino acids (glycine, treonine, tyrosine) and other compounds
352 (glycerol, urea, lactic acid). These results suggest that some metabolites shown to be
353 specifically secreted by *E. mutabilis in vitro*, may be synthesized by this protist *in situ*
354 allowing us to propose a model of the organic matter contribution of *E. mutabilis* at the
355 Carnoulès site (Figure 4).

356

357 **Discussion**

358 Our proteo-metabolomic data revealed that *E. mutabilis* may have a crucial role on the
359 functioning of the Carnoulès AMD ecosystem. Two main activities may be highlighted
360 according to our results: the ability of *E. mutabilis* to produce O₂ through photosynthetic
361 metabolism on the one hand, and its specific organic matter synthesis on the other hand.

362

363 **Adaptation of *E. mutabilis* to AMD and photosynthetic metabolism**

364 AMD ecosystems are usually described as extreme environments for microbial
365 communities. Our biochemical characterization of the Carnoulès ecosystem (Gard, France)
366 supports the existence of multiple abiotic stresses in accordance with previous observations
367 (Bruneel *et al.*, 2011; Casiot *et al.*, 2004b; Casiot *et al.*, 2003). In particular, osmotic stress,
368 low pH and oxidative stress caused by heavy metals and metalloids may explain the reduction

369 in eukaryotic and prokaryotic diversity observed in such ecosystems (Johnson and Hallberg,
370 2003). Despite these conditions, *E. mutabilis* is frequently observed in AMDs (Olaveson and
371 Nalewajko, 2000), which suggests that this protist is able to cope with most of these
372 environmental stresses. Indeed, our proteo-metabolomic approach highlighted the expression
373 of multiple defense mechanisms expressed *in situ*. For instance, the expression of proton efflux
374 system, the accumulation of antioxidant proteins and metabolites (ascorbic acid), and the
375 synthesis of trehalose may help *E. mutabilis* to cope with acidic, oxidative and osmotic
376 stresses, respectively.

377 In addition, the photosynthetic metabolism of *E. mutabilis* revealed by our proteomic
378 analysis may have an important impact on the Carnoulès ecosystem through several ways.
379 First, it has been shown that the ‘oversaturation’ in dissolved oxygen (i.e. >200% saturation
380 with the atmosphere) facilitates inorganic co-precipitation of arsenic and iron colloids (Brake
381 et al., 2001a; Brake et al., 2001b). Second, an increased concentration in dissolved oxygen may
382 favor the development of aerobic species such as *Thiomonas* or *Acidothiobacillus* spp., both
383 known to be dominant and involved in the bioattenuation processes at the Carnoulès AMD
384 (Bertin *et al.*, 2011; Bruneel *et al.*, 2003; Duquesne *et al.*, 2003). The photosynthetic
385 metabolism of *E. mutabilis* may therefore affect directly and indirectly the metal precipitation
386 processes observed in AMDs. Interestingly, its photosynthetic metabolism allows this protist to
387 produce organic compounds that may be substrates for the Carnoulès bacterial community.

388

389 **Metabolic interactions between *E. mutabilis* and bacterial community**

390 The *in situ* functional genomics analyses highlighted the ability of *E. mutabilis* to grow,
391 divide and metabolize under harsh environmental conditions. As this protist is auxotroph for
392 cobalamine, our observations suggest that this protist gets this cofactor the AMD waters. A
393 previous analysis performed on the prokaryotic community at the same study site revealed that
394 bacteria belonging to the novel phylum *Candidatus Fodinabacter communicans* express *in*

395 *situ* genes involved in cobalamine biosynthesis (Bertin *et al.*, 2011) . It is therefore tempting to
396 speculate that these bacteria may help *E. mutabilis*, at least in part, to grow and colonize the
397 Carnoulès ecosystem and that metabolic interactions may occur between the bacterial
398 community and *E. mutabilis*.

399 Conversely, the comparison between metabolites found *in situ* and those secreted by
400 the protist *in vitro* strongly suggest that *E. mutabilis* may be at the origin of the accumulation
401 of primary metabolites in AMD waters. This assertion is supported by our proteomic data
402 revealing that *Euglena* express anabolic pathways involved in fatty acid, amino acid, sugar and
403 nucleotide biosynthesis. Surprisingly, the metabolite accumulation in the medium observed *in*
404 *vitro* strongly differs according to their nature. Indeed, sugars, amino acids, urea and
405 polyamine compounds (putrescine, ornithine) are among the more actively secreted unlike
406 some organic acids and fatty acids (Figure 3). Interestingly, most of those secreted compounds
407 contain nitrogen, which may have a key role in the ecosystems where *E. mutabilis* is found.
408 Indeed, acidic lakes and peat bogs contain both very low available N concentrations and this
409 element has been shown to constitute the growth limiting factor for sulfate-reducing bacteria in
410 AMDs (Beamud *et al.*, 2010; Crowley and Bedford, 2011; Waybrant *et al.*, 2002). Therefore,
411 in the Carnoulès AMD, some of these secreted compounds may be used as nutrients by other
412 members of the microbial community. Similarly, some organic compounds may be used as an
413 energy source by bacteria. Indeed, previous studies revealed that *Thiomonas* spp. isolated from
414 Carnoulès carry genes involved in the transport and degradation of glucose, fructose, glycerol,
415 lactate and several amino acids and are able to use fructose as a unique carbon source (Arsène-
416 Ploetze *et al.*, 2010; Slyemi *et al.*, 2011). Moreover, the genome of *Thiomonas* strains contains
417 a complete operon involved in urea degradation possibly involved in acidic stress tolerance
418 (Arsène-Ploetze *et al.*, 2010). These observations suggest that *E. mutabilis* may have an
419 important role in the development of such *Thiomonas* strains in the Carnoulès AMD.
420 Interestingly, another dominant bacterial species present at this site, i.e. *Candidatus*

421 Fodinabacter communicans, has been shown to contain and express many genes involved in
422 amino acid transport and recycling (Bertin *et al.*, 2011). *E. mutabilis* may thus participate in
423 the feeding of these bacteria which in turn produce cobalamine useful for the protist (Bertin *et*
424 *al.*, 2011). These observations further support the existence of strong trophic interactions
425 within the microbial community.

426

427 In conclusion, our results revealed that *E. mutabilis* has a significant and selective role
428 in organic matter production in the partly oligotrophic ecosystem of Carnoulès. Although we
429 cannot rule out that other microorganisms are also involved in this process, the expression of
430 multiple anabolic pathways detected in *E. mutabilis* by the proteomic approach suggest an
431 important role of this protist in organic compound production. Consequently, *E. mutabilis* may
432 have an indirect role in the bioattenuation of AMD by feeding bacterial strains and maintaining
433 aerobic conditions *via* its photosynthetic metabolism. This specific organic matter production
434 may also have an important impact in other acidic environments where *E. mutabilis* is found,
435 especially in N-starved ecosystems such as peat bogs or volcanic lakes.

436 **Acknowledgements**

437 The authors are indebted to Tao Ye, Frédéric Plewniak and Olivier Poch, from the Plate-
438 forme Bio-informatique de Strasbourg, for updating the Inpact database. The study and
439 Florence Chollet were financed by the ANR 07-BLANC-0118 project (Agence Nationale de
440 la Recherche). Sébastien Gallien and David Halter were supported by a grant from the French
441 Ministry of Education and Research. This work was performed within the framework of the
442 research network “Arsenic Metabolism in Micro-organisms” (GDR2909-CNRS).

443

444 **References**

445

446 Arsène-Ploetze F, Koechler S, Marchal M, Coppée JY, Chandler M, Bonnefoy V *et al* (2010).
447 Structure, function, and evolution of the *Thiomonas* spp. genome. *PLoS Genetics* **6**.

448

449 Baker BJ, Lutz MA, Dawson SC, Bond PL, Banfield JF (2004). Metabolically active
450 eukaryotic communities in extremely acidic mine drainage. *Applied and Environmental*
451 *Microbiology* **70**: 6264-6271.

452

453 Baker BJ, Tyson GW, Goosherst L, Banfield JF (2009). Insights into the diversity of
454 eukaryotes in acid mine drainage biofilm communities. *Applied and Environmental*
455 *Microbiology* **75**: 2192-2199.

456

457 Beamud SG, Diaz MM, Pedrozo FL (2010). Nutrient limitation of phytoplankton in a
458 naturally acidic lake (Lake Caviahue, Argentina). *Limnology* **11**: 103-113.

459

460 Bednar AJ, Garbarino JR, Ranville JF, Wildeman TR (2002). Preserving the distribution of
461 inorganic arsenic species in groundwater and acid mine drainage samples. *Environmental*
462 *Science and Technology* **36**: 2213-2218.

463

464 Bertin PN, Heinrich-Salmeron A, Pelletier E, Goulhen-Chollet F, Arsène-Ploetze F, Gallien S
465 *et al* (2011). Metabolic diversity among main microorganisms inside an arsenic-rich
466 ecosystem revealed by meta- and proteo-genomics. *The ISME Journal*.

467

468 Bertin PN, Médigue C, Normand P (2008). Advances in environmental genomics: towards an
469 integrated view of micro-organisms and ecosystems. *Microbiology* **154**: 347-59.

470

471 Brake SS, Dannelly HK, Connors KA (2001a). Controls on the nature and distribution of an
472 alga in coal mine-waste environments and its potential impact on water quality.
473 *Environmental Geology* **40**: 458-469.

474

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

478

479 Bruneel O, Duran R, Casiot C, Elbaz-Poulichet F, Personné JC (2006). Diversity of
480 microorganisms in Fe-As-rich acid mine drainage waters of Carnoulès, France. *Applied and*
481 *Environmental Microbiology* **72**: 551-556.

482

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

486

487 Bruneel O, Volant A, Gallien S, Chaumande B, Casiot C, Carapito C *et al* (2011).
488 Characterization of the Active Bacterial Community Involved in Natural Attenuation
489 Processes in Arsenic-Rich Creek Sediments. *Microbial Ecology*: 1-18.

490

491 Canovas CR, Hubbard CG, Olias M, Nieto JM, Black S, Coleman ML (2008). Hydrochemical
492 variations and contaminant load in the Rio Tinto (Spain) during flood events. *Journal of*
493 *Hydrology* **350**: 25-40.
494

495 Casiot C, Bruneel O, Personné JC, Leblanc M, Elbaz-Poulichet F (2004a). Arsenic oxidation
496 and bioaccumulation by the acidophilic protozoan, *Euglena mutabilis*, in acid mine drainage
497 (Carnoules, France). *Sci Total Environ* **320**: 259-67.
498

499 Casiot C, Bruneel O, Personné JC, Leblanc M, Elbaz-Poulichet F (2004b). Arsenic oxidation
500 and bioaccumulation by the acidophilic protozoan, *Euglena mutabilis*, in acid mine drainage
501 (Carnoulès, France). *Science of the Total Environment* **320**: 259-267.
502

503 Casiot C, Morin G, Juillot F, Bruneel O, Personné JC, Leblanc M *et al* (2003). Bacterial
504 immobilization and oxidation of arsenic in acid mine drainage (Carnoulès creek, France).
505 *Water Research* **37**: 2929-2936.
506

507 Crowley KF, Bedford BL (2011). Mosses influence phosphorus cycling in rich fens by
508 driving redox conditions in shallow soils. *Oecologia*: 1-12.
509

510 Das BK, Roy A, Koschorreck M, Mandal SM, Wendt-Potthoff K, Bhattacharya J (2009).
511 Occurrence and role of algae and fungi in acid mine drainage environment with special
512 reference to metals and sulfate immobilization. *Water Research* **43**: 883-894.
513

514 Duquesne K, Lebrun S, Casiot C, Bruneel O, Personné JC, Leblanc M *et al* (2003).
515 Immobilization of Arsenite and Ferric Iron by *Acidithiobacillus ferrooxidans* and Its
516 Relevance to Acid Mine Drainage. *Applied and Environmental Microbiology* **69**: 6165-6173.
517

518 Elias JE, Gygi SP (2007). Target-decoy search strategy for increased confidence in large-scale
519 protein identifications by mass spectrometry. *Nat Methods* **4**: 207-14.
520

521 Halter D, Casiot C, Heipieper HJ, Plewniak F, Marchal M, Simon S *et al* (2011). Surface
522 properties and intracellular speciation revealed an original adaptive mechanism to arsenic in
523 the acid mine drainage bio-indicator *Euglena mutabilis*. *Applied Microbiology and*
524 *Biotechnology*: 1-10.
525

526 Johnson DB, Hallberg KB (2003). The microbiology of acidic mine waters. *Research in*
527 *Microbiology* **154**: 466-473.
528

529 Laemmli UK (1970). Cleavage of structural proteins during the assembly of the head of
530 bacteriophage T4. *Nature* **227**: 680-5.
531

532 Leblanc M, Achard B, Othman DB, Luck JM, Bertrand-Sarfati J, Personné JC (1996).
533 Accumulation of arsenic from acidic mine waters by ferruginous bacterial accretions
534 (stromatolites). *Applied Geochemistry* **11**: 541-554.
535

536 LeBlanc M, Casiot C, Elbaz-Poulichet F, Personné C. (2002). *Geological Society Special*
537 *Publication*, pp 267-274.
538

539 Lengfelder E, Elstner EF (1979). Cyanide insensitive iron superoxide dismutase in *Euglena*
540 *gracilis*. Comparison of the reliabilities of different test systems for superoxide dismutases.
541 *Zeitschrift fur Naturforschung. Section C: Biosciences* **34 C**: 374-380.

542
543 Noctor G, Bergot G, Mauve C, Thominet D, Lelarge-Trouverie C, Prioul JL (2007). A
544 comparative study of amino acid measurement in leaf extracts by gas chromatography-time of
545 flight-mass spectrometry and high performance liquid chromatography with fluorescence
546 detection. *Metabolomics* **3**: 161-174.
547
548 Olaveson MM, Nalewajko C (2000). Effects of acidity on the growth of two *Euglena* species.
549 *Hydrobiologia* **433**: 39-56.
550
551 Olaveson MM, Stokes PM (1989). Responses of the acidophilic alga *Euglena mutabilis*
552 (*Euglenophyceae*) to carbon enrichment at pH 3. *Journal of Phycology* **25**: 529-539.
553
554 Pentecost A (1982). The distribution of *Euglena mutabilis* in Sphagna, with reference to the
555 Malham Tarn North Fen. *Field Studies* **5**: 591-606.
556
557 Quéméneur M, Hamelin J, Latrille E, Steyer JP, Trably E (2010). Development and
558 application of a functional CE-SSCP fingerprinting method based on [Fe-Fe]-hydrogenase
559 genes for monitoring hydrogen-producing *Clostridium* in mixed cultures. *International*
560 *Journal of Hydrogen Energy* **35**: 13158-13167.
561
562 Scanlan PD, Marchesi JR (2008). Micro-eukaryotic diversity of the human distal gut
563 microbiota: Qualitative assessment using culture-dependent and -independent analysis of
564 faeces. *The ISME Journal* **2**: 1183-1193.
565
566 Seaver LC, Imlay JA (2001). Alkyl hydroperoxide reductase is the primary scavenger of
567 endogenous hydrogen peroxide in *Escherichia coli*. *Journal of Bacteriology* **183**: 7173-7181.
568
569 Shigeoka S, Nakano Y, Kitaoka S (1980). Metabolism of hydrogen peroxide in *Euglena*
570 *gracilis* Z by L-ascorbic acid peroxidase. *Biochemical Journal* **186**: 377-380.
571
572 Sittenfeld A, Mora M, Ortega JM, Albertazzi F, Cordero A, Roncel M *et al* (2002).
573 Characterization of a photosynthetic *Euglena* strain isolated from an acidic hot mud pool of a
574 volcanic area of Costa Rica. *FEMS Microbiology Ecology* **42**: 151-161.
575
576 Slyemi D, Moinier D, Brochier-Armanet C, Bonnefoy V, Johnson DB (2011). Characteristics
577 of a phylogenetically ambiguous, arsenic-oxidizing *Thiomonas* sp., *Thiomonas*
578 *arsenitoxydans* strain 3As(T) sp. nov. *Archive of microbiology* **6**: 439-449.
579
580 Tichy R, Rulkens WH, Grotenhuis JTC, Nydl V, Cuypers C, Fajtl J (1998). Bioremediation of
581 metals from soils or sediments. *Water Science and Technology* **37**: 119-127.
582
583 Waybrant KR, Ptacek CJ, Blowes DW (2002). Treatment of mine drainage using permeable
584 reactive barriers: Column experiments. *Environmental Science and Technology* **36**: 1349-
585 1356.
586
587 Weiss S, Carapito C, Cleiss J, Koechler S, Turlin E, Coppee J-Y *et al* (2009). Enhanced
588 structural and functional genome elucidation of the arsenite-oxidizing strain *Herminiimonas*
589 *arsenicoxydans* by proteomics data. *Biochimie* **91**: 192-203.
590
591
592

593 **Titles and legends to Figures**

594

595 **Table 1.** Exo-metabolites identified in the interstitial water of the sampling site. Each
596 metabolite was quantified in triplicates by using ribitol as an internal standard (AU =
597 Arbitrary Unit).

598

599 **Figure 1.** Diversity of the eukaryotic and prokaryotic microbial communities based on CE-
600 SSCP 18S/16S rRNA analysis. The microbial community at the sampling site presented a low
601 diversity (A) and seemed to be dominated by *E. mutabilis* (marked by the symbol *). For the
602 18S rRNA amplification, the signal specific to the *Euglena* 18S rRNA sequence raised from
603 40% of the total fluorescence before the fractionation step (A) to more than 83% after the
604 density gradient fractionation step (B). For the 16S rRNA sequence, these values are
605 respectively of less than 0.1% (A) to more than 60% (B).

606

607 **Figure 2.** Functional categories of the proteins identified *in situ* in *E. mutabilis* cells. Proteins
608 involved in primary production and organic matter biosynthesis are colored in green. The
609 number following the functional categories represents the number of proteins identified within
610 each of them. The list of the corresponding proteins is presented in Supplementary Table 2 and
611 3).

612

613 **Figure 3.** Heat map and hierarchical classification of endo- and exo-metabolites synthesized
614 and secreted by *E. mutabilis* after one or seven day incubation in minimal liquid media in
615 laboratory conditions (ac. = acid). Only the metabolites showing a significant (p value <0.05)
616 accumulation pattern after 7 day were represented. The analysis was performed in triplicates
617 and the color of each cell indicates the relative content of the corresponding metabolite. The
618 clustering analysis gathers metabolites with a similar variation pattern. Metabolites marked by

619 the symbol (*) were identified in the water at the sampling site The whole list of the
620 metabolites identified is presented in the Supplementary Table 1 and in the Supplementary
621 Figure 1.

622

623 **Figure 4.** Functional overview of *E. mutabilis* main metabolic traits determined from
624 proteomic and metabolomic data. For each metabolic pathway shown on the figure, examples
625 of metabolites are indicated (blue). Enzymes involved in those metabolic pathways are
626 represented by a red line when identified by the proteomic approach, a blue line otherwise. In
627 the extracellular medium, only the metabolites found in the water at the study site and shown
628 to be secreted by *E. mutabilis in vitro* were represented. Those possibly used as nutrient by
629 *Thiomonas* sp. and *Candidatus Fodinabacter communicans* according to our previous
630 analysis are represented in green and in purple, respectively (Bertin *et al.* 2011).

631

632

633 **Supplementary Figure 1.** *In vitro* metabolite variation in the presence of *Euglena* cells after
634 one week incubation in liquid mineral medium. The metabolite variations in the cell and in the
635 medium are represented in orange and in blue respectively (ac. = acid).

636

637 **Supplementary Figure 2.** One dimension SDS-PAGE of the proteins extracted from *Euglena*
638 cells after Nycodenz[®] density gradient fractionation step. Bands were systematically cut every
639 1.5 mm and analyzed individually by MS to facilitate the identification of the proteins.

640

641 **Supplementary Table 1.** Metabolites identified in this study. The relative amount of each
642 metabolite was calculated by using ribitol as an internal standard from triplicates.

643

644 **Supplementary Table 2.** Functional classes of the proteins identified in *Euglena* cells
645 recovered by a density gradient fractionation step.

646

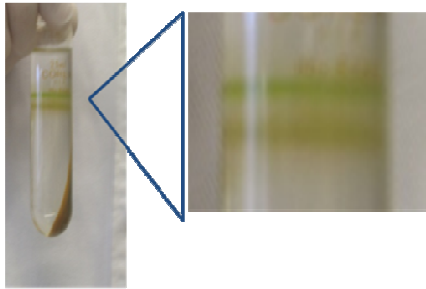
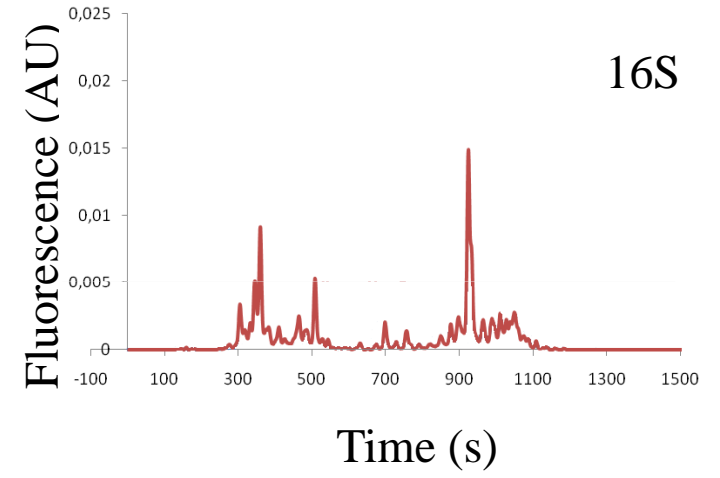
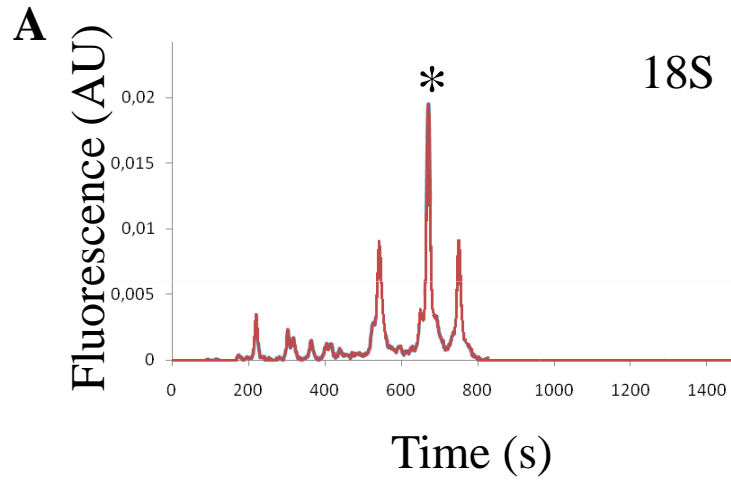
647 **Supplementary Table 3.** MS/MS identification results of separated proteins synthesized by *E.*
648 *mutabilis* in the Carnoulès creek.

Table 1

Metabolite	Quantity A.U.	+/- SD
Glycerol	11.079	2.699
Urea	4.130	0.947
Palmitic acid	1.909	0.597
Lactic acid	1.600	0.375
Glycolic acid	0.901	0.121
Putrescine	0.747	0.122
Stearic acid	0.477	0.160
Tetradecanoic acid	0.314	0.056
Lauric acid	0.246	0.043
Glycine	0.171	0.043
Octanoic acid	0.103	0.006
Tyramine	0.096	0.026
p-Hydroxybenzoic acid	0.084	0.016
Pipecolic acid	0.072	0.025
Decanoic acid	0.071	0.003
2-Furancarboxylic acid	0.054	0.009
Sucrose	0.052	0.050
Levogluconan	0.050	0.011
Maleic acid	0.047	0.014
Glucose	0.046	0.009
Galactosylglycerol	0.042	0.017
Fructose	0.034	0.005
Mannitol	0.031	0.004
Succinic acid	0.030	0.006
Trehalose	0.030	0.007
Fructose	0.029	0.007
Threonine	0.026	0.029



Environmental sample



Euglena mutabilis
after Nycodenz[®]
gradient

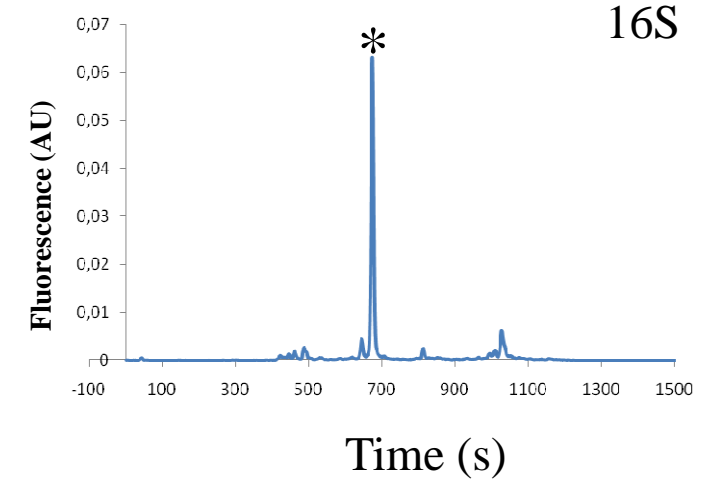
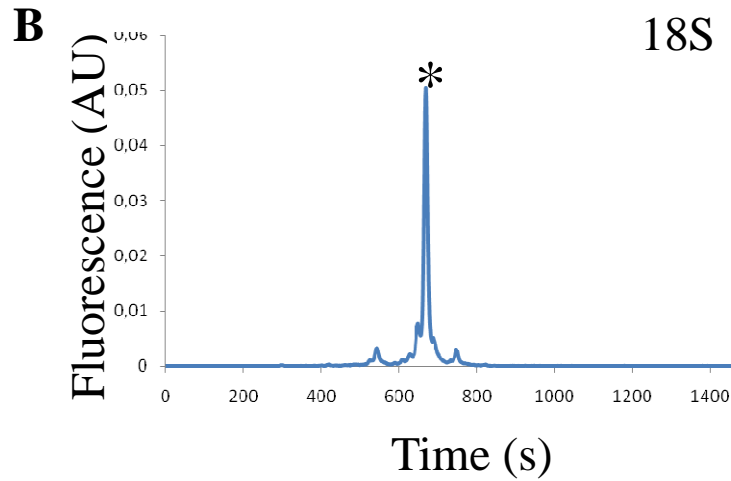


Figure 1

Figure 2

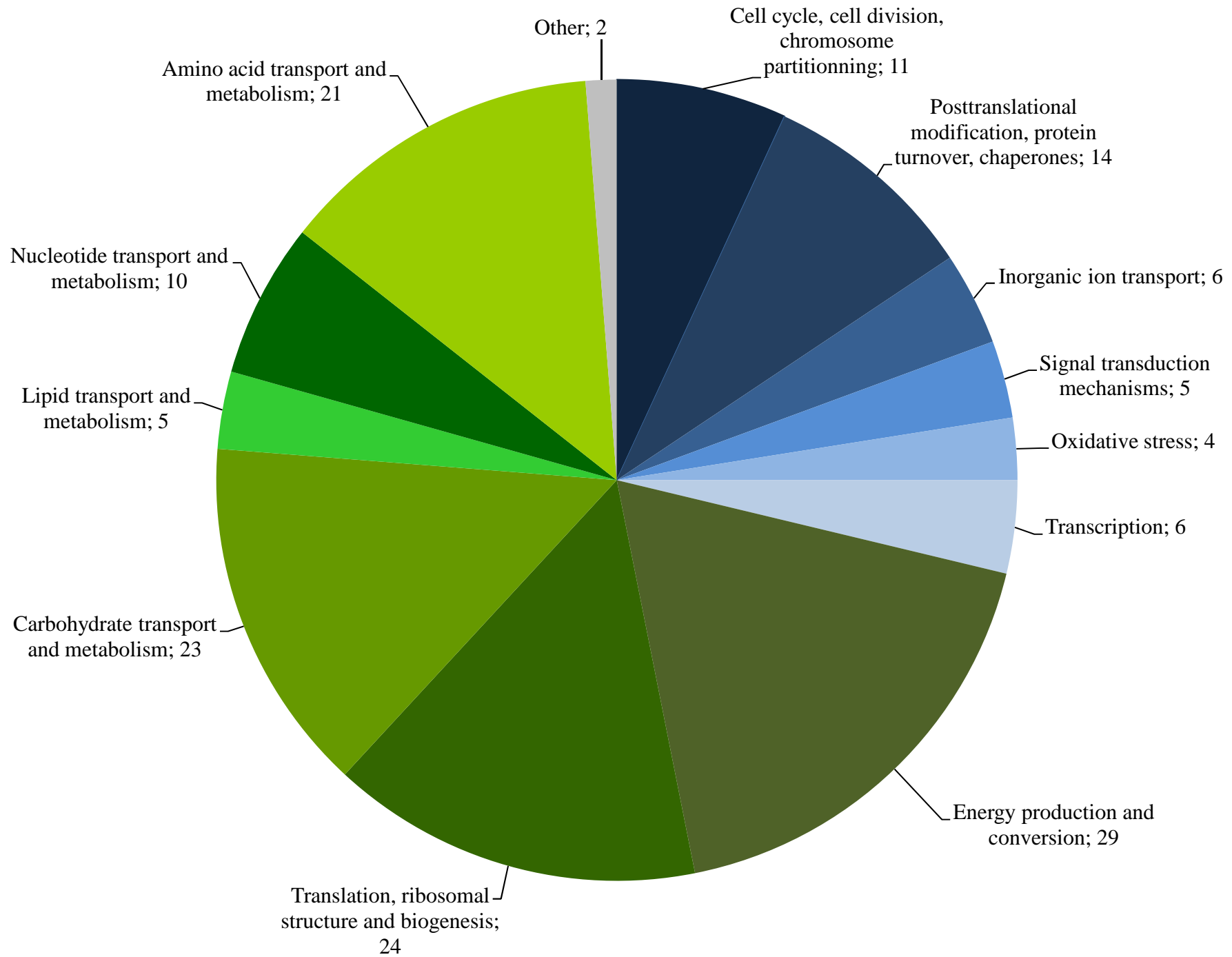
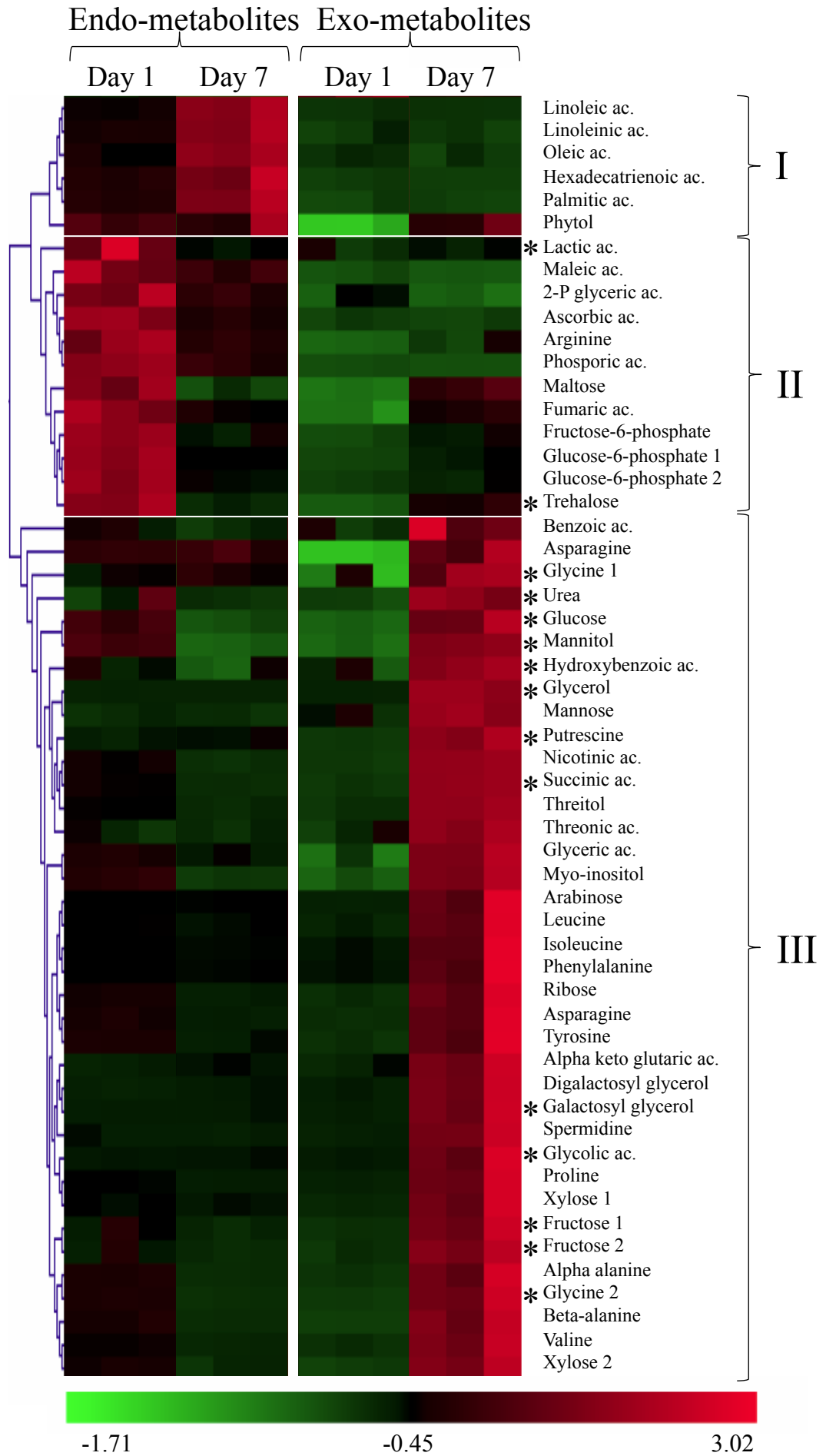


Figure 3



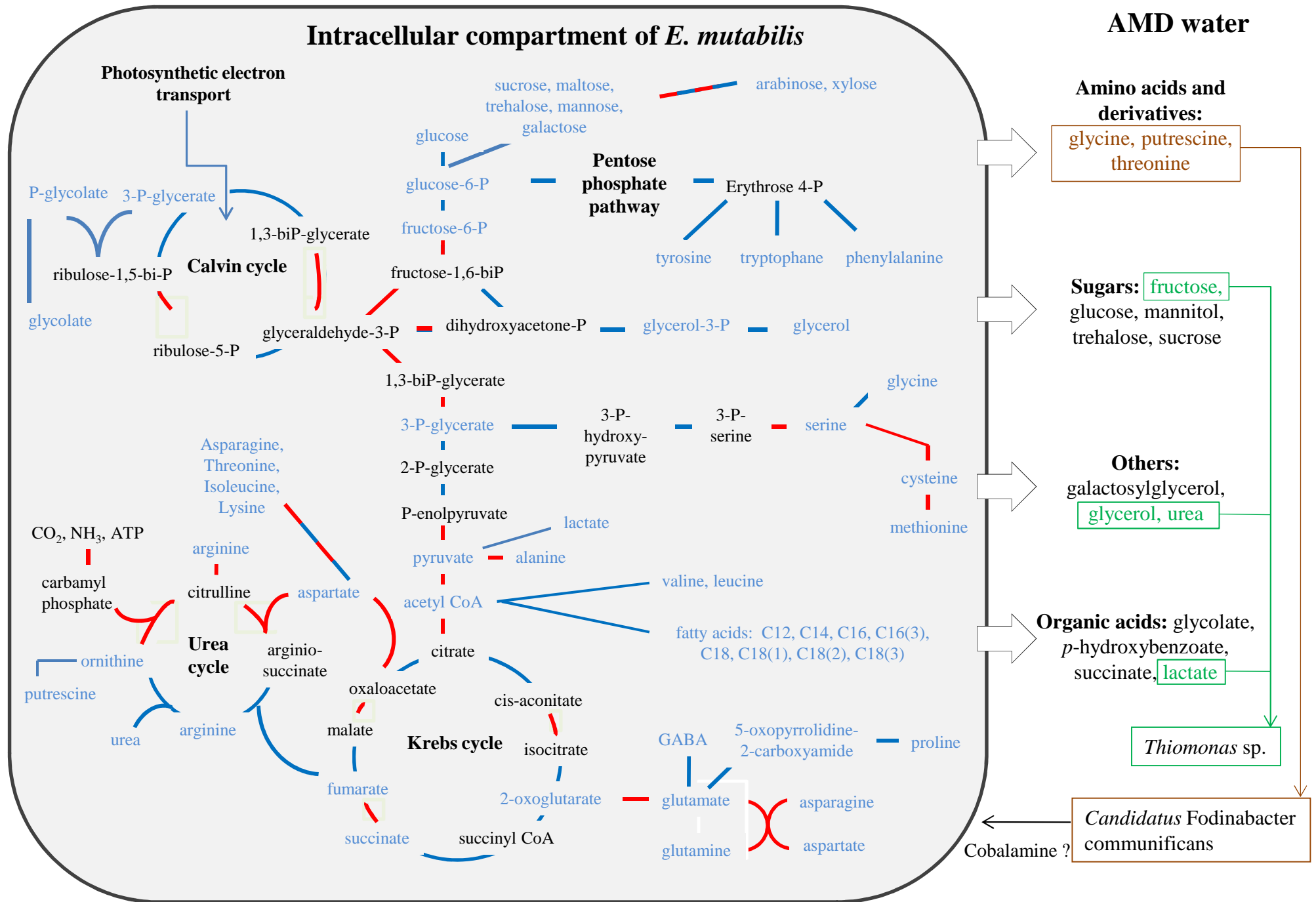
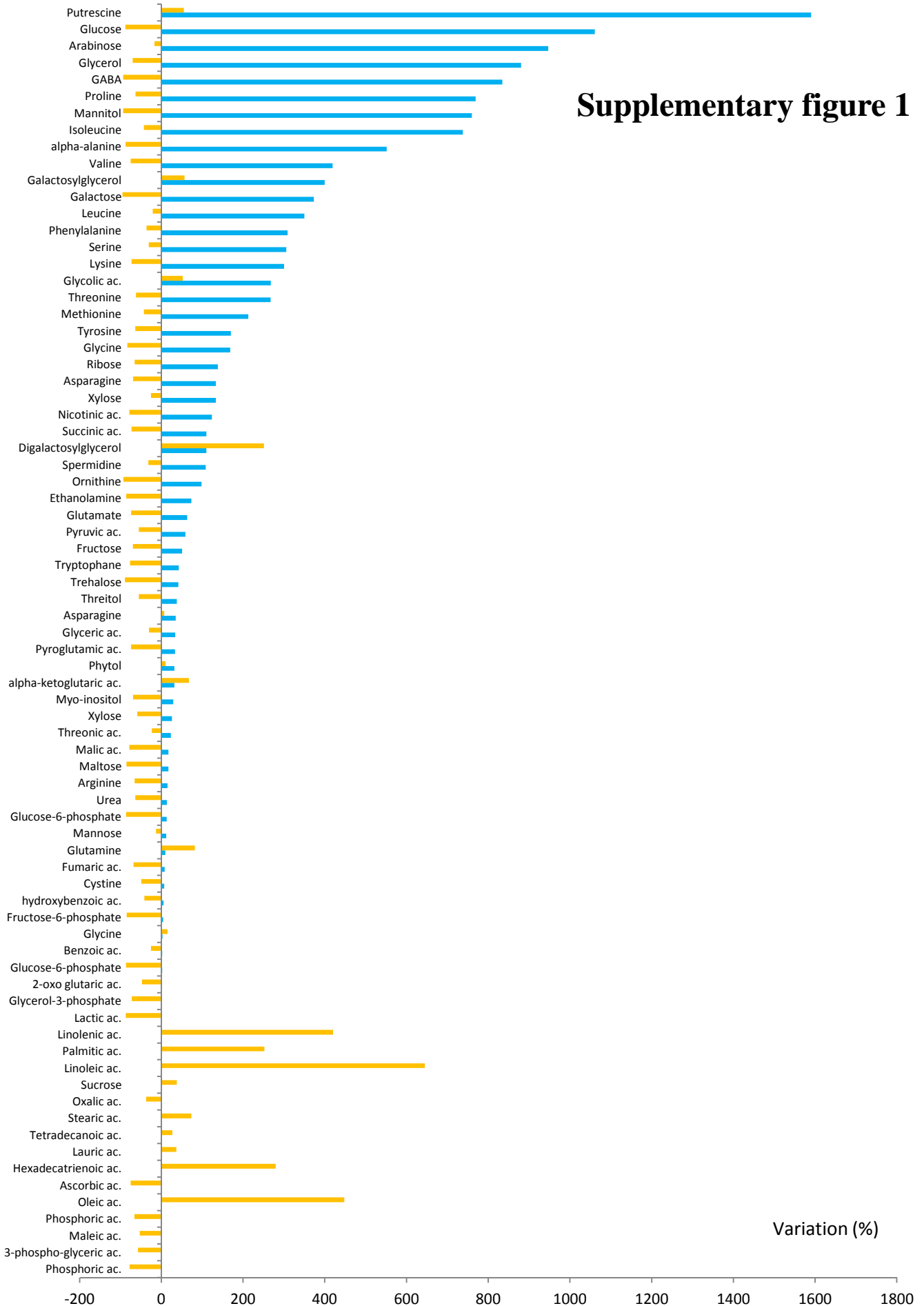
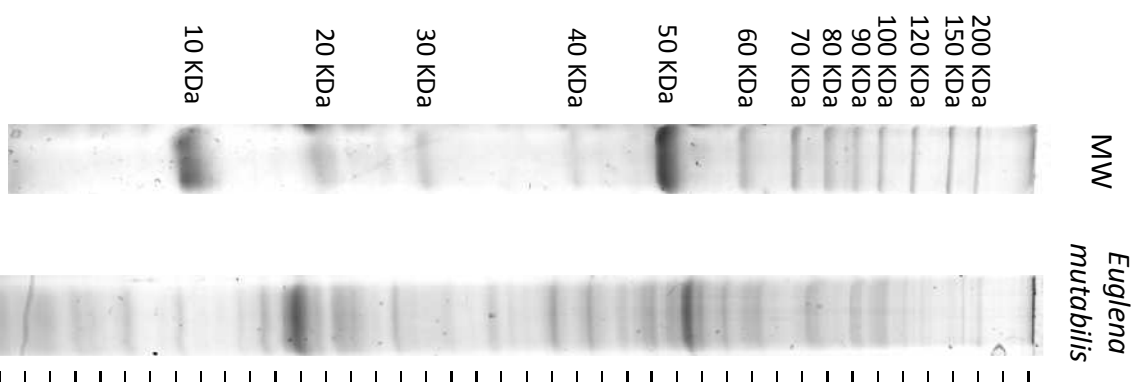


Figure 4

Supplementary figure 1



Supplementary figure 2



Supplementary table 1

metabolite	exo-metabolite <i>in situ</i>	endo-metabolite <i>in situ</i>	exo-metabolite <i>in vitro</i> (T=7j)	endo-metabolite <i>in vitro</i> (T=0)	endo-metabolite <i>in vitro</i> (T=7j)
2-Furancarboxylic acid	5,40E-05	0,00E+00	0,00E+00	0,00E+00	0,00E+00
2-Hydroxyglutaric acid	0,00E+00	0,00E+00	0,00E+00	3,07E-05	1,11E-04
3-Phosphoglyceric acid	0,00E+00	0,00E+00	0,00E+00	1,22E-04	2,17E-05
5-oxopyrrolidine 2 carboxamide	0,00E+00	3,95E-04	0,00E+00	0,00E+00	0,00E+00
Alpha-Alanine	0,00E+00	0,00E+00	0,00E+00	1,00E-02	4,08E-01
Alpha-KetoGlutaric acid	0,00E+00	0,00E+00	1,92E-06	7,79E-05	1,15E-03
Arabinose	0,00E+00	2,69E-02	2,45E-02	1,37E-03	1,14E-03
Arginine monohydrochloride	0,00E+00	2,10E-03	6,91E-04	5,34E-03	1,86E-03
Ascorbic acid	0,00E+00	9,53E-04	0,00E+00	7,25E-04	1,81E-04
Asparagine	0,00E+00	1,21E-03	1,55E-03	9,47E-04	1,01E-03
Aspartic acid	0,00E+00	1,17E-03	9,51E-03	1,65E-03	5,08E-04
Benzoic acid	0,00E+00	3,89E-04	2,24E-04	1,24E-04	9,27E-05
Beta-alanine	0,00E+00	1,53E-03	8,28E-03	1,76E-03	4,23E-04
Cystine	0,00E+00	2,14E-04	4,19E-04	3,57E-04	1,81E-04
Decanoic acid	7,12E-05	5,13E-04	0,00E+00	0,00E+00	0,00E+00
Digalactosylglycerol	0,00E+00	5,68E-04	6,97E-03	5,21E-05	1,83E-04
Ethanolamine	0,00E+00	2,18E-02	2,12E-02	1,82E-02	2,54E-03
Fructose 1	3,38E-05	2,90E-04	1,94E-03	3,06E-04	9,27E-05
Fructose 2	2,93E-05	2,96E-04	1,58E-03	2,27E-04	7,20E-05
Fructose-6-P	0,00E+00	0,00E+00	0,00E+00	1,47E-04	1,46E-04
Fumaric acid	0,00E+00	0,00E+00	0,00E+00	2,31E-04	2,71E-04
GABA	0,00E+00	2,19E-02	9,85E-02	6,56E-02	4,38E-03
Galactose	0,00E+00	6,47E-02	2,99E-02	5,12E-02	2,62E-03
Galactosylglycerol	4,17E-05	4,45E-03	5,64E-02	7,48E-04	1,17E-03
Glucose 1	4,64E-05	1,07E-02	1,21E-01	6,46E-02	7,71E-03
Glucose 2	0,00E+00	6,24E-03	4,38E-02	1,31E-02	1,36E-03
Glucose-6-P 1	0,00E+00	0,00E+00	0,00E+00	3,18E-04	2,61E-04
Glucose-6-P 2	0,00E+00	0,00E+00	0,00E+00	6,97E-05	5,44E-05
Glutamic acid	0,00E+00	2,56E-03	5,14E-03	2,33E-03	6,01E-04
Glutamine	0,00E+00	2,03E-03	4,33E-03	7,98E-04	1,46E-03
Glyceric acid	0,00E+00	0,00E+00	0,00E+00	2,19E-04	9,76E-04
Glycerol 3 phosphate	0,00E+00	1,47E-03	1,15E-03	1,67E-03	4,59E-04
Glycerol	1,11E-02	8,10E-03	4,29E-01	1,52E-03	4,56E-04
Glycine	1,71E-04	4,67E-03	1,72E-02	3,68E-03	6,17E-04
Glycolic acid	9,01E-04	0,00E+00	9,42E-03	9,84E-05	1,50E-04
Hexadecatrienoic acid	0,00E+00	1,86E-02	0,00E+00	7,01E-04	2,66E-03
Isoleucine	0,00E+00	1,79E-02	6,14E-02	2,26E-03	1,30E-03
Lactic acid	1,60E-03	1,81E-03	2,38E-05	3,05E-03	3,98E-04
Lauric acid	2,46E-04	4,17E-03	0,00E+00	4,14E-04	5,66E-04
Leucine	0,00E+00	0,00E+00	0,00E+00	3,86E-03	6,46E-02
Levoglucofan	4,99E-05	0,00E+00	0,00E+00	0,00E+00	0,00E+00
Linoleic acid	0,00E+00	3,07E-04	0,00E+00	3,15E-04	2,35E-03
Linolenic acid	0,00E+00	1,58E-03	0,00E+00	2,16E-04	1,13E-03
Lysine	0,00E+00	1,54E-02	1,87E-02	1,26E-02	3,43E-03
Maleic acid	4,67E-05	6,68E-04	0,00E+00	4,14E-04	1,95E-04
Malic acid	0,00E+00	0,00E+00	0,00E+00	2,28E-04	8,06E-04
Maltose	0,00E+00	0,00E+00	0,00E+00	1,11E-04	4,25E-04
Mannitol	3,07E-05	2,82E-02	1,64E-01	9,56E-02	6,46E-03
Mannose	0,00E+00	0,00E+00	4,73E-05	3,80E-05	7,15E-04

Methionine	0,00E+00	5,81E-03	3,45E-03	3,71E-04	2,12E-04
Myo-inositol	0,00E+00	8,32E-04	1,34E-03	5,27E-04	1,64E-04
Nicotinic acid	0,00E+00	6,44E-04	3,79E-03	5,72E-04	1,25E-04
Octadecanoic acid	0,00E+00	7,43E-04	0,00E+00	0,00E+00	0,00E+00
Octanoic acid	1,03E-04	2,36E-04	0,00E+00	0,00E+00	0,00E+00
Oleic acid	0,00E+00	0,00E+00	0,00E+00	5,75E-04	3,31E-05
Ornithine	0,00E+00	5,87E-03	1,45E-02	1,85E-02	1,31E-03
Oxalic acid	0,00E+00	0,00E+00	0,00E+00	8,41E-04	9,56E-04
Palmitic acid	1,91E-03	5,42E-02	1,97E-04	1,26E-02	4,44E-02
Phenylalanine	0,00E+00	6,79E-03	2,18E-02	7,25E-04	4,63E-04
Phosphoric acid	0,00E+00	1,60E-03	0,00E+00	2,64E-02	5,86E-03
Phosphoric acid	0,00E+00	0,00E+00	0,00E+00	2,78E-04	8,10E-06
p-Hydroxybenzoic acid	8,39E-05	2,41E-04	1,59E-04	5,73E-05	3,35E-05
Phytol	0,00E+00	2,15E-03	7,42E-04	8,15E-04	9,00E-04
Pipecolic acid	7,22E-05	7,99E-04	0,00E+00	0,00E+00	0,00E+00
Proline	0,00E+00	3,92E-02	1,84E-01	8,67E-03	3,19E-03
Putrescine	7,47E-04	2,99E-02	2,39E-01	1,56E-02	2,42E-02
Pyroglutamic acid	0,00E+00	5,99E-04	1,26E-03	6,65E-04	1,74E-04
Pyruvic acid	0,00E+00	0,00E+00	-6,51E-04	3,04E-04	1,31E-03
Ribose	0,00E+00	3,51E-03	4,47E-03	7,22E-04	2,50E-04
Serine	0,00E+00	8,18E-03	1,36E-02	1,28E-03	8,81E-04
Spermidine	0,00E+00	2,48E-04	4,20E-03	1,03E-04	7,03E-05
Stearic acid	4,77E-04	5,38E-03	0,00E+00	3,53E-03	6,11E-03
Succinic acid	3,03E-05	1,06E-03	1,05E-02	1,41E-03	3,85E-04
Sucrose	5,22E-05	0,00E+00	7,85E-06	1,37E-04	1,88E-04
Tetradecanoic acid	3,14E-04	6,08E-03	3,88E-06	8,68E-04	1,10E-03
Threitol	0,00E+00	0,00E+00	0,00E+00	3,62E-05	6,95E-04
Threonic acid	0,00E+00	0,00E+00	2,97E-06	3,54E-05	4,87E-04
Threonine	2,63E-05	5,59E-03	1,74E-02	1,01E-03	3,80E-04
trans ou cis Sinapinic acid	0,00E+00	2,73E-04	0,00E+00	0,00E+00	0,00E+00
Trehalose	3,01E-05	2,44E-03	1,06E-03	3,96E-03	4,42E-04
Tryptophan	0,00E+00	4,20E-03	2,11E-03	5,29E-03	1,24E-03
Tyramine	9,63E-05	0,00E+00	0,00E+00	0,00E+00	0,00E+00
Tyrosine	0,00E+00	1,35E-02	6,86E-03	1,40E-03	5,11E-04
Urea	4,13E-03	9,27E-04	1,66E-03	5,78E-04	2,10E-04
Valine	0,00E+00	4,83E-02	1,78E-01	2,12E-02	5,34E-03
Xylose 1	0,00E+00	5,53E-03	4,35E-03	2,98E-04	2,23E-04
Xylose 2	0,00E+00	3,02E-03	6,54E-04	1,42E-04	5,84E-05

Supplemenatry table 2

Cell cycle, cell division, chromosome partitioning	Lipid transport and metabolism
actin	acetyl CoA acetyltransferase
alpha-tubulin	dihydrolipoyl dehydrogenase
articulin	farnesyl pyrophosphate synthetase
beta-tubulin	L-3-hydroxyacyl-CoA dehydrogenase
cell division cycle 48 ATPase	zeta-carotene desaturase
cell division protease ftsH	
chromosome segregation protein SMC - <i>Paracoccus denitrificans</i>	Posttranslational modification, protein turnover, chaperones
fibrillarin	20S proteasome component alpha 1
histone H2A	20S proteasome component alpha 2
histone H3 - <i>Arabidopsis thaliana</i>	60 kDa chaperonin 1 - <i>Nitrobacter hamburgensis</i>
kinesin K39	ADP-ribosylation factor
S-phase kinase-associated protein	alanyl-tRNA synthetase
tubulin binding protein	ATP-dependent metalloproteinase
	chaperonin CPN60, mitochondrial precursor
Replication, recombination and repair	heat Shock Protein 70
ATP-dependent helicase HrpA - <i>Vibrio splendidus</i>	heat shock protein 90
DNA ligase, NAD-dependent - <i>Shewanella woodyi</i>	heat shock protein, 74 kDa
thidylate kinase - <i>Stappia aggregata</i>	peptidyl-prolyl cis-trans isomerase/cyclophilin
	proteasome regulatory ATPase subunit cc118.3
Nucleotide transport and metabolism	small ubiquitin protein
adenylate kinase	ubiquitin-conjugating enzyme
ATP/ADP translocator	
ferredoxin-nadp reductase	Energy production and conversion
GTPase	ATPase alpha subunit
GTP-binding protein	ATPase beta subunit
NAD(P) transhydrogenase, mitochondrial precursor	ATPase gamma subunit precursor
NAD+ synthase	carbonic anhydrase
NADH dehydrogenase subunit B	Chloro. carbonic anhydrase
plastidic ATP/ADP transporter	Chloro. fructose-1,6-bisphosphatase
pyridine nucleotide transhydrogenase beta subunit	Chloro. light-harvesting cpx I protein precursor Lhca1
	Chloro. light-harvesting cpx II protein precursor Lhcbm1
Translation, ribosomal structure and biogenesis	Chloro. light-harvesting cpx II protein precursor Lhcbm3
30S ribosomal protein S7 - <i>Neisseria meningitidis</i>	Chloro. light-harvesting cpx II protein precursor Lhcbm5
40S ribosomal protein S14	Chloro. light-harvesting cpx II protein precursor Lhcbm6
40S ribosomal protein S15A	Chloro. phosphoglycerate kinase precursor
40S ribosomal protein S23	Chloro. phosphoribulokinase
40S ribosomal protein S5	Cytochrome b559 subunit alpha
40S ribosomal protein S9	Cytochrome b6-f cpx
50S ribosomal protein L1 - <i>Legionella pneumophila</i>	electron transfer flavoprotein beta-subunit
60S ribosomal protein L10	electron transfer flavoprotein, alpha subunit

60S ribosomal protein L12	Ferredoxin--NADP reductase, Chloro. precursor
60S ribosomal protein L34	NADH:protochlorophyllide oxidoreductase
60S ribosomal protein L7	oxygen-evolving enhancer protein 1 precursor
60S ribosomal putative protein RPL23A	PS I P700 chlorophyll a apoprotein A1
elongation factor 1-alpha	PS I reaction center subunit II (PsaD)
Elongation factor 2 (Fragment)	PS II 44 kDa reaction center protein
elongation factor 2	PS II CP47 protein (Fragment) - <i>Ephedra sinica</i>
elongation factor EF-3	PS II D2 protein (Fragment) - <i>Philesia magellanica</i>
elongation factor TS	PS II oxygen evolving enhancer
elongation factor Tu	PS II chlorophyll A apoprotein - <i>Nephroselmis olivacea</i>
eukaryotic initiation factor 4	PS Q(B) protein - <i>Euglena mutabilis</i>
non-ribosomal peptide synthase - <i>Bradyrhizobium</i>	porphobilinogen deaminase, Chloro. precursor
ribosomal DEAD box protein	trimethylamine-N-oxide reductase TorA
ribosomal protein 17 50S large ribosomal subunit	uroporphyrinogen decarboxylase
60S ribosomal protein L11	Ribulose biphosphate carboxylase large chain
60S ribosomal protein L9	RuBisCO small chain (Fragment) - <i>Acetabularia cliftonii</i>
40S ribosomal protein S13	RuBisCO activase
40S ribosomal protein S18	
RplB - <i>Clostridium kluyveri</i>	Signal transduction mechanisms
rubber elongation factor protein - <i>Hevea brasiliensis</i>	calmodulin
T-cpx protein 1	CAMP-dependent protein kinase catalytic subunit
translation elongation factor G - <i>Wolbachia</i> endosymbiont of <i>Drosophila ananassae</i>	family 14-3-3 [Signal transduction mechanisms]
	glucose-regulated protein 78 - <i>Trypanosoma cruzi</i>
Transcription	protein kinase
ATP-dependent RNA helicase dhh1	α subunit of photoactivated adenylyl cyclase - <i>Colacium sideropus</i>
ATP-dependent RNA helicase-like protein DB10	B subunit of photoactivated adenylyl cyclase - <i>Euglena stellata</i>
DEAD-box ATP-dependent RNA helicase 46	serine/threonine protein kinase
DEAD-box ATP-dependent RNA helicase 8 - <i>Arabidopsis thaliana</i>	
polyadenylate-binding protein	Inorganic ion transport and metabolism
prohibitin 1-like protein (ISS)	F-ATPase family transporter: protons (vacuolar)
RNA-binding region-containing protein	H+-PPase family transporter: proton
	magnesium chelatase ATPase subunit D
stress	magnesium chelatase, subunit H
alkyl hydroperoxide reductase	magnesium-chelatase subunit chlI
ascorbate peroxidase	MgtA, Cation transport ATPase
superoxide dismutase	H+ translocating pyrophosphatase – <i>Ostreococcus sp.</i>
thioredoxin - <i>Escherichia coli</i>	rusticyanin precursor - <i>Thiobacillus ferrooxidans</i>
thioredoxin peroxidase	
	amino acid transport and metabolism
carbohydrate transport and metabolism	5-methyltetrahydropteroyltriglutamate--homocysteine S-methyltransferase - <i>Marinobacter aquaeolei</i>

acetyl CoA carboxylase	adenosylhomocysteinase A
acetyl-CoA acetyltransferase	aminotransferase
aconitate hydratase 3	aminotransferase, class V
acyl-CoA dehydrogenase	arginine deiminase
aldo-keto oxidoreductase	argininosuccinate synthase
cytosolic phosphoglycerate kinase	asparagine synthetase A
cytosolic triosephosphate isomerase	aspartate aminotransferase
enolase [Euglena gracilis]	carbamate kinase
fructose-1,6-bisphosphate aldolase	Glu/Leu/Phe/Val dehydrogenase <i>Comamonas</i> sp.
GDP-mannose 4,6 dehydratase	glutamate dehydrogenase
glyceraldehyde-3-phosphate dehydrogenase	glutamate-1-semialdehyde aminotransferase
ketol-acid reductoisomerase	glutamine synthetase type III
malate deshydrogenase	glycine cleavage system H protein
malate synthase-isocitrate lyase	glycine dehydrogenase
NAD dependent malic enzyme	hydroxymethylglutaryl-CoA synthase
phosphoenolpyruvate carboxykinase - <i>Beggiatoa</i> sp.	methionine synthase II
phosphoglycerate kinase - <i>Gramella forsetii</i>	ornithine carbamoyltransferase family protein
plastid transketolase [Euglena gracilis]	phosphoserine aminotransferase
methylmalonyl-CoA decarboxylase <i>Streptomyces</i> sp.	saccharopine dehydrogenase
fructose 6-phosphate-1-phosphotransferase	S-adenosylhomocysteine hydrolase
pyruvate dehydrogenase, mitochondrial precursor	S-adenosyl-L-methionine synthetase
pyruvate kinase	serine hydroxymethyltransferase
pyruvate: NADP+ oxidoreductase	threonine synthase - <i>Sinorhizobium saheli</i>
sedoheptulose-1,7-bisphosphatase	
succinate dehydrogenase flavoprotein	Other
UDP-glucose dehydrogenase	pyridoxine biosynthesis protein
	symbiosis-related protein

CHAPITRE III :

LA REPONSE ADAPTATIVE
***D'EUGLENA MUTABILIS* FACE AU**
STRESS ARSENIE PAR DES
APPROCHES DE PHYSIOLOGIE
COMPARATIVE

A : INTRODUCTION

Les analyses de protéomique effectuées *in situ* sur *E. mutabilis* (chapitre II) ont permis de mettre en évidence de multiples mécanismes de défense face au stress oxydatif, ce qui pourrait expliquer, au moins en partie, l'adaptation de ce protiste à ce métalloïde décrit pour induire de telles contraintes (Jin *et al.*, 2008; Wysocki & Tamas, 2010). Cependant, l'expression de protéines impliquées dans la réponse au stress oxydatif ne permet pas à elle seule d'expliquer l'hyper-tolérance à l'arsenic chez *E. mutabilis*. En effet, le caractère toxique de cet élément ne se limite pas seulement au stress oxydatif qui lui est associé, mais est également lié à l'analogie structurale de l'arséniate avec le phosphate et à la très forte affinité de l'arsénite pour des fonctions thiols des protéines (Berhe *et al.*, 1995).

Afin de comprendre plus en détail la réponse à l'arsenic de ce protiste, des observations physiologiques ont été réalisées en conditions de laboratoire, et comparées à celles d'un organisme phylogénétiquement apparenté mais non retrouvé dans les DMA, *E. gracilis*. Cette espèce est en effet utilisée depuis plusieurs décennies en laboratoire pour étudier la transition entre le métabolisme autotrophe et hétérotrophe chez les euglènes et, d'une manière plus générale, sur le métabolisme photosynthétique. La physiologie de cet organisme est donc relativement bien connue et des données sur le métabolisme de l'arsenic ont été récemment décrites dans la littérature scientifique. Ainsi, des analyses XANES (X Ray Absorption Near Edge Spectra) permettant de déterminer la nature des liaisons de coordination au sein des molécules ont souligné le fait que l'arsenic localisé dans le compartiment intracellulaire d'*E. gracilis* était complexé à du soufre (Miot *et al.*, 2008). Ces observations suggèrent que des composés tels le glutathion ou les phytochélatines pourraient être impliqués dans la défense face à l'arsenic chez ce protiste. Cette hypothèse est d'ailleurs appuyée par une plus grande sensibilité d'*E. gracilis* à l'arsenic en présence de buthionine sulfoximine, un inhibiteur de la synthèse de glutathion. D'autre part, l'arséniate semble être chez *E. gracilis* préférentiellement accumulé au niveau des membranes tandis que l'arsénite est plutôt localisé dans le cytoplasme des cellules (Miot *et al.*, 2009). Les différentes localisations subcellulaires de ces deux formes inorganiques de l'arsenic pourraient être liées à leurs propriétés physico-chimiques spécifiques. En effet, l'arséniate pourrait remplacer les phosphates de phospholipides membranaires, tandis que l'arsénite, présentant une très forte affinité pour les fonctions thiols, serait plutôt associé aux fonctions thiols des protéines et du glutathion au sein du cytoplasme (Miot *et al.*, 2009). Ces observations soulignent que l'arsenic, en interagissant spécifiquement avec différents constituants intracellulaires, pourrait perturber le fonctionnement global de la cellule ce

qui expliquerait la très forte inhibition de croissance d'*E. gracilis* en présence d'arsenic (42% de la valeur de croissance maximale à 250 ppm d'arsenic)(Miot *et al.*, 2008). Cependant, alors qu'*E. mutabilis* est retrouvé fréquemment dans les DMA, *E. gracilis* ne colonise pas ces environnements suggérant que des mécanismes de défense différents et/ou plus adaptés ont été développés par *E. mutabilis* pour pallier le stress arsénié très marqué dans le DMA de Carnoulès.

Des études comparatives de la réponse face au stress arsénié ont donc été réalisées chez *E. mutabilis* et *E. gracilis*. En particulier, la tolérance à l'arsenic, l'accumulation de ce métalloïde au sein des cellules, sa spéciation intracellulaire et extracellulaire, son impact sur les membranes cellulaires et sur la mobilité des cellules ont été étudiés. Les principaux résultats de cette étude sont présentés sous forme de manuscrit publié dans *Applied Microbiology and Biotechnology* (voir fin de chapitre).

B. RESULTATS

Les observations physiologiques réalisées sur *E. mutabilis* et *E. gracilis* ont souligné d'importantes différences au niveau de la réponse à l'arsenic entre ces deux espèces. En particulier, la mobilité par gliding a été montrée comme étant spécifiquement induite en présence de ce métalloïde chez *E. mutabilis* mais pas chez *E. gracilis*. Cette observation est à mettre en regard de l'augmentation de la fluidité membranaire montrée comme étant positivement corrélée avec la concentration extracellulaire d'arsenic et résultant d'une plus forte proportion d'acides gras polyinsaturés dans les membranes.

Par ailleurs, des analyses réalisées sur les membranes des euglènes ont souligné le caractère très hydrophobe des propriétés de surface d'*E. mutabilis* en comparaison à celles d'*E. gracilis*. Cette spécificité pourrait jouer un rôle dans la tolérance aux ions métalliques, et en particulier à l'arsénite. En effet, *E. mutabilis* présente une très forte tolérance à ce composé mais se révèle être, en comparaison à *E. gracilis*, beaucoup plus sensible à l'oxyde de phénylarsine, composé arsénié hydrophobe. La surface cellulaire hydrophobe pourrait ainsi expliquer, au moins en partie, la moindre accumulation de composés hydrophile comme l'arsénite dans le compartiment intracellulaire d'*E. mutabilis* en comparaison à *E. gracilis*. Par ailleurs, les deux espèces d'euglène ont visiblement développé des métabolismes de l'arsenic distincts, *E. gracilis* accumulant des formes arséniées méthylées qui ne sont pas retrouvées chez *E. mutabilis*.

C : DISCUSSION ET CONCLUSION

Les observations au niveau physiologique soulignent un métabolisme de l'arsenic original chez *E. mutabilis* par rapport à *E. gracilis*. L'induction de la mobilité observée chez *E. mutabilis* en présence d'arsenic pourrait faciliter la réorganisation de sa population sous forme de biofilm tel que cela est observé sur le site de Carnoulès et ainsi jouer un rôle important dans la tolérance à ce composé. D'autre part, si les propriétés hydrophobes de surface d'*E. mutabilis* semblent jouer un rôle dans les processus de défense face aux ions métalliques, ce protiste a probablement dû développer des mécanismes lui permettant d'accéder aux composés hydrophiles indispensables à sa croissance (sulfate, phosphate,...).

Afin de comprendre plus en détails les mécanismes moléculaires conduisant à ces phénotypes ainsi que la réponse à l'arsenic des euglènes d'une manière plus globale, des approches de génomique fonctionnelle ont été développées.

D: MANUSCRIT

L'intégralité de ces résultats ont conduit à une publication scientifique dans « Applied Microbiology and Biotechnology » dont le manuscrit correspondant est présenté ici.

Surface properties and intracellular speciation revealed an original adaptive mechanism to arsenic in the acid mine drainage bio-indicator *Euglena mutabilis*

David Halter · Corinne Casiot · Hermann J. Heipieper · Frédéric Plewniak · Marie Marchal · Stéphane Simon · Florence Arsène-Ploetze · Philippe N. Bertin

Received: 9 June 2011 / Revised: 5 July 2011 / Accepted: 15 July 2011
© Springer-Verlag 2011

Abstract *Euglena mutabilis* is a protist ubiquitously found in extreme environments such as acid mine drainages which are often rich in arsenic. The response of *E. mutabilis* to this metalloid was compared to that of *Euglena gracilis*, a protist not found in such environments. Membrane fatty acid composition, cell surface properties, arsenic accumulation kinetics, and intracellular arsenic speciation were determined. The results revealed a modification in fatty acid composition leading to an increased membrane fluidity in both *Euglena* species under sublethal arsenic concentrations exposure. This increased membrane fluidity correlated to an induced gliding motility observed in *E. mutabilis* in the presence of this metalloid but did not affect the flagellar dependent motility of *E. gracilis*. Moreover, when compared to *E. gracilis*, *E. mutabilis* showed highly hydrophobic cell surface properties and a higher tolerance to water-soluble arsenical compounds but not to hydropho-

bic ones. Finally, *E. mutabilis* showed a lower accumulation of total arsenic in the intracellular compartment and an absence of arsenic methylated species in contrast to *E. gracilis*. Taken together, our results revealed the existence of a specific arsenical response of *E. mutabilis* that may play a role in its hypertolerance to this toxic metalloid.

Keywords Arsenic speciation · *Euglena mutabilis* · *Euglena gracilis* · Hydrophobicity · Fatty acid · Unsaturation index · AMD

Introduction

Euglena mutabilis is a photosynthetic protist ubiquitously found in extreme environments such as acid mine drainages (AMD, Brake et al. 2001; Bray et al. 2008; Wolowski et al.

D. Halter · F. Plewniak · M. Marchal · F. Arsène-Ploetze · P. N. Bertin (✉)
Laboratoire de Génétique Moléculaire, Génomique et Microbiologie, Département Microorganismes, Génomes, Environnement UMR7156 Université de Strasbourg/CNRS, 28 rue Goethe, 67083 Strasbourg Cedex, France
e-mail: philippe.bertin@unistra.fr

D. Halter
e-mail: david.halter@gem.u-strasbg.fr

F. Plewniak
e-mail: f.plewniak@unistra.fr

M. Marchal
e-mail: marchal.m@unistra.fr

F. Arsène-Ploetze
e-mail: ploetze@unistra.fr

C. Casiot
Laboratoire Hydrosociences Montpellier, UMR 5569 (CNRS, IRD, UM1, UM2), Université Montpellier 2, Place E. Bataillon, 34095 Montpellier Cedex 05, France
e-mail: casiot@msem.univ-montp2.fr

H. J. Heipieper
Department Environmental Biotechnology, Helmholtz Centre for Environmental Research—UFZ, Permoserstr. 15, Leipzig, Germany
e-mail: hermann.heipieper@ufz.de

S. Simon
Groupement de Recherche Eau Sol Environnement Faculté des Sciences et Techniques 123, Avenue Albert Thomas, 87060 Limoges, France
e-mail: stephane.simon@unilim.fr

2008) and may be considered as a biological indicator of AMD-impacted sites (Valente and Gomes 2007). These sites often exhibit acidic pH and high concentrations of metals like arsenic, making them hostile for most of the living cells. The most common defense mechanisms against arsenic stress found in eukaryotic cells are the sequestration as arsenic–glutathione or arsenic–phytochelatin complexes and the modification of arsenic species to less toxic or bioavailable forms (Pawlik-Skowronska et al. 2004; Qin et al. 2009; Tsai et al. 2009). For example, *Cyanidioschyzon*, an alga often present in arsenic polluted environments, is able to oxidize, reduce, and methylate arsenical compounds, the latter process leading to the generation of arsenic volatile species such as trimethylarsine oxide (TMAO).

Metal often detected in AMD-exposed ecosystems such as chromium and cadmium has been shown to induce cellular damage in *Euglena gracilis*, especially regarding fatty acid composition and membrane organization (Einicker-Lamas et al. 1996; Rocchetta et al. 2006) or by affecting its flagellar dependent motility (Stallwitz and Hader 1994). In this protist, arsenic detoxification pathways involve probably its complexation to low-molecular weight compounds such as glutathione or phytochelatin (Miot et al. 2008). Unlike *E. mutabilis*, *E. gracilis* has not been found in AMD and is less tolerant to metal exposure.

E. mutabilis has been described as the dominant eukaryotic species in an AMD-exposed ecosystem, i.e., the Carnoulès ancient mining exploitation (southern France). Mining activities left 1.5 Mt of tailing materials containing As-rich sulfides. These materials were deposited over the spring of the small Reigous creek, leading to a water acidification down to pH 3 and enrichment in metals and arsenic as a result of sulfide oxidation (Casiot et al. 2003, 2009). At the spring of Reigous creek, *E. mutabilis* formed in this environment green mats on the riverbed (Bruneel et al. 2006) where arsenic concentration reached 4.66 mM, mainly in the reduced form arsenite As(III) (Casiot et al. 2003). To get insight into the adaptive responses of *E. mutabilis* to this arsenical stress, this protist was isolated and grown in vitro in arsenite-rich medium. The cell surface properties, kinetics of arsenite accumulation, and speciation of arsenical compounds in both the intracellular compartment and the media were monitored during cell growth and compared to the results obtained from similar experiments performed on *E. gracilis*.

Materials and methods

Biological material and sampling processing

Twenty grams of sediments was sampled in the Reigous creek, 40 m downstream of the spring, which corresponds

to the main growth zone of *E. mutabilis* (Casiot et al. 2004). This sampling site was characterized by acid pH (2.4–5.7), extremely high Fe (12–25 mM), SO₄ (10–70 mM), and As (0.9–3.5 mM) concentrations. Both reduced arsenite As(III) and oxidized arsenate As(V) species were found (Egal et al. 2010). Sediments were scraped with a spatula, placed in sterile plastic-capped vials, and kept cold until analysis. *E. gracilis* var. *bacillaris* cells were provided by Sorbalab (Etampes, France).

E. mutabilis isolation and culture conditions

A suspension of the microbial community was obtained by adding 10 ml of sterile NaCl 0.9% solution to 10 g of riverbed sediments. After agitation, 7 ml of this suspension was poured on 17.5 ml of Nycodenz[®] solution (65%) and submitted to centrifugation (2 h at 4°C and 10,000×g). A green zone corresponding to the *E. mutabilis* cells was recovered by pipetting. Cells were washed twice in sterile NaCl 0.9% solution and used for inoculation of adapted solid culture medium (Olaveson and Stokes 1989). Cells were then transferred several times onto solid medium until a pure culture was obtained. Further analyses were performed in triplicates on the same medium for both *Euglena* species, and arsenic tolerance was determined under increasing concentrations of sodium arsenite (NaAsO₂; 0, 5.33, 10.67, 16, 21.33, 26.66, and 32 mM) at pH 3.2 (Prolabo). The tolerance to phenyl arsine oxide was performed in the same conditions (0–400 μM). All cultures were incubated at 25°C, with a 16 h/8 h light–dark photoperiod and 45 μmol m⁻² s⁻¹ photon flux density.

DNA isolation, amplification, and sequencing of 16SrDNA and 18SrDNA of both *Euglena* species

Total genomic DNA of *Euglena* cultures was extracted using Wizard Genomic DNA extraction kit (Promega, USA) according to the manufacturer instructions. 16SrDNA and 18SrDNA genes were amplified as previously described (Bruneel et al. 2006; Scanlan and Marchesi 2008). The amplified 16SrDNA and 18SrDNA genes were sequenced by Millgen[®] and analyzed using BLAST program (Altschul et al. 1990).

Motility test

Motility was assessed on semisolid medium (0.3% agarose) containing an arsenite concentration gradient ranging from 0 to 4 mM. Ten microliters of *Euglena* cell suspension was dropped in triplicates on each plate extremity and incubated for 2 weeks. The gliding motility was evaluated by measuring the size variation of the motility ring after 2 weeks incubation.

Determination of the cell surface properties

Cell surface hydrophobicity was determined according to the microbial adhesion to hydrocarbons (MATH) test (Rosenberg et al. 1980). After 15 min of settlement at room temperature, the ODs of the aqueous cell suspensions were determined at 600 nm. Results are given as follows: hydrophobicity (in percent) = $100 (1 - OD_{\text{final}}/OD_{\text{initial}})$. The fatty acids were extracted with chloroform–methanol–water according to Bligh and Dyer (1959), and fatty acid methyl esters were prepared by 15 min incubation at 95°C in boron trifluoride–methanol (Morrison and Smith 1964) and further extracted with hexane. The fatty acid identification was performed by GC-FID according to Pepi et al. (2008). The fluidity of the membrane lipids was expressed as the fatty acid unsaturation index (UI). This was calculated as follows: fatty acid UI = $(\text{percent C16:1} + \text{C18:1}) + (\text{percent C18:2} \times 2) + (\text{percent C18:3} \times 3)/100$, as previously described (Heipieper et al. 2000).

Arsenic accumulation kinetics and speciation determination

E. mutabilis and *E. gracilis* were grown on solid medium as described in the “*E. mutabilis* isolation and culture conditions” section. After 2 weeks, 10 ml of 0.9% NaCl solution containing 1.33 mM of arsenite adjusted to pH 3 was added to the medium. *Euglena* cells cultures were recovered after 1, 4, 6, 12, 48, and 72 h incubation in the presence of arsenite and washed three times in 20 ml of 0.9% NaCl and one time in MilliQ water. Triplicates were obtained for each condition. After centrifugation, pellets were dried, weighted up, and stored at –80°C for further use. Total arsenic was quantified by graphite furnace atomic absorption spectrometry and brought to the pellet weight. Intracellular arsenic speciation was determined in *E. mutabilis* cells incubated 1 and 72 h in the presence of arsenite. Extraction was performed according to Raab et al. (2004) designed to extract arsenic peptides such as As triglutathione from plant material (Raab et al. 2005; Bluemlein et al. 2008, 2009). Briefly, 1 ml of 1% formic acid was added to the frozen cell pellet and the mixture was sonicated (Vibra cell Bioblock scientific, 750 W, 20 kHz, amplitude 20%) three times for 30 s each in an ice bath to avoid sample heating and cooled for 30 s between each ultrasonic step. Arsenic speciation was determined in the culture medium after 72 h incubation with *Euglena* cells in the presence of arsenite. Cells were harvested, and the supernatant was filtered through a 0.22- μm membrane (Millipore®). Samples were kept at –80°C until analysis. Arsenic speciation analysis in all samples was performed by HPLC-ICP-MS after thawing at ambient temperature and centrifugation (10,000 \times g, 1 min, 4°C). Two chromatography modes were used: reversed phase chromatography

with a Spherisorb S5 ODS2 column (250 \times 4.6 mm) and a linear gradient of eluent A (1% formic acid in water) and eluent B (methanol; 0 to 20 min up to 13% B, then 10 min 100% A) at 1 ml/min flow rate, and strong anion exchange chromatography with Hamilton PRP-X100 column (250 \times 4.6 mm) and a two-step elution mode with eluent C (5 mM $\text{NH}_4\text{H}_2\text{PO}_4$) and eluent D (30 mM $\text{NH}_4\text{H}_2\text{PO}_4/(\text{NH}_4)_2\text{HPO}_4$ buffer, pH 8; 0–4 min eluent C, then 4–10 min eluent D and finally 10–20 min eluent C) at 1 ml/min flow rate. An ICP-MS Thermo X7 Series equipped with a Babington nebulizer and Peltier cooled spray chamber was used for element specific detection of arsenic. The instrument settings were checked daily for arsenic sensitivity and optimized when needed. Arsenic triglutathione was synthesized according to previously published procedures (Raab et al. 2004). Other arsenic standards were prepared from NaAsO_2 (Prolabo), sodium arsenate ($\text{Na}_2\text{HAsO}_4 \cdot 7\text{H}_2\text{O}$, Sigma), dimethylarsinic acid ($(\text{CH}_3)_2\text{As}(=\text{O})\text{ONa}$, Merck), and monomethylarsonic acid ($\text{CH}_3\text{AsO}_3\text{Na}_2$, Carlo Erba). All samples and standards were prepared in 1% formic acid prior to analysis by HPLC-ICP-MS.

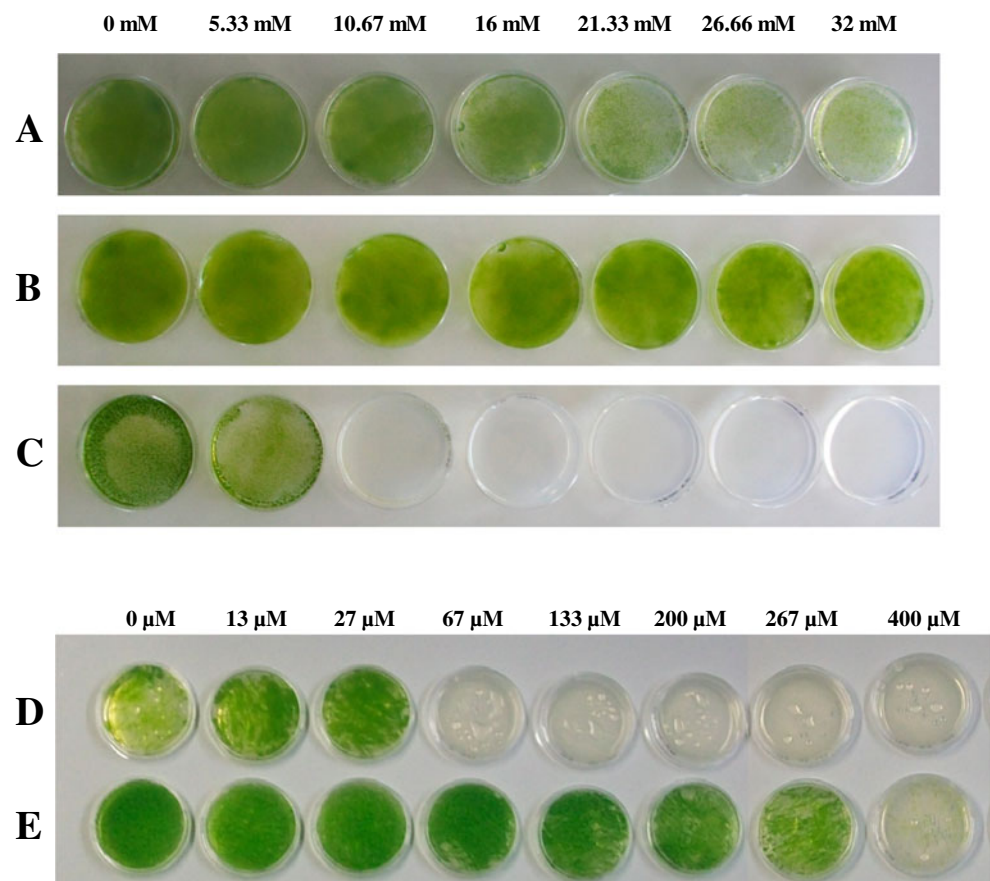
Results

Euglena isolation and behavior in the presence of arsenic

The first step in the *E. mutabilis* purification procedure consisted of the separation of *Euglena* cells from the sediment and the remaining part of the microbial community using a Nycodenz® density gradient. A pure culture of *Euglena* cells was next obtained by streaking the cells several times on minimal culture medium. The absence of any biological contaminant was confirmed by microscopic observations and 18S rDNA and 16S rDNA sequencing (data not shown), which also revealed that the isolated microorganism has more than 99% similarities with *E. mutabilis* 18S rDNA from Rio Tinto, Spain. Optical microscopy observations revealed that *Euglena* cells were 80–150 μm long and 25–30 μm wide and were characterized by the absence of any visible flagellum in contrast to *E. gracilis* displaying flagellar dependent motility.

When cultivated under laboratory conditions, *E. mutabilis* was able to grow in the presence of 32 mM of arsenite, i.e., eight times the average concentration measured at the Carnoulès site, while *E. gracilis* did not grow in media containing more than 6.66 mM of arsenite (Fig. 1). Both *Euglena* species showed an increase in the generation time and/or the lag time in parallel to arsenite concentration, revealing that this metalloid affects the growth of both protists. In addition, as metals have been shown to affect the flagellar motility of *E. gracilis*

Fig. 1 *E. mutabilis* and *E. gracilis* growth on agar plates containing increasing concentrations of arsenical compounds. Growth of *E. mutabilis* in the presence of arsenite was observed 9 days (a) or 13 days (b) after inoculation. Results were compared to the growth of *E. gracilis* in the presence of arsenite 13 days after inoculation (c). The growth of *E. mutabilis* (d) and *E. gracilis* (e) was also observed after 13 days in the presence of phenyl arsine oxide



(Stallwitz and Hader 1994), we measured the effect of arsenite on the flagellar and gliding motility in *E. gracilis* and *E. mutabilis*, respectively (Fig. 2). No positive or

negative chemotactism was observed, but an induction of gliding motility was observed in *E. mutabilis* in the presence of arsenite after 2 weeks. In contrast, the flagellar dependent motility of *E. gracilis* was not affected under arsenite exposure, suggesting a specific effect of this metalloid on *E. mutabilis* gliding motility.

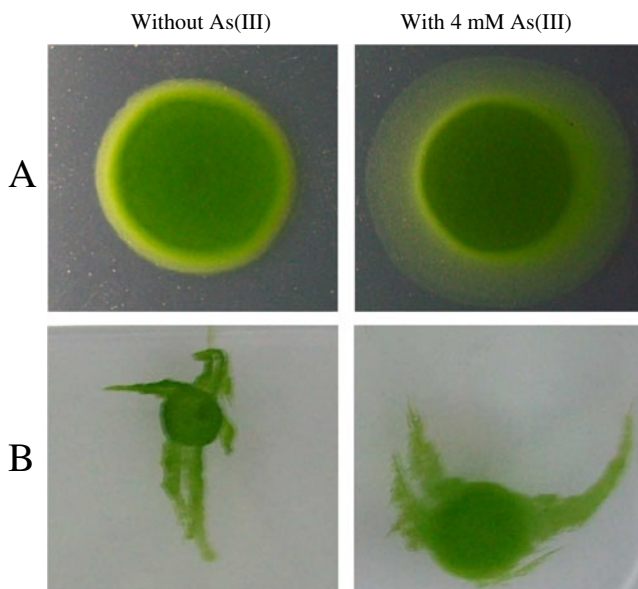


Fig. 2 Motility test performed on semisolid agar plates. *E. mutabilis* (a) showed an induction of the gliding motility in the presence of arsenite, but no significant modification in the flagellar dependent motility was observed in *E. gracilis* (b)

Cell surface properties and fatty acid composition

The induced gliding motility observed in *E. mutabilis* in the presence of arsenic could be linked to modifications of cell surface properties as previously observed in other species (Nichols et al. 1986). In addition, previous observations have revealed that extracellular stresses, in particular metal exposure, affect the fatty acid membrane composition of *E. gracilis* (Einicker-Lamas et al. 1996; Rocchetta et al. 2006). In order to determine whether arsenic modifies the fatty acid composition of *Euglena* species, we analyzed the membrane fatty acid composition in the presence of various arsenite concentrations ranging from 0 to 26.66 mM, i.e., from sublethal to lethal arsenic concentrations. The major phospholipid fatty acids present in both species was shown to be quite similar and mainly composed of myristic acid (C14:0), palmitic acid (C16:0), palmitoleic acid (C16:1), margaric acid (C17:0), oleic acid (C18:1), linoleic acid (C18:2), and linolenic acid (C18:3).

The concentrations of the four most important fatty acids, i. e., oleic acid, linolenic acid, palmitic acid, and linoleic acid, represented more than 90% of the total lipid fatty acids (Fig. 3), and the overall fatty acid profiles were in agreement with previous results obtained in *E. gracilis* (Rocchetta et al. 2006). Nevertheless, *E. mutabilis* membrane contained a higher proportion of C16:0 and C18:2 and a lower proportion of C18:3 (MANOVA test; p value=0.001165, 0.0005384, and 0.002625, respectively). In addition, arsenite was shown to affect the fatty acid composition in a dose-dependent manner. Indeed, the unsaturation index of both *Euglena* species, which was expressed as the sum parameter of polyunsaturated fatty acids, increased parallelly with arsenite concentration as long as this concentration did not alter cell growth (Fig. 3; linear regression analysis, p value=7.826e-06). For arsenite concentrations higher than the minimal inhibitory concentration, the UI decreased again suggesting that the modification of the fatty acid content is an active process. These modifications of the UI were mainly linked to variations in palmitic acid (C16:0), linoleic acid (C18:2), and linolenic acid (C18:3) proportions. These results suggest the existence of active modifications of the fatty acid composition in the presence of sublethal arsenic concentrations.

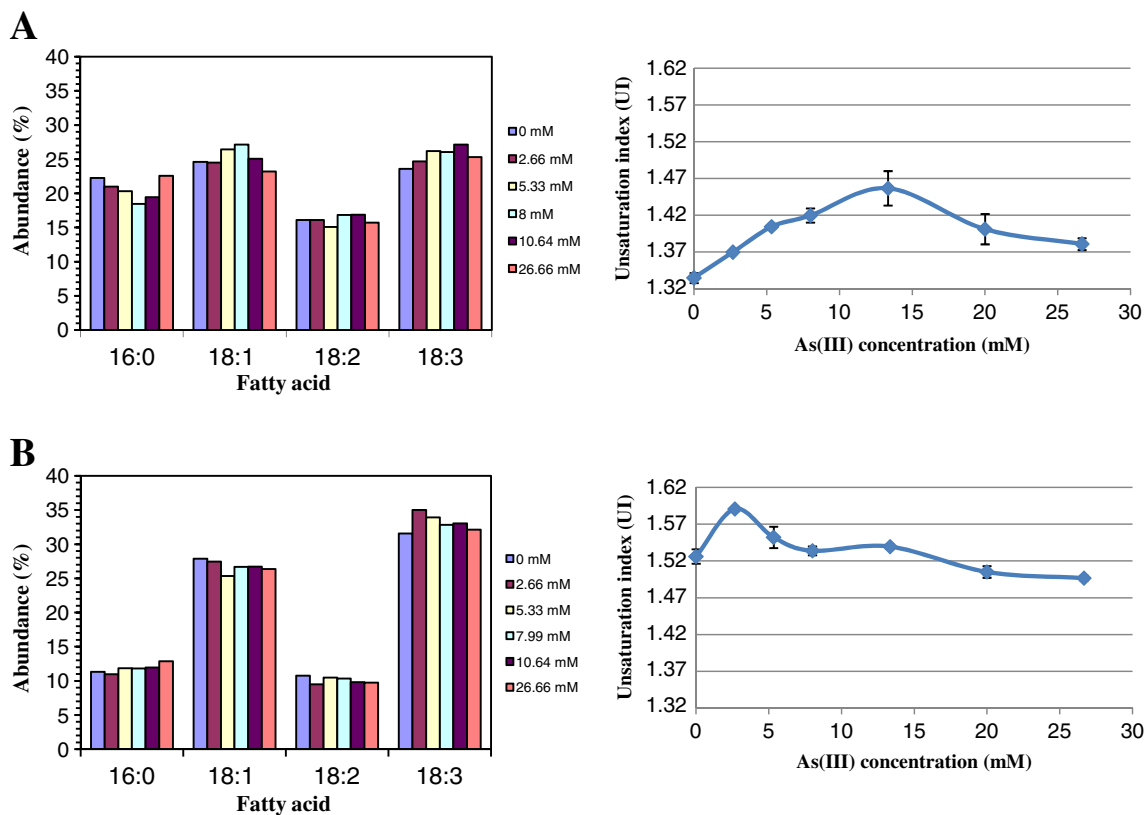


Fig. 3 Fatty acid composition and unsaturation index in both *Euglena* species under increasing concentrations of arsenite. *E. mutabilis* (a) and *E. gracilis* (b) showed a UI increasing in parallel with the arsenite

Cell surface properties were also investigated by measuring cell surface hydrophobicity according to the MATH test (Rosenberg et al. 1980, Fig. 4). These results revealed that the overall hydrophobicity of the cell surface was not affected by arsenite but that *E. mutabilis* cell surface is more hydrophobic than those of *E. gracilis* (Student test, p value=0.00158).

Arsenite accumulation kinetics and metabolism

The highly hydrophobic cell surface properties of *E. mutabilis* may limit the exposure of the cell to water-soluble toxic compounds such as arsenite. In order to test this hypothesis, we measured the tolerance of both species to a hydrophobic arsenical compound, i.e., phenylarsine oxide (Fig. 1d, e). *E. mutabilis* was about 10 times more sensible to this hydrophobic compound than *E. gracilis* suggesting that the tolerance of *E. mutabilis* to arsenite could be linked to a lower permeability to this compound. Arsenite quantification in both *Euglena* species revealed an increase in the intracellular arsenic content in parallel to the exposure time, but this content was low in regard to the extracellular concentration (Fig. 5). In *E. gracilis*, the total arsenic content was approximately four times higher than in *E. mutabilis*

concentration until the minimal inhibitory concentration was reached. For higher arsenical concentration, the UI decreased

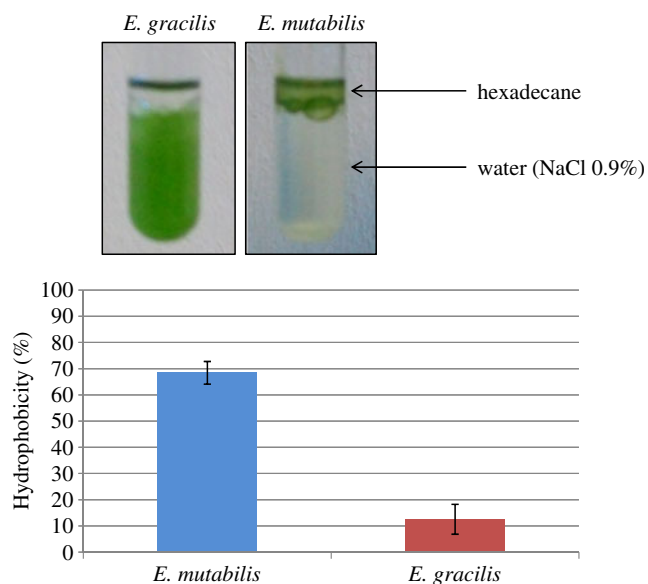


Fig. 4 Hexadecane/water (NaCl 0.9%) repartition coefficient of *E. mutabilis* and *E. gracilis*. Results are given as follows: hydrophobicity (in percent)=100 (1-OD_{final}/OD_{initial})

(covariance analysis, p value=2.74e-05), whatever the exposure time (covariance analysis, p value=5.06e-11). These results suggest that *E. mutabilis* accumulates less arsenic into the cell, which could be explained by its peculiar cell surface properties preventing its entry, an active extrusion system and/or by a specific arsenical metabolism.

To get insight into the arsenical metabolism in both species, the arsenic speciation was determined in the medium and in the *Euglena* cell intracellular compartments after 1 and 72 h exposure to a sublethal arsenite concentration, i.e., 4 mM. No

significant change in arsenic speciation was observed in the extracellular compartment for both *Euglena* species (data not shown). In contrast, the intracellular arsenic speciation revealed significant differences. Although the sum of arsenic species identified in formic acid extracts did not represent as much as 10% of the total arsenic content measured in *Euglena* cells, these results highlighted the presence of oxidized and reduced form of inorganic arsenic in both species, i.e., arsenate As(V) and arsenite As(III), as well as methylated forms, i.e., monomethylarsonic acid MMAs(V) and dimethylarsonic acid DMAs(V) in *E. gracilis*. Moreover, some unidentified arsenical compounds were also found in both species (Fig. 6).

Discussion

E. mutabilis is a protist considered as a biological indicator of AMD-exposed ecosystems often characterized by high concentrations of arsenic. Our analysis performed under in vitro conditions revealed that *E. mutabilis* is able to grow in media containing very high arsenite concentrations corresponding to at least eight times those measured at the sampling site. This hypertolerance to arsenite suggests that *E. mutabilis* has developed specific defense mechanisms and may colonize environments even more contaminated with this toxic element. In order to determine the mechanisms involved in the *E. mutabilis* arsenic hypertolerance, its response to this stress was compared to that of *E. gracilis*, a protist used as a model microorganism in laboratory conditions and not found in arsenic-contaminated environments.

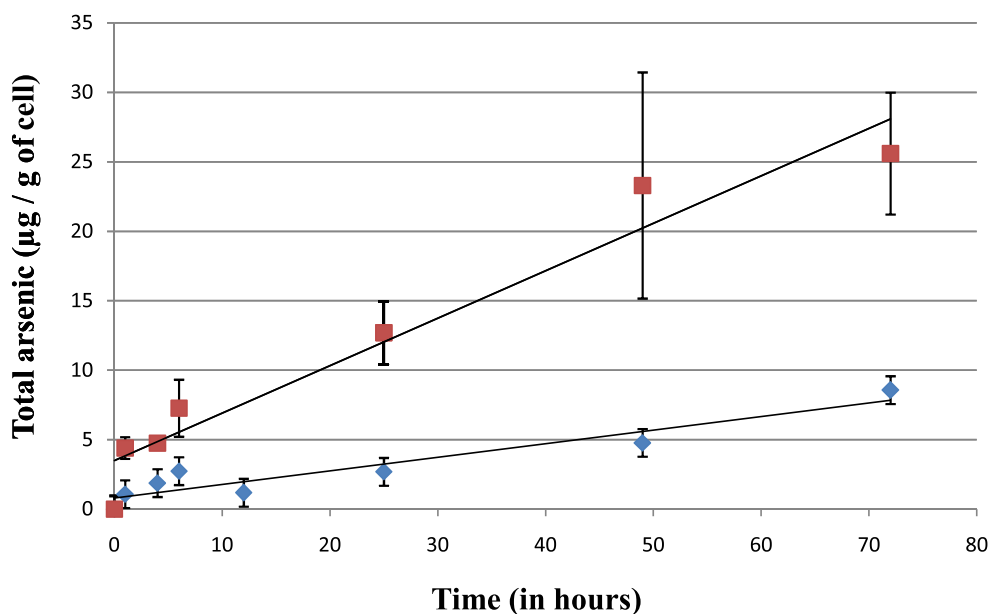
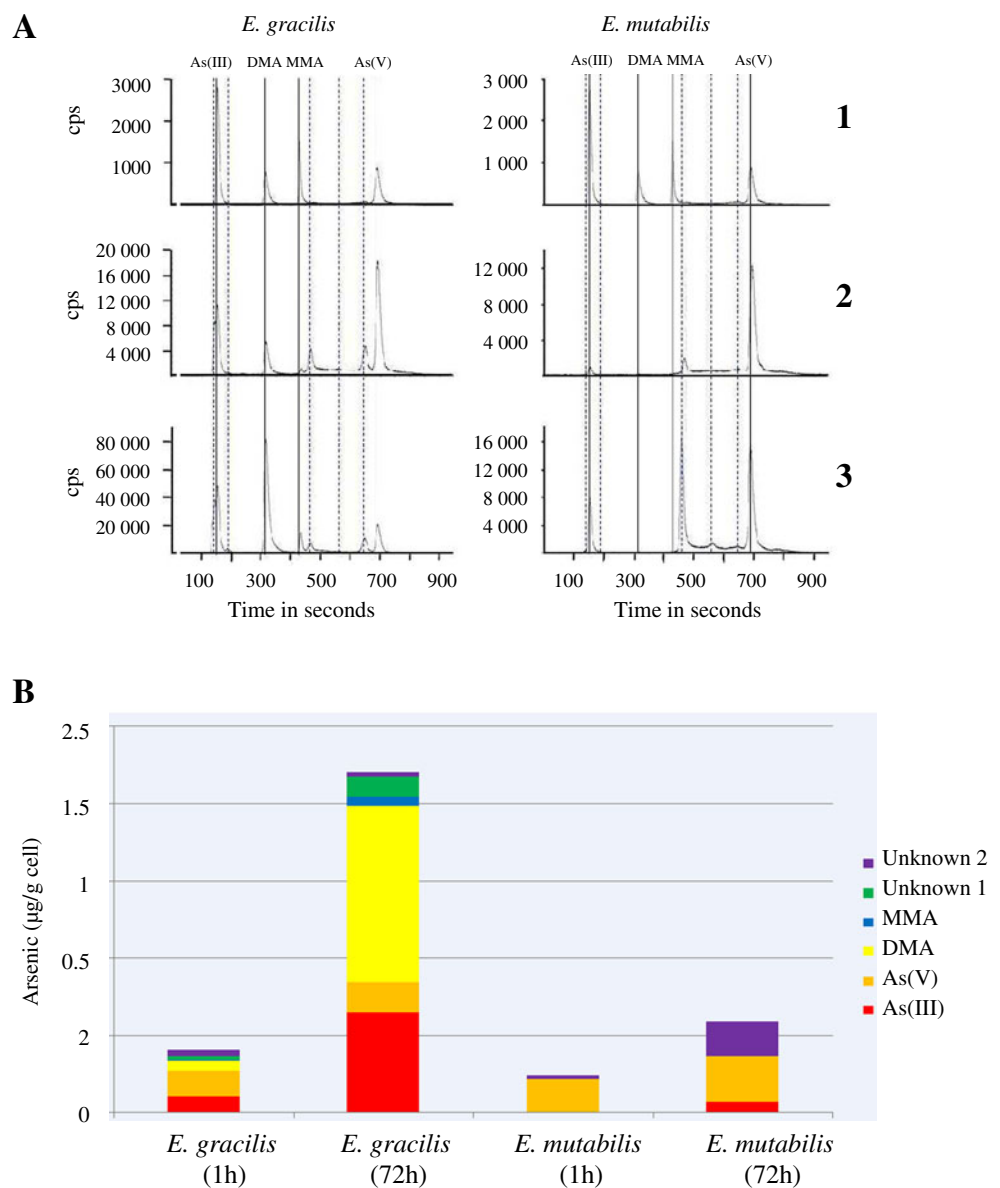


Fig. 5 Kinetics of total arsenic intracellular concentration in *Euglena* cells exposed to arsenite (1.33 mM). Results are given for both *E. gracilis* (filled square) and *E. mutabilis* (filled diamond)

Fig. 6 Speciation of arsenical compounds extracted from *E. gracilis* and *E. mutabilis* cells. The chromatograms (a) present the profile of pure arsenical compounds (lane 1) and the arsenical species extracted from *Euglena* cells after 1 h (lane 2) or 72 h (lane 3) exposure to arsenite. The arsenical species are represented by histograms (b) highlighting the accumulation of free arsenical methylated forms in *E. gracilis* after 72 h, in contrast to *E. mutabilis*



Fatty acid modification and gliding motility under arsenite exposure

Under arsenite exposure, both *Euglena* species exhibited a growth lag time and/or a slower growth rate suggesting a toxic effect of this metalloid, which is supported by the modification of the fatty acid composition in an arsenic dose-dependent manner. This fatty acid modification led to an increased UI of the membrane, known to affect the membrane fluidity and was due to a higher incorporation of polyunsaturated fatty acids such as linolenic acid (C18:3) under arsenic growth subinhibitory concentrations. Such an observation suggests that the modification of the fatty acid composition is an active adaptative process relying either on de novo fatty acid synthesis or on a modification of some fatty acids already present in the membrane such as

C16:0. The precise effect of the increased UI of the membrane under arsenical stress in both species remains to be determined. However, in *E. mutabilis* this process could facilitate the gliding motility observed under arsenite exposure by increasing the membrane fluidity, as suggested in other microorganisms (Nichols et al. 1986). Such gliding activity may be important for *E. mutabilis* in the AMD-exposed environment, where this protist groups together and forms green dense mats at the sediment–water interface.

Hydrophobicity and arsenic accumulation kinetic

The analysis of the pellicle of both *Euglena* cells in the presence or the absence of arsenic also revealed highly hydrophobic cell surface properties of *E. mutabilis* in

comparison to those of *E. gracilis* (Fig. 4). Previous experiments performed on yeast cells revealed that the amount of heavy metal biosorption and the cell surface hydrophobicity negatively correlated (Kordialik-Bogacka 2011). This hydrophobic cell surface properties measured in *E. mutabilis* may play an important role in passively protecting this protist from water-soluble toxic compounds such as arsenite As(III). Such a hypothesis is further supported by two observations. First, *E. mutabilis* showed a higher sensitivity to hydrophobic arsenical compounds such as phenyl arsine oxide and a higher tolerance to hydrophilic arsenical compounds such as arsenite in comparison to *E. gracilis* (Fig. 1). These results suggest that the arsenic tolerance observed in *Euglena* cells may rely, at least in part, on the hydrophobicity of arsenical elements. Second, as compared to *E. gracilis*, *E. mutabilis* accumulated a lower amount of total arsenic in the cell when exposed to water-soluble arsenite As(III) revealing a lower exposure of its intracellular compartment to this toxic element. Even though we cannot exclude the existence of arsenic extrusion process in *E. mutabilis* that would be absent in *E. gracilis*, our observations suggest that the highly hydrophobic cell surface properties of *E. mutabilis* may act as a barrier against arsenite. As *E. mutabilis* has been shown to be more tolerant to iron, cadmium, nickel, aluminum, and zinc in comparison to *E. gracilis* (Olaveson and Nalewajko 2000), such a barrier may constitute general resistance mechanisms to other metals often found in AMD-exposed ecosystems.

Arsenite biotransformation

The biotransformation of arsenite to less toxic or less bioavailable compounds is another general defense mechanism in both prokaryotic and eukaryotic microorganisms. One of the most common biotransformation processes observed in arsenite-tolerant prokaryotic cells is the oxidation of arsenite As(III) into arsenate As(V), a less toxic form that can lead to the formation of insoluble precipitates with iron and sulfate. This latter process has been described at the Carnoulès site from which *E. mutabilis* was isolated and has been shown to rely, at least in part, on arsenite oxidation activity mediated by *Thiomonas* sp. (Casiot et al. 2003; Bertin et al. 2011). Although, previous results have presented *E. mutabilis* as able to oxidize arsenite As(III) into arsenate As(V) (Casiot et al. 2004), our results did not reveal any significant arsenite oxidation activity mediated by *E. mutabilis* under in vitro conditions. Even though we cannot rule out the existence of distinct *E. mutabilis* strains in this ecosystem differing from each other according to their arsenite oxidation capacities, such a discrepancy may also be explained by a contamination of the previous *E. mutabilis* culture (Casiot et al. 2004) with bacteria related to the *Thiomonas* genus.

The extraction protocol used for arsenic speciation in the cell allowed us to measure only 10% of the total arsenical content. It is probable that the remaining part of the arsenical species is not soluble in the 1% formic acid extract or remains bound to the membranes as described previously in *E. gracilis* (Miot et al. 2009). Nevertheless, newly synthesized arsenical species accumulated in the cells revealed the existence of active biotransformation processes. Indeed, arsenate As(V) was found in both *Euglena* species exposed to arsenite As(III) and corresponds to the majority of the arsenical species identified in *E. mutabilis* after 1 h exposure. Such an observation may be explained by a biological arsenite oxidation occurring in the cells or to the uptake of abiotic oxidized arsenate As(V) from the medium. Even though any arsenite oxidation activity of both *Euglena* cells may not be excluded, this activity would obviously be too weak to significantly affect the medium arsenic speciation in our analysis.

The accumulation of unknown arsenical compounds in both species was observed in the intracellular compartment (Fig. 6). As a previous analysis has revealed that arsenic is found in complex with sulfuric compounds in *E. gracilis* (Miot et al. 2009), it is tempting to speculate that some of these unknown compounds may correspond to glutathione and/or phytochelatin complexes. Despite the use of adapted extraction procedures (Raab et al. 2005; Bluemlein et al. 2008, 2009), such complexes were not identified in our work. In contrast, the accumulation of free arsenic methylated forms, i.e., MMAs(V) and DMAs(V) was observed in *E. gracilis* but not in *E. mutabilis*. Arsenic methylation was until recently considered as a detoxification mechanism because it can lead to the formation of a volatile end product, i.e., TMAO (Aposhian 1997). Nevertheless, this assumption was more recently brought into question since some intermediates generated during arsenic methylation, i.e., monomethylarsenite MMA(III) and dimethylarsenite DMA(III), have been shown to be highly toxic (Akter et al. 2006). The absence of free arsenic methylated forms in *E. mutabilis* may consequently be part of the higher tolerance of this protist to arsenic.

Arsenic has a toxic effect on both species, even though *E. mutabilis* is much more tolerant to this metalloid. This higher tolerance may be explained by (1) the induction of gliding motility under arsenite exposure, (2) the highly hydrophobic cell surface properties of *E. mutabilis* that may explain the lower accumulation of arsenic in the cells in comparison to *E. gracilis*, and (3) a specific arsenical metabolism of *E. mutabilis* as compared to *E. gracilis* preventing the accumulation of free arsenic methylated forms in the cells. Additional defense mechanisms acting in parallel such as glutathione biosynthesis or active extrusion of arsenic from the cell may not be excluded. Therefore comparative analysis between both *Euglena* species based

on differential transcriptomic approaches will be performed in the future to get insight into the underlying molecular mechanisms involved in the particular metabolic and physiological traits of *E. mutabilis* towards arsenic.

Acknowledgments The study was financed by the ANR 07-BLANC-0118 project (Agence Nationale de la Recherche). David Halter was 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

- Akter KF, Owens G, Davey DE, Naidu R (2006) Arsenic speciation and toxicity in biological systems. *Rev Environ Contam Toxicol* 184:97–149
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410
- Aposhian HV (1997) Enzymatic methylation of arsenic species and other new approaches to arsenic toxicity. *Annu Rev Pharmacol Toxicol* 37:397–419
- Bertin PN, Heinrich-Salmeron A, Pelletier E, Goulhen-Chollet F, Arsène-Ploetze F, Gallien S, Lauga B, Casiot C, Calteau A, Vallenet D, Bonnefoy V, Bruneel O, Chane-Woon-Ming B, Cleiss-Arnold J, Duran R, Elbaz-Poulichet F, Fonknechten N, Giloteaux L, Halter D, Koechler S, Marchal M, Mormico D, Schaeffer C, Smith AAT, Van Dorselaer A, Weissenbach J, Médigue C, Le Paslier D (2011) Metabolic diversity among main microorganisms inside an arsenic-rich ecosystem revealed by meta- and proteo-genomics. *ISME J*. doi:10.1038/ismej.2011.51
- Bligh EG, Dyer WJ (1959) A rapid method of total lipid extraction and purification. *Can J Biochem Physiol* 37:911–917
- Bluemlein K, Raab A, Feldmann J (2009) Stability of arsenic peptides in plant extracts: off-line versus on-line parallel elemental and molecular mass spectrometric detection for liquid chromatographic separation. *Anal Bioanal Chem* 393:357–366
- Bluemlein K, Raab A, Meharg AA, Charnock JM, Feldmann J (2008) Can we trust mass spectrometry for determination of arsenic peptides in plants: comparison of LC-ICP-MS and LC-ES-MS/ICP-MS with XANES/EXAFS in analysis of *Thunbergia alata*. *Anal Bioanal Chem* 390:1739–1751
- Brake SS, Dannelly HK, Connors 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
- Bray JP, Broady PA, Niyogi DK, Harding JS (2008) Periphyton communities in New Zealand streams impacted by acid mine drainage. *Mar Freshwat Res* 59:1084–1091
- Bruneel O, Duran R, Casiot C, Elbaz-Poulichet F, Personné JC (2006) Diversity of microorganisms in Fe-As-rich acid mine drainage waters of Carnoulès, France. *Appl Environ Microbiol* 72:551–556
- Casiot C, Bruneel O, Personné JC, Leblanc M, Elbaz-Poulichet F (2004) Arsenic oxidation and bioaccumulation by the acidophilic protozoan, *Euglena mutabilis*, in acid mine drainage (Carnoulès, France). *Sci Total Environ* 320:259–267
- Casiot C, Egal M, Elbaz-Poulichet F, Bruneel O, Bancon-Montigny C, Cordier MA, Gomez E, Aliaume C (2009) Hydrological and geochemical control of metals and arsenic in a Mediterranean river contaminated by acid mine drainage (the Amous River, France): preliminary assessment of impacts on fish (*Leuciscus cephalus*). *Appl Geochem* 24:787–799
- Casiot C, Morin G, Juillot F, Bruneel O, Personné JC, Leblanc M, Duquesne K, Bonnefoy V, Elbaz-Poulichet F (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). *Appl Geochem* 25(12):1949–1957
- Einicker-Lamas M, Soares MJ, Soares MS, Oliveira MM (1996) Effects of cadmium on *Euglena gracilis* membrane lipids. *Braz J Med Biol Res* 29:941–948
- Heipieper HJ, Isken S, Saliola M (2000) Ethanol tolerance and membrane fatty acid adaptation in adh multiple and null mutants of *Kluyveromyces lactis*. *Res Microbiol* 151:777–784
- Kordialik-Bogacka E (2011) Surface properties of yeast cells during heavy metal biosorption. *Cent Eur J Chem* 9:348–351
- Miot J, Morin G, Skouri-Panet F, Féraud C, Aubry E, Briand J, Wang Y, Ona-Nguema G, Guyot F, Brown GE (2008) XAS study of arsenic coordination in *Euglena gracilis* exposed to arsenite. *Environ Sci Technol* 42:5342–5347
- Miot J, Morin G, Skouri-Panet F, Féraud C, Poitevin A, Aubry E, Ona-Nguema G, Juillot F, Guyot F, Brown GE Jr (2009) Speciation of arsenic in *Euglena gracilis* cells exposed to As (V). *Environ Sci Technol* 43:3315–3321
- Morrison WR, Smith LM (1964) Preparation of fatty acid methyl esters and dimethylacetate from lipids with boron fluoride-methanol. *J Lipid Res* 5:600608
- Nichols P, Stulp BK, Jones JG, White DC (1986) Comparison of fatty acid content and DNA homology of the filamentous gliding bacteria *Vitreoscilla*, *Flexibacter*, *Filibacter*. *Arch Microbiol* 146:1–6
- Olaveson MM, Nalewajko C (2000) Effects of acidity on the growth of two *Euglena* species. *Hydrobiologia* 433:39–56
- Olaveson MM, Stokes PM (1989) Responses of the acidophilic alga *Euglena mutabilis* (Euglenophyceae) to carbon enrichment at pH 3. *J Phycol* 25:529–539
- Pawlik-Skowronska B, Pirszel J, Kalinowska R, Skowronski T (2004) Arsenic availability, toxicity and direct role of GSH and phytochelatin in As detoxification in the green alga *Stichococcus bacillaris*. *Aquat Toxicol* 70:201–212
- Pepi M, Heipieper HJ, Fischer J, Ruta M, Volterrani M, Focardi SE (2008) Membrane fatty acids adaptive profile in the simultaneous presence of arsenic and toluene in *Bacillus* sp. ORAs2 and *Pseudomonas* sp. ORAs5 strains. *Extremophiles* 12:343–349
- Qin J, Lehr CR, Yuan C, Le XC, McDermott TR, Rosen BP (2009) Biotransformation of arsenic by a Yellowstone thermoacidophilic eukaryotic alga. *Proc Natl Acad Sci U S A* 106:5213–5217
- Raab A, Feldmann J, Meharg AA (2004) The nature of arsenic-phytochelatin complexes in *Holcus lanatus* and *Pteris cretica*. *Plant Physiol* 134:1113–1122
- Raab A, Schat H, Meharg AA, Feldmann J (2005) Uptake, translocation and transformation of arsenate and arsenite in sunflower (*Helianthus annuus*): formation of arsenic-phytochelatin complexes during exposure to high arsenic concentrations. *New Phytol* 168:551–558
- Rocchetta I, Mazzuca M, Conforti V, Ruiz L, Balzaretto V, De Molina MDCR (2006) Effect of chromium on the fatty acid composition of two strains of *Euglena gracilis*. *Environ Pollut* 141:353–358
- Rosenberg M, Gutnick D, Rosenberg E (1980) Adherence of bacteria to hydrocarbons: a simple method for measuring cell-surface hydrophobicity. *FEMS Microbiol Lett* 9:29–33
- Scanlan PD, Marchesi JR (2008) Micro-eukaryotic diversity of the human distal gut microbiota: qualitative assessment using culture-dependent and -independent analysis of faeces. *ISME J* 2:1183–1193

- Stallwitz E, Hader DP (1994) Effects of heavy metals on motility and gravitactic orientation of the flagellate, *Euglena gracilis*. Eur J Protistol 30:18–24
- Tsai SL, Singh S, Chen W (2009) Arsenic metabolism by microbes in nature and the impact on arsenic remediation. Curr Opin Biotechnol 20:659–667
- Valente TM, Gomes CL (2007) The role of two acidophilic algae as ecological indicators of acid mine drainage sites. J Iberian Geol 33:283–294
- Wolowski K, Turnau K, Henriques FS (2008) The algal flora of an extremely acidic, metal-rich drainage pond of Sao Domingos pyrite mine (Portugal). Cryptogam Algal 29:313–324

CHAPITRE IV :

LA REPONSE A L'ARSENIC CHEZ *E.*
***GRACILIS* ET *E. MUTABILIS* PAR DES**
APPROCHES DE GENOMIQUE
FONCTIONNELLE

A : INTRODUCTION

Les observations physiologiques réalisées sur *E. mutabilis* et *E. gracilis* ont permis de révéler une multitude de différences phénotypiques en présence d'arsenic chez ces deux espèces. Cependant, afin d'appréhender plus en détail les mécanismes moléculaires sous jacents, des approches de protéomique et de transcriptomique différentielle ont été développées en parallèle.

En effet, bien que l'accumulation différentielle de certaines protéines puisse être appréhendée par des approches de protéomique bi-dimensionnelle, le séquençage du transcriptome offre en plus l'avantage d'explorer la partie codante du génome des euglènes dans son intégralité en incluant les transcrits dont l'expression n'est pas affectée par l'arsenic. Si le transcriptome d'*E. gracilis* est déjà présent dans les bases de données, les informations relatives à l'organisation génétique d'*E. mutabilis* font cruellement défaut. Ainsi, seule les séquences d'une vingtaine de transcrits d'*E. mutabilis* connues, principalement des séquences d'ADN codant pour les sous-unités ribosomiques, n'apportant que peu d'information sur les potentialités fonctionnelles de cet organisme.

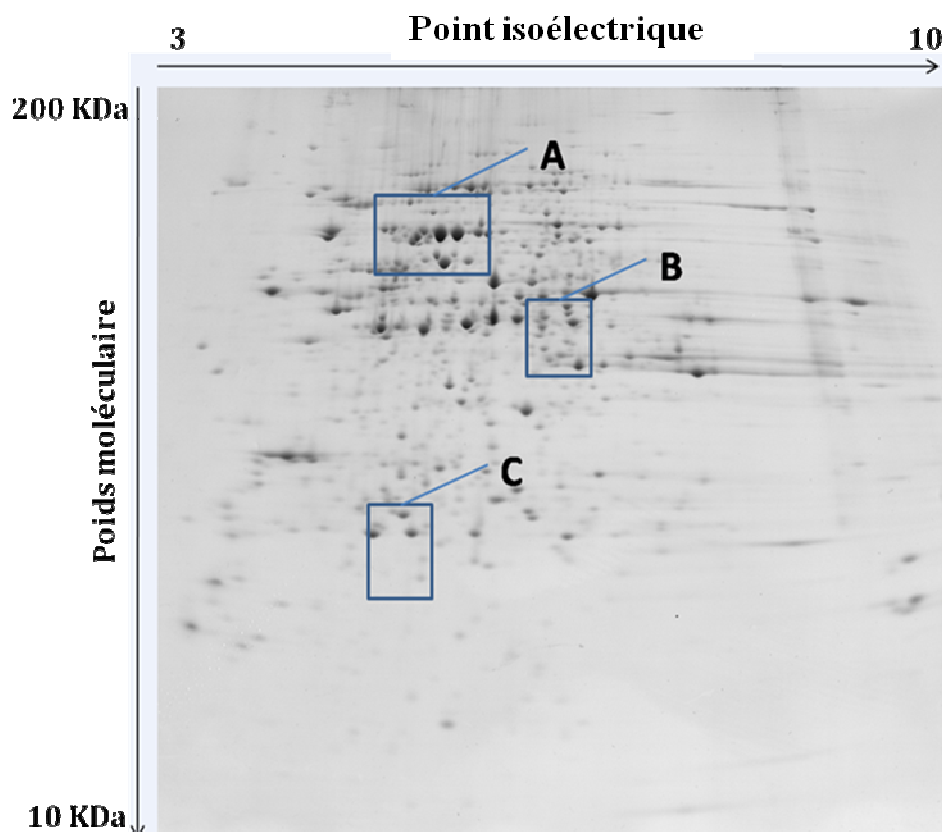
Les approches de transcriptomique ayant profité de l'essor remarquable des techniques de séquençage de seconde génération, le séquençage des transcriptomes des euglènes a pu être entrepris à la fois en présence et en absence d'arsenic pour chacune des deux espèces. Cette étude a fait l'objet d'un projet pilote soumis au Génoscope et à pour objectif d'apporter à la fois des informations sur la réponse à l'arsenic chez *E. gracilis* et *E. mutabilis*, mais également de permettre des études de génomique comparative plus globales. Les résultats de ces travaux sont présentés dans le chapitre suivant et sont complétés dans certains cas par des analyses physiologiques ciblées afin de valider les hypothèses émises sur base des analyses de génomique fonctionnelle.

Certaines parties de ces travaux nécessitent encore d'être validées par des approches qui sont décrites sous forme de perspectives à court terme et n'ont par conséquent pas encore fait l'objet d'une publication scientifique.

B : RESULTATS

1) approche de protéomique différentielle

Une approche de protéomique différentielle a été développée sur des cultures d'*E. mutabilis* réalisées en quadruplicat en présence de 300 ppm d'AsIII, 300 ppm d'AsV ou non exposées à l'arsenic. Les euglènes ont été lysées par sonication (Vibra cell Bioblock scientifique, 750 W, 20 KHz, amplitude 20%, 5 x 30 secondes dans la glace) puis les protéines ont été récupérées et purifiées par précipitation dans une solution d'acide trichloroacétique à 10% dans l'acétone. Après dosage Bradford, 300µg de protéines ont été séparés en fonction de leur point isoélectriques sur des strips présentant un gradient de pH non linéaire allant de 3 à 10 puis les protéines ont été séparées dans un second temps en fonction de leurs masses moléculaires sur gel de polyacrylamide à 12% (Figure 4.1). Les gels ont été réalisés en quadruplicats pour chacune des 3 conditions puis analysés avec le logiciel ImageMaster 7.0. Etonnamment, seuls 7 spots ont été identifiés comme ayant un niveau d'accumulation significativement différent entre les 3 conditions testées.



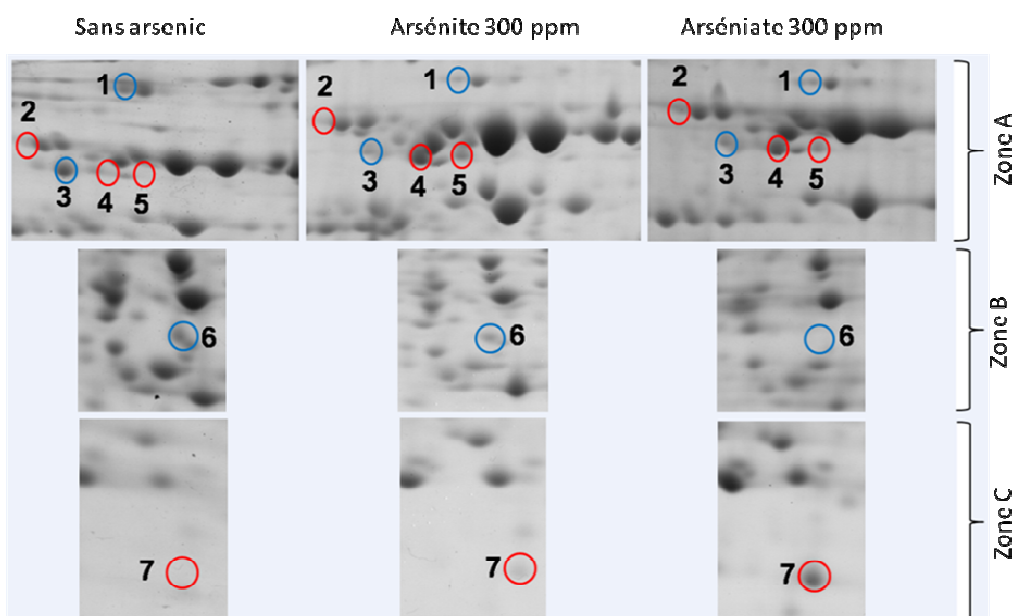


Figure 4.1 : Exemple de gel de polyacrylamide bi-dimensionnel obtenu avec les protéines extraites d'*E. mutabilis*. La figure A présente un gel de polyacrylamide dans son intégralité et la figure B détaille plus spécifiquement les zones où des protéines à accumulation différentielle ont été détectées (les protéines plus ou moins accumulées en présence d'arsenic sont entourées en rouge et en bleu, respectivement).

Les spots ont été excisés puis les protéines correspondantes ont été analysées par spectrométrie de masse au sein de l'Equipe d'Alain Van Dorsselaer par Christine Carapito grâce à un système couplant chromatographie liquide (nanoACQUITY Ultra-Performance-Liquide Chromatography, Waters) et un système de spectrométrie de masse en tandem (SYNAPT hybrid quadrupole orthogonal acceleration time-of-flight tandem mass spectrometer). Les protéines ayant le meilleur score par rapport aux données de spectrométrie de masse obtenues pour chacun des spots sont présentées ci-dessous (Table 4.1).

AC	Protéine	Taxonomie	Masse (Da)	pI	Couv.	RMS (+/- ppm)	Score	Nb pep.
gi-209882040	DnaK family protein	<i>Cryptosporidium muris</i>	72 188	6,14	5	87	188	2
gi-730308	Phospho-glucomutase	<i>Agrobacterium tumefaciens</i>	57 766	5,3	1	55	48	1
gi-113474864	Phospho-glucomutase	<i>Trichodesmium erythraeum</i>	59 298	5,2	2	53	67	2
gi-730308	Phospho-glucomutase	<i>Agrobacterium tumefaciens</i>	57 766	5,3	4	50	49	1
gi-870796	glyceraldehyde-3-phosphate dehydrogenase	<i>Euglena gracilis</i>	48 569	8,11	8	96	164	3

gi-33694250	heat shock protein 70	<i>Euglena gracilis</i>	50 799	6,47	28	37	463	8
gi-62287512	Trans-2-enoyl-CoA reductase	<i>Euglena gracilis</i>	57 266	6.08	5	60	127	3
gi-33694248	heat shock protein 90	<i>Euglena gracilis</i>	73 852	5.35	19	48	572	12
gi-11467034	ATP synthase CF1 beta subunit	<i>Euglena gracilis</i>	52 045	5.13	38	50	1071	13
gi-170293981	ATP synthase beta subunit	<i>Raphidonema nivale</i>	39 636	5.02	17	61	247	4
gi-66806375	cell division cycle protein 48	<i>Dictyostelium discoideum</i>	88 500	5.19	13	32	474	6

Table 4.1 : Identification des protéines présentant une accumulation différentielle chez *E. mutabilis* en présence d'arsenic. Les numéros d'accèsion (AC), le nom de la protéine ayant le meilleur score, la taxonomie de l'organisme où cette protéine a été identifiée, la masse et le point isoélectrique théorique de la protéine correspondante, la proportion (en %) de la protéine recouverte par les peptides identifiés (couv.), le seuil de tolérance pour la masse des peptides identifiés (RMS), le score et le nombre de peptides ayant permis l'identification de la protéine sont présentés.

Ces analyses ont souligné que des protéines chaperons (Hsp70 et 90) et des protéines liées à la synthèse d'ATP étaient plus accumulées en présence d'arsenic. Ces observations pourraient être liées au fait que l'arsenic est connu pour perturber le repliement tri-dimensionnel des protéines et la phosphorylation oxydative (Jin *et al.*, 2008; Wysocki & Tamas, 2010; Zhang *et al.*, 2007). En effet, l'accumulation de chaperons facilitant le repliement tri-dimensionnel des protéines ainsi que l'augmentation de la production d'ATP permettrait de compenser, au moins en partie, les perturbations engendrées par l'arsenic. D'autre part, une protéine du métabolisme intermédiaire, la glycéraldéhyde 3 phosphate deshydrogénase, semble également être plus accumulée en présence d'arsenic. Cette observation, déjà réalisée dans d'autres études soulignant une augmentation de l'activité de cette protéine sous stress oxydatif, pourrait probablement être liée au fait que la réaction catalysée par cette enzyme conduit à la synthèse d'agents réducteurs du type NADPH+H⁺ utilisés pour compenser le stress oxydatif lié à la présence d'arsenic (Bustos *et al.*, 2008).

Par opposition, des protéines impliquées dans la synthèse d'acides gras polyinsaturés (trans-2-enoyl-CoA reductase) et la division cellulaire (cell division cycle protein 48) semblent être moins accumulées sous stress arsénié. Si la répression de la division cellulaire est en accord avec les observations physiologiques soulignant un temps de génération un peu plus long chez *E. mutabilis* en présence d'arsenic, la moindre accumulation de la trans-2-enoyl-CoA reductase est plus difficile à interpréter. En effet, l'étude des acides gras a révélé chez *E. mutabilis* une augmentation de la

proportion d'acide linoléique (18:3) sous stress arsénié ce qui conduit à une apparente contradiction des résultats. Il n'est cependant pas à exclure que la trans-2-enoyl-CoA reductase pourrait être régulée par des modifications post traductionnelles fractionnant sa localisation sur la carte protéomique et entraînant par conséquent une baisse apparente de son niveau d'accumulation pour un spot donné. Cette hypothèse, qui permettrait d'expliquer des résultats en apparente contradiction, reste néanmoins à confirmer par d'autres analyses, notamment en étudiant l'expression des transcrits correspondant.

Enfin, la phosphoglucomutase a été identifiée à la fois dans des spots plus intenses et moins intenses en présence d'arsenic. Ces spots, localisés très près les uns des autres sur la carte protéomique (spots 3, 4 et 5 dans la zone A, Figure 2.4), pourraient correspondre à la même protéine mais ayant subi une modification post-traductionnelle affectant légèrement son point isolélectrique mais pas sa masse apparente. Le rôle de cette protéine dans la réponse au stress arsénié reste à déterminer, même si une augmentation de l'accumulation de cette protéine a également été observée chez une autre algue exposée au cuivre (Contreras *et al.*, 2009).

Au vu du faible nombre de protéines identifiées par l'approche de protéomique différentielle, une analyse de transcriptomique différentielle a été développée dans un second temps.

2) Approche de transcriptomique différentielle

a. Préparation des échantillons et analyse des données

Les ARNm utilisés pour les analyses de transcriptomique différentielle ont été extraits de 10 cultures indépendantes d'euglènes pour chacune des 4 conditions (*Euglena mutabilis* et *Euglena gracilis* avec et sans arsenic) afin de limiter les variations biologiques non spécifiquement induites par ce métalloïde. Le protocole retenu pour l'extraction est celui décrit par Gonzales *et al.* (2009), basé sur l'utilisation de SDS, permettant à la fois une lyse efficace des cellules et une inactivation rapide des RNAses. Les ARN totaux extraits des 10 cultures correspondant à chacune des 4 conditions ont par la suite été regroupés avant d'en purifier les ARNm sur base de la présence d'une queue polyA (illustra mRNA Purification Kit, GE Healthcare). La qualité et la quantité des ARNm ainsi que le degré de contamination par des ADN et des ARNr résiduels ont été évalués pour chacune des 4 conditions (Experion RNA HighSens Analysis Kit, BioRad)(Figure 4.2).

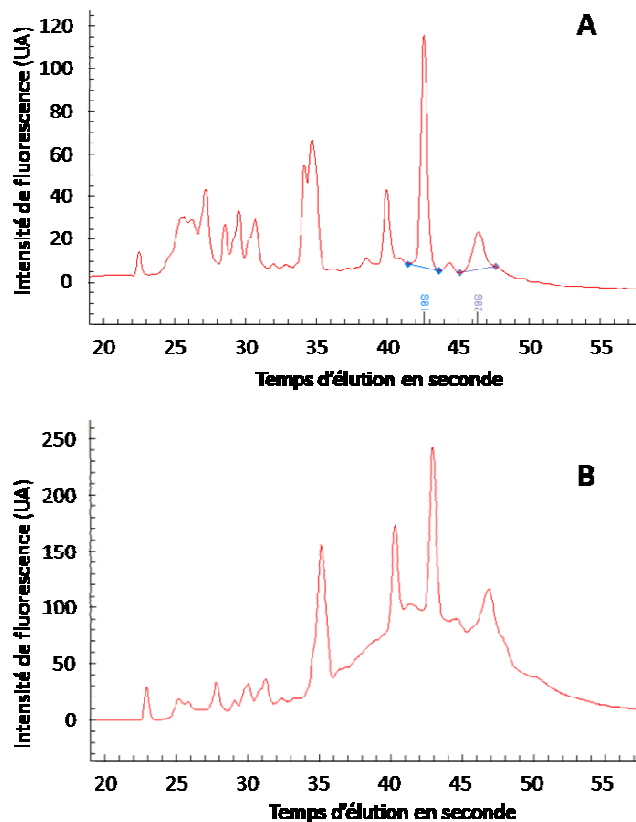


Figure 4.2 : Exemple du profil des ARN extraits des euglènes avant (A) et après (B) enrichissement des ARNm. Les euglènes présentent 18 ARN ribosomiques différents correspondant aux 18 pics. Après enrichissement (B), les ARNm sont visibles et forment un large pic éluant entre 35 et 50 secondes. Les ARNr résiduels ont été éliminés dans un second temps par une étape de rétro-transcription utilisant des amorces polyT ciblant spécifiquement les ARNm poly- adénylés.

Les ARNm correspondant à chacune de ces 4 conditions de culture ont été rétro-transcrits en cDNA double brins grâce au « SuperScript™ Double-Stranded cDNA Synthesis Kit » (Invitrogen). Aucun génome de référence n'étant disponible pour les euglènes, la technique de séquençage retenue pour l'étude des transcrits est la technique 454 GS flex de Titanium permettant de concilier le séquençage haut débit et la génération de séquences nucléotidiques ayant une taille relativement longue (environ 330 pb).

Ces deux caractéristiques permettent de faciliter l'assemblage des séquences générées et d'accéder ainsi plus facilement à la séquence du transcrit dans son intégralité, mais également d'avoir une couverture suffisamment importante pour permettre des analyses comparatives statistiquement significatives. Le séquençage 454 ainsi que l'assemblage des séquences obtenues (logiciel Newbler) ont été effectués par la plateforme bioinformatique du Génoscope en collaboration avec Julie Poulain et Corinne DaSilva. Les principales caractéristiques des séquences obtenues sont présentées ci-dessous (Table 4.2).

	EM sans arsenic	EM avec arsenic	EG sans arsenic	EG avec arsenic
Nb de nucléotides séquencés	66 572 920	80 401 536	51 086 259	71 709 822
Taille totale après assemblage (pb)	6 600 095	8 615 937	2 778 693	3 396 167
Nombre de contigs	10,062	12,287	5,167	4,837
Taille moyenne des contigs	655 ± 378 bp	701 ± 413	537 ± 378	702 ± 478 bp
Contig le plus long	6199 nt	4808 nt	4525 nt	5452 nt

Table 4.2 : Principales caractéristiques des contigs obtenus pour les transcriptomes d'*E. mutabilis* et d'*E. gracilis* en présence ou en absence d'arsenic. Sont présentés le nombre de nucléotides séquencés avant et après assemblage avec Newbler, le nombre de contigs obtenus pour chaque condition, la taille moyenne de ces contigs et la taille du plus long et du plus court des contigs (limite fixée à 100 nt). EM = *E. mutabilis*, EG = *E. gracilis*.

Le rapport entre le nombre de nucléotides séquencés et le nombre de nucléotides des contigs assemblés suggère que la couverture moyenne du transcriptome est de l'ordre de 10 X. Cependant, cette observation cache une très grande disparité entre les différents contigs. En effet, les transcrits ne présentent pas nécessairement tous un niveau d'expression similaire ce qui conduit inévitablement à des degrés de couverture différents. Les données relatives à l'assemblage des séquences révèlent par ailleurs que la taille moyenne des séquences assemblées est du même ordre de grandeur que celle des transcrits eucaryotes en général, suggérant une qualité d'assemblage relativement bonne. Il est à noter que pour *E. mutabilis* le nombre de contigs est plus important que pour *E. gracilis* (Table 4.2), et ceci malgré un effort de séquençage identique, voire sensiblement plus important. Cette dernière observation suggère que le transcriptome d'*E. mutabilis* est plus complexe que celui d'*E. gracilis*, conduisant à un nombre de contigs après assemblage plus important.

Les contigs obtenus après assemblage pour chacun des 4 transcriptomes ont été annotés grâce à un système d'annotation automatisé en libre accès, MG-RAST (pour « Meta Genom Rapid Annotation using Subsystem Technology »). Ce système permet d'étudier phylogénétiquement et fonctionnellement un lot de séquences nucléotidiques obtenues par différentes techniques de séquençage (Sanger, 454 ou Solexa si les séquences générées par cette dernière technique ont été préalablement assemblées sous forme de contigs).

Les séquences soumises à l'analyse sont dans un premier temps nettoyées en supprimant les séquences strictement similaires considérées comme des artefacts liés aux technologies de séquençage (Meyer *et al.*, 2008). Les séquences nucléotidiques sont dans un second temps annotées automatiquement en réalisant un BlastX, c'est-à-dire en traduisant la séquence nucléotidique dans tous les cadres de lecture puis en recherchant les similarités dans les bases de données protéiques.

L'annotation est effectuée grâce à l'utilisation de bases de données protéiques non redondantes annotées d'une manière experte (FigFam, SEED-NR) ainsi que d'autres bases de données généralistes (NCBI...) ou plus spécialisées telles que les bases de données ribosomales (GreenGenes, RDP-II, European 16S RNA database) ou d'éléments génétiques mobiles (ACLAME). Les algorithmes et critères de significativité utilisés pour l'annotation des séquences sont spécifiques à chacune des bases de données, mais une e-value de 0,01 est généralement retenue pour les annotations fonctionnelles dans les bases de données généralistes.

Une fois annotées, les fonctions identifiées sont réparties dans des « sous-systèmes » regroupant les séquences impliquées dans une même fonction biologique (métabolisme du soufre, des acides aminés, ...) ou un même complexe structural (flagelle, ...). Cette organisation des données en sous-systèmes permet de comparer deux lots de séquences entre eux, en se basant sur le nombre de séquences associées à un sous-système donné pour chacun des lots. Cette comparaison peut être plus ou moins globale, chaque sous-système étant divisé lui-même en sous-catégories. Ainsi, le sous système « métabolisme des protéines » est-t-il sous divisé en « bio-synthèse des protéines, dégradation des protéines, séléno-protéines, modification des protéines et repliement des protéines », cette dernière sous-catégorie étant elle-même sous-divisée en « protéines chaperons, GroEL/GroES et peptidyl cis-trans isomérase ».

Une vision globale de différentes étapes réalisées durant l'analyse automatisée par MG-RAST est présentée ci-dessous (Figure 4.3) :

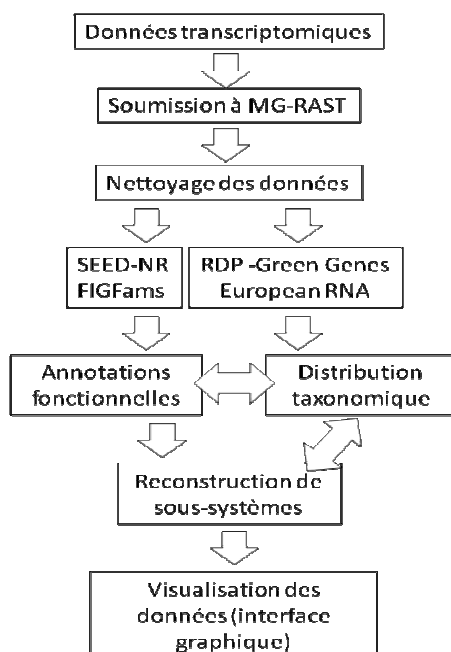


Figure 4.3 : Principales étapes aboutissant à la répartition des séquences soumises à MG-RAST en grands groupes fonctionnels. (Meyer et al., 2008)

L'interface graphique du serveur MG-RAST permet de présenter les résultats sous forme d'histogrammes soulignant rapidement les grandes fonctions biologiques entre deux jeux de données. Dans le cas de l'étude transcriptomique différentielle des euglènes, MG-RAST offre donc la possibilité d'effectuer des analyses comparatives entre les 4 lots de séquences (*E. mutabilis* et *E. gracilis* cultivées en présence ou en absence d'arsenic). Les réponses globales de ces deux protistes face un stress arsénié peuvent donc être étudiées individuellement, mais également comparées entre elles afin de déterminer les points qui pourraient spécifiquement expliquer l'hyper-tolérance à l'arsenic observée chez *E. mutabilis* par rapport à *E. gracilis*.

b. Mécanismes induits chez *E. gracilis* en présence d'arsenic

L'analyse comparative du transcriptome d'*E. gracilis* a permis d'identifier des fonctions biologiques qui n'étaient retrouvées qu'en présence ou qu'en l'absence d'arsenic (Table supplémentaire 4.3 en fin de chapitre). Ces données ont conduit à résumer la réponse globale à l'arsenic d'*E. gracilis* en 3 points:

Synthèse de biofilm : d'importantes modifications des propriétés de surface sont observées en présence d'arsenic. En effet, la synthèse de glycannes contenant du rhamnose, l'utilisation de N-acétyl glucosamine ainsi que la synthèse de d-DTP rhamnose suggèrent fortement la synthèse active par *E. gracilis* d'une matrice extracellulaire sous stress arsénié. Ce mécanisme de défense est susceptible de limiter l'exposition à l'arsenic et semble avoir été développé par ce protiste en réponse à ce métalloïde.

Diminution du métabolisme et inhibition de la division cellulaire : L'exposition à l'arsenic semble également induire une diminution du métabolisme cellulaire chez *E. gracilis* ce qui est soutenu à la fois par une répression de la réplication de l'ADN et par l'induction de gènes homologues à ceux impliqués dans la sporulation chez *Bacillus subtilis* (SpoVS). En effet, *E. gracilis* est décrit comme pouvant entrer en dormance en produisant des cystes lorsque les conditions environnementales sont défavorables (Triemer, 1980). Cependant, des observations en microscopie optique n'ont pour l'instant pas permis de mettre en évidence de telles structures en présence d'arsenic chez *E. gracilis*. D'autre part il est à noter que de multiples mécanismes conduisant à la production d'oxyde nitrique (NO) sont spécifiquement induits en présence d'arsenic. Si le rôle exact du NO n'a pas été étudié en détail chez les euglènes, ce composé conduit chez certaines diatomées à une diminution de leur croissance et à l'induction de protéines liées à la mort cellulaire programmée (Foresi *et al.*, 2010). Par conséquent, une extrapolation de ces observations à *E. gracilis* pourrait aller dans le sens d'une diminution de son métabolisme en présence d'arsenic, hypothèse qui est par

ailleurs confortées par l'augmentation du temps de génération des euglènes en présence d'arsenic (chapitre III). La diminution du métabolisme d'*E. gracilis* pourrait par conséquent constituer un mécanisme de défense passif face à l'arsenic chez cette espèce.

Réponse au stress oxydatif : L'arsenic induit la transcription du gène codant pour la superoxyde dismutase chez *E. gracilis*. D'autre part, le gène codant pour la glutathion-S-transférase, enzyme impliquée dans la chélation de composés toxiques par le glutathion, est réprimé en présence d'arsenic suggérant que le contenu en glutathion de la cellule est plutôt orienté vers la lutte contre le stress oxydatif que vers la séquestration de ce métalloïde. D'autre part, aucune induction de la synthèse du glutathion n'a pu être mise en évidence.

L'étude du transcriptome différentiel d'*E. gracilis* a permis de mettre en évidence l'induction de mécanismes de défense au stress arsénié qui pourraient expliquer le fait que sa tolérance à ce métalloïde, bien que moindre que celle d'*E. mutabilis*, atteint des niveaux relativement élevés (environ 4 fois la concentration minimal inhibitrice de *Saccharomyces cerevisiae*). Cependant, *E. gracilis* semble orienter sa réponse à l'arsenic principalement vers des mécanismes de défense passifs tels que la dormance et une diminution de son métabolisme.

c. Mécanismes induits chez *E. mutabilis* en présence d'arsenic

Afin de comprendre plus en détail les mécanismes qui sont spécifiquement impliqués dans la tolérance à l'arsenic chez *E. mutabilis*, l'analyse des fonctions induites ou réprimées en présence de ce métalloïde a été entreprise par une approche similaire à celle utilisée chez *E. gracilis*. Ces fonctions biologiques dont l'expression semble être affectée par l'arsenic sont présentées en fin de chapitre (Table supplémentaire 5.3) et ont conduit à définir 3 mécanismes principaux de défense à l'arsenic chez *E. mutabilis*.

Mécanisme de réparation de l'ADN : de multiples mécanismes de réparation de l'ADN sont spécifiquement induits sous stress arsénié chez *E. mutabilis* et pourraient jouer un rôle très important dans sa tolérance à l'arsenic. En effet, ce métalloïde, en induisant un stress oxydatif permanent et, en interagissant avec les mécanismes de réparation de l'ADN, est susceptible d'induire de nombreuses mutations au sein du génome et d'affecter par la même occasion la survie de l'organisme (Flora, 2011; Jin *et al.*, 2008). La présence de mécanismes redondants de réparation de l'ADN limiterait donc indirectement l'effet toxique de l'arsenic sur la cellule.

Métabolisme du glutathion : l'arsenic activerait le métabolisme du glutathion chez *E. mutabilis*, hypothèse soutenue à la fois par l'induction du cycle du gamma glutamyl et par l'assimilation de cystine, deux composés impliqués dans la synthèse de ce composé. D'autre part, la glutathion-S-transférase, enzyme utilisant comme substrat le glutathion pour sa complexation avec des éléments toxiques tels l'arsenic, est réprimée en présence de ce métalloïde. Ces observations suggèrent donc une synthèse active par *E. mutabilis* de glutathion sous stress arsénié. Cependant, au même titre que pour *E. gracilis*, ce composé sert visiblement davantage à lutter contre le stress oxydatif associé à l'arsenic plutôt qu'à séquestrer directement ce métalloïde sous forme de complexe. Cette hypothèse permettrait d'expliquer le fait que la spéciation de l'arsenic dans le compartiment intracellulaire n'a pas permis de mettre en évidence de forme arséniées couplées à du triglutathion (chapitre III).

Diminution de l'exposition à l'arsenic : *E. mutabilis* semble, comme *E. gracilis*, limiter l'exposition du compartiment intracellulaire à l'arsenic en induisant la synthèse de biofilm. En effet, la synthèse d'exo-polysaccharides couplée à l'activation du système de sécrétion de type IV (impliqués dans la sécrétion de constituants majeurs de la matrice extracellulaire tels que les protéines et les nucléoprotéines), pourraient jouer un rôle important dans ce processus. La diminution de l'exposition à l'arsenic chez *E. mutabilis* est également soutenue par l'induction du système de régulation à deux composants impliqué dans l'expression du transporteur spécifique du phosphate, limitant l'incorporation d'arséniate dans la cellule. Enfin, le métabolisme des sucres est également étonnamment affecté par la présence d'arsenic. Cette observation pourrait être liée au fait que l'arsenic est décrit comme pouvant entrer dans la cellule par le biais de transporteurs du glucose, et que par conséquent, l'utilisation d'autres sources de carbone sont favorisées par ce protiste en présence de ce métalloïde.

Cette étude souligne de multiples mécanismes de défense mis en œuvre par *E. mutabilis* pour pallier le stress arsénié. La plupart de ces mécanismes prédits par l'approche transcriptomique semblent spécifique à cette espèce en comparaison à *E. gracilis*. Cependant, afin de souligner plus clairement les différences entre les deux espèces, le transcriptome de ces deux organismes a été comparé entre eux.

d. Etude transcriptomique comparative entre *E. mutabilis* et *E. gracilis*

Dans un premier temps, les séquences annotées pour chacun des deux protistes ont été regroupées en 28 grandes catégories fonctionnelles. Ainsi, à chacune de ces 28 catégories, peut être associée une valeur qui est d'autant plus grande que le nombre de fonctions annotées qu'elle renferme est élevé (Figure 4.4).

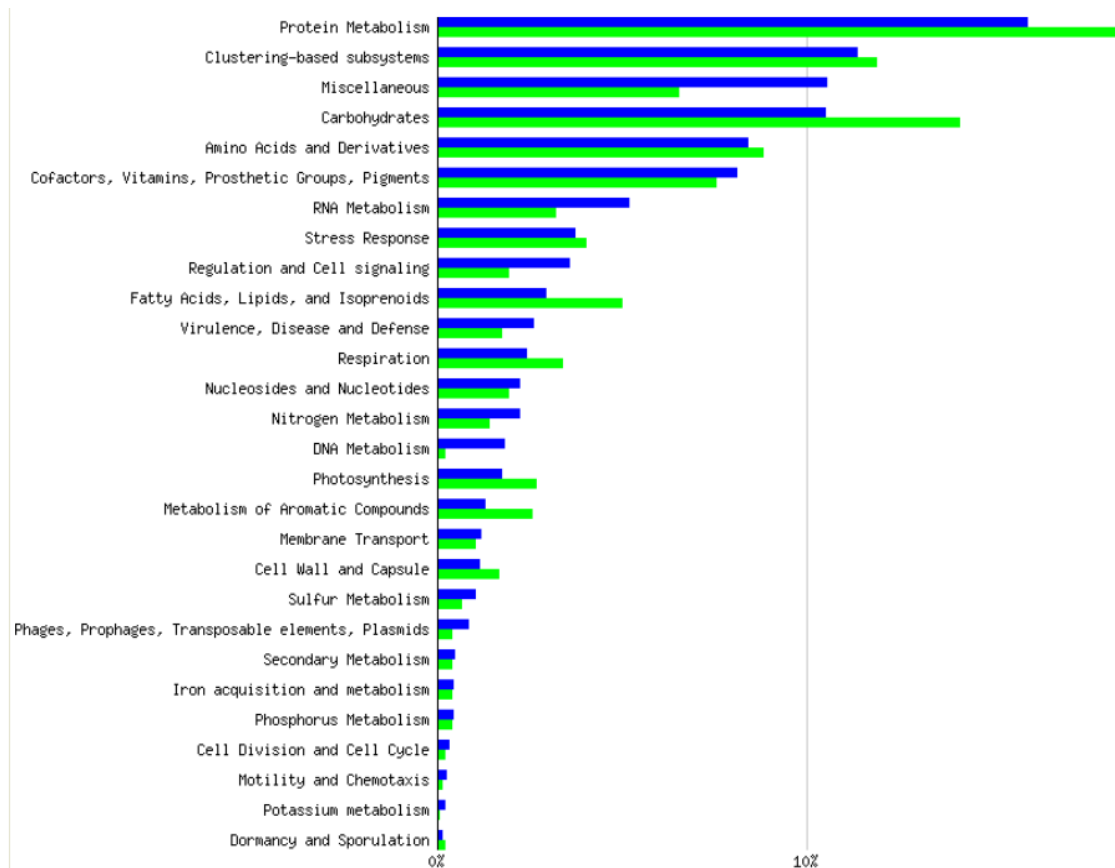


Figure 4.4: Répartition en 28 grands groupes fonctionnels des séquences annotées par MG-RAST correspondant aux séquences d'*E. gracilis* et d'*E. mutabilis* en présence d'arsenic (en vert et en bleu respectivement). L'axe des abscisses représente la proportion relative des séquences dans chaque catégorie fonctionnelle.

Cette représentation graphique permet de souligner que certaines classes fonctionnelles semblent être surreprésentées chez *E. mutabilis* (métabolisme des ARNs, régulation et signalisation cellulaire, métabolisme de l'ADN) tandis que d'autres le sont plus chez *E. gracilis* (métabolisme des sucres, métabolisme des acides gras, photosynthèse, respiration, métabolisme des composés aromatiques). Afin de réaliser des analyses plus poussées, chacune des 28 catégories fonctionnelles a été étudiée plus en détail. Ceci a conduit à l'identification de plusieurs fonctions propres à l'un ou

l'autre protiste (Tables supplémentaire 5.4) et à définir 3 points principaux qui pourraient expliquer les adaptations physiologiques spécifiques d'*E. mutabilis* lui offrant la possibilité de résister à l'arsenic :

Synthèse d'exo-polysaccharides spécifiques : Les analyses transcriptomiques ont révélé que les deux espèces d'euglènes étaient potentiellement capables de synthétiser une matrice extracellulaire, processus qui pourrait être spécifiquement régulé chez *E. mutabilis* par des mécanismes de quorum sensing et de chimiotactisme induits en présence d'arsenic. Cependant, la matrice extracellulaire synthétisée par ces deux espèces est visiblement de nature différente. Ainsi, *E. mutabilis* semble avoir développé un métabolisme des alginates, polymères dont la structure est présentée ci-dessous (Figure 4.5). Si le rôle précis des constituants du biofilm spécifiquement retrouvés chez *E. mutabilis* reste à déterminer, il est probable que ces derniers puissent expliquer ses propriétés de surface originales, en particulier sa forte hydrophobicité conduisant à une moindre exposition à l'arsenic (chapitre III).

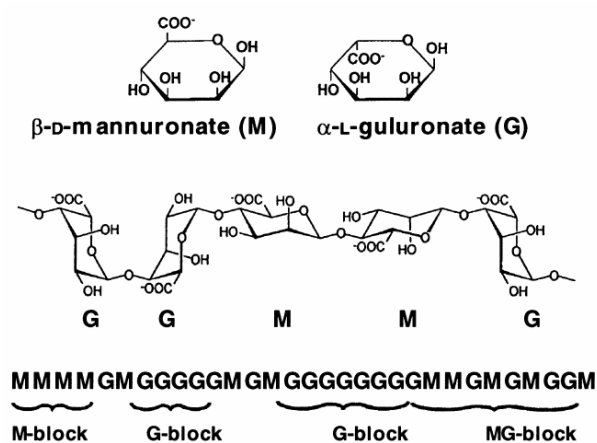


Figure 4.5 : Caractéristiques structurales des alginates et des monomères qui les constituent. Ces monomères sont regroupés sous forme de blocs plus ou moins longs au sein des polysaccharides, ce qui peut influencer leurs propriétés physico-chimiques (Draget et al., 2005).

La synthèse d'une matrice extracellulaire par *E. mutabilis* a par ailleurs été confirmée par des observations microscopiques, appuyant les données obtenues par l'analyse transcriptomique (figure 4.6).

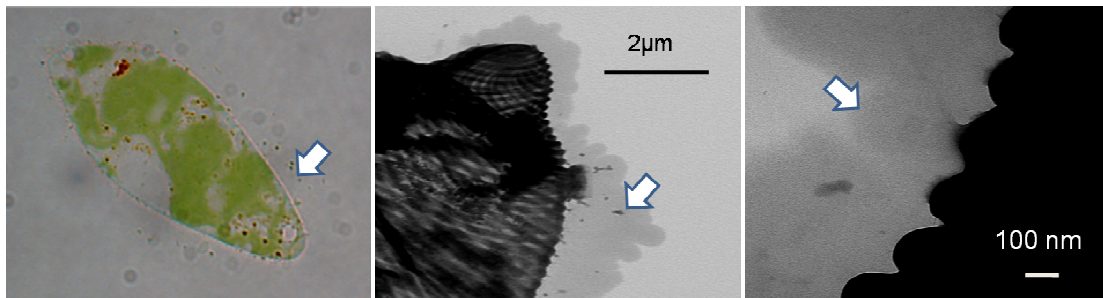


Figure 4.6 : Images de microscopie optique et électronique montrant la synthèse d'une matrice extracellulaire (flèches) par *E. mutabilis*.

Transport des métaux et autres composés inorganiques : L'analyse comparative du transcriptome des deux espèces d'euglènes a révélé de nombreux mécanismes spécifiques à *E. mutabilis* impliqués dans le transport et d'expulsion de métaux. En particulier, des pompes d'efflux du cuivre et de l'arsenic ont été identifiées chez cette espèce ce qui pourrait expliquer, en partie, la moindre concentration d'arsenic dans ses cellules par rapport à *E. gracilis* (chapitre III). Par ailleurs, des transporteurs pour des composés inorganiques indispensables tels que le fer, le molybdène, le manganèse, le sulfate ainsi que des transporteurs à haute affinité pour le phosphate sont retrouvés spécifiquement chez *E. mutabilis*. Enfin, ce protiste exprime également des pompes d'efflux de protons, point qui avait déjà été mis en avant par les analyses de protéomique *in situ* (chapitre II), mais qui s'est révélé ici être spécifique à cette espèce. La présence de ces multiples mécanismes de transport suggère qu'*E. mutabilis* contrôlerait finement la nature et la quantité des composés inorganiques transportés à travers sa pellicule, lui offrant la possibilité de prélever dans l'environnement les éléments indispensables à sa croissance tout en limitant l'entrée de composés toxiques.

Mécanismes de réparation et de maintien de l'intégrité cellulaire : Les multiples mécanismes de réparation et de maintien de l'intégrité cellulaire ont également été montrés comme étant spécifiquement retrouvés chez *E. mutabilis*. En effet, des mécanismes de réparation de l'ADN, de stabilisation de la structure tri-dimensionnelle des protéines et un métabolisme du glutathion orienté vers les réactions d'oxydo-réduction ont été identifiés uniquement chez ce protiste. Ces fonctions biologiques pourraient jouer un rôle important dans la tolérance au stress arsénisé chez *E. mutabilis* dans la mesure où ces fonctions ont été montrées comme étant induites en présence de ce métalloïde chez cette espèce.

C : DISCUSSION

1) approche protéomique

Les analyses de protéomique différentielle permettent d'appuyer dans un certain nombre de cas ces observations physiologiques et/ou des données de la littérature sur le stress arsénié. Cependant, le faible nombre de protéines à accumulation différentielle identifiées ainsi que l'absence de protéines directement impliquées dans le métabolisme de l'arsenic n'offrent que peu d'information sur les mécanismes moléculaires expliquant l'hyper-tolérance à l'arsenic chez *E. mutabilis*. Différentes hypothèses peuvent être émises pour expliquer la faible modification du contenu protéique de ce protiste sous stress arsénié.

Premièrement, le contenu protéique d'*E. mutabilis* pourrait réellement n'être que légèrement affecté en présence d'arsenic. Selon cette hypothèse, cet organisme disposerait majoritairement de mécanismes de résistance constitutifs qui ne seraient, par définition, pas induits. Si cette hypothèse ne peut pas être exclue, elle reste néanmoins peu probable au vu des nombreuses manifestations physiologiques induites en présence de ce métalloïde et qui impliquent un minimum de réorganisation du contenu protéique. Une deuxième hypothèse repose sur le fait que la protéomique différentielle ne permette pas d'accéder à l'intégralité du contenu protéique de la cellule. En effet, seules les protéines les plus solubles dans le tampon d'extraction sont accessibles pour l'analyse et la sensibilité limitée des techniques de coloration des gels ne permet d'accéder qu'aux protéines les plus accumulées dans le compartiment intracellulaire. Les protéines membranaires et/ou faiblement accumulées sont donc généralement exclues de l'analyse protéomique différentielle bien qu'elles puissent jouer un rôle important dans la réponse au stress arsénié. De plus, il n'est pas à exclure que le protocole d'extraction des protéines utilisé dans cette analyse ne soit pas optimal. En effet, bien que la séparation bi-dimensionnelle des protéines provenant des différentes cultures d'*E. mutabilis* conduite à des résultats très reproductibles, le nombre de spots par gel de situe aux alentours de 650 ce qui pourrait sans doute être amélioré. Plusieurs protocoles d'extraction ont été testés et certains d'entre eux, notamment ceux basés sur l'utilisation de détergents forts comme le Sodium Dodécyl Sulfate (SDS), se sont montrés plus efficaces pour lyser les cellules et/ou l'extraction des protéines. Cependant, ces protocoles sont incompatibles avec la technique de protéomique bi-dimensionnelle car le SDS, en se fixant sur les protéines, modifie à la fois leurs masses apparentes et leurs points isoélectriques.

Ces résultats ont conduit à développer une autre approche de génomique fonctionnelle pour comprendre les mécanismes moléculaires induits chez les euglènes sous stress arsénié et pouvant expliquer les différences physiologiques observées.

2) Approche transcriptomique : bilan et perspectives

L'étude comparative des transcriptomes d'*E. mutabilis* et d'*E. gracilis* en présence ou en absence d'arsenic a révélé un certain nombre de mécanismes qui pourraient expliquer la différence de sensibilité à l'arsenic chez ces deux espèces. En particulier, de nombreux mécanismes de réparation de l'ADN et de maintien de l'intégrité cellulaire, une entrée sélective des métaux dans la cellule par le biais de transporteurs spécifiques, des pompes d'efflux de l'arsenic, une incorporation active de fer, une composition du biofilm originale et de nombreux mécanismes anti-oxydants semblent être les points qui distinguent le plus *E. mutabilis* par rapport à *E. gracilis* au niveau de la réponse au stress arsénié. Ces caractéristiques pourraient conduire à une tolérance accrue à l'arsenic par différents mécanismes.

Incorporation du fer : l'incorporation du fer joue visiblement un rôle important chez *E. mutabilis* sous stress arsénié. En effet, certains mécanismes d'incorporation du fer sont retrouvés uniquement chez *E. mutabilis* qui semble activer par ailleurs l'incorporation de ce composé en présence d'arsenic. Cette hypothèse est soutenue par des observations microscopiques montrant, après 2 mois d'incubation dans un milieu minéral contenant du fer ferrique, une accumulation importante de ce composé dans les cellules sous forme de granules orangés, en particulier chez *E. mutabilis* (Figure 4.7).

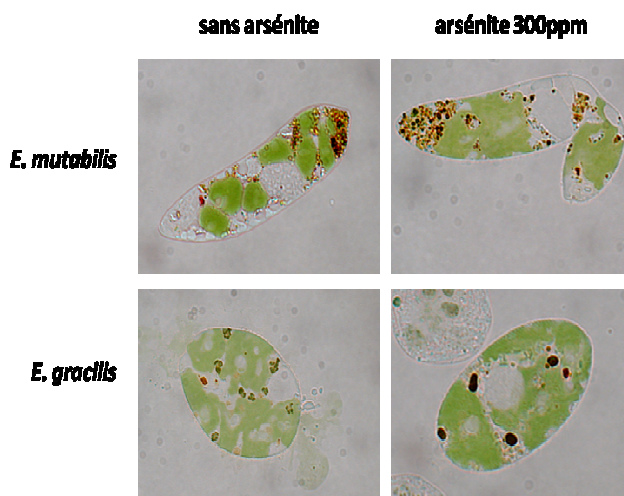


Figure 4.7 : Observations en microscopie optique d'*E. mutabilis* et d'*E. gracilis* après une incubation de deux mois dans un milieu liquide contenant du fer et en présence ou en absence d'arsenic.

Le fer ayant une forte affinité avec l'arsenic, ces structures pourraient être impliquées dans l'immobilisation de ce métalloïde au sein de la cellule et donc constituer un mécanisme de défense à part entière. Pour être validée, cette hypothèse nécessite d'identifier la nature des éléments présents dans ces granules. Un protocole visant à déterminer la composition de ces particules par micro-analyse X a donc été développé en collaboration avec Mathieu Erhardt (IBMP, Strasbourg) et Jacques Faerber (IPCMS, Cronenbourg) (Figure 4.8). Cette analyse repose sur l'excitation des atomes par un rayon X balayant l'échantillon puis par l'analyse spectrale des photons émis en chaque point de l'échantillon, reflétant la nature des éléments excités. Cette approche permet d'avoir une cartographie à l'échelle micrométrique de la nature des éléments constituant l'euglène, et en particulier celle des granules de fer. Afin de permettre ces analyses, les échantillons ont dans un premier temps été fixés puis ont fait l'objet de coupes ultra-fines (100 nm). Les analyses préliminaires ont révélés que les raies d'émissions d'un agent fixateur, le tétr oxyde d'osmium, se superposaient aux raies de l'arsenic ce qui perturbe l'interprétation des résultats. Un protocole alternatif de préparation des échantillons est donc en cours de développement afin de permettre les analyses ultérieures.

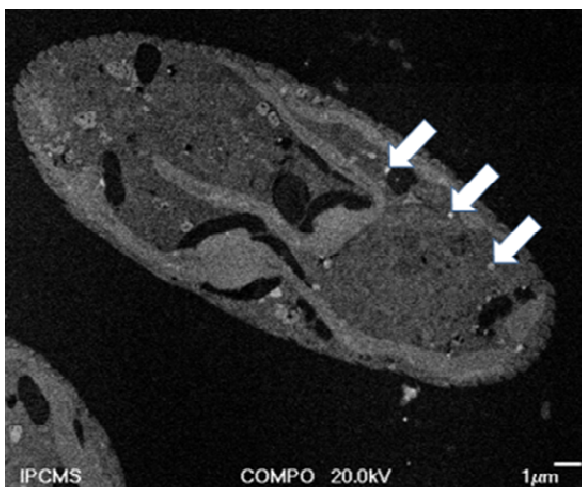


Figure 4.8 : Coupe ultra-fine réalisée sur les euglènes permettant de mettre en évidence la présence des particules (flèches) dont la composition reste à déterminer.

Les analyses transcriptomiques offrent donc un aperçu original des mécanismes potentiels intervenant dans l'hyper-tolérance à l'arsenic observés chez *E. mutabilis*. Si certains de ces mécanismes ont déjà été validés expérimentalement (accumulation d'arsenic, synthèse de biofilm, formation de granules de fer,...), des analyses plus poussées restent à être finalisées pour tirer des conclusions définitives sur ces mécanismes de tolérance à l'arsenic.

Maintien de l'intégrité du génome : l'étude comparative du transcriptome d'*E. gracilis* et d'*E. mutabilis* a fait ressortir que cette dernière disposait de multiples mécanismes impliqués dans la réparation d'ADN et de lutte contre le stress oxydatif. Ces caractéristiques pourraient jouer un rôle

primordial dans les DMA où les fortes concentrations en ions métalliques et en arsenic conduisent à de telles contraintes susceptibles d'induire des mutations de l'ADN et des déstabilisations de la structure tri-dimensionnelle des protéines (Flora, 2011). Par conséquent, l'adaptation d'*E. mutabilis* aux conditions extrêmes des DMA ne serait pas seulement liée à des mécanismes de défense visant à limiter son exposition à l'arsenic et aux autres métaux, mais reposerait également sur le maintien très actif de son intégrité cellulaire. Il est à noter que des observations similaires ont été réalisées chez *Deinococcus radiodurans*, bactérie où de nombreux mécanismes de réparation de l'ADN et de maintien de l'intégrité cellulaire permettent à cet organisme de faire face à des doses très élevées de radiations ionisantes génératrices de radicaux libres (Slade & Radman, 2011). Afin de confirmer la présence d'un mécanisme similaire chez *E. mutabilis*, un protocole visant à évaluer et comparer le nombre de cassures de l'ADN occasionnées par le stress oxydatif arséno-dépendant chez les deux espèces d'euglènes est en cours de développement. Le protocole retenu est décrit ci-dessous et repose sur le test des comètes mis au point historiquement par Osting et Johanson (1984)(Figure 4.9).

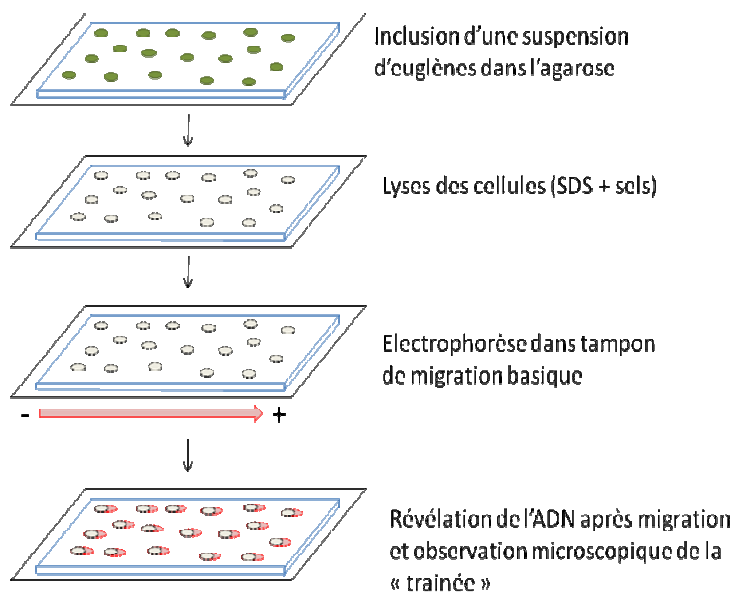


Figure 4.9 : Détermination de l'effet de l'arsenic sur l'intégrité du génome d'*E. gracilis* et d'*E. mutabilis* selon le « test des comètes ». Après avoir été incluses dans l'agarose, les euglènes cultivées en présence ou en absence d'arsenic sont lysées et l'ADN génomique est séparé par électrophorèse dans un tampon basique ($pH = 13$) pour dénaturer l'ADN sous forme simple brin. Dans le cas de cassures liées au stress oxydatif arséno-dépendant, les fragments d'ADN générés forment une trainée d'autant plus longue que l'ADN est dégradé.

Le test des comètes offre l'avantage d'être relativement sensible et permet de détecter environ 0,2 à 2 coupures par 10^6 paires de bases. Cependant, le protocole de lyse des euglènes incluses dans l'agarose nécessite encore d'être optimisé pour faciliter l'interprétation des résultats.

Matrice extracellulaire : en comparaison à d'autres polysaccharides, les alginates potentiellement produits par *E. mutabilis* ont comme principale caractéristique de lier les cations multivalents de manière sélective (Draguet et al., 2005). En effet, certains alginates extraits d'une algue brune du genre *Sargassum* ont été décrits comme liant faiblement le cadmium et le calcium alors qu'ils fixent avec une forte affinité le sodium et le magnésium (Davis et al., 2003). Par conséquent, *E. mutabilis* pourrait, en synthétisant des alginates, se protéger sélectivement de certains cations toxiques tout en laissant une plus grande bio-disponibilité pour d'autres composés. Afin de déterminer si la composition, visiblement différente, des matrices extracellulaires synthétisées par les deux espèces d'euglène peut affecter l'exposition de la cellule à l'arsenic des études plus poussées sont envisagées. En particulier, les biofilms peuvent dans un premier temps être soumis à une digestion enzymatique (par la glucanase par exemple) pour analyser indépendamment les monomères qui les constituent par des approches de métabolomique. Dans un second temps, des tests d'affinité entre ces matrices extracellulaires et différents métaux permettraient de mettre en évidence une affinité différente des ces structures pour certains métaux (Figure 4.10).

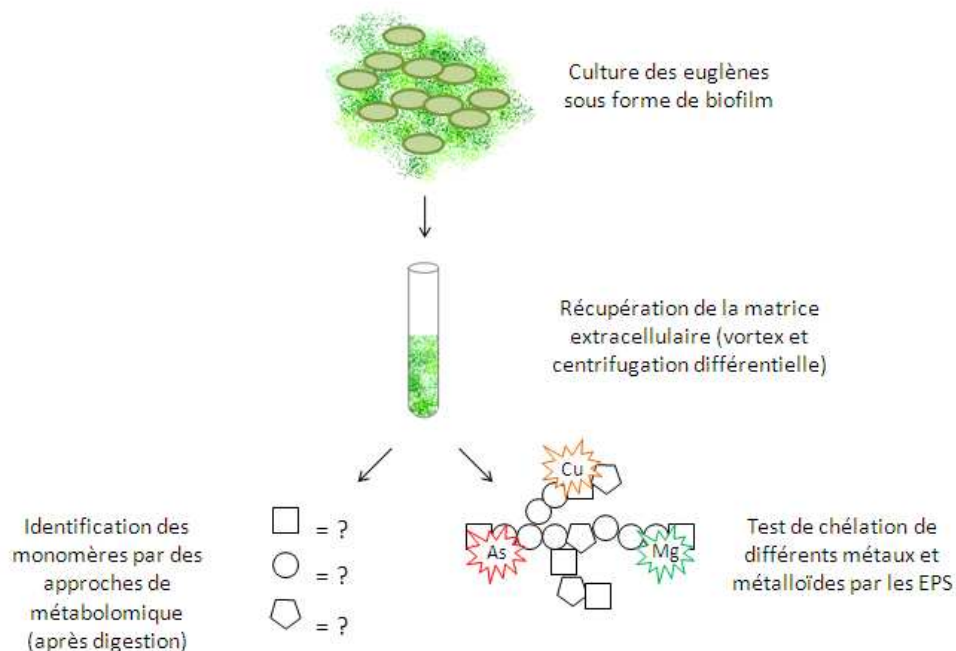


Figure 4.10 : Principales étapes d'un protocole visant à étudier les caractéristiques physico-chimiques des biofilms produit par les euglènes,

Ces analyses permettraient à la fois de préciser si la matrice extracellulaire produite par les euglènes constitue un mécanisme de défense spécifique, mais également si ces organismes pourraient être utilisés pour chélater sélectivement certains métaux et ainsi jouer un rôle éventuel dans des procédés de bio-remédiation.

Ces résultats offrent donc des pistes prometteuses dans la compréhension de la réponse au stress arsénié chez ces deux espèces, mais également dans la compréhension de leur physiologie dans leur globalité.

D : TABLES SUPPLEMENTAIRES

Table 4.3 : Fonctions biologiques induites ou réprimées en présence d'arsenic chez *E. gracilis*.

		<i>E. gracilis</i> (plus As)	<i>E. gracilis</i> (sans As)
Capsule et membrane cellulaire	-	Synthèse de Glycanes contenant du rhamnose Synthèse de dTDP-rhamnose Métabolisme de l'acide sialique Biosynthèse EPS Assemblage et biosynthèse de polysaccharides de la capsule Synthèse d'heptose pour la capsule	
Dormance sporulation	-	Induction de gènes homologues à SpoVS	
Métabolisme l'ADN	-		Réplication de l'ADN
Métabolisme l'azote	-	Synthèse d'oxyde nitrique (NO)	
Métabolisme sucres	-	Utilisation de lactate (lactate deshydrogenase) Biosynthèse de tréhalose Utilisation de ribose Utilisation de N acétylgalactosamine	Utilisation de melbiose Utilisation de raffinose Utilisation du mannose
Métabolisme soufre	-		Assimilation de soufre organique (cystine) Métabolisme des sulfatides et galactocéramides
Métabolisme photosynthétique	-		protéorhodopsine
Métabolites secondaires	-	Synthèse de phényl propanoïdes (phytoalexine, flavone et acide caféique)	
Phage, prophage et éléments transposables	-	Induction d'un phage homologue à R1t d'un streptocoque	
Régulation signalisation cellulaire	-	Voie de la MAP kinase P38 Régulation de la protéolyse (chaperon grp78) Métabolisme de (p)ppGpp	
Réponse aux stress	-	Manganèse superoxide dismutase Métabolisme du diméthyl-arginine (synthèse de NO)	Glutathion S transférase (réaction non redox)
Transport membranaire	-	Système de sécrétion de type III Transport du Zinc	Système de sécrétion de type IV (protéines et nucléoprotéines)

Table 4.4 : Fonctions biologiques induites ou réprimées par *E. mutabilis* en présence d'arsenic.

		<i>E. mutabilis</i> (plus As)	<i>E. mutabilis</i> (sans As)
Capsule et membrane cellulaire	et	- Synthèse d'EPS - Modification de la capsule (adenylyl sulfate kinase)	- Synthèse d'acide colanique - Synthèse d'heptose pour la capsule
Métabolisme de l'ADN	de	- Recombinaison - Réparation de l'ADN (homo. au système bactérien RecFOR) - Réparation de l'ADN (homo. à UvrD et aux hélicases associées) - Réparation de l'ADN (homo. au système RecBCD bactérien)	
Métabolisme de l'azote	de	- Fixation de l'azote	
Metabolisme des ARNs	des	- Induction d'un facteur de transcription homologue à RpoD	
Métabolisme des sucres	des	- Utilisation de sucrose - Utilisation de galacturonate - Utilisation de xylose - Utilisation de fructose	- Isobutyryl CoA to propionyl CoA - Utilisation de fucose - Utilisation de ribose
Métabolisme du fer		- Transport du fer	
Métabolisme du phosphate	du	- Système de régulation à deux composantes (PhoR / PhoB)	
Métabolisme du soufre	du	- Assimilation de soufre organique (cystine)	
Métabolites secondaires		- Synthèse d'octadécanoïdes et de cannabinoïdes	
Mobilité chimiotactisme	et	- Chiomotactisme homologue au système bactérien	
Phage, prophage et éléments transposables		- Régulation de l'expression gènes phagiques (homologue au phage phi-MU50B) - Réplication de phage	
Réponse aux stress		- Biosynthèse du glutathion et cycle du gamma-glutamyle	- Glutathion S transférase (réaction non redox)
Transport membranaire		- Système de sécrétion de type IV (protéines et nucléo-protéines)	

Table 4.5 : Principales différences révélées par les analyses de transcriptomique différentielle effectuées entre *E. gracilis* et *E. mutabilis* exposées à l'arsenic.

		<i>E. mutabilis</i>	<i>E. gracilis</i>
Capsule et membrane cellulaire		- Métabolisme de l'alginate - Biosynthèse d'EPS homologues à ceux de <i>Vibrio</i> - Modification de la capsule (adenylyl sulfate kinase)	- Synthèse d'une capsule à base d'acide hyaluronique - Biosynthèse d'acide colanique et d'EPS - Synthèse d'heptose pour la capsule
Division cellulaire		- Partition des chromosomes	
Métabolisme des acides aminés		- Dégradation de la thréonine - Synthèse de la lysine	- Synthèse de tryptophane
Métabolisme de l'ADN		- Réparation et recombinaison de l'ADN - Réparation de l'ADN (système UvrABC) - Réparation de l'ADN (homo. au système bactérien RecFOR) - Réparation de l'ADN (homo. au système RecBCD bactérien) - Réparation de l'ADN (homo. à UvrD et aux hélicases associées) - Réplication de l'ADN - Topoisomérase I (ATP indépendant) - Topoisomérase II (ATP dépendant)	

Métabolisme de l'azote	- Fixation d'azote - Ammonification de nitrates et de nitrites - Assimilation d'ammonium	- Synthèse d'oxyde nitrique
Métabolisme des composés aromatiques	- Dégradation et transport du benzoate - Catabolisme d'amines aromatiques	
Métabolisme des protéines	- Repliement des protéines : Peptidyl-prolylcis-trans isomérase - Repliement des protéines : peptide methionine sulfoxide réductase - Biosynthèse de lipoprotéines	
Métabolisme des sucres	- Utilisation du lactose (beta galactosidase) - Utilisation du sucrose (fructokinase) - Synthèse de tréhalose	
Métabolisme du fer	- Assemblage et transport de sidérophore - Transport du fer	
Métabolisme du phosphate	- Transporteur du phosphate à haute affinité - Système de régulation à deux composantes (PhoR / PhoB)	- Phosphoenolpyruvate phosphomutase
Métabolisme du potassium	- Homéostasie du potassium - Système d'efflux du potassium régulé par le glutathion	
Métabolisme du soufre	- Assimilation de soufre inorganique (sulfate perméase) - Sulfite réductase - Métabolisme des sulfatides et galactocéramides	
Métabolisme photosynthétique	- Protéorhodopsine	
Métabolites secondaires	- Biosynthèse de phényl-propanoïdes (subérine) - Synthèse d'octadécanoïdes	
Mobilité et chimiotactisme	- Chiomotactisme homologue au système bactérien	
Phage, prophage et éléments transposables	- Réplication de phage - Synthèse d'ADN phagique - Régulation de l'expression de gènes phagiques homologues au phage phi-MU50B	
Réponse au stress osmotique	- Synthèse et transport de choline et de bétaïne - Synthèse de glucanes (beta 1-2 glucanes) périplasmiques osmomorégulés	
Réponse au stress oxydatif	- Biosynthèse du glutathion et cycle du gamma-glutamyle - Glutarédoxine - Superoxyde dismutase - Protéine de stress thermique (homologue à DnaK) - Protéine de choc froid (famille CspA)	- Glutathion S transférase (réaction non Redox) - Manganèse superoxyde dismutase
Résistance au stress acide	- Glutamate décarboxylase	
Signalisation cellulaire	- Quorum sensing et formation de biofilms (synthèse d'un homologue de l'autoinducer 2)	
Synthèse de cofacteur	- Synthèse de cobalamine	
Transcription	- Facteur d'initiation de la transcription homologue à RpoD	
Transports membranaires	- System de transport TAT Twin arginine translocating system) - Transport du manganèse - Transport du molybdène - Système de transport Ton et Tol - Système de transport de type IV homologue (Single Strand Binding protein et Topoisomérase I)	- Système de sécrétion de type II (pilus de type IV) - Système de sécrétion de type III
Virulence et mécanisme de défense	- Mécanismes de défense à l'arsenic (pompe d'efflux) - Pompe d'efflux du cuivre - Cluster de résistance multidrogues homologue à Mdt ABCD - Beta lactamase	- Protéine de résistance au Cobalt/zinc/cadmium

CONCLUSION GENERALE ET PERSPECTIVES

Les approches de génomique fonctionnelle (métabolomique, transcriptomique, protéomique) mises en œuvre durant ce travail de thèse ont permis de préciser considérablement le fonctionnement de la communauté microbienne dans le DMA de Carnoulès. Ainsi, des analyses de protéo-génomiques ont souligné les principales potentialités fonctionnelles de la communauté bactérienne et ont validé l'expression *in situ* de certaines fonctions impliquées directement dans les processus de bio-remédiation. D'autre part, ces approches ont également conduit à lever un voile sur les multiples interactions trophiques siégeant au sein de cet écosystème et ont souligné le rôle clé d'*E. mutabilis* dans la sécrétion active de métabolites. Prises dans leur ensemble, les analyses effectuées *in situ* nous ont permis de proposer un modèle du fonctionnement de cet écosystème microbien en conditions environnementales, en incluant certaines espèces non cultivables ou non cultivées comme CARN 1 et 4 appartenant à un nouveau phylum bactérien.

Cependant, de telles analyses présentent souvent un caractère descriptif et nécessitent d'être validées par des observations en condition de laboratoire. Ainsi, l'isolement *in vitro* d'*E. mutabilis* a permis de définir précisément quels types de métabolites étaient susceptibles d'être libérés spécifiquement par ce protiste en conditions environnementales tout en soulignant sa réponse à l'arsenic. Les analyses comparatives avec *E. gracilis* ont mis en évidence le fait que l'hyper-tolérance à l'arsenic observée chez *E. mutabilis* était multifactorielle et impliquait très probablement à la fois les propriétés hydrophobes de sa surface, des mécanismes de transport de l'arsenic, de nombreux mécanismes de maintien de l'intégrité cellulaire et un mode de développement sous forme de biofilm. Certaines expérimentations physiologiques restent cependant à réaliser pour valider ces résultats et constituent des perspectives à court terme. Elles offrent néanmoins dès maintenant des pistes prometteuses dans la compréhension des mécanismes adaptatifs de ce protiste face à un stress arsénié. Ces travaux étant focalisés sur un micro-organisme considérés comme un bio-indicateur de DMA, les observations réalisées sur le site de Carnoulès sont susceptibles d'être extrapolées, au moins en partie, à d'autres environnements similaires offrant d'une manière plus générale un intérêt dans la modélisation de ces écosystèmes.

A moyen et plus long terme, d'autres approches peuvent être envisagées pour comprendre plus en détail les fonctions impliquées dans la tolérance à l'arsenic chez *E. mutabilis* et pour valider les interactions trophiques qui siègent au sein de la communauté microbienne sur le site de Carnoulès. Deux exemples d'approches sont proposés ci-dessous et pourraient faire l'objet de travaux ultérieurs.

Modélisation des flux trophiques au sein de la communauté microbienne :

Une approche pour étudier en détail les flux de matière organique au sein de la communauté microbienne du site de Carnoulès pourrait reposer sur l'utilisation de traceurs radioactifs en microcosme (Figure 5.1). Ainsi, les euglènes isolées en culture pure et cultivées en présence de $C^{13}O_2$ pourraient être co-cultivées dans un second temps avec la communauté bactérienne récupérée par gradient de densité. Après différentes durée de co-cultures, les espèces bactériennes pourraient être séparées les unes des autres grâce à l'usage de techniques couplant marquage FISH et cytométrie en flux. Cette approche pourrait avantageusement tirer profit des données générées par l'approche méta-génomique pour élaborer des sondes spécifiques des différentes espèces/populations bactériennes. Un suivi des métabolites marqués dans les différentes espèces bactériennes au cours du temps permettrait de préciser quels métabolites sont prélevés par quels organismes et ainsi de valider un certain nombre d'hypothèses émises à partir des observations réalisées *in situ*.

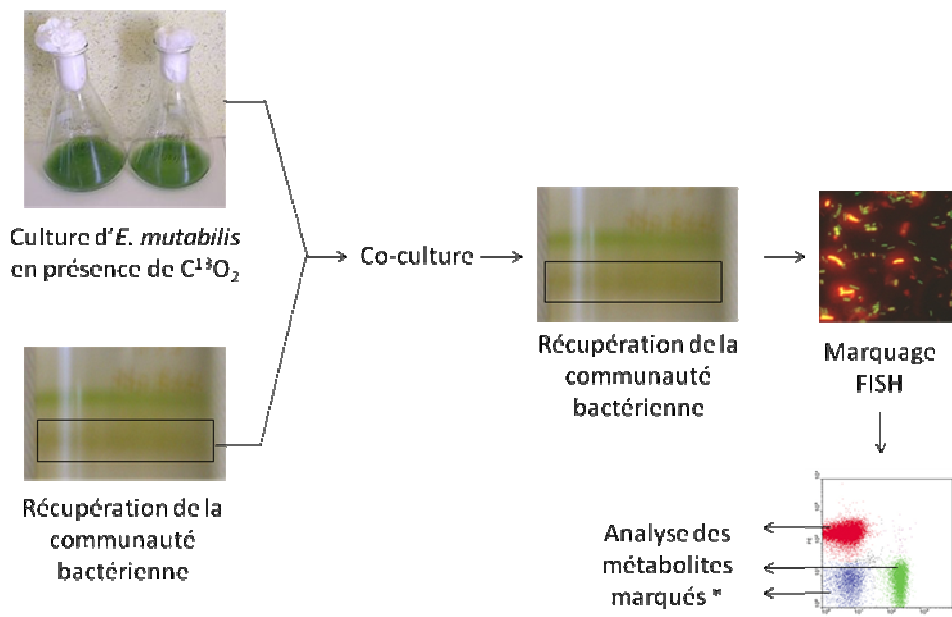


Figure 5.1 : Perspective d'étude des flux métaboliques entre *E. mutabilis* et la communauté bactérienne de Carnoulès par des approches couplant marquage isotopique, marquage FISH, cytométrie en flux et analyses métabolomiques. Les données méta-génomiques pourraient être utilisées pour dessiner des sondes spécifiques aux différentes espèces bactériennes du site.

Identification de nouvelles fonctions impliquées dans la résistance aux stress associés aux

DMA :

L'étude des données de génomique fonctionnelle repose sur des annotations obtenues à partir de bases de données et ne permettent par conséquent généralement pas d'identifier des fonctions inconnues. Afin de contourner ce problème, des criblages de banque de mutants ou des systèmes d'expression hétérologue sont souvent utilisés. Cependant, ces approches ne peuvent que difficilement être appliquées aux micro-organismes eucaryotes du site de Carnoulès. En effet, les protocoles de mutagenèse aléatoire par transposons ne fonctionnent pas en raison de l'organisation particulière des pellicules de ces organismes. D'autre part, le criblage d'une banque d'ADN génomique des euglènes pose le problème des mécanismes d'épissage qui ne sont pas nécessairement les mêmes entre ces organismes et ceux où les gènes sont exprimés. Afin de contourner ces problèmes, des systèmes d'expression hétérologue de cDNA correspondants aux ARNm extraits de cultures d'euglènes pourraient être envisagés. En effet, les ARNm sont déjà épissés et sont donc susceptibles de coder directement pour une protéine active lorsqu'ils sont exprimés dans un autre organisme procaryotique (*E. coli*) ou eucaryotique (*S. cerevisiae*). Un protocole expérimental résumant les principales étapes est proposé ci-dessous en prenant l'exemple d'*E. coli* comme vecteur d'expression (Figure 5.2). Les inserts conférant une résistance à l'arsenic peuvent être identifiés à partir des bases de données dans le cas où les séquences correspondantes y sont présentes ou faire l'objet d'analyses plus poussées dans le cas inverse (recherche de domaines protéiques, purification de la protéine correspondante, recherche de partenaires moléculaires, ...)

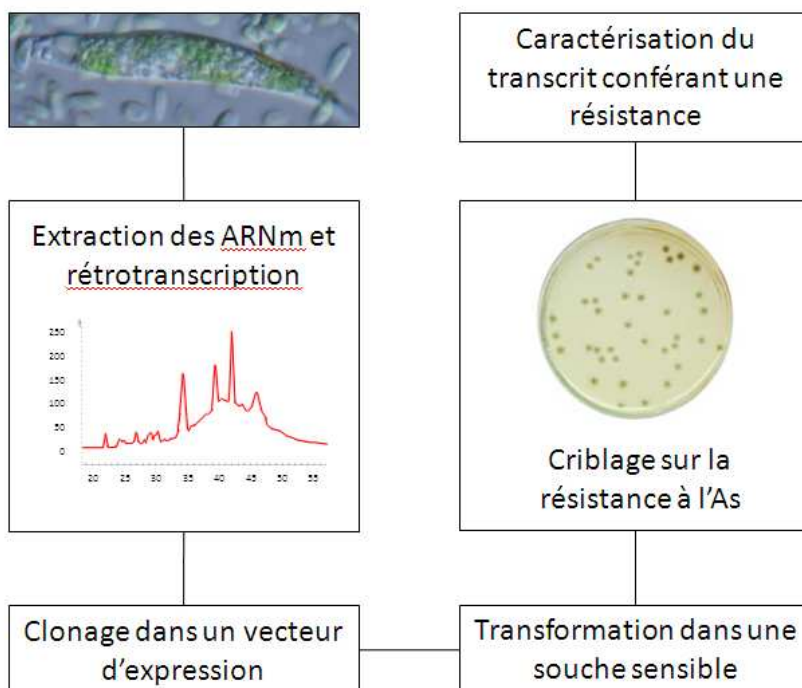


Figure 5.2 :
Principales étapes d'un protocole visant à cribler les transcrits conférant une tolérance à l'arsenic chez *E. mutabilis*.

D'autre part, bien qu'*E. mutabilis* soit un indicateur biologique des DMA et soit très abondant sur le site de Carnoulès, d'autres micro-organismes eucaryotes sont également retrouvés dans cet écosystème. Ainsi, les procédures d'isolement d'*E. mutabilis* réalisées durant ma thèse ont conduit à isoler en parallèle une algue verte unicellulaire à partir des sédiments du site de Carnoulès (Figure 5.3). Des observations microscopiques couplées au séquençage du gène codant pour l'ARN 18S ont permis d'identifier cette algue comme appartenant à l'espèce *Chlamydomonas acidophila*. Cet organisme est couramment retrouvé dans les DMA et est considéré au même titre qu'*E. mutabilis* comme un indicateur biologique de ces éco-systèmes. Cependant, contrairement à *E. mutabilis*, cette algue semble avoir acquis ses chloroplastes par endosymbiose primaire ce qui suggère une histoire évolutive radicalement différente.

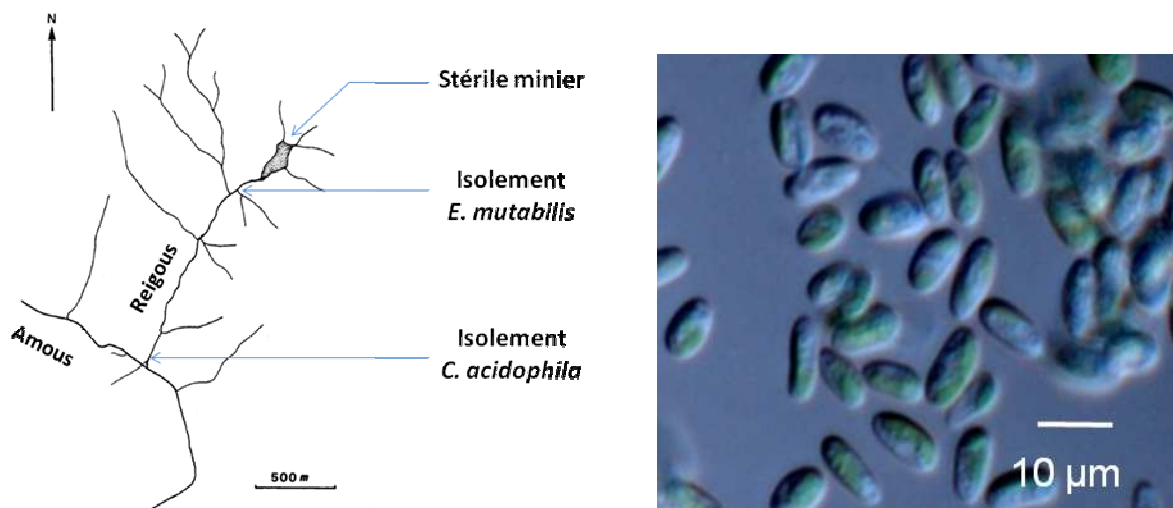


Figure 5.3 : Topologie du site de Carnoulès précisant la localisation du stérile minier et des points de prélèvement des sédiments conduisant à l'isolement d'*E. mutabilis* et de *C. acidophila* (photo), respectivement à 30 m (CowG) et à 1 500 m (confluence) en aval des stériles miniers.

Les expériences préliminaires réalisées durant ma thèse ont permis de souligner que *C. acidophila* est également hyper-tolérante à l'arsenic et présente une croissance à plus de 3000 ppm d'arsenic. Cependant, cette algue se distingue d'*E. mutabilis* à la fois par ses propriétés de surfaces hydrophiles, par une composition en acides gras qui n'est pas affectée par la présence d'arsenic et une absence d'induction de sa mobilité en présence de ce métalloïde. Bien que n'apportant pas d'information sur les mécanismes adaptatifs de cette algue face au stress arsénié, ces résultats soulignent une réponse originale de *C. acidophila* par rapport à *E. mutabilis* face à ce métalloïde. Afin de comprendre plus en détail cette réponse, des analyses de génomique fonctionnelle couplant

protéomique différentielle et séquençage du transcriptome par Solexa en présence ou en absence d'arsenic ont été réalisées et sont actuellement en cours d'interprétation.

Ainsi, *C. acidophila* partage avec *E. mutabilis* un même mode trophique et une même niche écologique malgré une histoire évolutive radicalement différente et une réponse à l'arsenic visiblement distincte. Ces deux organismes ont par conséquent su développer des mécanismes adaptatifs originaux pour faire face aux contraintes associées aux DMA et offrent des modèles d'études très intéressants pour comprendre dans quelle mesure ces mécanismes sont conservés entre différentes espèces. La comparaison de la réponse à l'arsenic chez cette algue et chez ce protiste photosynthétique pourrait par conséquent apporter un éclairage sur une éventuelle convergence évolutive et/ou sur des mécanismes adaptatifs distincts conduisant tous deux à une hyper-tolérance à l'arsenic chez les eucaryotes photosynthétiques.

D'une manière plus générale, nos travaux ont démontré que l'application à différentes communautés microbiennes de méthodes d'investigation couplant les outils les plus récents de la génomique descriptive et fonctionnelle, à des tests physiologiques et un criblage de fonctions biologiques en conditions de laboratoire permet d'accéder à l'incroyable diversité fonctionnelle des mécanismes adaptatifs chez les micro-organismes, en particulier dans des environnements souvent considérés, parfois à tort, comme hostiles à la vie.

ANNEXES

Lors de mon stage de master, mes travaux se sont focalisés sur le site de Sainte Marie-aux-Mines en Alsace qui est modérément mais significativement contaminé à l'arsenic. Ces travaux, bien que réalisés durant le stage de master ont conduit à deux publications qui ont été rédigées durant ma thèse et sont par conséquent présentés dans ce manuscrit de thèse. Le site d'échantillonnage qui a été retenu se localise à la sortie d'une ancienne exploitation minière au niveau d'un cours d'eau situé 30 m en aval. La teneur en arsenic dans les sédiments est de l'ordre de 300 mg.kg^{-1} mais la présence de ce métalloïde s'explique surtout par un fond géo-chimique naturellement riche en ce composé plutôt qu'à un processus de drainage minier. Ceci explique d'ailleurs le fait que les eaux de ruissellement contiennent des teneurs en arsenic basses, habituellement sous le seuil de potabilité fixé à $10 \mu\text{g.L}^{-1}$.

Les travaux réalisés sur le site de Sainte Marie-aux-Mines avaient pour objectif de déterminer la diversité et la composition de la communauté microbienne dans les sédiments modérément contaminés à l'arsenic, mais également d'évaluer la présence et la diversité du gène codant pour l'arsénite oxydase, considéré comme un marqueur génétique de sites contaminés par ce métalloïde.

Pour répondre à ces objectifs, plusieurs approches ont été réalisées en parallèle. La diversité de la communauté bactérienne a été appréhendée par 1) des approches de séquençage des gènes codant pour les ARNr16S et par 2) l'identification de protéines hautement conservées pouvant être utilisées comme marqueurs phylogénétiques des bactéries physiologiquement actives *in situ*. La composante génétique (séquençage des gènes codant pour les ARNr 16S et analyses de raréfaction) a été réalisée par Audrey Cordi au sein du Laboratoire des Interactions Écotoxicologie Biodiversité Ecosystèmes de Metz, tandis que l'analyse protéomique a été réalisée par mes soins au sein du laboratoire de Génétique Moléculaire, Génomique, Microbiologie de Strasbourg. Ces résultats ont conduit à un manuscrit publié dans *Research in Microbiology*, et constituent le premier papier présenté en annexe.

Dans un second temps, la diversité des gènes codant pour l'arsénite oxydase a été entreprise par amplification des gènes correspondants grâce à l'utilisation d'amorces dégénérées et par l'étude de souches bactériennes résistantes à l'arsenic isolées à partir des sédiments. Ma part expérimentale effective dans cette étude était de réaliser une banque de gènes *aox* à partir de l'ADN extrait des sédiments. Ces résultats ont permis de confirmer l'adaptation de la communauté bactérienne au stress arsénié mais ont également souligné la diversité très importante et insoupçonnée de gènes *aox* sur ce site. Ces travaux ont fait l'objet d'une publication dans *Applied and Environmental Microbiology*, et constituent le second papier présenté en annexe.



INSTITUT PASTEUR

Research in Microbiology xx (2011) 1–11



www.elsevier.com/locate/resmic

Taxonomic and functional prokaryote diversity in mildly arsenic-contaminated sediments

David Halter^a, Audrey Cordi^b, Simonetta Gribaldo^c, Sébastien Gallien^d,
 Florence Goulhen-Chollet^a, Audrey Heinrich-Salmeron^a, Christine Carapito^d,
 Christophe Pagnout^b, Didier Montaut^e, Fabienne Seby^e, Alain Van Dorsselaer^d,
 Christine Schaeffer^d, Philippe N. Bertin^a, Pascale Bauda^b, Florence Arsène-Ploetze^{a,*}

^aLaboratoire de Génétique Moléculaire, Génomique, Microbiologie, Département Microorganismes, Génomes, Environnement, UMR7156 Université de Strasbourg/CNRS, 28 rue Goethe, 67083 Strasbourg Cedex, France

^bLaboratoire des Interactions Ecotoxicologie Biodiversité Ecosystèmes (LIEBE), UMR7146, CNRS, Université Paul Verlaine, Campus Bridoux, rue du Général Delestraint, 57070 Metz, France

^cInstitut Pasteur, Département de Microbiologie, Unité de Biologie Moléculaire chez les Extrémophiles (BMGE), 5, rue du Docteur Roux 75724 Paris cedex 15, France

^dLaboratoire de Spectrométrie de Masse Bio-organique, Institut Pluridisciplinaire Hubert Curien, UMR7178 Université de Strasbourg/CNRS, 25 rue Becquerel, 67087 Strasbourg, France

^eLaboratoire Ultra-Traces Analyse Aquitaine, HélioParc Pau-Pyrénées, 2 avenue du Président Angot, 64053 Pau cedex 9, France

Received 13 October 2010; accepted 9 May 2011

Abstract

Arsenic-resistant prokaryote diversity is far from being exhaustively explored. In this study, the arsenic-adapted prokaryotic community present in a moderately arsenic-contaminated site near Sainte-Marie-aux-Mines (France) was characterized, using metaproteomic and 16S rRNA-encoding gene amplification. High prokaryotic diversity was observed, with a majority of Proteobacteria, Acidobacteria and Bacteroidetes, and a large archaeal community comprising Euryarchaeota and Thaumarchaeota. Metaproteomic analysis revealed that Proteobacteria, Planctomycetes and Cyanobacteria are among the active bacteria in this ecosystem. Taken together, these results highlight the unsuspected high diversity of the arsenic-adapted prokaryotic community, with some phyla never having been described in highly arsenic-exposed sites.

© 2011 Institut Pasteur. Published by Elsevier Masson SAS. All rights reserved.

Keywords: Environmental genomics; Arsenic; Mining; Biodiversity; Metaproteomics

1. Introduction

Arsenic is distributed ubiquitously throughout the world: its abundance on earth is around 1.5–3 mg kg⁻¹ and it is the 20th most abundant element. Arsenic levels present in soil and sediments usually range from 0.1 to 50 mg kg⁻¹, giving a mean value of 5–6 mg kg⁻¹ (Mandal and Suzuki, 2002). Arsenic concentrations present in soils and sediments actually vary considerably from one geographical region to another depending on the geochemical characteristics of the soil. For example, arsenic contents have been reported to range from 10 to 196 mg kg⁻¹ in West Bengali sediments (India), from 0.01

* Corresponding author.

E-mail addresses: David.halter@gem.u-strasbg.fr (D. Halter), Audrey.cordi@univ-metz.fr (A. Cordi), simonetta.gribaldo@pasteur.fr (S. Gribaldo), sgallien@chimie.u-strasbg.fr (S. Gallien), Florence.Chollet@gem.u-strasbg.fr (F. Goulhen-Chollet), audrey.salmeron@gem.u-strasbg.fr (A. Heinrich-Salmeron), ccarapito@chimie.u-strasbg.fr (C. Carapito), pagnout@univ-metz.fr (C. Pagnout), didier.montaut@univ-pau.fr (D. Montaut), fabienne.seby@univ-pau.fr (F. Seby), vandors@chimie.u-strasbg.fr (A. Van Dorsselaer), cschaeff@chimie.u-strasbg.fr (C. Schaeffer), philippe.bertin@unistra.fr (P.N. Bertin), bauda@univ-metz.fr (P. Bauda), ploetze@unistra.fr (F. Arsène-Ploetze).

0923-2508/\$ - see front matter © 2011 Institut Pasteur. Published by Elsevier Masson SAS. All rights reserved.

doi:10.1016/j.resmic.2011.06.001

Please cite this article in press as: Halter, D., et al., Taxonomic and functional prokaryote diversity in mildly arsenic-contaminated sediments, *Research in Microbiology* (2011), doi:10.1016/j.resmic.2011.06.001

to 626 mg kg⁻¹ in soil and sediments from China and from 1 to 72 mg kg⁻¹ in soil originating from the United States (Mandal and Suzuki, 2002). High levels of arsenic can be due either to natural contamination by the parent rock (India) or human activities (anthropogenic forms of contamination) (Mandal and Suzuki, 2002). In natural waters, less than 10 µg l⁻¹ is usually present, but arsenic levels of 5000 µg l⁻¹ have been observed in contaminated waters in the USA, South America, India and Bangladesh (Smedley and Kinniburgh, 2002). Arsenic-rich waters occur naturally in geothermal regions or because of the geochemical characteristics of the soils. Anthropogenic contaminations are observed principally in mining regions due to acid mine drainages (AMDs) efflux characterized by high sulfate, iron and other metal concentrations (Coupland et al., 2004; Hallberg and Johnson, 2005). In these aquatic environments, arsenic occurs mainly in the form of the inorganic species arsenate (As(V)) and arsenite (As(III)), which are generally more toxic than the organic forms to living organisms (Sharma and Sohn, 2009). As(III) is more soluble, and usually thought to be more toxic than As(V).

Several studies have shown that bacterial metabolic processes such as arsenic oxidation and reduction play an important role in arsenic speciation in the environment and affect its bioavailability (Lièvreumont et al., 2009). However, knowledge available about arsenic-resistant bacteria was restricted until quite recently to results of studies on bacteria grown *in vitro* (Lièvreumont et al., 2009; Tsai et al., 2009). To explore the diversity of arsenic-metabolizing prokaryotes, several laboratories have developed environmental genomic approaches such as meta-genomic, meta-transcriptomic and

meta-proteomic approaches (Bertin et al., 2008; Deneff et al., 2010). Functional studies have been performed on highly arsenic-contaminated AMDs (VerBerkmoes et al., 2009; Deneff et al., 2010; Mueller et al., 2010; Bruneel et al., *in press*). However, arsenic-tolerant prokaryote diversity is far from being explored, since most of those previous studies focused on highly contaminated sites or on AMD-exposed sites which combine several abiotic stresses such as low pH and high heavy metals concentration, where a low level of bacterial diversity was observed.

The aim of this study was to further investigate arsenic-adapted prokaryote diversity. For this purpose, a study site located in the vicinity of the Gabe-Gottes mine, which was worked until 1940 in the Sainte-Marie-aux-Mines Valley (France), was chosen (Fig. 1). The material extracted initially consisted of silver ore and, more recently, of arsenic ore. Currently, waters are percolating from the mine walls and accumulate in a creek. Preliminary experiments showed that the arsenic present in the creek sediments amounts to about 500 mg kg⁻¹, which constitutes a moderate but significant level of contamination. In this study, the physical and chemical characteristics of this site, the microbial community inhabiting the creek sediments and the main proteins expressed were analyzed using environmental genomic/proteomic approaches.

2. Materials and methods

2.1. *In situ* sampling procedure and chemical analyses

The sampling site was at a creek formed by the waters from a former mine located 40 m from the mine entrance, at

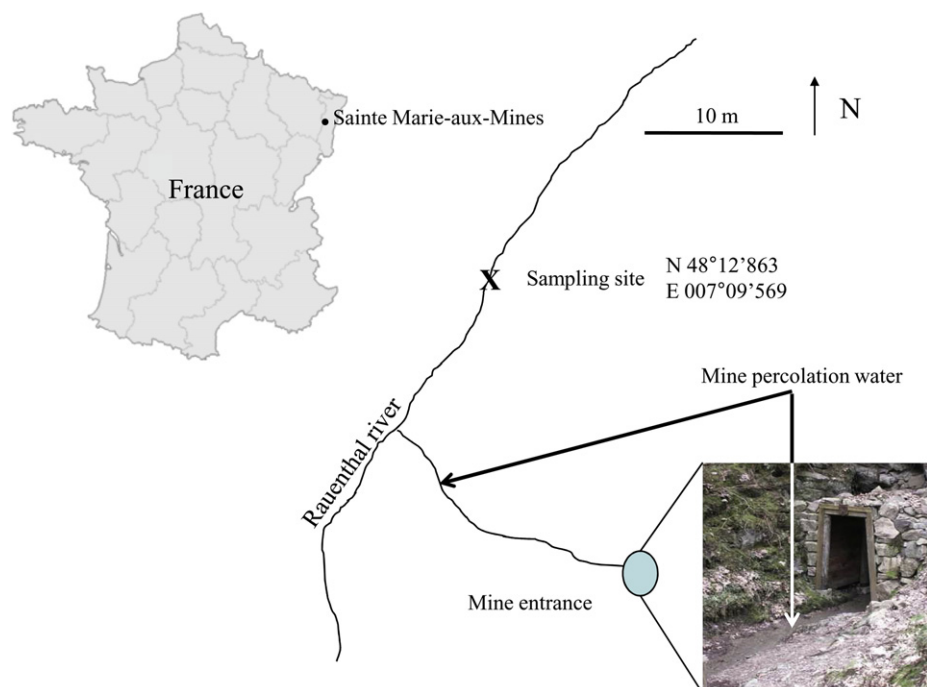


Fig. 1. Localization of the sampling site near the site of Sainte-Marie-aux-Mines (France). Samples were collected in sediment of the Rauenthal River, 40 m downstream from a mine entrance in waters percolating through the ancient mining exploitation.

Sainte-Marie-aux-Mines in France (Fig. 1). One liter of water was filtered through a 0.22 μm sterile nucleopore filter and stored at 4 °C for chemical analysis, which was performed within 48 h. The 5 cm most superficial sediments from the creek were collected in December 2007 with a plastic shovel at several points along the creek and homogenized. After being sampled, water and sediments were separated, sediments were stored at -80 °C for DNA extraction or in glycerol (25%, w/v) at -80 °C for protein extraction.

Water pH was measured using a microprocessor pH meter (pH 3000, WTW) and acid-neutralizing capacity (ANC) was determined using Gran's titration method. Conductivity was measured with a Metrohm Herisau Conductometer E518 (Herisau, Switzerland) at 25 °C. Mineralization of sediments (0.5 g dry weight) was performed in line with the French standard AFNOR NFX 31-151 prior to performing metal analysis using atomic absorption spectrophotometry methods. Organic content was determined from the loss of mass determined after treatment at 550 °C for 4 h. Metal and metalloid concentrations (after acidification with HNO_3) were determined by performing atomic absorption spectrophotometry (Analyst 100; Perkin-Elmer and Varian SpectraAA-300) and Br^- , Cl^- , SO_4^{2-} , NO_3^- and NO_2^- concentrations by performing ion chromatography (Dionex 1500i with a AS 4 A SC column; Sunnyvale, CA, USA). The suspended solids content, total phosphorus (TP), total Kjeldahl nitrogen (TKN), COD (chemical oxygen demand), BOD5 (biological oxygen demand after 5 days), and TSS (total suspended solids) were determined according to French standards relating to water analysis (AFNOR, 1994). Total phosphorus was determined in keeping with the French norm AFNOR NF T90-023. Chlorophyllic pigments were quantified spectrophotometrically (Specord 205) at 750 and 755 nm.

In order to analyze arsenic speciation in the interstitial water, part of the frozen sediment was centrifuged. The supernatant was collected, filtered (0.45 μm) and further quantified using standard additions to prevent occurrence of matrix effects with an HPLC-ICP-MS apparatus (PerkinElmer SCIEX ELAN 6100 DRC, Concord, ON, Canada) fitted with a Meinhard nebulizer and a cyclonic spray chamber. Briefly, arsenic species were separated with an anion-exchange column PRPX-100 (100 mm \times 4.1 mm, Hamilton, Reno, NV, USA). Eluent was delivered online at a flow rate of 1 ml min^{-1} with a high performance liquid chromatography pump (DIONEX ICS 3000, Sunnyvale, CA, USA) coupled with the ICP-MS. Samples were injected using an autosampler (DIONEX AS50, Sunnyvale, CA, USA) through a 100 μl PEEK injection loop attached to an injection valve. The outlet of the HPLC column was directly connected to the nebulizer of the ICP-MS via 20 cm of PEEK capillary tubing (0.17 mm i.d.). Concentration values and uncertainties involved were calculated from values determined in duplicate. To measure arsenic species concentrations in sediments, some of the frozen samples were lyophilized and ground prior to the extraction process. Extractions were performed in duplicate. The extraction method used was adapted from (Thomas et al., 1997). 10 mL of H_3PO_4 1 M were added to 0.4 g of sample in a Teflon vessel. The mixture was heated with a closed microwave system (CEM model MARS, Charlotte, NC,

USA) at 120 °C for 20 min. The remaining solution was diluted to 50 ml with ultrapure water and then analyzed by performing HPLC-ICP-MS.

2.2. Metagenomic DNA extraction and 16S rRNA gene analysis of the whole community

DNA was extracted directly from the sediment community using a PowerMax soil DNA isolation kit in line with the manufacturer's recommendations (MoBio Laboratories, Inc.), concentrated by precipitation in 2.5 v/v 100% ethanol and 0.1 v/v sodium acetate (3 M, pH 5.6) and stored at -20 °C until further processing.

1.5 kb bacterial and 571 bp archaeal 16S rRNA-encoding genes were amplified using metagenomic DNA as a matrix and universal bacterial and archaeal primers W01/W02 (5'-AGAGTTTGATCMTGGCTC-3' and 5'-GNTACCTTGTTACGACTT-3') and ARC344F/ARC915R (5'-ACGGGGYG CAGCAGGCGCGA-3' and, 5'-GTGCTCCCCGCCAATTCCT-3'), respectively (Stahl and Amann, 1991; Raskin et al., 1994). All PCR reaction mixtures contained 50 ng DNA template, 5 μl PCR buffer (5 PRIME buffer 10 \times), 1 μl of deoxynucleotide mix (Mix 5 PRIME, 10 mM of each dNTP) 0.2 mM of each primer and 2.5 U Taq DNA polymerase (5 PRIME) in 50 μl volume. Bacterial 16S rRNA gene PCR amplification reactions were carried out with a Mastercycler gradient (Eppendorf) with an initial denaturation step at 95 °C for 5 min, followed by 35 cycles of denaturation (40 s at 95 °C), a 30 s annealing step at 52 °C, a 45 s elongation step at 68 °C and a final 10 min denaturation step at 68 °C. Archaeal 16S rRNA gene cycling conditions consisted of a 5 min denaturation step at 94 °C followed by 35 denaturation cycles of 1 min at 94 °C, a 1 min annealing step at 71–61 °C with a 0.5 °C decrement per cycle during the first 20 cycles and a 1 min elongation step at 72 °C, followed by a final 10 min elongation step at 72 °C.

The fragments obtained were cloned into plasmids pCR[®]2.1 using the TOPO[®] TA cloning kit in line with the manufacturer's recommendations (Invitrogen Laboratories). 240 clones containing bacterial 16S rRNA-encoding genes and 83 clones containing archaeal 16S rRNA encoding genes were sequenced using primers ARC344F and W01, respectively (Stahl and Amann, 1991; Raskin et al., 1994; Achour et al., 2007). Sequences harboring more than 600 bp in the case of bacterial 16S rRNA genes and more than 300 bp in that of archaeal 16S rRNA genes were used for phylogenetic analysis. Indeed, these sequences were compared with those in the GenBank database by performing BLAST searches (Altschul et al., 1997) and those in the RDP database by performing SEQ MATCH searches (Cole et al., 2009). Chimeras were checked using pintail online (Ashelford et al., 2005). Sequences that corresponded to chimera were removed for further analysis.

Archaeal 16S rRNA gene homologs were collected from the nr database at NCBI, using the BLAST program (Altschul et al., 1997) with default parameters. The same procedure was used on the environmental database at NCBI, using each clone as a seed in order to enrich the samples. Sequences were

aligned using the Muscle program (Edgar, 2004) with default parameters. Based on preliminary phylogenetic trees, only a few sequences representative of archaeal diversity were selected, giving a dataset of 152 sequences for final analysis. For similarity search, 445 unambiguously aligned positions were selected. A maximum likelihood tree was obtained using the PhyML program (Guindon and Gascuel, 2003) with a GTR model, 4 evolutionary rates, a calculated proportion of invariant sites and calculated nucleotide frequencies (default parameters). Statistical likelihood at nodes was calculated via a likelihood-ratio test (Anisimova and Gascuel, 2006).

Partial prokaryotic sequences were aligned using ClustalW (Thompson et al., 2002) separately on bacterial and archaea samples, and the respective distance matrices were obtained using DNADIST from the PHYLIP package (<http://evolution.genetics.washington.edu/phylip.html>). The resulting matrices were fed into DOTUR (Schloss and Handelsman, 2005) in order to calculate the number of OTUs at 97% and 85% level. These percentage identities OTUs served for generating rarefaction curves (Fig. 2).

2.3. Metaproteomic analyses

In order to identify the main proteins expressed in situ by the sediment community, prokaryotic cells were recovered from sediment as follows: 8×12.5 g of frozen sediment were washed with 8×12.5 mL of salt solution ($0.15 \text{ g l}^{-1} \text{ Na}_2\text{SO}_4$, $0.45 \text{ g l}^{-1} (\text{NH}_4)_2\text{SO}_4 \cdot 10\text{H}_2\text{O}$, $0.05 \text{ g l}^{-1} \text{ KCl}$, $0.5 \text{ g l}^{-1} \text{ MgSO}_4 \cdot 7\text{H}_2\text{O}$, $0.05 \text{ g l}^{-1} \text{ KH}_2\text{PO}_4$, $0.014 \text{ g l}^{-1} \text{ Ca}(\text{NO}_3)_2 \cdot 4\text{H}_2\text{O}$) and gently

shaken overnight at 4°C . After a 30 min decantation step, the supernatant (8×7.5 ml) was added to Nycodenz solution without any mixing (8×17.5 mL, 65%, w/v) (Axis-Shield, Dundee, Scotland), and centrifuged for 3 h at 10 000 g. The cellular fraction was removed and washed by adding 3 volumes of NaCl 0.9% and centrifuged for 1 h at 10 000 g at 4°C . Cells were resuspended in NaCl 0.9% and an aliquot was stained using the LIVE/DEAD BacLight Bacterial Viability kit (L13152, Molecular Probes, Invitrogen) as recommended by the manufacturer, in order to assess the proportion of dead cells. The preparation was examined using an epifluorescent microscope Nikon Eclipse TE2000 S. Images were taken with an integrated Hamamatsu digital camera (C4742-95; Nikon UK).

Proteins were extracted from these cells and separated by 1D-SDS-PAGE using Laemmli's method (Laemmli, 1970) with a 12% gradient slab gel (PROTEAN II, Bio-Rad laboratories). Electrophoresis was carried out at 200 V. Proteins were stained with Coomassie Brilliant Blue, bands were systematically cut every 1.5 mm and the 29 gel pieces were stored at -20°C before performing mass spectrometry analysis.

In-gel digestion of gel bands was performed as previously described (Weiss et al., 2009). The resulting peptides were analyzed by nanoLC-MS/MS on nanoACQUITY Ultra-Performance-LC (UPLC, Waters, Milford, MA, USA) coupled to a SYNAPT hybrid quadrupole orthogonal acceleration time-of-flight tandem mass spectrometer (Waters, Milford, MA, USA). The capillary voltage was set at 3500 V and the cone voltage at 35 V. In tandem MS experiments, the system operated with automatic switching between MS and MS/MS modes. The 3 most abundant peptides, preferably doubly and triply charged ions, were selected from each MS spectrum for further isolation and CID fragmentation with 2 energies set using collision energy profile methods. The system was entirely controlled by MassLynx 4.1 (SCN 566, Waters, Milford, MA). Raw data collected during nanoLC-MS/MS analyses were processed and converted with a ProteinLynx Browser 2.3 (23, Waters, Milford, MA) into .pkl peak list format.

MS/MS data were analyzed using the MASCOT 2.2.0 algorithm (Matrix Science, London, UK). Spectra were searched for with a mass tolerance of 30 ppm for MS and 0.1 Da for MS/MS data, allowing a maximum of one missed trypsin cleavage site and taking carbamidomethylation of cysteines and oxidation of methionines as variable modifications. Spectra were first searched for against a target-decoy version of a subset of the NCBI nr database restricted to the bacteria found in 16S rRNA gene community analysis (see Results). Common contaminants (keratins and trypsin) were added to the database. Protein identification was confirmed when at least two peptides with high quality MS/MS spectra (less than 20 points below Mascot's threshold identity score at 95% confidence level) were detected. All spectra that did not satisfy these criteria were exported using Scaffold software program 2.2.0 (Proteome Software, Portland, USA) and used to search using the same database. Protein identification was confirmed when one peptide with a very high quality MS/MS spectrum (more than 5 points above Mascot's threshold identity score at 95% confidence level) was detected. After removing common contaminants,

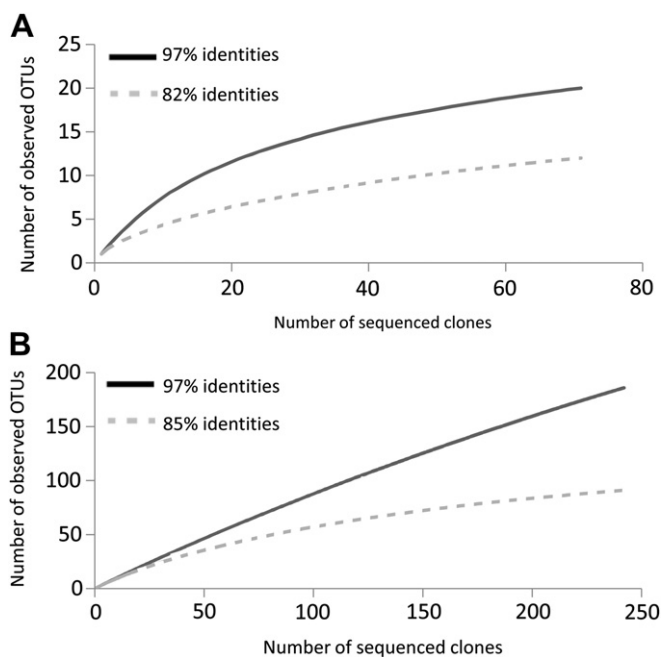


Fig. 2. Rarefaction analysis of the (A) archaeal 16S rRNA and (B) bacterial sequences from Sainte-Marie-aux-Mines sediments. The total number of sequenced clones is plotted against the number of OTUs observed in the same library. OTUs were defined at the 97% identity level (species level) for both and at 82% and 85% (class level) for Archaea and Bacteria, respectively. For archaeal analysis, the DOTUR program did not calculate the 85%, but the 82% value.

these thresholds led to protein identification with a false discovery rate of less than 3%.

3. Results

3.1. Chemical characterization of the sampling site

The physico-chemical composition of sediment and water samples from the Gabe-Gottes mining site is shown in Table 1. The pH of the water was neutral. The low biological oxygen demand (BOD) and chemical oxygen demand (COD) values as well as the low amounts of suspended solids detected suggested that organic matter was not abundant in this water.

Table 1
Physical and chemical characteristics of water and sediment samples.

	Water	Sediments
General characteristics		
pH	6.9	NT
Density (kg l ⁻¹)	NT	2.07
Moisture 20 °C (%)	NT	18
Moisture 105 °C (%)	NT	19.3
Fraction <2 mm (%)	NT	57.8
Fraction >2 mm (%)	NT	42.2
OM on <2 mm (%)	NT	2
Conductivity (μS cm ⁻¹ at 25 °C)	50.4	NT
Alkalinity (meq l ⁻¹)	0.18	NT
TSS (mg l ⁻¹)	0.7	NT
VSS	100%	NT
Chemical characteristics		
Total P (mg l ⁻¹ in water; % in sediments)	0.03	0.08
PO ₄ (mg P l ⁻¹)	0.02	NT
COD (mg O l ⁻¹)	5	NT
BOD5 (mg O l ⁻¹)	0.4	NT
NH ₄ (mg N l ⁻¹)	0.02	NT
NO ₂ (mg N l ⁻¹)	<0.01	NT
NO ₃ (mg N l ⁻¹)	1.2	NT
TKN (mg N l ⁻¹)	<0.1	NT
Cl ⁻ (mg l ⁻¹)	2.1	NT
SO ₄ (mg l ⁻¹)	6.2	NT
Br (mg l ⁻¹)	0.004	NT
Chlo a (μg l ⁻¹)	0.3	NT
Pheo (μg l ⁻¹)	0.2	NT
Metals and metalloids (μg l ⁻¹ in water, μg g ⁻¹ in sediments)		
Al	98	15 150
As	6	320 ^a
Cd	ND	0.3
Cr	ND	40
Cu	1	103
Fe	ND	23 745
Mn	ND	510
Ni	ND	23
Pb	ND	49
Zn	2	88

ND: not detected; NT: not tested; OM: organic matter; TKN: total Kjeldahl nitrogen; COD: chemical oxygen demand; BOD5: biological oxygen demand after 5 days; Chlo a: chlorophyll a; Pheo: pheophytine; TSS: total suspended solids; VSS: volatile suspended solids.

^a As(III) and As(V) concentrations measured in interstitial water were 14.6 and 135 μg l⁻¹, respectively.

In addition, the organic content measured in the sediments themselves was only 2%. Sulfate levels and water conductivity were also low, which suggest that ionic species were not abundant. The stream trophic state was assessed by measuring total nitrogen, phosphorus and chlorophyll content, which showed the presence of photosynthetic organisms (Dodds et al., 1998). Nitrogen and phosphorus compound contents were in the same range as levels detected in oligotrophic streams (<700 and 25 μg l⁻¹, respectively) (Dodds et al., 1998). The chlorophyll a (Chlo a) content was lower in these samples than values found in mesotrophic streams (<10 μg l⁻¹ chlorophyll) (Dodds et al., 1998), which suggests that photosynthetic organisms were present but not very abundant. Based on these characteristics, this system was considered to be an oligotrophic ecosystem.

Concentrations of Pb, Al and Fe in these sediment samples were 5, 190 and 500 times higher, respectively, than the average metal concentrations in the surrounding soils (<http://www.stats.environnement.developpement-durable.gouv.fr/acces-thematique/sol/le-sol/la-contamination-des-sols-par-les-elements-traces.html>). Total arsenic contamination measured in the water column was lower (6 μg l⁻¹) than the 10 μg l⁻¹ limit recommended for drinking water. On the other hand, in sedimentary interstitial water, significantly higher As(III) and As(V) concentrations were measured (14.6 (±0.3) μg l⁻¹ and 135 (±5) μg l⁻¹, respectively), whereas none of the methylated forms searched for, i.e. DMA and MMA, were detected. The arsenic concentration observed in the sediment samples (320 mg kg⁻¹ dry weight) was six times higher than the average geochemical content in this region (48 mg kg⁻¹ in the Vosges, <http://infoterre.brgm.fr>) and the As(V) species was 10 times more concentrated than the As(III) species. All in all, the physical and chemical characteristics determined here showed that this site contained neutral oligotrophic water with a moderately high level of heavy metal contamination.

3.2. Structure of the prokaryote community in sediments based on 16S rRNA genes

An inventory of the whole bacterial and archaeal community was drawn up by performing 16S rRNA-based community analysis. Total DNA extracted from the sediment was used as template for 16S rRNA gene amplification. Partial sequencing of 240 bacterial and 83 archaeal 16S rRNA genes and RPD analysis (see Materials and Methods) led to the identification of 193 and 20 OTUs (calculated with 97% identity) belonging to 14 bacterial and 2 archaeal phyla, respectively. Concerning the bacterial library, rarefaction curves calculated at the class level (85% identities) reached a plateau (Fig. 2). As the class level is the rank usually used for representing the bacterial community with less ambiguity, we limited our analysis to these clones, even though the rarefaction curves calculated at the species level (97% identity) did not reach a plateau. The rarefaction curves produced for archaeal sequence libraries reached a plateau which confirmed that the archaeal clone library was sufficiently sampled.

The bacterial 16S rRNA gene sequences were found to be affiliated at the order, family or genus level, and considerable diversity was observed (Fig. 3). The bacterial community was composed mainly of Bacteroidetes (13%), Betaproteobacteria (15%), Acidobacteria (17%) and Deltaproteobacteria (12%). Several phyla, such as Alphaproteobacteria, Gammaproteobacteria, Actinobacteria, Nitrospira and Planctomycetes, accounted for 5–10% of the bacteria. Lastly, Firmicutes, Verrucomicrobia, Cyanobacteria, Gemmatimonadetes and Deinococcus/Thermus accounted for only a small proportion of the bacteria (less than 5%) (Fig. 3). At the order level, Sphingobacteriales, Rhizobiales, Burkholderiales, Myxococcales and Planctomycetaceae were present. Interestingly, 12% of the clones were related to Acidobacteriaceae, the most abundant family. 14 clones amounting to 6% of the clones were related to the *Nitrospira* genus. This genus was therefore the most abundant one identified in this ecosystem. Lastly, 4 clones could not be definitively attributed to any known phylum, which suggests that unknown bacteria account for at least 1% of this community. Archaeal phylogenetic analysis (Fig. 4) showed that this population included Euryarchaeota (29 clones, corresponding to 38% of the sample) and Thaumarchaeota (also known as group I Archaea) (DeLong, 1992; Fuhrman et al., 1992; Brochier-Armanet et al., 2008) (48 clones, corresponding to 62% of the sample). The Euryarchaeota detected in this study site fell into four clusters (Fig. 4, I–IV), three of which (I–III) were found to be affiliated with methanogenic Archaea and were possibly from the anaerobic part of the sediment. One cluster (Fig. 4, IV) was more divergent and not closely related to any known species, but was affiliated with several 16S rDNA clones previously reported (Rudolph et al., 2004; Galand et al., 2006). A large number of clones from Thaumarchaeota were also identified (Fig. 4). Three of them (clones 64 HQ664164, 19 HQ664132, 20 HQ664133) were affiliated with the species “*Candidatus*

Nitrososphaera gargensis”, a moderately thermophilic Thaumarchaeon recently isolated from a terrestrial hot spring (Hatzenpichler et al., 2008). The others were affiliated with environmental sequences originating from the soil which likely represent uncultivated lineages of Thaumarchaeota (Sliwinski and Goodman, 2004; Nicol et al., 2006; Valenzuela-Encinas et al., 2008).

3.3. Metaproteomic analysis

Complementary to 16S RNA analysis, a metaproteomic approach was used to identify the main functional traits of the bacterial community. Microorganisms were separated from sediments and interstitial water by performing Nycodenz density gradient. An aliquot of these cells was stained with a live/dead BacLight kit (Invitrogen) and results showed that most cells extracted from the sediments were viable (data not shown). This finding showed that the gradient density-based bacterial recovery step did not affect cell membrane integrity. After SDS-PAGE electrophoresis, proteins were identified using an NCBI nr database restricted to bacteria obtained with 16S rRNA-based community analysis (Table 2). Only 23 proteins were identified, half of them are involved in biological processes conserved within the prokaryotic kingdom, such as replication and transcription (elongation factors, transcriptional regulator), translation (ribosomal proteins), energy metabolism (ATP synthases), regulation (histone family, transcriptional regulator) and stress responses (Hsp90, GroEL). As these proteins are involved in widespread and highly conserved biological processes, they could be assigned to specific taxons. These proteins originated from Betaproteobacteria and Deltaproteobacteria, which were abundant in this community, as well as from Alphaproteobacteria, and Planctomycetes, each of which accounted for less than 10% of the community. Indeed, several of these proteins were

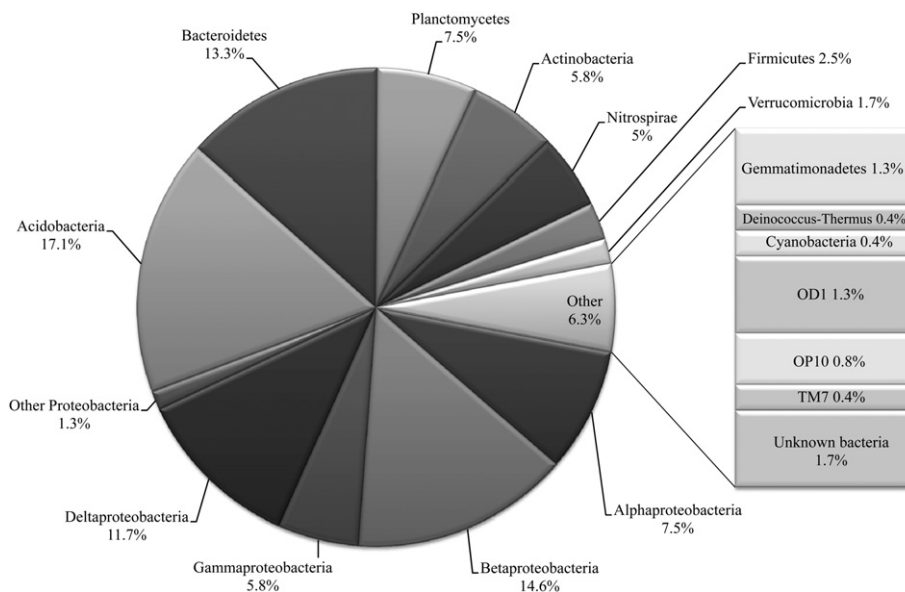


Fig. 3. Phylogenetic RDP affiliation of bacterial 16S rRNA genes (Class rank for Proteobacteria, phyla rank for the others). The most abundant phyla are shown in the circular diagram and the less abundant ones on the right.

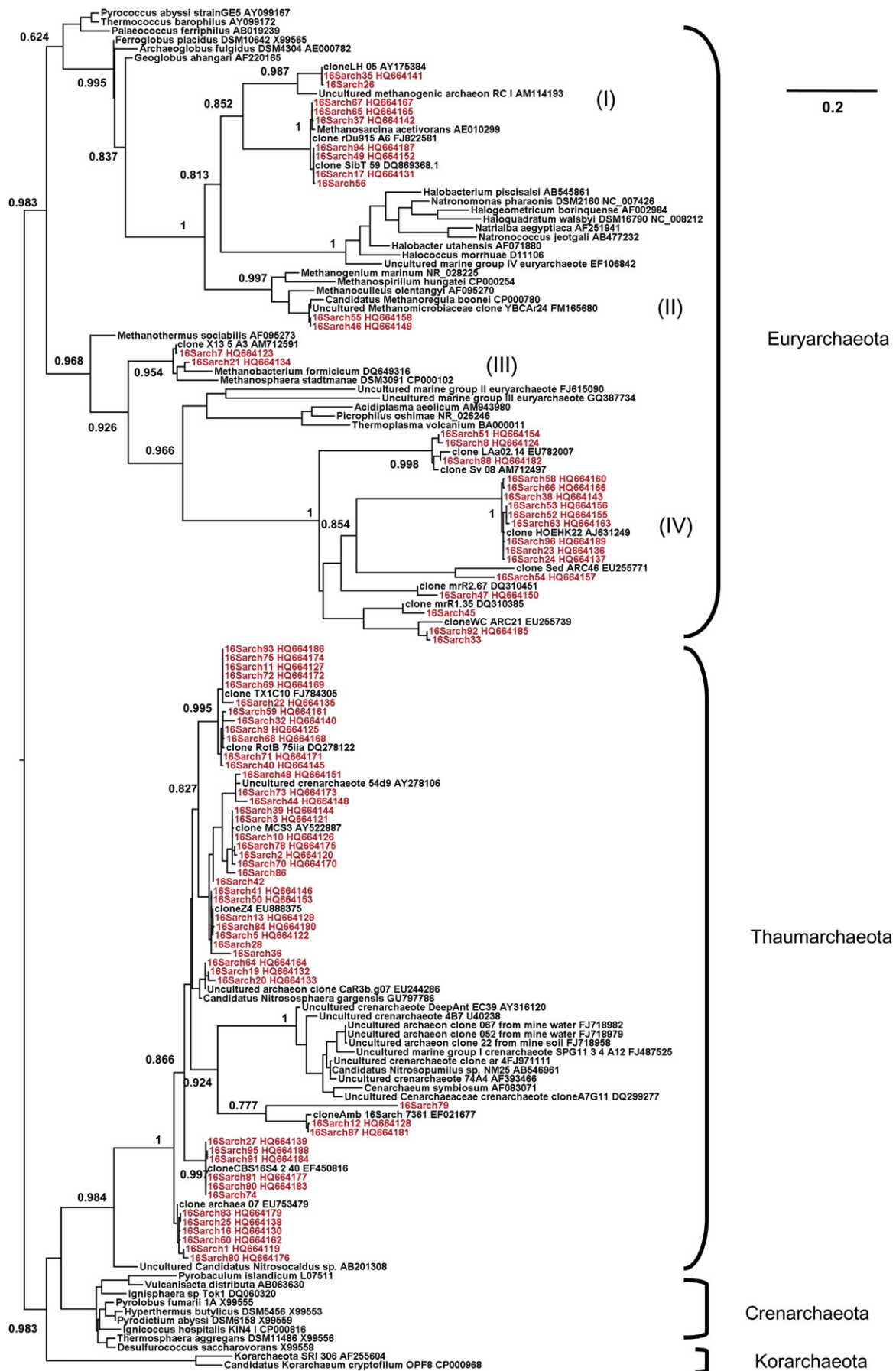


Fig. 4. Maximum likelihood tree of 16S rRNA gene homologs from the archaeal clones (in red font) along with a selection of representatives of archaeal diversity. The four main archaeal phyla are indicated. Numbers at nodes indicate aLTR (approximate likelihood-ratio test) branch support as computed by PhyML. The scale bar gives the average number of substitutions per site.

expressed by bacteria affiliated with *Hyphomicrobium*, *Parvimicrobium*, *Bradyrhizobium*, *Sphingopyxis*, *Sideroxydans*, *Desulfovibrio* and *Planctomyces*. The metaproteomic approach also revealed the presence of 6 proteins involved in the photosynthetic process and expressed by Cyanobacteria, although these photosynthetic bacteria accounted for less than 1% of the bacterial community.

4. Discussion

The aim of this study was to characterize, both chemically and microbiologically, a moderately arsenic-contaminated environment and to identify the arsenic-tolerant prokaryotic bacterial species. The bacterial and archaeal diversity observed at our study site is greater than that observed at other more severely contaminated sites (Baker and Banfield, 2003; Johnson and Hallberg, 2003), at the Carnoulès AMD, for example, where less than 10 phyla were identified (Bruneel et al., 2005, 2006, 2008) and in meta-genomic studies on AMD biofilm (4–6 bacterial species) (Tyson et al., 2004). Nor was such diversity observed in geothermal spring waters (6 phyla (Hamamura et al., 2009)).

This high level of diversity makes the meta-proteomic approach rather difficult to apply, which explains why only 23 proteins were identified despite efficient bacteria and protein extraction procedures. A low number of identified proteins was also observed previously in soil in which a high level of diversity was observed (Benndorf et al., 2007; Williams et al., 2010). As pointed out by the authors of recent reviews (VerBerkmoes et al., 2009; Deneff et al., 2010), meta-proteomic studies are successful when applied to communities with low levels of diversity. When a high level of diversity is observed, each protein is diluted in a complex preparation and only the most abundant proteins are therefore likely to be identified. Moreover, a large proportion of the bacteria forming this community have never been studied thus far in vitro, and their genome sequences and hence their protein sequences, which are required for MS identification purposes, are not available in public databases. For example, contrary to Proteobacteria, only a few Acidobacteria protein sequences are available in the existing protein databases. Likewise, no archaeal proteins were identified, probably because only a few protein sequences of Archaea are available. On the other hand, the meta-proteomic approach has some advantages compared to other approaches. Indeed, as proteins are more stable than RNA (especially those originating from prokaryotes), the metaproteome content was presumed to be less affected by the extraction procedure than the transcriptome and probably gives better insight into the biological functions expressed in situ. Moreover, the meta-proteomic approach gives taxonomic information that was not observed with the 16S rRNA gene-based approach. First, the diversity estimated with the 16S rRNA gene-based approach might be underestimated because of DNA extraction and PCR biases. Second, the meta-proteomic approach highlights expression of proteins involved in conserved biological processes that could be assigned to a specific taxon at the genus or species level, whereas RDP analysis allowed us to affiliate only 28% at the genus level

despite the use of almost all 16S rRNA gene sequence lengths. Such a limit was in agreement with previous observations establishing that it is not always possible to affiliate certain bacteria using only 16S rRNA gene sequences (Schleifer, 2009). Indeed, some microorganisms showing very similar 16S rRNA gene sequences turned out to belong to different taxons when other phylogenetic markers were used. For these reasons, meta-proteomic data provide a valid complementary tool for use with 16S rRNA gene-based taxonomy.

Several Thermoplasmatales related to our clones (Fig. 4, cluster IV) were previously identified in arsenic-rich environments such as that present in Carnoulès (Bruneel et al., 2008) as, for example, several *Ferroplasma* (Baker and Banfield, 2003; Deneff et al., 2010). On the other hand, various Thaumarchaeota present here are affiliated with environmental sequences originating from soil (Sliwinski and Goodman, 2004; Nicol et al., 2006; Valenzuela-Encinas et al., 2008). To our knowledge, none of them have been previously observed in arsenic-contaminated environments. Concerning the bacterial phyla, Bacteroidetes have rarely been observed in arsenic-rich environments (except for a few *Flavobacteria*) (Honschopp et al., 1996), whereas other phyla detected here were previously described in arsenic-contaminated samples, as for example *Sideroxydans*, *Rhodoferrax*, *Variovorax*, *Flavobacterium* and *Microbacterium* (Baker and Banfield, 2003; Macur et al., 2004; Bruneel et al., 2006, 2011; Casiot et al., 2006; Achour et al., 2007; Achour-Rokbani et al., 2010). Such bacteria could impact arsenic speciation in this studied site. Bacteria related to *Sideroxydans lithotrophicus* were active in this ecosystem, as revealed by the metaproteomic approach. Such microorganisms are known to be preferentially present in a low oxygen concentration environment where they efficiently oxidize iron and may consequently have an indirect impact on the soluble arsenic concentration. Indeed, Fe(III) produced by these bacteria may co-precipitate with arsenate, leading to a decrease in soluble arsenite. Similarly, other species belonging to *Rhodoferrax* are known to reduce iron under anaerobic conditions; their metabolism may therefore lead to an increased concentration of soluble arsenic compounds. Finally, some Betaproteobacteria identified using our proteogenomic approach may directly change arsenic speciation. Indeed, *Variovorax* sp. are arsenic-resistant bacteria known to reduce arsenate (Rathinasabapathi et al., 2006) and bacteria related to *Flavobacterium* were shown to methylate arsenic (Honschopp et al., 1996). Since arsenic was mainly found in an oxidized form (AsV) in this environment, it is tempting to speculate that both arsenic reduction and methylation activities were not important, in contrast to oxidation activity. Although we cannot exclude that abiotic processes also influence speciation of arsenic in this environment, arsenite oxidation may be catalyzed by bacteria affiliated with *Microbacterium*, as previously shown (Mokashi and Paknikar, 2002). Finally, other bacteria may also play a significant role in the functioning of this ecosystem. Identification of proteins originating from Planctomycetes, Cyanobacteria and Proteobacteria such as Rhizobiales revealed that these bacteria were not only present, but also viable, and physiologically active despite their low abundance, as suggested by the 16S rDNA metagenomic approach (less than 5% of the bacterial

Table 2
Main proteins expressed by the community.

Phylum	Class	Order; family; genus; species	Protein name	Protein molecular weight (Da)	Number of unique peptides	Percentage sequence coverage (%)	Peptide sequence		
Proteobacteria	Unknown	Unknown	Thioredoxin	11 540.5	1	8.33	GIPTLILFK		
	Alphaproteobacteria	Rhodobacterales or Rhizobiales	F ₀ F ₁ -ATP synthase subunit beta	51 052.6	3	7.97	FTQAGSEV SALLGR IGLFGG AGVGK TVLIMELI NNVAK VAWTF STGVLR		
				Pyrrolo-quinoline quinone	65 645.4	1	1.82	STGVLR	
			Conserved hypothetical protein	21 210.5	2	4.35	GPVSLPR RGPVSLPR		
			Histone family protein DNA-binding protein	12 014.1	2	16.80	LTGLGILQVR		
			30S Ribosomal protein S4	23 808.7	1	4.37	NPATGEAIK LSDYGVQLR		
			Heat shock protein 90	68 916.8	1	2.24	GVIDSEDL PLNISR		
			30S Ribosomal protein S3	27 155.2	1	3.32	LGGAEIAR		
			30S Ribosomal protein S12	13 973.1	1	6.50	PTINQLVR		
			Betaproteobacteria	Gallionellales; Gallionellaceae; Sideroxydans; Sideroxydans lithotrophicus ES-1	Chaperonin GroEL	57 287.8	2	7.85	AAVEEGIVPG GGVALLR
						AGRPLLIIE DVDGEALA TLVVNNIR			
	Elongation Factor Tu	43 021.5				1	3.28	QVGVPYII VFLNK	
	Enolase	45 992.7			2	5.61	AAAEEES GLPLYR AAVPSG ASTGSR		
		5-Methyl-tetrahydro-pteroyl-triglutamate/homocysteine S-methyltransferase			86 464.0	1	1.28	VLSIGIVDGR	
	50S ribosomal protein L2	30 162.6			1	4.01	SAGTSAT LLAR		
	Deltaproteobacteria	Desulfovibrionales; Desulfovibrionaceae; Desulfovibrio; Desulfovibrio salexigens			Periplasmic binding protein/ LacI transcriptional regulator	30 060.5	1	4.81	VIQLEGLA GTSAAAR
Cyanobacteria or Firmicutes	Unknown	Unknown	Elongation Factor Tu	43 262.6	1	3.28	TTLTAAITT VLAK		

(continued on next page)

Table 2 (continued)

Phylum	Class	Order; family; genus; species	Protein name	Protein molecular weight (Da)	Number of unique peptides	Percentage sequence coverage (%)	Peptide sequence
Cyanobacteria	Unknown	Unknown	Photosystem II reaction centre protein PsbD/D2	39 443.4	2	5.40	AYDFVSQLER
			Photosystem Q(B) protein 1	29 295.2	2	8.68	NILLNEGIR LIFQYAS FNNSR
			Allophycocyanin alpha chain	17 413.2	2	11.80	VINTWADIINR SIVNADAEAR YLSPGELDR
	Chroococcales	<i>Synechococcus</i>	Hypothetical protein S7335_4928	7758.3	1	10.30	LATVVVTPR
	Nostocales	Nostocaceae	Photosystem Q(B) protein 1 (32 kDa thylakoid membrane protein 1); Photosystem II protein D1	39 531.8	1	6.39	TIDTWAD LLNR
Planctomycetes	Planctomycetacia	Planctomycetales; Planctomycetaceae; <i>Planctomyces</i> ; <i>Planctomyces limnophilus</i>	ATP synthase F1 subcomplex beta subunit	51 974.8	2	4.99	FSQAGSEV SALLGR TVILQELIAR
			F ₀ F ₁ -ATP synthase subunit beta	52 729.3	1	4.97	TVILTELIAR FSQAGSEVS ALLGR
		Planctomycetales; Planctomycetaceae; <i>Planctomyces</i> ; <i>Planctomyces maris</i>					

clones). As Cyanobacteria and Rhizobiales are known to be involved in carbon and nitrogen fixation, these organisms could play an important role in such oligotrophic ecosystems.

In conclusion, by combining proteomic and genomic approaches in a mildly arsenic-contaminated environment, the diversity of arsenic-adapted prokaryotic communities was enlarged. Such approaches can thus serve to explore the diversity and the main functional traits of a bacterial community, including those having a high level of diversity.

Acknowledgments

Audrey Cordi was supported by grants from the French Ministry of Research and the “Zone Atelier Moselle” (ZAM) project. Audrey Heinrich-Salmeron, David Halter and Sébastien Gallien were supported by grants from the French Ministry of Research and the Université de Strasbourg. Florence Goulhen-Chollet was supported by the Agence Nationale de la Recherche (ANR), RARE project (Reactivity of an Arsenic-Rich Ecosystem). This study was financed by the EC2CO program (Institut National des Sciences de l’Univers, CNRS) and BRG. This work was performed in the framework of the Groupement de Recherche “Métabolisme de l’Arsenic chez les micro-organismes: de la résistance à la détoxification” (GDR2909-CNRS).

References

Achour-Rokbani, A., Cordi, A., Poupin, P., Bauda, P., Billard, P., 2010. Characterization of the *ars* gene cluster from extremely arsenic-resistant *Microbacterium* sp. strain A33. Appl. Environ. Microbiol. 76, 948–955.

Achour, A.R., Bauda, P., Billard, P., 2007. Diversity of arsenite transporter genes from arsenic-resistant soil bacteria. Res. Microbiol. 158, 128–137.

AFNOR, 1994. Qualité de l’Eau, seconde ed Paris.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucl. Acids Res. 25, 3389–3402.

Anisimova, M., Gascuel, O., 2006. Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. Syst. Biol. 55, 539–552.

Ashelford, K.E., Chuzhanova, N.A., Fry, J.C., Jones, A.J., Weightman, A.J., 2005. At least 1 in 20 16S rRNA sequence records currently held in public repositories is estimated to contain substantial anomalies. Appl. Environ. Microbiol. 71, 7724–7736.

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

Benndorf, D., Balcke, G.U., Harms, H., von Bergen, M., 2007. Functional metaproteome analysis of protein extracts from contaminated soil and groundwater. ISME J. 1, 224–234.

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

Brochier-Armanet, C., Boussau, B., Gribaldo, S., Forterre, P., 2008. Mesophilic Crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota. Nat. Rev. Microbiol. 6, 245–252.

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

Bruneel, O., Duran, R., Koffi, K., Casiot, C., Fourçans, A., Elbaz-Poulichet, F., Personné, J.C., 2005. Microbial diversity in a pyrite-rich tailings impoundment (Carnoules, France). Geomicrobiol. J. 22, 249–257.

Bruneel, O., Pascault, N., Egal, M., Bancon-Montigny, C., Goni-Urriza, M.S., Elbaz-Poulichet, F., Personné, J.C., Duran, R., 2008. Archaeal diversity in a Fe–As rich acid mine drainage at Carnoules (France). Extremophiles 12, 563–571.

Bruneel, O., Volant, A., Gallien, S., Chaumande, B., Casiot, C., Carapito, C., Bardil, A., Morin, G., Brown Jr., G.E., Personné, J.C., Le Paslier, D., Schaeffer, C., Van Dorselaer, A., Bertin, P.N., Elbaz-Poulichet, F., Arsène-Ploetz, F., 2011. Characterization of the active bacterial community involved

- in natural attenuation processes in arsenic-rich creek sediments. *Microb. Ecol.* 61, 793–810.
- Casiot, C., Pedron, V., Bruneel, O., Duran, R., Personne, J.C., Grapin, G., Drakides, C., Elbaz-Poulichet, F., 2006. A new bacterial strain mediating As oxidation in the Fe-rich biofilm naturally growing in a groundwater Fe treatment pilot unit. *Chemosphere* 64, 492–496.
- Cole, J.R., Wang, Q., Cardenas, E., Fish, J., Chai, B., Farris, R.J., Kulam-Syed-Mohideen, A.S., McGarrell, D.M., Marsh, T., Garrity, G.M., Tiedje, J.M., 2009. The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucl. Acids Res.* 37, D141–D145.
- Coupland, K., Battaglia-Brunet, F., Hallberg, K.B., Dictor, M.C., Garrido, F., Johnson, D.B., 2004. Oxidation of iron, sulfur and arsenic in mine waters and mine wastes: an important role of novel *Thiomonas* spp. In: Tsezos, A. H.M., Remondaki, E. (Eds.), *Biohydrometallurgy: A Sustainable Technology in Evolution*. National Technical University of Athens, Zografou, Greece, pp. 639–646.
- DeLong, E.F., 1992. Archaea in coastal marine environments. *Proc. Natl. Acad. Sci. U S A* 89, 5685–5689.
- Denef, V.J., Mueller, R.S., Banfield, J.F., 2010. AMD biofilms: using model communities to study microbial evolution and ecological complexity in nature. *ISME J.* 4, 599–610.
- Dodds, W.K., Jones, J.R., Welch, E.B., 1998. Suggested classification of stream trophic state: distributions of temperate stream types by chlorophyll, total nitrogen, and phosphorus. *Water Res.* 32, 1455–1462.
- Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucl. Acids Res.* 32, 1792–1797.
- Fuhrman, J.A., McCallum, K., Davis, A.A., 1992. Novel major archaeobacterial group from marine plankton. *Nature* 356, 148–149.
- Galand, P.E., Lovejoy, C., Vincent, W.F., 2006. Remarkably diverse and contrasting archaeal communities in a large arctic river and the coastal Arctic Ocean. *Aquat. Microb. Ecol.* 44, 115–126.
- Guindon, S., Gascuel, O., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* 52, 696–704.
- Hallberg, K.B., Johnson, D.B., 2005. Microbiology of a wetland ecosystem constructed to remediate mine drainage from a heavy metal mine. *Sci. Total Environ.* 338, 53–66.
- Hamamura, N., Macur, R.E., Korf, S., Ackerman, G., Taylor, W.P., Kozubal, M., Reysenbach, A.L., Inskeep, W.P., 2009. Linking microbial oxidation of arsenic with detection and phylogenetic analysis of arsenite oxidase genes in diverse geothermal environments. *Environ. Microbiol.* 11, 421–431.
- Hatzenpichler, R., Lebedeva, E.V., Spieck, E., Stoecker, K., Richter, A., Daims, H., Wagner, M., 2008. A moderately thermophilic ammonia-oxidizing crenarchaeote from a hot spring. *Proc. Natl. Acad. Sci. U S A* 105, 2134–2139.
- Honschopp, S., Brunken, N., Nehrhorn, A., Breunig, H.J., 1996. Isolation and characterization of a new arsenic methylating bacterium from soil. *Microbiol. Res.* 151, 37–41.
- Johnson, D.B., Hallberg, K.B., 2003. The microbiology of acidic mine waters. *Res. Microbiol.* 154, 466–473.
- Laemmli, U.K., 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227, 680–685.
- Lièvremon, D., Bertin, P.N., Lett, M.C., 2009. Arsenic in contaminated waters: biogeochemical cycle, microbial metabolism and biotreatment processes. *Biochimie* 91, 1229–1237.
- Macur, R.E., Jackson, C.R., Botero, L.M., McDermott, T.R., Inskeep, W.P., 2004. Bacterial populations associated with the oxidation and reduction of arsenic in an unsaturated soil. *Environ. Sci. Technol.* 38, 104–111.
- Mandal, B.K., Suzuki, K.T., 2002. Arsenic round the world: a review. *Talanta* 58, 201–235.
- Mokashi, S.A., Paknikar, K.M., 2002. Arsenic (III) oxidizing *Microbacterium lacticum* and its use in the treatment of arsenic contaminated groundwater. *Lett. Appl. Microbiol.* 34, 258–262.
- Mueller, R.S., Denef, V.J., Kalnejais, L.H., Suttle, K.B., Thomas, B.C., Wilmes, P., Smith, R.L., Nordstrom, D.K., McCleskey, R.B., Shah, M.B., Verberkmoes, N.C., Hettich, R.L., Banfield, J.F., 2010. Ecological distribution and population physiology defined by proteomics in a natural microbial community. *Mol. Syst. Biol.* 6, 374.
- Nicol, G.W., Tschirko, D., Chang, L., Hammesfahr, U., Prosser, J.I., 2006. Crenarchaeal community assembly and microdiversity in developing soils at two sites associated with deglaciation. *Environ. Microbiol.* 8, 1382–1393.
- Raskin, L., Stromley, J.M., Rittmann, B.E., Stahl, D.A., 1994. Group-specific 16S rRNA hybridization probes to describe natural communities of methanogens. *Appl. Environ. Microbiol.* 60, 1232–1240.
- Rathinasabapathi, B., Raman, S.B., Kertulis, G., Ma, L., 2006. Arsenic-resistant proteobacterium from the phyllosphere of arsenic-hyperaccumulating fern (*Pteris vittata* L.) reduces arsenate to arsenite. *Can. J. Microbiol.* 52, 695–700.
- Rudolph, C., Moissl, C., Henneberger, R., Huber, R., 2004. Ecology and microbial structures of archaeal/bacterial strings-of-pearls communities and archaeal relatives thriving in cold sulfidic springs. *FEMS Microbiol. Ecol.* 50, 1–11.
- Schleifer, K.H., 2009. Classification of Bacteria and Archaea: past, present and future. *Syst. Appl. Microbiol.* 32, 533–542.
- Schloss, P.D., Handelsman, J., 2005. Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. *Appl. Environ. Microbiol.* 71, 1501–1506.
- Sharma, V.K., Sohn, M., 2009. Aquatic arsenic: toxicity, speciation, transformations, and remediation. *Environ. Int.* 35, 743–759.
- Sliwinski, M.K., Goodman, R.M., 2004. Spatial heterogeneity of crenarchaeal assemblages within mesophilic soil ecosystems as revealed by PCR-single-stranded conformation polymorphism profiling. *Appl. Environ. Microbiol.* 70, 1811–1820.
- Smedley, P.L., Kinniburgh, D.G., 2002. A review of the source, behaviour and distribution of arsenic in natural waters. *Appl. Geochem.* 17, 517–568.
- Stahl, D.A., Amann, R., 1991. Development and application of nucleic acid probes. In: Stackebrandt, E., Goodfellow, M. (Eds.), *Nucleic Acid Techniques in Bacterial Systematics*. John Wiley & Sons Inc., New York, pp. 205–248.
- Thomas, P., Finnie, J.K., Williams, J.G., 1997. Feasibility of identification and monitoring of arsenic species in soil and sediment samples by coupled high-performance liquid chromatography – inductively coupled plasma mass spectrometry. *J. Anal. Spectrom.* 12, 1367–1372.
- Thompson, J.D., Gibson, T.J., Higgins, D.G., 2002. Multiple Sequence Alignment Using ClustalW and ClustalX. *Curr. Protoc. Bioinformatics* Chapter 2, Unit 2.3.
- Tsai, S.L., Singh, S., Chen, W., 2009. Arsenic metabolism by microbes in nature and the impact on arsenic remediation. *Curr. Opin. Biotechnol.* 20, 659–667.
- Tyson, G.W., Chapman, J., Hugenholtz, P., Allen, E.E., Ram, R.J., Richardson, P.M., Solovyev, V.V., Rubin, E.M., Rokhsar, D.S., Banfield, J. F., 2004. Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* 428, 37–43.
- Valenzuela-Encinas, C., Neria-Gonzalez, I., Alcantara-Hernandez, R.J., Enriquez-Aragón, J.A., Estrada-Alvarado, I., Hernandez-Rodriguez, C., Dendooven, L., Marsch, R., 2008. Phylogenetic analysis of the archaeal community in an alkaline-saline soil of the former lake Texcoco (Mexico). *Extremophiles* 12, 247–254.
- Verberkmoes, N.C., Denef, V.J., Hettich, R.L., Banfield, J.F., 2009. Systems biology: functional analysis of natural microbial consortia using community proteomics. *Nat. Rev. Microbiol.* 7, 196–205.
- Weiss, S., Carapito, C., Cleiss, J., Koechler, S., Turlin, E., Coppee, J.Y., Heymann, M., Kugler, V., Stauffert, M., Cruveiller, S., Médigue, C., Van Dorsselaer, A., Bertin, P.N., Arsène-Ploetze, F., 2009. Enhanced structural and functional genome elucidation of the arsenite-oxidizing strain *Herminiimonas arsenicoxydans* by proteomics data. *Biochimie* 91, 192–203.
- Williams, T.A., Codoner, F.M., Toft, C., Fares, M.A., 2010. Two chaperonin systems in bacterial genomes with distinct ecological roles. *Trends Genet.* 26, 47–51.

Unsuspected Diversity of Arsenite-Oxidizing Bacteria as Revealed by Widespread Distribution of the *aoxB* Gene in Prokaryotes^{∇†}

Audrey Heinrich-Salmeron,^{1‡} Audrey Cordi,^{2‡} Céline Brochier-Armanet,³ David Halter,¹
Christophe Pagnout,² Elham Abbaszadeh-fard,² Didier Montaut,⁴ Fabienne Seby,⁴
Philippe N. Bertin,¹ Pascale Bauda,² and Florence Arsène-Plöetze^{1*}

Laboratoire de Génétique Moléculaire, Génomique, Microbiologie, Département Microorganismes, Génomes, Environnement, UMR7156, Université de Strasbourg/CNRS, 28 Rue Goethe, 67083 Strasbourg Cedex, France¹; Laboratoire des Interactions Ecotoxicologie Biodiversité Ecosystèmes (LIEBE), UMR7146, CNRS, Université Paul Verlaine, Campus Bridoux, Rue du Général Delestraint, 57070 Metz, France²; Université de Provence (Aix-Marseille Université), Laboratoire de Chimie Bactérienne (CNRS, UPR9043), Institut de Microbiologie de la Méditerranée (IFR88), 31 Chemin Joseph Aiguier, 13402 Marseille Cedex 20, France³; and Laboratoire Ultra-Traces Analyse Aquitaine, Hélioparc Pau-Pyrénées, 2 Avenue du Président Angot, 64053 Pau Cedex 9, France⁴

Received 9 December 2010/Accepted 3 May 2011

In this study, new strains were isolated from an environment with elevated arsenic levels, Sainte-Marie-aux-Mines (France), and the diversity of *aoxB* genes encoding the arsenite oxidase large subunit was investigated. The distribution of bacterial *aoxB* genes is wider than what was previously thought. AoxB subfamilies characterized by specific signatures were identified. An exhaustive analysis of AoxB sequences from this study and from public databases shows that horizontal gene transfer has likely played a role in the spreading of *aoxB* in prokaryotic communities.

Arsenic, which is one of the most toxic metalloids, is distributed ubiquitously but not uniformly around the world. Levels of arsenic differ considerably from one geographical region to another, depending on the geochemical characteristics of the soil (natural contamination) and the industrial activities carried out in the vicinity (anthropogenic contamination) (22). In aquatic environments, arsenic occurs mainly in the form of the inorganic species arsenate [As(V)] and arsenite [As(III)]; the latter species, which is more bioavailable, is usually thought to have more-toxic effects on prokaryotes than As(V) (34). As(III) oxidation leads to the formation of the less available form As(V), which can either precipitate with iron [Fe(III)] or be adsorbed by ferrihydrite. The oxidation process may be mediated by microbial activities, which contribute to the natural remediation processes observed in contaminated environments (21, 26, 27, 34). Consequently, bioprocesses for the treatment of arsenic-contaminated waters have been developed based on the precipitation or adsorption of the As(V) produced by bacteria (4, 9, 21). Some well-known prokaryotes oxidize As(III) into As(V) under aerobic (e.g., *Herminiimonas arsenicoxydans*, *Thiomonas* spp., or *Rhizobium* sp. strain NT26) or anaerobic (e.g., *Alkalilimnicola ehrlichii*) conditions as part of a detoxification process (12, 17, 31, 32, 39). Some chemolithotrophs also use arsenite as an electron donor (e.g., *Rhizobium* sp. strain NT26 or *Thiomonas arsenivorans*) (5, 32). The

aerobic arsenite oxidases involved in such processes are heterodimers consisting of a large subunit with a molybdenum center and a [3Fe-4S] cluster (AroA, AsoA, and AoxB) and a small subunit containing a Rieske-type [2Fe-2S] cluster (AroB, AsoB, and AoxA) (1, 13). The large subunit in these enzymes is similar to that found in other members of the dimethyl sulfoxide (DMSO) reductase family of molybdenum enzymes but is clearly phylogenetically divergent from the respiratory arsenate reductases (ArrA) or other proteins of the DMSO reductase family of molybdenum oxidoreductases, such as the new arsenite reductase described recently for *Alkalilimnicola ehrlichii* (25, 31, 40).

aox genes have been identified in 25 bacterial and archaeal genera isolated from various arsenic-rich environments, most of which belong to the *Alpha*-, *Beta*-, or *Gamma*proteobacteria phylum (7, 10, 12, 14, 23, 25, 29, 32, 37). Recent studies based on environmental DNA extracted from soils, sediments, and geothermal mats with different chemical characteristics and various levels of arsenic contamination have suggested that the distribution and the diversity of arsenite-oxidizing microorganisms may be greater than previously suggested (6, 10, 14–16, 18, 28, 29). The aim of this study was to investigate the presence of the *aox* gene in bacteria other than the *Proteobacteria* in order to improve our knowledge about the phylogeny and evolution of AoxB. With this aim, an environment with elevated arsenic levels showing considerable prokaryote diversity was chosen. The *aoxB* sequences studied here were amplified from the DNA of bacteria isolated from sediments collected at this site and from the DNA extracted directly from these sediments.

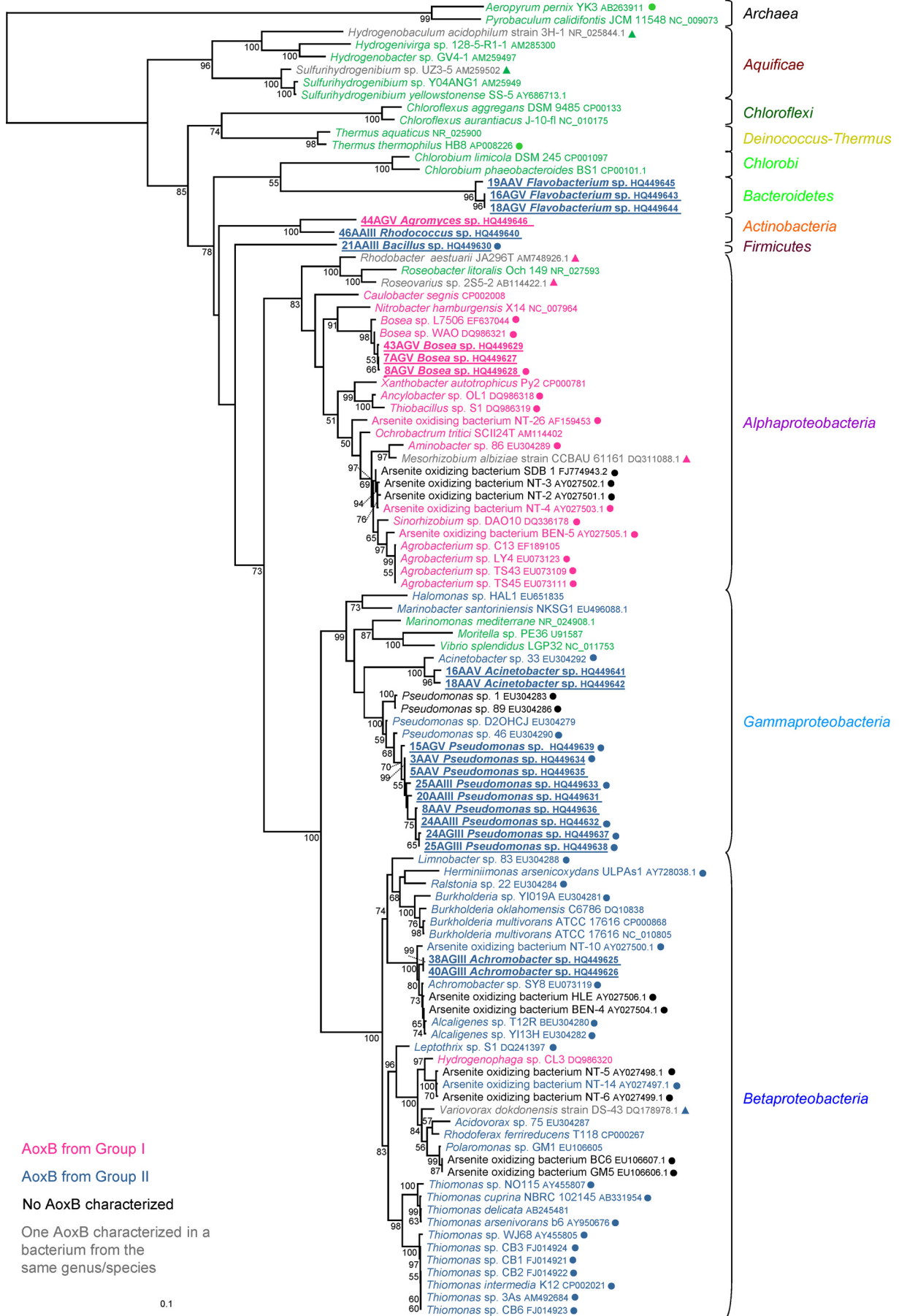
The study site is located in the vicinity of the Gabe-Gottes mine in the Sainte-Marie-aux-Mines valley (France) (48°12'51"N, 007°09'34"E) (15a). Water percolating from the mine walls accumulates in a creek, where sediments were collected in

* Corresponding author. Mailing address: Laboratoire de Génétique Moléculaire, Génomique, Microbiologie, Département Microorganismes, Génomes, Environnement, UMR7156 Université de Strasbourg/CNRS, 28 Rue Goethe, 67083 Strasbourg Cedex, France. Phone: 33 3 90 24 18 12. Fax: 33 3 90 24 20 28. E-mail: ploetze@unistra.fr.

‡ These authors contributed equally to this work.

† Supplemental material for this article may be found at <http://aem.asm.org/>.

∇ Published ahead of print on 13 May 2011.



December 2007, 40 m downstream of the entrance to the mine, from a 30-m-long part of the creek. In these sediment samples, where the total arsenic level was 320 mg kg⁻¹, the As(V) species concentration was 10 times greater than the As(III) species (15a). The main arsenic species observed in sedimentary interstitial water were As(III) and As(V), in concentrations of 14.6 (±0.3) µg liter⁻¹ and 135 (±5) µg liter⁻¹, respectively. Such concentrations correspond to moderately high arsenic levels; our samples were either more severely contaminated than or similarly contaminated relative to most of previously studied sites at which arsenic oxidase-encoding genes have been detected (see Table S1 in the supplemental material) (3, 6, 8, 10, 14–16, 18, 29, 37). Moreover, the sediments at this site were previously found to show considerable bacterial diversity (15a). These characteristics made these samples suitable for investigation of the diversity of *aoxB* genes.

In the present study, *aoxB* genes were amplified from the total sediment DNA and from DNA extracted from several arsenic-resistant isolates. These bacterial isolates were grown for 18 days at 20°C on DR2A medium, for which the composition was as follows (per liter): 0.05 g each of yeast extract, peptone, acid hydrolysate of casein, dextrose, and soluble starch, 0.03 g each of dipotassium phosphate and sodium pyruvate, and 0.0024 g of magnesium sulfate (36) amended with either As(III) or As(V) [isolates named with the suffix III or V refer to the presence of As(III) or As(V), respectively] under aerobic or anaerobic conditions (isolates named with the prefix A refer to aerobic conditions) using the Anaerocult P system (Grosseron) (the middle letter, A or G, corresponds to the gelling agent, i.e., agar or gellan gum, respectively). Among the 220 isolates obtained and tested using BM1-2F/BM3-1R primers (see Table S2 in the supplemental material) to detect the presence of the *aoxB* gene, positive PCR amplification was obtained for 22 aerobic strains (see Table S3 in the supplemental material) and no *aoxB* amplification was observed for anaerobe isolates. 16S rRNA genes from these isolates were amplified and sequenced using primers W01/W02 (see Table S2). These 22 isolates belong to eight genera (*Achromobacter*, *Pseudomonas*, *Agromyces*, *Rhodococcus*, *Flavobacterium*, *Bosea*, *Acinetobacter*, and *Bacillus*), representing four bacterial phyla (*Firmicutes*, *Actinobacteria*, *Bacteroidetes*, and *Proteobacteria*) (Fig. 1 [underlined, bold names]; see also Table S3). To our knowledge, this is the first time *aoxB* genes have been reported to occur in *Bacillus*, *Rhodococcus*, *Agromyces*, and *Flavobacterium* spp. Positive arsenite oxidase

activity was detected with the AgNO₃ method (35) using *H. arsenicoxydans* ULPAs1 as a positive control (39) in the case of 8 isolates, 6 of which belong to the *Pseudomonas* genus, 1 to the *Bacillus* genus, and 1 to the *Bosea* genus (Fig. 1 [underlined, bold names highlighted with circles]). The absence of activity in the other strains may be due to a lack of *aoxB* gene expression under the conditions tested, the lack of the small subunit, or a gene coding for a nonfunctional enzyme. Arsenite oxidation activity has been observed in previous studies in the *Pseudomonas*, *Acinetobacter*, *Achromobacter*, and *Bosea* genera (7, 28–30) (Fig. 1 [indicated with circles]).

The phylogenetic analysis of the 22 deduced AoxB amino acid sequences and their homologues present in the nonredundant protein database at the NCBI was performed. A preliminary phylogenetic analysis (data not shown) helped us to select only aerobic arsenite oxidase (AoxB, AroA, or AsoA) sequences and to exclude respiratory arsenate reductases (ArrA) or other proteins of the DMSO reductase family of molybdenum oxidoreductases, such as the new arsenite reductase recently described for *Alkalilimnicola ehrlichii*, which are clearly phylogenetically divergent from the AoxB/AsoA/AroA proteins, as previously shown (25, 40). The phylogenetic analysis of these AoxB amino acid sequences showed that most of them belonged to two major groups (see Fig. S1 in the supplemental material), as recently described (29). Group I sequences were found mainly in *Alphaproteobacteria*, whereas group II was composed mainly of betaproteobacterial sequences (Fig. 2 and 3; see also Fig. S1). This pattern suggests that these two groups probably originated from these two proteobacterial divisions. AoxB sequences belonging to none of these two groups were found in a few representatives of various bacterial divisions and a few archaeal phyla (see Fig. S1). Among the 22 AoxB characterized here, four were found to belong to group I. More specifically, three of them were amplified from isolates classified as *Bosea* sp. (*Alphaproteobacteria*) and one as *Agromyces* sp. (*Actinobacteria*) (Fig. 2). The 18 remaining AoxB sequences belong to group II. These sequences were amplified from isolates affiliated with *Betaproteobacteria* (*Achromobacter* sp.), *Gammaproteobacteria* (*Pseudomonas* sp.), *Actinobacteria* (*Rhodococcus* sp.), *Firmicutes* (*Bacillus* sp.), and *Bacteroidetes* (*Flavobacterium* sp.) (Fig. 3).

Comparisons between the 16S rRNA and AoxB phylogenies showed the existence of some striking inconsistencies between organism and gene evolutionary histories (Fig. 1, 2, and 3; see also Fig. S1 in the supplemental material). These can be ex-

FIG. 1. Maximum likelihood phylogenetic tree of the 16S rRNA sequences of our 22 isolates showing positive *aoxB* PCR amplification. This tree also includes 16S rRNA sequences deposited in public databases, corresponding to the organisms found (or close relatives) to carry the *aoxB* gene or closely related organisms. The tree was rooted in between *Archaea* and *Bacteria* domains. Phylogenetic analyses were performed on the 1,014 unambiguously aligned nucleic acid positions by using Treefinder (19). The evolutionary model GTR + Γ4 was used for this purpose, as suggested by the “propose model” tool available in Treefinder. Numbers at branches are bootstrap values obtained using the nonparametric bootstrap approach implemented in Treefinder (based on 100 replicates of the original data set). Only bootstrap values above 50% are shown. Bacteria harboring *aoxB* sequences corresponding to phylogenetic group I or group II or neither of these two groups are given in pink, dark blue, or green, respectively. The 22 bacterial isolates of this study are in boldface and underlined. All bacteria experimentally shown to be able to oxidize arsenite are labeled with a circle at the end of the name. Bacteria which were previously shown to be able to oxidize arsenite or harbor an *aoxB* gene, but for which no 16S sequence was available, were therefore not included in this tree. However, in this case, we included the closest relative of these strains for which a 16S gene was available (in gray), and the triangle color at the end of these names corresponds to the group to which the *aoxB* gene harbored by the strain belongs (pink, dark blue, and green corresponding to group I, group II, and neither of these groups, respectively). The names of bacteria which were described in the literature as able to oxidize arsenite but for which *aoxB* sequences were not available are presented in black. The scale bar represents the average number of substitutions estimated per site.



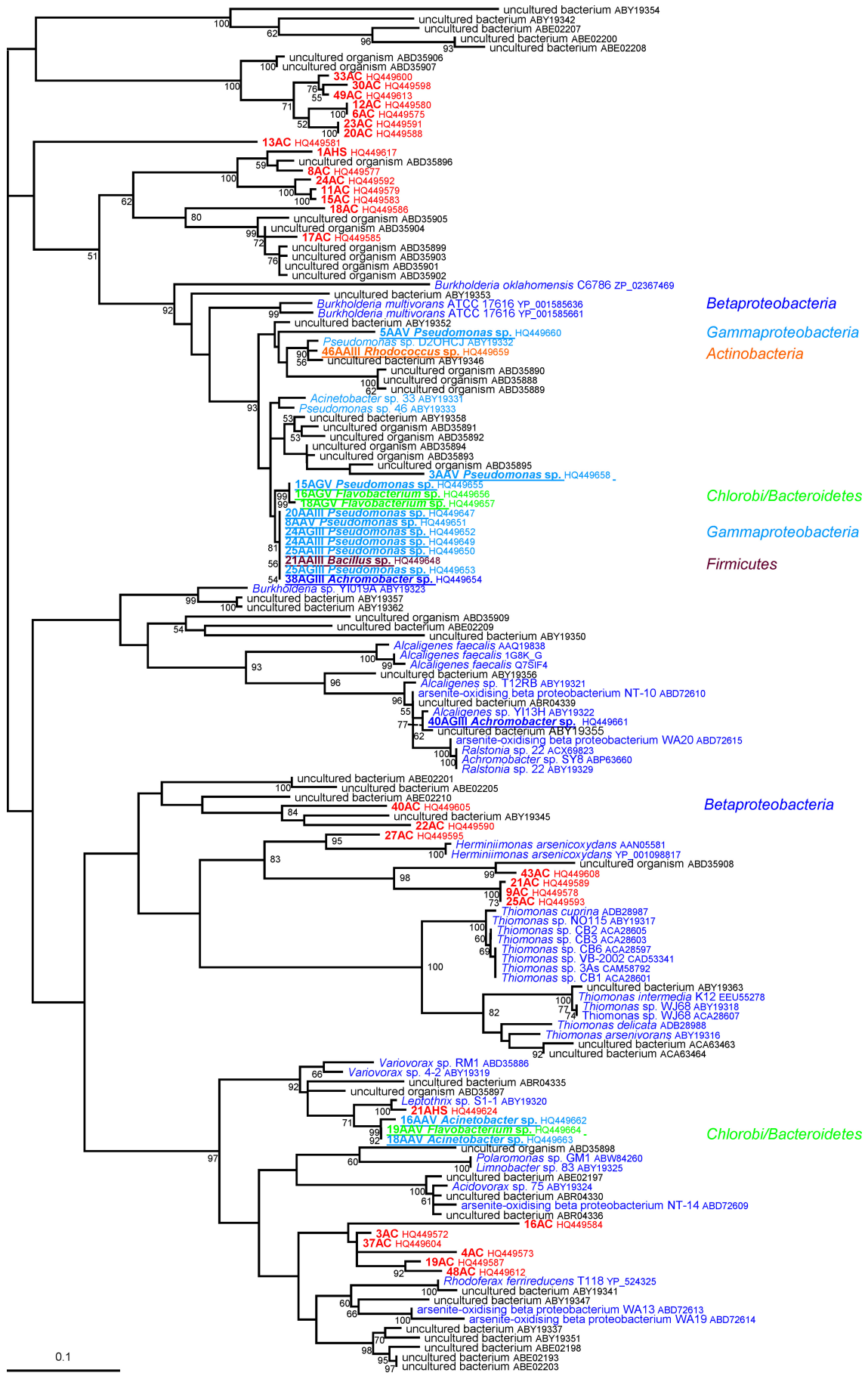
plained by four hypotheses: (i) hidden paralogies (i.e., gene duplications followed by differential gene losses), (ii) allele sorting, (iii) horizontal gene transfer (HGT), and (iv) tree reconstruction artifacts. In the case of AoxB, a few HGT events represent the most likely hypothesis to explain the presence of identical (or nearly identical) AoxB sequences in some distant lineages, such as group I AoxB sequences found in *Agromyces* and *Bosea*, two unrelated bacteria that belong respectively to *Actinobacteria* and *Alphaproteobacteria* (Fig. 2). Moreover, in addition to including sequences from *Alphaproteobacteria*, group I included one betaproteobacterial AoxB sequence (*Hydrogenophaga* sp. strain CL3) (Fig. 2), which indicates that this bacterium has very likely acquired its *aoxB* gene from *Alphaproteobacteria* via HGT. A similar process may account for the sequence observed in the *Agromyces* sp. isolated in this study (Fig. 2). In group II, some *Pseudomonas*, *Marinobacter*, *Halomonas*, and also *Firmicutes* and *Chlorobi/Bacteroidetes* AoxB sequences were intermixed with those from *Betaproteobacteria* (Fig. 3). This observation strongly suggests that the former acquired their *aoxB* gene from the latter by HGT. More convincingly, one of the two actinobacterial sequences characterized in this study belonged to group I, whereas the second sequence belonged to group II (Fig. 2 and 3), although the strains from which these sequences were amplified are closely related (Fig. 1). This suggests that these two *aoxB* genes were acquired by HGT from two unrelated proteobacterial donors. The great similarity observed between some AoxB sequences, such as those from *Bacillus* (isolate 21AIII), *Flavobacterium* (isolates 16AGV and 18AGV) and *Pseudomonas*, indicates that the HGT events probably occurred quite recently (Fig. 3). HGT in prokaryotic communities has been proposed previously to explain the strong similarities observed between AoxB proteins from *Ralstonia* spp. and *Achromobacter* sp. strain SY8 (GenBank accession numbers ACX69823 and ABP63660) (20). In that previous study, it was suggested that *Ralstonia* has acquired the corresponding gene cluster by HGT from another betaproteobacterium (20). The proximity of these AoxB proteins in the AoxB phylogenetic tree (Fig. 3) strengthens this proposal to explain the presence of nearly identical AoxB sequences in these two *Betaproteobacteria*. Interestingly, the present work revealed for the first time that such HGT may have occurred between bacteria belonging to different classes or phyla.

Genomic islands (GEI) or plasmids may be involved in such transfers. This is, for instance, the case for *aoxB* genes found in *Thiomonas* (Fig. 3) that are localized in one genomic island (2). We searched for other examples of genomic islands containing *aoxB* belonging to group I (*Nitrobacter hamburgensis* X14 [*Alphaproteobacteria*]) (Fig. 2) or group II (*Thiomonas* strain 3As, *Thiomonas intermedia* K12, and *H. arsenicoxydans* [three *Betaproteobacteria*]) (Fig. 3) or outside these two groups

(*Vibrio splendidus* [*Gammaproteobacteria*], *Aeropyrum pernix*, and *Pyrobaculum calidifontis* [*Archaea*], and *Thermus thermophilus* [*Thermus/Deinococcus*]) (see Fig. S1 in the supplemental material). With this aim, we used the RGPfinder (for regions of genomic plasticity [RGP]) tool of the Microscope platform (<https://www.genoscope.cns.fr/agc/microscope/>) (38). Interestingly, *aoxB* genes were present in RGP in *H. arsenicoxydans* as previously proposed (24), in *V. splendidus* and *P. calidifontis* (see Fig. S2 in the supplemental material). In addition, *aoxB* genes were carried on plasmids in *N. hamburgensis* X14 (plasmid 2, NC_007960) and *T. thermophilus* HB8 (pTT27, NC_006462). In the case of this microorganism, a transposase-encoding gene was found in the vicinity of the *aoxAB* genes (in green in Fig. S2 in the supplemental material). Finally, whereas the *aox* operon is present in a GEI in *Thiomonas* strain 3As (2), the orthologous genes were not found in an RGP in *Thiomonas intermedia*. In fact, the synteny between these two strains is restricted to the *aoxABD* and *ptxB-aoxRS* operons. More generally, the synteny of *aoxAB* genes is conserved among all bacteria for which a genome sequence was available (see Fig. S2). In contrast, synteny with the other *aox* genes, *aoxC* and *aoxD* or *aoxRS*, was observed only in few bacteria belonging to group II (*Thiomonas* sp. strain 3As and *H. arsenicoxydans*), whose *aoxB* genes are related (Fig. 3), and in *Xanthobacter autotrophicus*. Synteny of *aoxAB* with *aoxR* or *aoxS* was conserved in several bacteria belonging to group II or to group I. These observations suggest that the synteny conservation is not linked to the *aoxB* phylogeny observed or to a specific energy/carbon metabolism (heterotrophs versus autotrophs).

The phylogenetic analyses support the hypothesis that HGTs have played a role in the widespread distribution of some *aoxB* genes in prokaryotes, meaning that the *aoxB* gene is not an efficient phylogenetic marker. Indeed, because of the discrepancies found to exist between the 16S rRNA and *aoxB* phylogenies, it is not possible to predict the taxonomic affiliation of strains based on the *aoxB* sequences alone. This precludes the use of approaches involving the direct amplification of *aoxB* in order to characterize the arsenite oxidizers from environmental samples. This is the case for the 56 *aoxB* sequences obtained in this study from the total DNA (shown in red in Fig. 2 and 3). As in the case of the sequences obtained from cultivated isolates, most of these deduced AoxB proteins belonged to group I (24 sequences) and group II (29 sequences). In contrast, only three of the characterized sequences were similar to sequences from *Chloroflexi* and *Thermus*, which branched outside these two groups (data not shown). Interestingly, most of our AoxB sequences belonging to groups I and II were found to be closely related to sequences from hitherto uncultured and/or unidentified bacteria or to constitute new subgroups (Fig. 2 and 3). This situation clearly shows that the diversity of AoxB is far from having been exhaustively investigated.

FIG. 2. Unrooted phylogenetic trees of the 86 group I AoxB sequences (152 unambiguously aligned positions). The 86 sequences were retrieved from the nonredundant database of the NCBI, deduced from *aoxB* genes from 4 isolated strains (in bold and underlined) or deduced from the 24 PCR products obtained from metagenomic DNA (in red). Phylogenetic analyses were performed using the maximum likelihood approach implemented in Treefinder (19). The LG + Γ 4 model was used for this purpose, as suggested by the "propose model" tool available in Treefinder. Numbers at branches are bootstrap values determined using the nonparametric bootstrap approach implemented in Treefinder (based on 100 replicates of the original data set). The scale bar represents the average number of substitutions estimated per site.



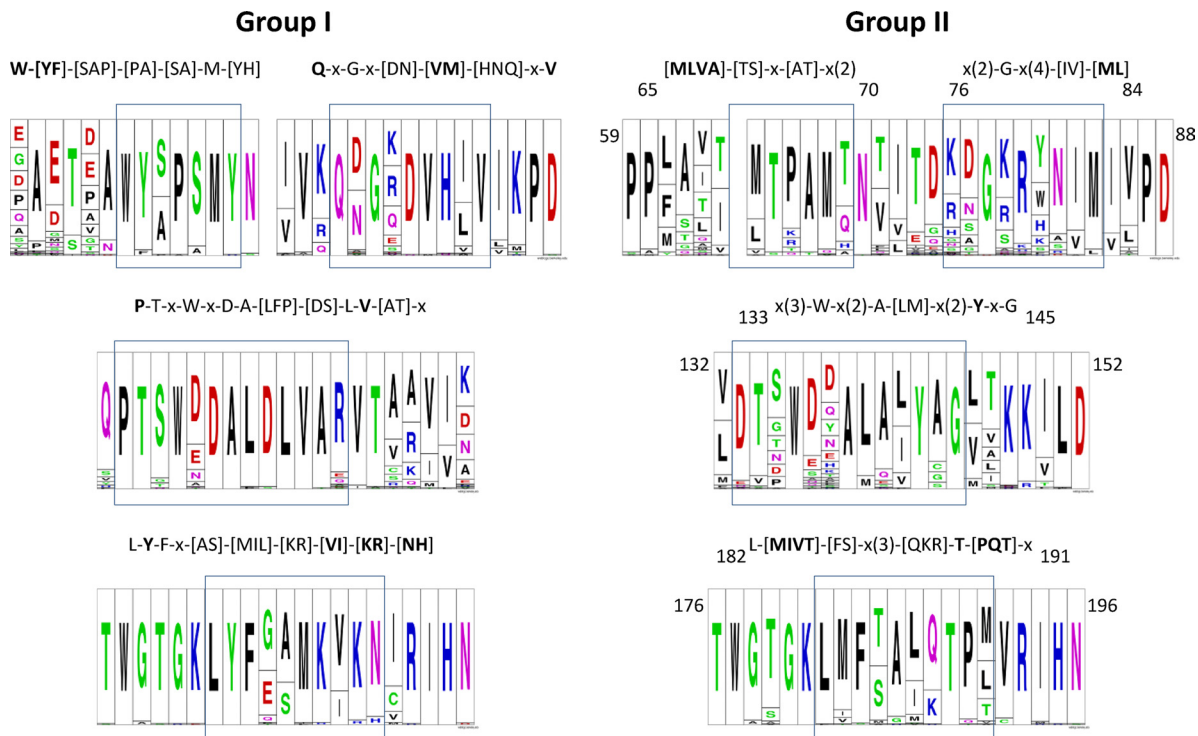


FIG. 4. Group I and group II-specific motifs deduced from AoxB sequence alignment with the WebLogo software program via the WebLogo website (<http://weblogo.berkeley.edu/>) (11, 33). Residues with which these motifs could be unambiguously defined are in bold. The absence of these motifs in AoxB sequences belonging to neither group I nor group II was checked visually. Amino acid numbering was based on the *Alcaligenes faecalis* protein sequence (UniProt accession number Q7SIF4).

The comparisons between 78 AoxB proteins characterized in this study, including the most divergent sequences, showed that residues present in the catalytic site [²¹C-X₂-²⁴C-X₃-²⁸C-X₇₀-⁹⁹S] according to the betaproteobacterium *Alcaligenes faecalis* NCIB 8687 protein sequence numbering (13) (UniProt accession number Q7SIF4), and the hydrophilic channel near the Mo centers that was proposed to be the substrate binding site (from amino acid positions 195 to 203 according to the *A. faecalis* AoxB) were conserved in 76 of the 78 sequences studied (data not shown; see also Fig. S3 in the supplemental material). This suggested that when they are expressed, these AoxB proteins might function like the *A. faecalis* arsenite oxidase, for which a crystal structure was determined (13). Close comparisons between all the AoxB sequences belonging to groups I and II showed that four motifs differ between these two groups (Fig. 4; see also Fig. S3 in the supplemental material). These four regions extend from amino acid positions 65 to 70, 76 to 84, 133 to 145, and 182 to 191 (according to *A. faecalis* AoxB protein sequence numbering). They contain the motifs W-[YF]-[SAP]-[PA]-[SA]-M-[YH] and [MLVA]-[TS]-X-[AT]-X₂, Q-X-G-X-[DN]-[VM]-[HNQ]-X-V and X₂-G-

X₄-[IV]-[ML], P-T-X-W-X-D-A-[LFP]-[DS]-L-V-[AT]-X and X₃-W-X₂-A-[LM]-X₂-Y-X-G or L-Y-F-X-[AS]-[MIL]-[KR]-[VI]-[KR]-[NH] and L-[MIVT]-[FS]-X₃-[QKR]-T-[PQT]-X in the case of group I sequences and group II sequences, respectively (according to the PROSITE pattern nomenclature) (Fig. 4). These motifs were not detected in other AoxB sequences and are therefore specific to each of these groups.

In conclusion, the results of these studies on a mildly arsenic-contaminated environment by using cell culture-dependent and -independent approaches show that the *aoxB* gene diversity is wider than described in previous studies. Comparisons between 78 AoxB bacterial sequences obtained from the Sainte-Marie-aux-Mines creek sedimentary microbial community and homologues previously deposited in public databases confirmed that most of the *aoxB* sequences detected so far belong to two subfamilies (groups I and II) harboring specific signatures. It is worth noting that the pattern of *aoxB* gene evolution was not strictly correlated with organism evolution and that AoxB was not found to be a suitable phylogenetic marker for studying the microbial diversity of arsenic-contaminated environments.

FIG. 3. Unrooted phylogenetic trees of the 97 group II AoxB sequences (260 unambiguously aligned positions). The 97 sequences were retrieved from the nonredundant database of the NCBI, deduced from 18 isolated strains (in bold and underlined) or deduced from the 29 PCR products obtained from metagenomic DNA (in red). Phylogenetic analyses were performed using the maximum likelihood approach implemented in Treefinder (19). The LG +Γ4 model was used for this purpose, as suggested by the “propose model” tool available in Treefinder. Numbers at branches are bootstrap values determined using the nonparametric bootstrap approach implemented in Treefinder (based on 100 replicates of the original data set). The scale bar represents the average number of substitutions estimated per site.

Nucleotide sequence accession numbers. Sequences were submitted to the GenBank database with the following accession numbers: HQ449625 to HQ449646 for 16S rRNA genes, HQ449647 to HQ449668 for *aoxB* from isolates, and HQ449570 to HQ449624 for *aoxB* from metagenomic DNA.

Audrey Cordi was supported by grants of the French Ministry of Research and the Lorraine Region (the Ecotoxicology Platform). Audrey Heinrich-Salmeron and David Halter were supported by grants of the French Ministry of Research and the University of Strasbourg. This study was financed by the EC2CO program (Institut National des Sciences de l'Univers, CNRS). It was performed in the framework of the Groupement de Recherche Métabolisme de l'Arsenic Chez les Micro-organismes: de la Résistance à la Détoxication and supported by the Centre National de la Recherche Scientifique (CNRS) (GDR2909). Céline Brochier-Armanet is supported by the CNRS (Action Thématique et Incitative sur Programme [ATIP]).

REFERENCES

- Anderson, G. L., J. Williams, and R. Hille. 1992. The purification and characterization of arsenite oxidase from *Alcaligenes faecalis*, a molybdenum-containing hydroxylase. *J. Biol. Chem.* **267**:23674–23682.
- Arsène-Ploetze, F., et al. 2010. Structure, function, and evolution of the *Thiomonas* spp. genome. *PLoS Genet.* **6**:e1000859.
- Bruneel, O., R. Duran, C. Casiot, F. Elbaz-Poulichet, and J.-C. Personné. 2006. Diversity of microorganisms in Fe-As-rich acid mine drainage waters of Carnoulès, France. *Appl. Environ. Microbiol.* **72**:551–556.
- Bruneel, O., et al. 2003. Mediation of arsenic oxidation by *Thiomonas* sp. in acid-mine drainage (Carnoulès, France). *J. Appl. Microbiol.* **95**:492–499.
- Bryan, C. G., et al. 2009. Carbon and arsenic metabolism in *Thiomonas* strains: differences revealed diverse adaptation processes. *BMC Microbiol.* **9**:127.
- Cai, L., G. Liu, C. Rensing, and G. Wang. 2009. Genes involved in arsenic transformation and resistance associated with different levels of arsenic-contaminated soils. *BMC Microbiol.* **9**:4.
- Cai, L., C. Rensing, X. Li, and G. Wang. 2009. 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.
- Casiot, C., et al. 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, C., et al. 2003. Bacterial immobilization and oxidation of arsenic in acid mine drainage (Carnoulès Creek, France). *Water Res.* **37**:2929–2936.
- Chang, J.-S., et al. 2010. Arsenic detoxification potential of *aox* genes in arsenite-oxidizing bacteria isolated from natural and constructed wetlands in the Republic of Korea. *Environ. Geochem. Health* **32**:95–105.
- Crooks, G. E., G. Hon, J.-M. Chandonia, and S. E. Brenner. 2004. WebLogo: a sequence logo generator. *Genome Res.* **14**:1188–1190.
- Duquesne, K., et al. 2008. Arsenite oxidation by a chemoautotrophic moderately acidophilic *Thiomonas* sp.: from the strain isolation to the gene study. *Environ. Microbiol.* **10**:228–237.
- Ellis, P. J., T. Conrads, R. Hille, and P. Kuhn. 2001. Crystal structure of the 100 kDa arsenite oxidase from *Alcaligenes faecalis* in two crystal forms at 1.64 Å and 2.03 Å. *Structure* **9**:125–132.
- Escalante, G., et al. 2009. Arsenic resistant bacteria isolated from arsenic contaminated river in the Atacama Desert (Chile). *Bull. Environ. Contam. Toxicol.* **83**:657–661.
- Fan, H., et al. 2008. Sedimentary arsenite-oxidizing and arsenate-reducing bacteria associated with high arsenic groundwater from Shanyin, northwestern China. *J. Appl. Microbiol.* **105**:529–539.
- Halter, D., et al. Taxonomic and functional prokaryote diversity in mildly arsenic contaminated sediments. *Res. Microbiol.*, in press.
- Hamamura, N., et al. 2009. Linking microbial oxidation of arsenic with detection and phylogenetic analysis of arsenite oxidase genes in diverse geothermal environments. *Environ. Microbiol.* **11**:421–431.
- Hoefl, S. E., et al. 2007. *Alkalilimnicola ehrlichii* sp. nov., a novel, arsenite-oxidizing haloalkaliphilic gammaproteobacterium capable of chemoautotrophic or heterotrophic growth with nitrate or oxygen as the electron acceptor. *Int. J. Syst. Evol. Microbiol.* **57**:504–512.
- Inskip, W. P., et al. 2007. Detection, diversity and expression of aerobic bacterial arsenite oxidase genes. *Environ. Microbiol.* **9**:934–943.
- Jobb, G., A. von Haeseler, and K. Strimmer. 2004. TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. *BMC Evol. Biol.* **4**:18.
- Lieutaud, A., et al. 2010. Arsenite oxidase from *Ralstonia* sp. 22. *J. Biol. Chem.* **285**:20433–20441.
- Lièvremon, D., P. N. Bertin, and M.-C. Lett. 2009. Arsenic in contaminated waters: biogeochemical cycle, microbial metabolism and biotreatment processes. *Biochimie* **91**:1229–1237.
- Mandal, B. K., and K. T. Suzuki. 2002. Arsenic round the world: a review. *Talanta* **58**:201–235.
- Muller, D., D. Lièvremon, D. D. Simeonova, J.-C. Hubert, and M.-C. Lett. 2003. Arsenite oxidase *aox* genes from a metal-resistant beta-proteobacterium. *J. Bacteriol.* **185**:135–141.
- Muller, D., et al. 2007. A tale of two oxidation states: bacterial colonization of arsenic-rich environments. *PLoS Genet.* **3**:e53.
- Oremland, R. S., C. W. Saltikov, F. Wolfe-Simon, and J. F. Stolz. 2009. Arsenic in the evolution of earth and extraterrestrial ecosystems. *Geomicrobiol. J.* **26**:522.
- Oremland, R. S., and J. F. Stolz. 2005. Arsenic, microbes and contaminated aquifers. *Trends Microbiol.* **13**:45–49.
- Oremland, R. S., and J. F. Stolz. 2003. The ecology of arsenic. *Science* **300**:939–944.
- Quéméneur, M., et al. 2010. Population structure and abundance of arsenite-oxidizing bacteria along an arsenic pollution gradient in waters of the Upper Isle River Basin, France. *Appl. Environ. Microbiol.* **76**:4566–4570.
- Quéméneur, M., et al. 2008. Diversity surveys and evolutionary relationships of *aoxB* genes in aerobic arsenite-oxidizing bacteria. *Appl. Environ. Microbiol.* **74**:4567–4573.
- Rhine, E. D., S. M. Ní Chadhain, G. J. Zylstra, and L. Y. Young. 2007. The arsenite oxidase genes (*araAB*) in novel chemoautotrophic arsenite oxidizers. *Biochem. Biophys. Res. Commun.* **354**:662–667.
- Richey, C., et al. 2009. Respiratory arsenate reductase as a bidirectional enzyme. *Biochem. Biophys. Res. Commun.* **382**:298–302.
- Santini, J. M., and R. N. vanden Hoven. 2004. Molybdenum-containing arsenite oxidase of the chemolithoautotrophic arsenite oxidizer NT-26. *J. Bacteriol.* **186**:1614–1619.
- Schneider, T. D., and R. M. Stephens. 1990. Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res.* **18**:6097–6100.
- Sharma, V. K., and M. Sohn. 2009. Aquatic arsenic: toxicity, speciation, transformations, and remediation. *Environ. Int.* **35**:743–759.
- Simeonova, D. D., et al. 2004. Microplate screening assay for the detection of arsenite-oxidizing and arsenate-reducing bacteria. *FEMS Microbiol. Lett.* **237**:249–253.
- Tamaki, H., et al. 2005. Comparative analysis of bacterial diversity in freshwater sediment of a shallow eutrophic lake by molecular and improved cultivation-based techniques. *Appl. Environ. Microbiol.* **71**:2162–2169.
- Valenzuela, C., V. L. Campos, J. Yañez, C. A. Zaror, and M. A. Mondaca. 2009. Isolation of arsenite-oxidizing bacteria from arsenic-enriched sediments from Camarones River, northern Chile. *Bull. Environ. Contam. Toxicol.* **82**:593–596.
- Vallenet, D., et al. 2009. MicroScope: a platform for microbial genome annotation and comparative genomics. *Database (Oxford)* **2009**:bap021.
- Weeger, W., et al. 1999. Oxidation of arsenite to arsenate by a bacterium isolated from an aquatic environment. *Biomaterials* **12**:141–149.
- Zargar, K., S. Hoefl, R. Oremland, and C. W. Saltikov. 2010. Identification of a novel arsenite oxidase gene, *arcA*, in the haloalkaliphilic, arsenite-oxidizing bacterium *Alkalilimnicola ehrlichii* strain MLHE-1. *J. Bacteriol.* **192**:3755–3762.

RÉFÉRENCES BIBLIOGRAPHIQUES

Abedin, M. J., Feldmann, J. & Meharg, A. A. (2002). Uptake kinetics of arsenic species in rice plants. *Plant Physiology* **128**, 1120-1128.

Ahmad, S., Kitchin, K. T. & Cullen, W. R. (2000). Arsenic species that cause release of iron from ferritin and generation of activated oxygen. *Archives of Biochemistry and Biophysics* **382**, 195-202.

Akter, K. F., Owens, G., Davey, D. E. & Naidu, R. (2006). Arsenic speciation and toxicity in biological systems. In *Reviews of Environmental Contamination and Toxicology*, pp. 97-149.

Amaral Zettler, L. A., Gomez, F., Zettler, E., Keenan, B. G., Amils, R. & Sogin, M. L. (2002). Microbiology: eukaryotic diversity in Spain's River of Fire. *Nature* **417**, 137.

Aposhian, H. V. (1997). Enzymatic methylation of arsenic species and other new approaches to arsenic toxicity. In *Annual Review of Pharmacology and Toxicology*, pp. 397-419.

Arsène-Ploetze, F., Koechler, S., Marchal, M. & other authors (2010). Structure, function, and evolution of the *Thiomonas* spp. genome. *PLoS Genetics* **6**.

Baffico, G. D., Diaz, M. M., Wenzel, M. T., Koschorreck, M., Schimmele, M., Neu, T. R. & Pedrozo, F. (2004). Community structure and photosynthetic activity of epilithon from a highly acidic (pH = 2) mountain stream in Patagonia, Argentina. *Extremophiles* **8**, 463-473.

Baker, B. J. & Banfield, J. F. (2003). Microbial communities in acid mine drainage. *FEMS Microbiology Ecology* **44**, 139-152.

Baker, B. J., Tyson, G. W., Goosherst, L. & Banfield, J. F. (2009). Insights into the diversity of eukaryotes in acid mine drainage biofilm communities. *Applied and Environmental Microbiology* **75**, 2192-2199.

Beech, I. B. & Sunner, J. (2004). Biocorrosion: Towards understanding interactions between biofilms and metals. *Current Opinion in Biotechnology* **15**, 181-186.

Berhe, A., Fristedt, U. & Persson, B. L. (1995). Expression and purification of the high-affinity phosphate transporter of *Saccharomyces cerevisiae*. *European Journal of Biochemistry* **227**, 566-572.

Bienert, G. P. & Jahn, T. P. (2010). Major intrinsic proteins and arsenic transport in plants: new players and their potential role. *Advances in experimental medicine and biology* **679**, 111-125.

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

Brake, S. S., Dannelly, H. K., Connors, K. A. & Hasiotis, S. T. (2001b). 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. *Applied Geochemistry* **16**, 1641-1652.

Bruneel, O., Personné, J. C., Casiot, C., Leblanc, M., Elbaz-Poulichet, F., Mahler, B. J., Le Flèche, A. & Grimont, P. A. D. (2003). Mediation of arsenic oxidation by *Thiomonas* sp. in acid-mine drainage (Carnoulès, France). *Journal of Applied Microbiology* **95**, 492-499.

Bruneel, O., 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. *Applied and Environmental Microbiology* **72**, 551-556.

Bruneel, O., Pascault, N., Egal, M., Bancon-Montigny, C., Goni-Urriza, M. S., Elbaz-Poulichet, F., Personné, J. C. & Duran, R. (2008). Archaeal diversity in a Fe-As rich acid mine drainage at Carnoulès (France). *Extremophiles* **12**, 563-571.

Bustos, D. M., Bustamante, C. A. & Iglesias, A. A. (2008). Involvement of non-phosphorylating glyceraldehyde-3-phosphate dehydrogenase in response to oxidative stress. *Journal of Plant Physiology* **165**, 456-461.

Caille, N., Zhao, F. J. & McGrath, S. P. (2005). Comparison of root absorption, translocation and tolerance of arsenic in the hyperaccumulator *Pteris vittata* and the nonhyperaccumulator *Pteris tremula*. *New Phytologist* **165**, 755-761.

Calatayud, M., Gimeno, J., Vélez, D., Devesa, V. & Montoro, R. (2011). Characterization of the intestinal absorption of arsenate, monomethylarsonic acid, and dimethylarsinic acid using the caco-2 cell line. *Chemical Research in Toxicology* **23**, 547-556.

Casiot, C., Morin, G., Juillot, F., Bruneel, O., Personné, J. C., Leblanc, M., Duquesne, K., Bonnefoy, V. & Elbaz-Poulichet, F. (2003). Bacterial immobilization and oxidation of arsenic in acid mine drainage (Carnoulès creek, France). *Water Research* **37**, 2929-2936.

Casiot, C., Bruneel, O., Personné, J. C., Leblanc, M. & Elbaz-Poulichet, F. (2004). Arsenic oxidation and bioaccumulation by the acidophilic protozoan, *Euglena mutabilis*, in acid mine drainage (Carnoulès, France). *Science of the Total Environment* **320**, 259-267.

Cheng, H., Hu, Y., Luo, J., Xu, B. & Zhao, J. (2009). Geochemical processes controlling fate and transport of arsenic in acid mine drainage (AMD) and natural systems. *Journal of Hazardous Materials* **165**, 13-26.

Christensen, T. (1964). The gross classification of algae. *Alga and Man*, 59-64.

Cobbett, C. & Goldsbrough, P. (2002). Phytochelatin and metallothionein: Roles in heavy metal detoxification and homeostasis. In *Annual Review of Plant Biology*, pp. 159-182.

Contreras, L., Moenne, A., Gaillard, F., Potin, P. & Correa, J. A. (2009). Proteomic analysis and identification of copper stress-regulated proteins in the marine alga *Scytosiphon gracilis* (Phaeophyceae). *Aquatic Toxicology*.

Costerton, J. W., Lewandowski, Z., Caldwell, D. E., Korber, D. R. & Lappin-Scott, H. M. (1995). Microbial biofilms. *Annual Review of Microbiology* **49**, 711-745.

Cullen, W. R. & Bentley, R. (2005). The toxicity of trimethylarsine: An urban myth. *Journal of Environmental Monitoring* **7**, 11-15.

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

Das, B. K., Roy, A., Singh, S. & Bhattacharya, J. (2009b). Eukaryotes in acidic Mine Drainage environments: Potential applications in bioremediation. *Reviews in Environmental Science and Biotechnology* **8**, 257-274.

Davis, T. A., Llanes, F., Volesky, B. & Mucci, A. (2003). Metal selectivity of *Sargassum* spp. and their alginates in relation to their L-guluronic acid content and conformation. *Environmental Science and Technology* **37**, 261-267.

Daviss, B. (2005). Growing pains for metabolomics. *Scientist* **19**, 25-28.

Devars, S., Avilés, C., Cervantes, C. & Moreno-Sanchez, R. (2000). Mercury uptake and removal by *Euglena gracilis*. *Archives of Microbiology* **174**, 175-180.

Diels, L., Van Roy, S., Taghavi, S. & Van Houdt, R. (2009). From industrial sites to environmental applications with *Cupriavidus metallidurans*. *Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology* **96**, 247-258.

Domon, B. & Aebersold, R. (2006). Mass spectrometry and protein analysis. *Science* **312**, 212-217.

Douglas, G. E., John, D. M., Williamson, D. B. & Reid, G. (1998). The aquatic algae associated with mining areas in Peninsula Malaysia and Sarawak: Their composition, diversity and distribution. *Nova Hedwigia* **67**, 189-211.

Douglas, S. & Beveridge, T. J. (1998). Mineral formation by bacteria in natural microbial communities. *FEMS Microbiology Ecology* **26**, 79-88.

Draguet, K. I., Smidsrod, O. & Skjak-break, G. (2005). Aginate from algae. Edited by P. a. p. Polysaccharides and polyamides in the Food industrie: Properties: Steinbuchel.

Drobna, Z., Walton, F. S., Paul, D. S., Xing, W., Thomas, D. J. & Styblo, M. (2010). Metabolism of arsenic in human liver: The role of membrane transporters. *Archives of Toxicology* **84**, 3-16.

Dupraz, C. & Visscher, P. T. (2005). Microbial lithification in marine stromatolites and hypersaline mats. *Trends in Microbiology* **13**, 429-438.

Duquesne, K., Lebrun, S., Casiot, C., Bruneel, O., Personné, J. C., Leblanc, M., Elbaz-Poulichet, F., Morin, G. & Bonnefoy, V. (2003). Immobilization of Arsenite and Ferric Iron by *Acidithiobacillus ferrooxidans* and Its Relevance to Acid Mine Drainage. *Applied and Environmental Microbiology* **69**, 6165-6173.

Fahey, R. C., Brown, W. C., Adams, W. B. & Worsham, M. B. (1978). Occurrence of glutathione in bacteria. *Journal of Bacteriology* **133**, 1126-1129.

Flora, S. J. S. (2011). Arsenic-induced oxidative stress and its reversibility. *Free Radical Biology and Medicine* **51**, 257-281.

Foresi, N., Correa-Aragunde, N., Parisi, G., Calo, G., Salerno, G. & Lamattina, L. (2010). Characterization of a nitric oxide synthase from the plant kingdom: NO generation from the green alga *Ostreococcus tauri* is light irradiance and growth phase dependent. *Plant Cell* **22**, 3816-3830.

Foster, S., Thomson, D. & Maher, W. (2008). Uptake and metabolism of arsenate by anoxic cultures of the microalgae *Dunaliella tertiolecta* and *Phaeodactylum tricornutum*. *Marine Chemistry* **108**, 172-183.

Garcia-Meza, J. V., Barrangue, C. & Admiraal, W. (2005). Biofilm formation by algae as a mechanism for surviving on mine tailings. *Environmental Toxicology and Chemistry* **24**, 573-581.

Ghosh, M., Shen, J. & Rosen, B. P. (1999). Pathways of As(III) detoxification in *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 5001-5006.

Gibbs, S., P. (1978). The chloroplasts of *Euglena* may have evolved from symbiotic green algae. *Canadian Journal of Botany* **56**, 2883-2889.

Goffeau, A., Barrell, G., Bussey, H. & other authors (1996). Life with 6000 genes. *Science* **274**, 546-567.

Gomes, D., Agasse, A., Thiébaud, P., Delrot, S., Geros, H. & Chaumont, F. (2009). Aquaporins are multifunctional water and solute transporters highly divergent in living organisms. *Biochimica et Biophysica Acta - Biomembranes* **1788**, 1213-1228.

Gonzalez-Mendoza, D., Morales-Trejo, A. & Brito-Vera, H. (2009). A rapid and efficient method for isolation of total RNA from *Euglena gracilis* (Euglenoidea). *Genetics and Molecular Research* **8**, 482-486.

Grispen, V. M. J., Irtelli, B., Hakvoort, H. W. J., Vooijs, R., Bliet, T., ten Bookum, W. M., Verkleij, J. A. C. & Schat, H. (2009). Expression of the *Arabidopsis* metallothionein 2b enhances arsenite sensitivity and root to shoot translocation in tobacco. *Environmental and Experimental Botany* **66**, 69-73.

Groen, A. J. & Lilley, K. S. (2010). Proteomics of total membranes and subcellular membranes. *Expert Review of Proteomics* **7**, 867-878.

Guedry, O., Lazard, M., Delort, F., Dauplais, M., Grigoras, I., Blanquet, S. & Plateau, P. (2003). Ycf1p-dependent Hg(II) detoxification in *Saccharomyces cerevisiae*. *European Journal of Biochemistry* **270**, 2486-2496.

Gumaelius, L., Lahner, B., Salt, D. E. & Banks, J. A. (2004). Arsenic hyperaccumulation in gametophytes of *Pteris vittata*. A new model system for analysis of arsenic hyperaccumulation. *Plant Physiology* **136**, 3198-3208.

Hader, D. P. & Melkonian, M. (1983). Phototaxis in the gliding flagellate, *Euglena mutabilis*. *Archives of Microbiology* **135**, 25-29.

Halliwell, B. & Gutteridge, J. M. C. (1984). Oxygen toxicity, oxygen radicals, transition metals and disease. *Biochemical Journal* **219**, 1-14.

Handelsman, J., Rondon, M. R., Brady, S. F., Clardy, J. & Goodman, R. M. (1998). Molecular biological access to the chemistry of unknown soil microbes: A new frontier for natural products. *Chemistry and Biology* **5**, R245-R249.

Harrison, J. J., Ceri, H. & Turner, R. J. (2007). Multimetal resistance and tolerance in microbial biofilms. *Nature Reviews Microbiology* **5**, 928-938.

Hassinen, V. H., Tervahauta, A. I., Schat, H. & Kärenlampi, S. O. (2011). Plant metallothioneins - metal chelators with ROS scavenging activity? *Plant Biology* **13**, 225-232.

Hegde, P. S., White, I. R. & Debouck, C. (2003). Interplay of transcriptomics and proteomics. *Current Opinion in Biotechnology* **14**, 647-651.

Heywood, J. L., Sieracki, M. E., Bellows, W., Poulton, N. J. & Stepanauskas, R. (2010). Capturing diversity of marine heterotrophic protists: One cell at a time. *ISME Journal* **5**, 674-684.

Hoffman, R. D. & Lane, M. D. (1992). Iodophenylarsine oxide and arsenical affinity chromatography: New probes for dithiol proteins. Application to tubulins and to components of the insulin receptor-glucose transporter signal transduction pathway. *Journal of Biological Chemistry* **267**, 14005-14011.

Huang, B., Whitchurch, C. B. & Mattick, J. S. (2003). FimX, a multidomain protein connecting environmental signals to twitching motility in *Pseudomonas aeruginosa*. *J Bacteriol* **185**, 7068-7076.

Jensen, L. T., Ajua-Alemanji, M. & Culotta, V. C. (2003). The *Saccharomyces cerevisiae* High Affinity Phosphate Transporter Encoded by PHO84 Also Functions in Manganese Homeostasis. *Journal of Biological Chemistry* **278**, 42036-42040.

Ji, G. & Silver, S. (1992). Reduction of arsenate to arsenite by the arsC protein of the arsenic resistance operon of *Staphylococcus aureus* plasmid pI258. *Proceedings of the National Academy of Sciences of the United States of America* **89**, 9474-9478.

Jian, F. M., Yamaji, N., Mitani, N., Xu, X. Y., Su, Y. H., McGrath, S. P. & Zhao, F. J. (2008). Transporters of arsenite in rice and their role in arsenic accumulation in rice grain. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 9931-9935.

Jiang, X., McDermott, J. R., Abdul Ajees, A., Rosena, B. P. & Liu, Z. (2010). Trivalent arsenicals and glucose use different translocation pathways in mammalian GLUT1. *Metallomics* **2**, 211-219.

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

Jing, Y., Dai, J., Chalmers-Redman, R. M. E., Tatton, W. G. & Waxman, S. (1999). Arsenic trioxide selectively induces acute promyelocytic leukemia cell apoptosis via a hydrogen peroxide-dependent pathway. *Blood* **94**, 2102-2111.

Jo, W. J., Loguinov, A., Wintz, H., Chang, M., Smith, A. H., Kalman, D., Zhang, L., Smith, M. T. & Vulpe, C. D. (2009). Comparative functional genomic analysis identifies distinct and overlapping sets of genes required for resistance to monomethylarsonous acid (MMAIII) and arsenite (AsIII) in yeast. *Toxicological Sciences* **111**, 424-436.

Johnson, D. B. & Rang, L. (1993). Effects of acidophilic protozoa on populations of metal-mobilizing bacteria during the leaching of pyritic coal. *Journal of General Microbiology* **139**, 1417-1423.

Johnson, D. B. & Hallberg, K. B. (2003). The microbiology of acidic mine waters. *Research in Microbiology* **154**, 466-473.

Johnson, D. B. & Hallberg, K. B. (2005). Acid mine drainage remediation options: A review. *Science of the Total Environment* **338**, 3-14.

Kaddurah-Daouk, R., Kristal, B. S. & Weinshilboum, R. M. (2008).Metabolomics: A global biochemical approach to drug response and disease. In *Annual Review of Pharmacology and Toxicology*, pp. 653-683.

Kageyama, Y., Ishibashi, K., Hayashi, T., Xia, G., Sasaki, S. & Kihara, K. (2001). Expression of aquaporins 7 and 8 in the developing rat testis. *Andrologia* **33**, 165-169.

Kamjunke, N., Gaedke, U., Tittel, J., Weithoff, G. & Bell, E. M. (2004). Strong vertical differences in the plankton composition of an extremely acidic lake. *Archiv fur Hydrobiologie* **161**, 289-306.

Kapfer, M. (1998). Assessment of the colonization and primary production of microphytobenthos in the littoral of acidic mining lakes in Lusatia (Germany). *Water, Air, and Soil Pollution* **108**, 331-340.

Kertulis, G. M., Ma, L. Q., MacDonald, G. E., Chen, R., Winefordner, J. D. & Cai, Y. (2005). Arsenic speciation and transport in *Pteris vittata* L. and the effects on phosphorus in the xylem sap. *Environmental and Experimental Botany* **54**, 239-247.

Kitchin, K. T. & Wallace, K. (2008). The role of protein binding of trivalent arsenicals in arsenic carcinogenesis and toxicity. *Journal of Inorganic Biochemistry* **102**, 532-539.

Kligerman, A. D. & Tennant, A. H. (2007). Insights into the carcinogenic mode of action of arsenic. *Toxicology and Applied Pharmacology* **222**, 281-288.

Kneer, R. & Zenk, M. H. (1997). The formation of Cd-phytochelatin complexes in plant cell cultures. *Phytochemistry* **44**, 69-74.

Kobayashi, I., Fujiwara, S., Shimogawara, K., Kaise, T., Usuda, H. & Tsuzuki, M. (2003). Insertional mutagenesis in a homologue of a Pi transporter gene confers arsenate resistance on *Chlamydomonas*. *Plant and Cell Physiology* **44**, 597-606.

Koehler, S., Cleiss-Arnold, J., Proux, C. & other authors (2010). Multiple controls affect arsenite oxidase gene expression in *Herminiimonas arsenicoxydans*. *BMC Microbiology* **10**.

Krause, F. (2006). Detection and analysis of protein-protein interactions in organellar and prokaryotic proteomes by native gel electrophoresis: (Membrane) protein complexes and supercomplexes. *Electrophoresis* **27**, 2759-2781.

Kuroda, M., Dey, S., Sanders, O. I. & Rosen, B. P. (1997). Alternate energy coupling of ArsB, the membrane subunit of the Ars anion-translocating ATPase. *Journal of Biological Chemistry* **272**, 326-331.

Lafaye, A., Junot, C., Pereira, Y., Lagniel, G., Tabet, J. C., Ezan, E. & Labarre, J. (2005). Combined proteome and metabolite-profiling analyses reveal surprising insights into yeast sulfur metabolism. *Journal of Biological Chemistry* **280**, 24723-24730.

Leander, B. S., Esson, H. J. & Breglia, S. A. (2007). Macroevolution of complex cytoskeletal systems in euglenids. *BioEssays* **29**, 987-1000.

Leblanc, M., Achard, B., Othman, D. B., Luck, J. M., Bertrand-Sarfati, J. & Personné, J. C. (1996). Accumulation of arsenic from acidic mine waters by ferruginous bacterial accretions (stromatolites). *Applied Geochemistry* **11**, 541-554.

LeBlanc, M., Casiot, C., Elbaz-Poulichet, F. & Personné, C. (2002). Arsenic removal by oxidizing bacteria in a heavily arsenic-contaminated acid mine drainage system (Carnoulès, France). In *Geological Society Special Publication*, pp. 267-274.

Lee, J., Shim, D., Song, W. Y., Hwang, I. & Lee, Y. (2004). *Arabidopsis* metallothioneins 2a and 3 enhance resistance to cadmium when expressed in *Vicia faba* guard cells. *Plant Molecular Biology* **54**, 805-815.

Lehr, C. R., Polishchuk, E., Radoja, U. & Cullen, W. R. (2003). Demethylation of methylarsenic species by *Mycobacterium neoaurum*. *Applied Organometallic Chemistry* **17**, 831-834.

Lessmann, D., Deneke, R., Ender, R., Hemm, M., Kapfer, M., Krumbeck, H., Wollmann, K. & Nixdorf, B. (1999). Lake Plessa 107 (Lusatia, Germany) - An extremely acidic shallow mining lake. *Hydrobiologia* **408-409**, 293-299.

Lessmann, D., Fyson, A. & Nixdorf, B. (2000). Phytoplankton of the extremely acidic mining lakes of Lusatia (Germany) with pH = 3. *Hydrobiologia* **433**, 123-128.

Li, R. Y., Ago, Y., Liu, W. J., Mitani, N., Feldmann, J., McGrath, S. P., Ma, J. F. & Zhao, F. J. (2009). The rice aquaporin Isi1 mediates uptake of methylated arsenic species. *Plant Physiology* **150**, 2071-2080.

Li, Z. S., Lu, Y. P., Zhen, R. G., Szczypka, M., Thiele, D. J. & Rea, P. A. (1997). A new pathway for vacuolar cadmium sequestration in *Saccharomyces cerevisiae*: YCF1-catalyzed transport of bis(glutathionato)cadmium. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 42-47.

Lin, Y. F., Walmsley, A. R. & Rosen, B. P. (2006). An arsenic metallochaperone for an arsenic detoxification pump. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 15617-15622.

Liu, S. X., Athar, M., Lippai, I., Waldren, C. & Hei, T. K. (2001). Induction of oxyradicals by arsenic: Implication for mechanism of genotoxicity. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 1643-1648.

Liu, W. J., Wood, B. A., Raab, A., McGrath, S. P., Zhao, F. J. & Feldmann, J. (2010). Complexation of arsenite with phytochelatins reduces arsenite efflux and translocation from roots to shoots in *Arabidopsis*. *Plant Physiology* **152**, 2211-2221.

Liu, Z., Shen, J., Carbrey, J. M., Mukhopadhyay, R., Agre, P. & Rosen, B. P. (2002). Arsenite transport by mammalian aquaglyceroporins AQP7 and AQP9. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 6053-6058.

Liu, Z., Boles, E. & Rosen, B. P. (2004). Arsenic Trioxide Uptake by Hexose Permeases in *Saccharomyces cerevisiae*. *Journal of Biological Chemistry* **279**, 17312-17318.

Liu, Z., Sanchez, M. A., Jiang, X., Boles, E., Landfear, S. M. & Rosen, B. P. (2006). Mammalian glucose permease GLUT1 facilitates transport of arsenic trioxide and methylarsonous acid. *Biochemical and Biophysical Research Communications* **351**, 424-430.

Lopez-Archilla, A. I., Marin, I. & Amils, R. (2001). Microbial community composition and ecology of an acidic aquatic environment: The Tinto River, Spain. *Microbial Ecology* **41**, 20-35.

Maciaszczyk-Dziubinska, E., Migdal, I., Migocka, M., Bocer, T. & Wysocki, R. (2010). The yeast aquaglyceroporin Fps1p is a bidirectional arsenite channel. *FEBS Letters* **584**, 726-732.

Mandal, B. K. & Suzuki, K. T. (2002). Arsenic round the world: A review. *Talanta* **58**, 201-235.

Mandl, M., Matulova, P. & Docekalova, H. (1992). Migration of arsenic(III) during bacterial oxidation of arsenopyrite in chalcopyrite concentrate by *Thiobacillus ferrooxidans*. *Applied Microbiology and Biotechnology* **38**, 429-431.

Maret, W. (2004). Zinc and Sulfur: A Critical Biological Partnership. *Biochemistry* **43**, 3301-3309.

Marin, B., Palm, A., Klingberg, M. & Melkonian, M. (2003). Phylogeny and taxonomic revision of plastid-containing euglenophytes based on SSU rDNA sequence comparisons and synapomorphic signatures in the SSU rRNA secondary structure. *Protist* **154**, 99-145.

McGinness, S. & Johnson, D. B. (1992). Grazing of acidophilic bacteria by a flagellated protozoan. *Microbial Ecology* **23**, 75-86.

Meharg, A. A., Naylor, J. & Macnair, M. R. (1994). Phosphorus nutrition of arsenate-tolerant and nontolerant phenotypes of velvetgrass. *Journal of Environmental Quality* **23**, 234-238.

Meharg, A. A. & Jardine, L. (2003). Arsenite transport into paddy rice (*Oryza sativa*) roots. *New Phytologist* **157**, 39-44.

Meister, A. & Tate, S. S. (1976). Glutathione and related gamma-glutamyl compounds: biosynthesis and utilization. *Annual Review of Biochemistry* **45**, 559-604.

Mendoza-Cozatl, D., Devars, S., Loza-Tavera, H. & Moreno-Sanchez, R. (2002). Cadmium accumulation in the chloroplast of *Euglena gracilis*. *Physiologia Plantarum* **115**, 276-283.

Mendoza-Cozatl, D. G. & Moreno-Sanchez, R. (2005). Cd²⁺ transport and storage in the chloroplast of *Euglena gracilis*. *Biochimica et Biophysica Acta - Bioenergetics* **1706**, 88-97.

Menezes, R. A., Amaral, C., Batista-Nascimento, L., Santos, C., Ferreira, R. B., Devaux, F., Eleutherio, E. C. A. & Rodrigues-Pousada, C. (2008). Contribution of Yap1 towards *Saccharomyces cerevisiae* adaptation to arsenic-mediated oxidative stress. *Biochemical Journal* **414**, 301-311.

Meng, Y. L., Liu, Z. & Rosen, B. P. (2004). As(III) and Sb(III) Uptake by GlpF and Efflux by ArsB in *Escherichia coli*. *Journal of Biological Chemistry* **279**, 18334-18341.

Metzker, M. L. (2010). Sequencing technologies the next generation. *Nature Reviews Genetics* **11**, 31-46.

Meyer, F., Paarmann, D., D'Souza, M. & other authors (2008). The metagenomics RAST server - A public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* **9**.

Miot, J., Morin, G., Skouri-Panet, F. & other authors (2008). XAS study of arsenic coordination in *Euglena gracilis* exposed to arsenite. *Environmental Science and Technology* **42**, 5342-5347.

Miot, J., Morin, G., Skouri-Panet, F. & other authors (2009). Speciation of arsenic in *Euglena gracilis* cells exposed to as(V). *Environmental Science and Technology* **43**, 3315-3321.

Morin, G., Juillot, F., Casiot, C., Bruneel, O., Personne, J. C., Elbaz-Poulichet, F., Leblanc, M., Ildefonse, P. & Calas, G. (2003). Bacterial formation of tooeleite and mixed arsenic(III) or arsenic(V)-iron(III) gels in the Carnoules acid mine drainage, France. A XANES, XRD, and SEM study. *Environ Sci Technol* **37**, 1705-1712.

Mukhopadhyay, R. & Rosen, B. P. (2002). Arsenate reductases in prokaryotes and eukaryotes. *Environmental Health Perspectives* **110**, 745-748.

Nagel, K., Adelmeier, U. & Voigt, J. (1996). Subcellular distribution of cadmium in the unicellular green alga *Chlamydomonas reinhardtii*. *Journal of Plant Physiology* **149**, 86-90.

Nakatsu, C. & Hutchinson, T. C. (1988). Extreme metal and acid tolerance of *Euglena mutabilis* and an associated yeast from Smoking Hills, Northwest Territories, and their apparent mutualism. *Microbial Ecology* **16**, 213-231.

Navarro, L., Torres-Marquez, M. E., Gonzalez-Moreno, S., Devars, S., Hernandez, R. & Moreno-Sanchez, R. (1997). Comparison of physiological changes in *Euglena gracilis* during exposure to heavy metals of heterotrophic and autotrophic cells. *Comparative Biochemistry and Physiology - C Pharmacology Toxicology and Endocrinology* **116**, 265-272.

Nicchia, G. P., Frigeri, A., Nico, B., Ribatti, D. & Svelto, M. (2001). Tissue distribution and membrane localization of aquaporin-9 water channel: Evidence for sex-linked differences in liver. *Journal of Histochemistry and Cytochemistry* **49**, 1547-1556.

Novis, P. & Harding, J. (2007). Extreme acidophiles :Freshwater Algae Associated with Acid Mine Drainage. *Extreme acidophiles*.

Oehler, S. (2011). Comment on "A bacterium that can grow by using arsenic instead of phosphorus". *Science* **332**.

Olaveson, M. M. & Stokes, P. M. (1989). Responses of the acidophilic alga *Euglena mutabilis* (Euglenophyceae) to carbon enrichment at pH 3. *Journal of Phycology* **25**, 529-539.

Oliver, S. G., Winson, M. K., Kell, D. B. & Baganz, F. (1998). Systematic functional analysis of the yeast genome. *Trends in Biotechnology* **16**, 373-378.

Ostling, O. & Johanson, K. J. (1984). Microelectrophoretic study of radiation-induced DNA damages in individual mammalian cells. *Biochemical and Biophysical Research Communications* **123**, 291-298.

Pawlik-Skowronska, B. (2001). Phytochelatin production in freshwater algae *Stigeoclonium* in response to heavy metals contained in mining water; effects of some environmental factors. *Aquatic Toxicology* **52**, 241-249.

Pawlik-Skowronska, B., Pirszel, J., Kalinowska, R. & Skowronski, T. (2004). Arsenic availability, toxicity and direct role of GSH and phytochelatins in As detoxification in the green alga *Stichococcus bacillaris*. *Aquatic Toxicology* **70**, 201-212.

Perales-Vela, H. V., Pena-Castro, J. M. & Canizares-Villanueva, R. O. (2006). Heavy metal detoxification in eukaryotic microalgae. *Chemosphere* **64**, 1-10.

Persson, B. L., Petersson, J., Fristedt, U., Weinander, R., Berhe, A. & Pattison, J. (1999). Phosphate permeases of *Saccharomyces cerevisiae*: Structure, function and regulation. *Biochimica et Biophysica Acta - Reviews on Biomembranes* **1422**, 255-272.

Petrick, J. S., Ayala-Fierro, F., Cullen, W. R., Carter, D. E. & Vasken Aposhian, H. (2000). Monomethylarsonous acid (MMA(III)) is more toxic than arsenite in Chang human hepatocytes. *Toxicology and Applied Pharmacology* **163**, 203-207.

Pickering, I. J., Gumaelius, L., Harris, H. H., Prince, R. C., Hirsch, G., Banks, J. A., Salt, D. E. & George, G. N. (2006). Localizing the biochemical transformations of arsenate in a hyperaccumulating fern. *Environmental Science and Technology* **40**, 5010-5014.

Qin, J., Lehr, C. R., Yuan, C., Le, X. C., McDermott, T. R. & Rosen, B. P. (2009). Biotransformation of arsenic by a yellowstone thermoacidophilic eukaryotic alga. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 5213-5217.

Quéméneur, M., Heinrich-Salmeron, A., Muller, D., Lièvremon, D., Jauzein, M., Bertin, P. N., Garrido, F. & Joulain, C. (2008). Diversity surveys and evolutionary relationships of *aoxB* genes in aerobic arsenite-oxidizing bacteria. *Applied and Environmental Microbiology* **74**, 4567-4573.

Raab, A., Schat, H., Meharg, A. A. & Feldmann, J. (2005). Uptake, translocation and transformation of arsenate and arsenite in sunflower (*Helianthus annuus*): Formation of arsenic-phytochelatin complexes during exposure to high arsenic concentrations. *New Phytologist* **168**, 551-558.

Raab, A., Williams, P. N., Meharg, A. & Feldmann, J. (2007). Uptake and translocation of inorganic and methylated arsenic species by plants. *Environmental Chemistry* **4**, 197-203.

Ramadan, D., Rancy, P. C., Nagarkar, R. P., Schneider, J. P. & Thorpe, C. (2009). Arsenic(III) species inhibit oxidative protein folding in Vitro. *Biochemistry* **48**, 424-432.

Redfield, R. J. (2011). Comment on "A bacterium that can grow by using arsenic instead of phosphorus". *Science* **332**.

Rodriguez-Zavala, J. S., Garcia-Garcia, J. D., Ortiz-Cruz, M. A. & Moreno-Sanchez, R. (2007). Molecular mechanisms of resistance to heavy metals in the protist *Euglena gracilis*. *Journal of Environmental Science and Health - Part A Toxic/Hazardous Substances and Environmental Engineering* **42**, 1365-1378.

Rose, P. D., Boshoff, G. A., Van Hille, R. P., Wallace, L. C. M., Dunn, K. M. & Duncan, J. R. (1998). An integrated algal sulphate reducing high rate ponding process for the treatment of acid mine drainage wastewaters. *Biodegradation* **9**, 247-257.

Rosen, B. P. (2002). Transport and detoxification systems for transition metals, heavy metals and metalloids in eukaryotic and prokaryotic microbes. *Comparative Biochemistry and Physiology - A Molecular and Integrative Physiology* **133**, 689-693.

Rousch, J. M. & Sommerfeld, M. R. (1999). Effect of manganese and nickel on growth of selected algae in pH buffered medium. *Water Research* **33**, 2448-2454.

Saltikov, C. W. & Newman, D. K. (2003). Genetic identification of a respiratory arsenate reductase. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 10983-10988.

Samikkannu, T., Chen, C. H., Yih, L. H., Wang, A. S. S., Lin, S. Y., Chen, T. C. & Jan, K. Y. (2003). Reactive oxygen species are involved in arsenic trioxide inhibition of pyruvate dehydrogenase activity. *Chemical Research in Toxicology* **16**, 409-414.

Schat, H., Llugany, M., Vooijs, R., Hartley-Whitaker, J. & Bleeker, P. M. (2002). The role of phytochelatins in constitutive and adaptive heavy metal tolerances in hyperaccumulator and non-hyperaccumulator metallophytes. *Journal of Experimental Botany* **53**, 2381-2392.

Shah, D., Shen, M. W. Y., Chen, W. & Da Silva, N. A. (2010). Enhanced arsenic accumulation in *Saccharomyces cerevisiae* overexpressing transporters Fps1p or Hxt7p. *Journal of Biotechnology* **150**, 101-107.

Shao, W., Fanelli, M., Ferrara, F. F. & other authors (1998). Arsenic trioxide as an inducer of apoptosis and loss of PML/RARalpha protein in acute promyelocytic leukemia cells. *Journal of the National Cancer Institute* **90**, 124-133.

Shen, Z. Y., Tan, L. J., Cai, W. J., Shen, J., Chen, C., Tang, X. M. & Zheng, M. H. (1999). Arsenic trioxide induces apoptosis of oesophageal carcinoma *in vitro*. *International Journal of Molecular Medicine* **4**, 33-37.

Shi, H., Shi, X. & Liu, K. J. (2004). Oxidative mechanism of arsenic toxicity and carcinogenesis. *Molecular and Cellular Biochemistry* **255**, 67-78.

Slade, D. & Radman, M. (2011). Oxidative stress resistance in *Deinococcus radiodurans*. *Microbiology and Molecular Biology Reviews* **75**, 133-191.

Slyemi, D., Moinier, D., Brochier-Armanet, C., Bonnefoy, V. & Johnson, D. B. (2011). Characteristics of a phylogenetically ambiguous, arsenic-oxidizing *Thiomonas* sp., *Thiomonas arsenitoxidans* strain 3As(T) sp. nov. *Archive of microbiology* **6**, 439-449.

Smith, A. H., Hopenhayn-Rich, C., Bates, M. N., Goeden, H. M., Hertz-Picciotto, I., Duggan, H. M., Wood, R., Kosnett, M. J. & Smith, M. T. (1992). Cancer risks from arsenic in drinking water. *Environmental Health Perspectives* **97**, 259-267.

Souza-Egipsy, V., Altamirano, M., Amils, R. & Aguilera, A. (2011). Photosynthetic performance of phototrophic biofilms in extreme acidic environments. *Environmental Microbiology*.

Steffens, J. C. (1990). The heavy metal-binding peptides of plants. *Annual Review of Plant Physiology and Plant Molecular Biology* **41**, 553-575.

Stewart, P. S. (1998). A review of experimental measurements of effective diffusive permeabilities and effective diffusion coefficients in biofilms. *Biotechnology and Bioengineering* **59**, 261-272.

Stohs, S. J. & Bagchi, D. (1995). Oxidative mechanisms in the toxicity of metal ions. *Free Radical Biology and Medicine* **18**, 321-336.

Su, Y. H., McGrath, S. P., Zhu, Y. G. & Zhao, F. J. (2008). Highly efficient xylem transport of arsenite in the arsenic hyperaccumulator *Pteris vittata*. *New Phytologist* **180**, 434-441.

Sumida, S., Lyman, H., Kiyohara, N. & Osafune, T. (2007). Mechanism of conversion from heterotrophy to autotrophy in *Euglena gracilis*. *Cytologia* **72**, 447-457.

Tanaka, S., Komatsu, K. & Iyobe, T. (2000). Distribution of *Euglena mutabilis* Schmitz and its Function on the Formation of Chemical Environment in Acid Mine Drainage (AMD). *Proceedings of the 3rd Japan –Taiwan Joint International Symposium on Environmental Science and Technology*.

Thomas, D. J., Styblo, M. & Lin, S. (2001). The cellular metabolism and systemic toxicity of arsenic. *Toxicology and Applied Pharmacology* **176**, 127-144.

Thomas, D. J., Li, J., Waters, S. B., Xing, W., Adair, B. M., Drobna, Z., Devesa, V. & Styblo, M. (2007). Arsenic (+3 oxidation state) methyltransferase and the methylation of arsenicals. *Experimental Biology and Medicine* **232**, 3-13.

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

Thorsen, M., Perrone, G. G., Kristiansson, E., Traini, M., Ye, T., Dawes, I. W., Nerman, O. & Tamas, M. J. (2009). Genetic basis of arsenite and cadmium tolerance in *Saccharomyces cerevisiae*. *BMC Genomics* **10**.

Tittel, J., Bissinger, V., Gaedke, U. & Kamjunke, N. (2005). Inorganic carbon limitation and mixotrophic growth in *Chlamydomonas* from an acidic mining lake. *Protist* **156**, 63-75.

Triemer, E. R. (1980). Role of Golgi apparatus in mucilage production and cyst formation in *Euglena gracilis*., pp. 46-52. *Journal of phycology*.

Tsai, S. L., Singh, S. & Chen, W. (2009). Arsenic metabolism by microbes in nature and the impact on arsenic remediation. *Current Opinion in Biotechnology* **20**, 659-667.

Tsukaguchi, H., Weremowicz, S., Morton, C. C. & Hediger, M. A. (1999). Functional and molecular characterization of the human neutral solute channel aquaporin-9. *American Journal of Physiology - Renal Physiology* **277**, F685-F696.

Tu, C., Ma, L. Q. & Bondada, B. (2002). Arsenic accumulation in the hyperaccumulator Chinese brake and its utilization potential for phytoremediation. *Journal of Environmental Quality* **31**, 1671-1675.

Uthus, E. O. (1992). Evidence for arsenic essentiality. *Environmental Geochemistry and Health* **14**, 55-58.

Valente, T. M. & Gomes, C. L. (2007). The role of two acidophilic algae as ecological indicators of acid mine drainage sites. *Journal of Iberian Geology* **33**, 283-294.

Van Hille, R. P., Boshoff, G. A., Rose, P. D. & Duncan, J. R. (1999). A continuous process for the biological treatment of heavy metal contaminated acid mine water. *Resources, Conservation and Recycling* **27**, 157-167.

Verb, R. G. & Vis, M. L. (2001). Macroalgal communities from an acid mine drainage impacted watershed. *Aquatic Botany* **71**, 93-107.

Wang, T. & Wu, M. (2006). An ATP-binding cassette transporter related to yeast vacuolar ScYCF1 is important for Cd sequestration in *Chlamydomonas reinhardtii*. *Plant, Cell and Environment* **29**, 1901-1912.

Weiss, S., Carapito, C., Cleiss, J. & other authors (2009). Enhanced structural and functional genome elucidation of the arsenite-oxidizing strain *Herminiimonas arsenicoxydans* by proteomics data. *Biochimie* **91**, 192-203.

Wilkins, M. R. (1996). From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis. *Bio/Technology* **14**, 61-65.

Wilmes, P. & Bond, P. L. (2006). Metaproteomics: Studying functional gene expression in microbial ecosystems. *Trends in Microbiology* **14**, 92-97.

Winterbourn, M. J., McDuffett, W. F. & Eppley, S. J. (2000). Aluminium and iron burdens of aquatic biota in New Zealand streams contaminated by acid mine drainage: Effects of trophic level. *Science of the Total Environment* **254**, 45-54.

Woelfl, S., Tittel, J., Zippel, B. & Kringel, R. (2000). Occurrence of an Algal Mass Development in an Acidic (pH 2.5), Iron and Aluminium-rich Coal Mining Pond. *Acta Hydrochimica et Hydrobiologica* **28**, 305-309.

Wolfe-Simon, F., Blum, J. S., Kulp, T. R. & other authors (2010). A Bacterium That Can Grow by Using Arsenic Instead of Phosphorus. *Science*, 1-6.

Wollmann, K., Deneke, R., Nixdorf, B. & Packroff, G. (2000). Dynamics of planktonic food webs in three mining lakes across a pH gradient (pH 2-4). *Hydrobiologia* **433**, 3-14.

Wu, J. & Rosen, B. P. (1991). The ArsR protein is a trans-acting regulatory protein. *Molecular Microbiology* **5**, 1331-1336.

Wu, J., Zhang, R. & Lilley, R. M. (2002). Methylation of arsenic in vitro by cell extracts from bentgrass (*Agrostis tenuis*): Effect of acute exposure of plants to arsenate. *Functional Plant Biology* **29**, 73-80.

Wykoff, D. D. & O'Shea, E. K. (2001). Phosphate transport and sensing in *Saccharomyces cerevisiae*. *Genetics* **159**, 1491-1499.

Wysocki, R., Chéry, C. C., Wawrzycka, D., Van Hulle, M., Cornelis, R., Thevelein, J. M. & Tamas, M. J. (2001). The glycerol channel Fps1p mediates the uptake of arsenite and antimonite in *Saccharomyces cerevisiae*. *Molecular Microbiology* **40**, 1391-1401.

Wysocki, R. & Tamas, M. J. (2010). How *Saccharomyces cerevisiae* copes with toxic metals and metalloids. *FEMS Microbiology Reviews* **34**, 925-951.

Xu, X. Y., McGrath, S. P. & Zhao, F. J. (2007). Rapid reduction of arsenate in the medium mediated by plant roots. *New Phytologist* **176**, 590-599.

Yamanaka, K., Hoshino, M., Okamoto, M., Sawamura, R., Hasegawa, A. & Okada, S. (1990). Induction of DNA damage by dimethylarsine, a metabolite of inorganic arsenics, is for the major part likely due to its peroxy radical. *Biochemical and Biophysical Research Communications* **168**, 58-64.

Yin, H., Cao, L., Xie, M., Chen, Q., Qiu, G., Zhou, J., Wu, L., Wang, D. & Liu, X. (2008). Bacterial diversity based on 16S rRNA and *gyrB* genes at Yinshan mine, China. *Systematic and Applied Microbiology* **31**, 302-311.

Yoshinaga, M., Cai, Y. & Rosen, B. P. (2011). Demethylation of methylarsonic acid by a microbial community. *Environmental Microbiology* **13**, 1205-1215.

Yuan, C., Lu, X., Jie, Q., Rosen, B. P. & Le, X. C. (2008). Volatile arsenic species released from *Escherichia coli* expressing the *asIII* S-adenosylmethionine methyltransferase gene. *Environmental Science and Technology* **42**, 3201-3206.

Zardoya, R., Ding, X., Kitagawa, Y. & Chrispeels, M. J. (2002). Origin of plant glycerol transporters by horizontal gene transfer and functional recruitment. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 14893-14896.

Zettler, L. A. A., Gomez, F., Zettler, E., Keenan, B. G., Amils, R. & Sogin, M. L. (2002). Eukaryotic diversity in Spain's River of Fire: This ancient and hostile ecosystem hosts a surprising variety of microbial organisms. *Nature* **417**, 137.

Zhang, X., Yang, F., Shim, J. Y., Kirk, K. L., Anderson, D. E. & Chen, X. (2007). Identification of arsenic-binding proteins in human breast cancer cells. *Cancer Letters* **255**, 95-106.

Zhao, F. J., Ma, J. F., Meharg, A. A. & McGrath, S. P. (2009). Arsenic uptake and metabolism in plants. *New Phytologist* **181**, 777-794.

Zheng, Y., Wu, J., Ng, J. C., Wang, G. & Lian, W. (2002). The absorption and excretion of fluoride and arsenic in humans. *Toxicology Letters* **133**, 77-82.

Zhigang, A., Cuijie, L., Yuangang, Z., Yejie, D., Wachter, A., Gromes, R. & Rausch, T. (2006). Expression of BjMT2, a metallothionein 2 from *Brassica juncea*, increases copper and cadmium tolerance in *Escherichia coli* and *Arabidopsis thaliana*, but inhibits root elongation in *Arabidopsis thaliana* seedlings. *Journal of Experimental Botany* **57**, 3575-3582.