UNIVERSITE DE STRASBOURG

ÉCOLE DOCTORALE DES SCIENCES DE LA VIE ET DE LA SANTE

UMR 7156 Génétique Moléculaire Génomique et Microbiologie

THÈSE présentée par :

Julien FARASIN

soutenue le : **16 décembre 2015**

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

Discipline : Microbiologie

Diversité génomique et fonctionnelle de bactéries du genre *Thiomonas* isolées du drainage minier acide de Carnoulès (Gard)

RAPPORTEURS :

M^me^ BATTAGLIA-BRUNET Fabienne	Ingénieur de Recherche, BRGM
M^me^ BONNEFOY Violaine	Directeur de Recherche, CNRS

AUTRES MEMBRES DU JURY :

M^me^ SCHALK Isabelle	Directeur de recherche, CNRS

THÈSE dirigée par :

M^me^ ARSENE-PLOETZE Florence	Maître de conférences, Université de Strasbourg

Remerciements

Je tiens tout d'abord à remercier Florence Arsène-Ploetze pour la confiance qu'elle m'a accordée, ainsi que pour ses nombreux conseils et critiques, et d'une manière générale pour sa gestion du bon déroulement de cette thèse.

Je remercie grandement Sandrine Koechler pour son soutien scientifique et psychologique, son efficacité, et sa bonne humeur en toute circonstance.

Je remercie Kelle Freel pour le travail réalisé durant notre courte collaboration, sa sympathie, ses cookies et autres pâtisseries (mais surtout ses cookies).

Un grand merci à Pauline Chaignaud pour noss grands moments de rédaction partagés.
Bonne continuation et beaucoup de réussite pour la suite Pauline !

Merci à Frédéric Plewniak, compagnon de bureau durant ces trois années, pour son aide, ses conseils et nos différents échanges.

Je remercie, et félicite par la même occasion, Martina Péres pour la qualité de son travail.

Merci à mes deux collègues thésards Antoine et Jérémy pour leur sympathie et les bons moments partagés.

Je remercie Olfa, Jin et Florian pour leur travail.

Je remercie Aline, Jessica, Marie-Claire, Didier, Maxime, Philippe, pour leur sympathie et leurs remarques constructives.

Merci pour tous ces bons moments et bon courage pour la suite aux "levures" et au "3ème étage" : Anastasie, Jing, Sabrina, Bruno, David, Esshan, Jean-Seb, Loulou, Ludo, Jackson et Téo.

D'une manière générale, je remercie l'ensemble du personnel de l'UMR 7156 pour leur accueil et leur sympathie.

Merci à David H. pour son analyse critique et son sens de l'accueil.

Merci à Yann Leva et Karine Jézéquel du "Laboratoire Vigne Biotechnologie et Environnement" (LVBE) de Colmar pour les expériences de dosage de l'arsenic.

Je remercie également Julien Deschamps et Romain Briandet de l'UMR 1319 (MICALIS) pour leur aide et leurs conseils concernant l'analyse des biofilms en microscopie confocale.

Je remercie Karine Labadie, Sophie Magenot, Stéphane Cruveiller, David Roche et Valérie Barbe du Laboratoire de Finition du Génoscope pour le séquençage et l'assemblage des génomes.

Merci à Caroline Proux, Marie-Agnès Dillies, Hugo Varet, Berndt Jagla et Jean-Yves Coppée de la plateforme "Transcriptome et Épigénome" de l'institut Pasteur pour l'expérience de transcriptomique en RNAseq.

Enfin, je remercie l'ensemble des membres du jury, Fabienne Battaglia-Brunet, Violaine Bonnefoy, Isabelle Schalk et Romain Briandet d'avoir accepté de juger ce travail.

Table des matières

Liste des figures

Liste des tableaux

Liste des abréviations

As(III) et As(V) : Arsénite et Arséniate

CAI : Index d'Adaptation des Codons (*Codon adaptation Index*)

CMI : Concentration Minimale Inhibitrice

CGH : Hybridation Génomique Comparative (*Comparative Genomic Hybridization*)

DMA : Drainage Minier Acide

EGM : Élément Génétique mobile

EPS : Substance Polymérique Extracellulaire (*Extracellular Polymeric Substance*)

GEI : Ilot Génomique (*Genomic Islands*)

HGT : Transfert Génétique Horizontal (*Horizontal Gene Transfer*)

HR : Recombinaison Homologue (*Homolog recombination*)

ICE : Élément Intégratif et Conjugatif (Integrative and Conjugative Element)

IS : Séquence d'Insertion (*Insertion Sequence*)

IME : Élément Intégratif et Mobilisable (Integrative and Mobilizable Element)

ORF : Cadre Ouvert de Lecture (*Open Reading Frame*)

RGP : Région de Plasticité Génomique (*Region of Genomic Plasticity*)

Avant propos

Cette thèse porte sur l'étude de la variabilité génomique et fonctionnelle de bactéries du genre *Thiomonas* isolées d'un drainage minier acide à Carnoulès. Je me suis intéréssé à la variabilité génomique de plusieurs souches très proches phylogénétiquement, et à l'impact de l'arsénite et du développement en biofilm sur cette variabilité. La partie introductive de cette thèse est constitutée d'une introduction bibliographique abordant la structure et l'évolution des génomes ainsi que différents éléments et facteurs influençant celles-ci. Une deuxième partie intitulée "contexte du sujet de thèse" aborde plus précisément la résistance à l'arsenic chez les bactéries, les drainages miniers acides et le genre *Thiomonas*.

INTRODUCTION BIBLIOGRAPHIQUE

I - Le génome des bactéries

1 - Généralités

Les connaissances concernant la structure, le fonctionnement, et l'évolution des génomes bactériens, sont intimement liées aux avancées technologiques dans le domaine. Les premières études reposaient principalement sur la détermination de profils de restriction permettant l'établissement d'une carte physique du chromosome (Riley and Anilionis, 1978). Le développement des techniques d'électrophorèses en champs pulsés ont ensuite facilité l'établissement de ces cartes (Casjens, 1998) et permis de déterminer facilement la taille et la topologie des chromosomes (ex. (Marin et al., 1997)). En 1995, ont été séquencés les deux premiers génomes bactériens : celui d'*Haemophilus influenza* (Fleischmann et al., 1995) et celui de *Mycoplasma genitalium* (Fraser et al., 1995). A partir de cette date, le développement des outils de séquençage associé à la réduction de leur coût ont entrainé une augmentation rapide du nombre de génomes disponibles (Figure 1). C'est notamment le cas à partir de 2005 avec l'apparition des outils de séquençage de seconde génération (NGS, *Next Generation Sequencing*) (Land et al., 2015).

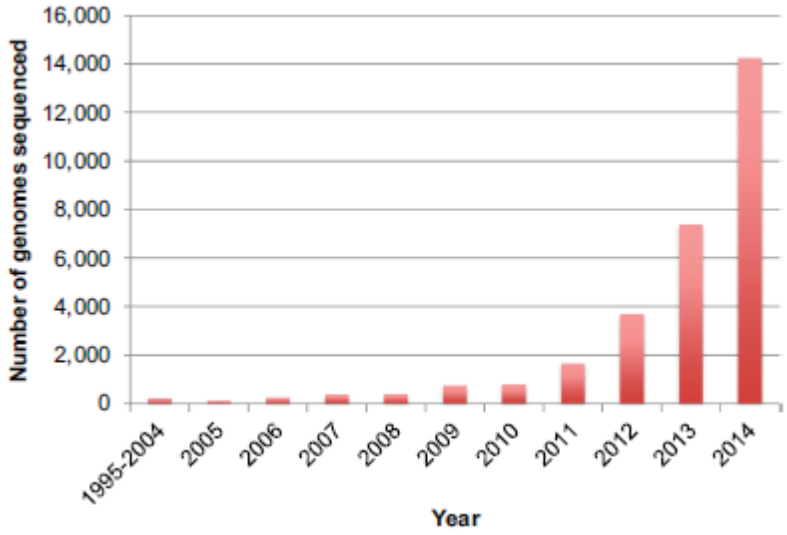


Figure 1 Nombre de génomes de bactéries et d'archées séquencés chaque année depuis le séquençage des deux premiers génomes de bactéries en 1995. Depuis les années 2010, ce chiffre augmente de façon exponentielle. D'après Land et al., 2015.

La plupart des chromosomes bactériens sont circulaires et généralement uniques, mais il a également été démontré l'existence de chromosomes linéaires, comme par exemple chez *Streptomyces* (Lin et al., 1993) ou *Borrelia burgdorferi* (Casjens and Huang, 1993). Leur taille varie, allant de moins 1 Mb comme par exemple dans le cas de *Mycoplasma genitalium* (Fraser et al., 1995) à plus de 10 Mb dans le cas de *Sorangium cellulosum* (Han et al., 2013). En plus de varier de par leur taille, les génomes varient également de par leur contenu génétique, et ce même entre deux souches de la même espèce. Les régions concernées sont généralement des fragments d'ADN discrets portant des fonctions non-essentielles au fonctionnement de la bactérie, mais pouvant être avantageuses dans certaines conditions environnementales. Ces régions, qualifiées d'îlots génomiques, résultent d'évènements de transferts génétiques horizontaux (HGTs), c'est-à-dire d'échanges de matériels génétiques entre bactéries, et/ou de recombinaisons intrachromosomiques. Les HGTs sont dans certains cas le résultat de l'activité d'éléments génétiques mobiles (EGMs). Ces éléments sont des séquences d'ADN codant des protéines permettant leur déplacement dans le génome ou leur transfert dans le génome d'une autre cellule. Les recombinaisons sont généralement liées à la présence de séquences d'ADN répétées, c'est-à-dire des fragments d'ADN identiques ou presque, qui de par leur homologie peuvent s'apparier.

Lors de la division cellulaire, le génome bactérien est répliqué de manière bidirectionnelle à partir d'une origine de réplication unique, le site *ori*. Ce site est constitué de motifs particuliers, c'est à dire de courtes séquences d'ADN non-codantes ayant un rôle biologique (Touzain et al., 2011). Ces motifs sont reconnus par la protéine DnaA qui s'y fixe pour former un complexe nucléoprotéique. Ceci entraîne le déroulement de la molécule d'ADN de cette région au niveau de laquelle peut ainsi se lier le réplisome, un complexe protéique comprenant notamment l'ADN polymérase III, responsable de la synthèse du brin d'ADN complémentaire. Deux complexes se déplacent de manière opposée depuis l'origine de réplication vers le site de terminaison de la réplication (*Ter*) en formant deux fourches de réplication. Ces deux parties du génome répliquées en parallèle constituent les réplichores, chacun correspondant environ à la moitié du génome (Figure 2 et Figure 3 page 16).

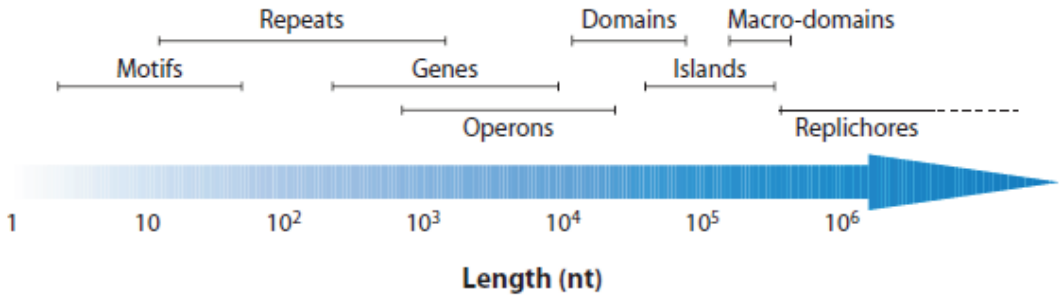


Figure 2 Différentes échelles d'organisation du génome bactérien. Les motifs, de courtes séquences d'ADN non-codantes ayant un rôle biologique, sont le plus petit niveau d'organisation. Certaines séquences, dont la taille est relativement variable, sont répétées (Repeats) et peuvent entrainer des recombinaisons et donc de la variabilité génomique. L'organisation des gènes en unité transcriptionelle, l'opéron, est très conservée chez les bactéries. Les domaines topologiques (Domains) et les macrodomaines sont deux niveaux d'organisation spatiale du chromosome. Certaines régions très variables différent entre souches proches et sont appelées "îlots génomiques" (Islands). Le génome est divisé quasi-symétriquement en deux réplichores, correspondants aux deux régions séparées par les sites d'origine et de terminaison de la réplication. D'après Rocha 2008.

Dans certains cas, le génome comporte également des éléments génétiques extra-chromosomiques, les plasmides. Le terme de "plasmide" a été introduit par Joshua Lederberg en 1952 (Lederberg, 1952). Ces éléments sont généralement formés d'un ADNdb circulaire (il existe aussi des formes linéaires) aux extrémités covalemment fermées, et capables de réplication autonome vis à vis du chromosome de la cellule hôte. Ces éléments sont aussi capables de contrôler leur nombre de copies par cellule et de se répartir lors de la division cellulaire. Contrairement aux chromosomes, ils ne sont pas indispensables au fonctionnement de la cellule, mais portent des gènes pouvant conférer un avantage dans une niche écologique donnée. Il existe cependant certains cas, notamment chez les α-Proteobacteries (Pinto et al., 2012), où la distinction entre plasmides et chromosomes surnuméraires ou chromides est difficile, car ces derniers partagent certaines propriétés avec les plasmides, mais portent des gènes essentiels (Pinto et al., 2012). Plus de 4600 plasmides sont répertoriés dans la base de données GenBank, et leur taille s'échelonne entre 744 pb et 2,58 Mpb avec une moyenne de 80 Kpb (Shintani et al., 2015). Dans les années 1970, les plasmides ont été classés par groupes d'incompatibilités (groupes Inc), c'est à dire par leur incapacité à partager un même système de partition et de réplication. Depuis, d'autres classifications ont été réalisées, basées sur la phylogénie des relaxases (Garcillan-Barcia et al., 2009), des protéines de formation du pore de conjugaison (Smillie et al., 2010), ou prenant en compte plusieurs de ces facteurs (Shintani et al., 2015).

Certaines caractéristiques de la structure des génomes sont relativement conservées au

sein des bactéries car elles sont sous la contrainte de processus cellulaires essentiels, que ce soit à un échelle locale (e.g transcription) ou globale (ex. réplication) (Rocha, 2008). Plusieurs niveaux d'organisation peuvent ainsi être distingués suivant l'échelle spatiale, et peuvent concerner l'organisation de l'information génétique ou l'organisation physique du chromosome.

2 - Différents niveaux d'organisation

a. Organisation de l'information génétique

i. Les opérons

Historiquement, le premier niveau d'organisation de l'information à avoir été décrit est l'organisation des gènes en opérons (Jacob and Monod, 1961), c'est-à-dire en unités transcriptionnelles. Bien que peu d'opérons soient conservés au sein des bactéries, le regroupement de plusieurs gènes sous le contrôle d'un promoteur unique est une caractéristique conservée (Koonin, 2009; Rocha, 2008). Les gènes d'un même opéron sont généralement impliqués dans un même processus cellulaire. Plusieurs modèles ont été proposés pour expliquer l'existence des opérons, les plus connus étant le "modèle régulateur" et l'"opéron égoïste" (Lawrence and Roth, 1996).

D'après le modèle régulateur, l'organisation des gènes en opérons s'expliquerait par des avantages conférés aux bactéries permettant une régulation plus efficace de l'expression génétique. Les gènes d'un opéron formant souvent un module fonctionnel, l'expression de leur totalité est nécessaire et leur regroupement sous un même promoteur semble ainsi avantageux. Dans le cas d'un ARNm codant différentes sous-unités d'une protéine devant être produite en quantité égale, leur co-transcription permet une stoechiométrie précise, et la dégradation d'un seul ARNm suffit à stopper la traduction de cette protéine. De plus, les séquences régulatrices des opérons sont souvent plus complexes que celles de gènes isolés, ce qui peut en plus permettre une régulation plus fine. D'après le modèle égoïste, l'organisation en opéron résulte d'une sélection au niveau des gènes et non de l'individu. La proximité et l'organisation de gènes en module fonctionnel s'expliquerait par la probabilité qu'ils soient transférés avec succès à d'autres bactéries. Ces gènes, s'ils ne sont pas transférés ensemble, ne procurent aucun avantage à la bactérie receveuse et ne sont donc pas sélectionnés. D'après ce modèle, le regroupement des gènes en opérons aurait lieu suite à des réarrangements génomiques ou des transferts génétiques horizontaux ; une grande importance est donc donnée à la variabilité génomique.

ii. Asymétrie du génome

Les génomes bactériens sont marqués par plusieurs asymétries entre le brin direct et le brin indirect. Le brin direct possède tout d'abord une densité de gènes plus importante. Cette différence s'expliquerait par le fait qu'elle diminue la probabilité de collisions entre l'ADN polymérase et l'ARN polymérase, qui progressent toutes deux dans le sens 5'→ 3'. Les gènes présents à proximité de l'origine de réplication sont généralement les plus exprimés, ce qui serait lié à leur surreprésentation pendant la réplication, comme c'est le cas chez *E. coli* lorsque plusieurs cycles de réplication sont engagés (Rocha, 2004). Le brin direct est également caractérisé par un taux de base G plus élevé. Ce biais de composition s'expliquerait par le fait que pendant sa réplication, le brin direct reste sous forme simple brin, ce qui favoriserait les désaminations de C et de A (C→T et A→G). Enfin, le nombre de sites de recombinaison *chi* (*Crossover Hot-spot Instigator*) est important (Figure 3). Les sites *chi* sont des motifs avec lesquels interagissent RecBCD, impliqué dans la réparation de l'ADN par recombinaison homologue. Cette enzyme reconnait et se lie à des coupures double brins de l'ADN, puis le dégrade jusqu'à éventuellement rencontrer un site *chi*. Dans ce cas, elle stoppe la dégradation et produit une extrémité 3' sortante, qui peut servir de substrat à des protéines impliquées dans la recombinaison homologue.

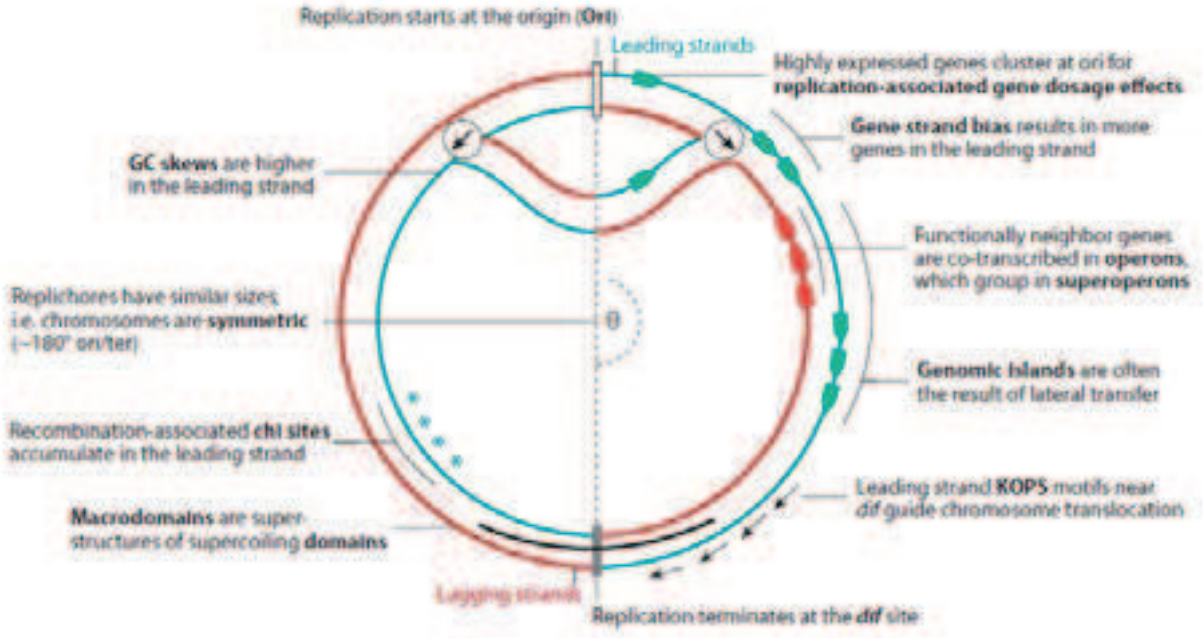


Figure 3 Schéma de la structure générale des génomes bactériens montrant les différentes asymétries entre les brins direct (bleu) et indirect (rouge). Les gènes proches de l'ori sont plus exprimés, la densité de gènes, le taux de G et le nombre de sites chi sont plus importants sur le brin direct. D'après Rocha 2008.

b. Structure du chromosome

Le chromosome bactérien a une taille qui varie entre 2 et 8 Mpb selon les espèces considérées, et est compacté plus de 1000 fois au sein d'une structure de moins de 100 µm, le nucléoïde (Wang et al., 2013). Cette compaction est organisée, hiérarchisée, et participe au bon déroulement des processus de réplication, d'expression génétique, de recombinaison et de réparation de l'ADN. Cette organisation est relativement mal connue, mais des progrès techniques ont permis des avancées dans ce domaine (Le and Laub, 2014).

i. Les domaines topologiques

Le principal mécanisme de compaction de l'ADN est le surenroulement négatif, c'est-à-dire dans le sens inverse de la double hélice d'ADN. Ces surenroulements ont pour conséquence la formation de boucles et de branches dites "plectonémiques", observables en microscopie électronique. En plus de condenser le chromosome, les surenroulements créent des domaines topologiques séparés les uns les autres, et dont la taille moyenne est estimée à une dizaine de Kpb (Figure 4). Un génome moyen de 4 Mpb peut ainsi contenir 400 domaines surenroulés indépendants. La séparation topologique de ces structures est maintenue par des éléments appelés domainines, de nature protéique ou ARN.

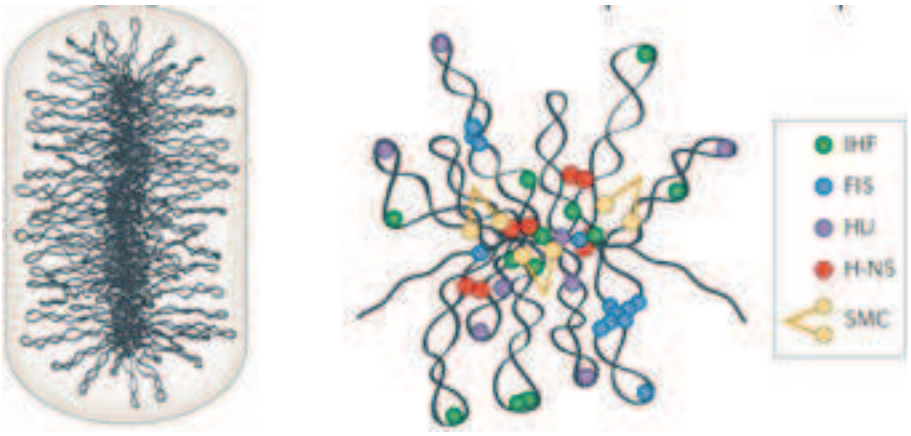


Figure 4 Schéma de la structure topologique du chromosome (gauche) et des protéines impliquées dans le maintien de la structure du chromosome (droite). D'après Wang et al. 2013.

Les domaines topologiques sont très dynamiques. L'équilibre entre surenroulement et relâchement est principalement médié par l'action de deux enzymes. L'ADN gyrase introduit des surenroulements négatifs et la topoisomerase I permet la relaxation de ces derniers. Les surenroulements de l'ADN n'expliquent pas à eux seuls le haut degré de compaction de l'ADN. Certaines protéines, à la manière des histones chez les eucaryotes, se lient de manière spécifique ou non-spécifique à l'ADN, le courbent pour le compacter puis créent des ponts entre les brins pliés afin de stabiliser la boucle ainsi constituée. Les principales protéines histone-like chez *E. coli* sont HU, IHF (*Integration Host Factor*), Fis et H-NS (Figure 4). D'autres protéines sont supposées assurer le maintien de l'organisation topologique du chromosome, comme par exemple le complexe condensine SMC (*Structural Maintenance of the Chromosome*) qui est également présent chez les eucaryotes et donc très conservée. Une étude récente menée chez *E. coli* rapporte l'existence d'une corrélation positive entre le taux de mutations et certains domaines structurés par Fis et HU (Foster et al., 2013). Ceci suggèrerait que le contenu génétique de certaines régions du chromosome soit plus à même à varier que d'autres.

ii. Les macrodomaines

Les macrodomaines constituent un degré d'organisation supérieur du chromosome, dont les régions mesurent de 800 Kpb à 1 Mpb. Chez *E. coli,* quatre macrodomaines ont été décrits, mais ces structures sont supposées être une caractéristique commune à toutes les bactéries. Le taux de recombinaisons au sein de ces domaines est plus important qu'entre les différents domaines. Cette organisation physique du chromosome n'est donc pas sans conséquence sur l'organisation de l'information génétique. Les domaines de *E. coli* sont les domaines *ori* (origin)

et *ter* (terminus) ainsi que deux domaines bordant le domaine *ter* et appelés macrodomaines *left* et *right*. En plus de ces macrodomaines, *ori* est entouré par deux régions flexibles apparemment sans organisation particulière. Bien que les mécanismes responsables de l'organisation en macro-domaines soient encore quasiment inconnus, l'implication de protéines se liant à l'ADN de manière séquence-spécifique est devenue évidente. A titre d'exemple, un motif nommé *matS* est notamment surreprésenté dans le domaine *ter*. Il a été montré que la protéine MatP se lie à ce motif *in vivo*. Ce motif semblerait organiser les sites *matS*, regroupant ainsi les boucles enroulées dans la région *ter*.

3 - Comparaison de la structure des génomes d'espèces proches

L'augmentation du nombre de données disponibles concernant les séquences de génomes bactériens a permis de mieux comprendre leur structure et leur évolution, et de revoir la notion d'espèce. Il est par exemple vite paru évident que ces génomes étaient caractérisés par une certaine "plasticité", plusieurs génomes de la même espèce pouvant avoir une organisation et un contenu génétique différents (Casjens, 1998). En conséquence, la structure génétique d'une espèce ne peut être définie à partir d'un seul génome et cette notion a été modifiée. La synténie, c'est-à-dire l'organisation des gènes dans le génome, est généralement peu conservée même entre des espèces proches (Koonin, 2009), et diffèrent d'autant plus que les espèces sont éloignées. Les cas les plus fréquemment observés sont des inversions symétriques par rapport à l'origine de réplication pouvant résulter de recombinaisons homologues au niveau de la fourche de réplication (Eisen et al., 2000). D'autres inversions peuvent être causées par des recombinaisons site-spécifique, et des éléments génétiques mobiles tels que les séquences d'insertions (IS, *Insertion Sequence*) peuvent également être à l'origine de translocations. En plus de modifier la structure des génomes, une autre conséquence des remaniements est la perte de certains gènes ou régions du génome. Ces remaniments, associés aux transferts horizontaux, ont pour conséquence des différences de contenu génétique entre les souches. L'étude de ces différences est aujourd'hui permise par le développement des outils génomiques, et a mené à l'introduction du concept de pan-genome et de core-genome.

a. Le pan-genome et core-genome

i. Définition

Le pan-génome est défini comme l'addition du core-genome, ensemble des gènes partagés par les souches comparées, et du génome flexible ou variable, composé des gènes partagés par un sous-groupe seulement. Ce concept a été introduit par Tettelin et al. en 2005, lors d'un travail pionnier de comparaisons des génomes de huit souches de *Streptococcus agalactiae* (Tettelin et al., 2005). Cependant, l'étude du pan-génome peut être réalisée à d'autres résolutions phylogénétiques que celle de l'espèce, la plus large à ce jour étant le super-phylum Bacteria (Lapierre and Gogarten, 2009). Lors de cette dernières étude, les gènes souche-spécifiques ou partagés par peu de souches représentaient 28 %, et le core-genome seulement 8 %. La notion de core-genome est donc relative : plus les organismes comparés sont éloignés, plus ce core-genome diminue. Suivant le contexte, l'étude du pan-génome peut servir à la définition du génome bactérien minimal (Lapierre and Gogarten, 2009), à estimer la diversité d'un ensemble d'organismes au sein d'une niche écologique, et/ou à fournir de nouveaux éléments pour la définition de l'espèce bactérienne. Deux "types" de pan-génome ont été décrits suivant les espèces. Il existe des cas où la taille du pan-génome a pu être défini, auquel cas il est qualifié de "fermé". Dans d'autre cas la taille du pan-génome a été estimée comme infinie (mathématiquement), et qualifiée d'"ouvert".

ii. Méthode

L'étude du pan-génome d'un groupe d'organismes consiste dans un premier temps à comparer leur génome afin de déterminer quels gènes sont communs, et constituent donc le core-génome. Les autres gènes sont donc non-essentiels ou encore souche-spécifiques et font partie du génome dit flexible ou variable. La taille de ce pan-génome est estimée en calculant combien de nouveaux gènes seraient ajoutés au pan-génome à chaque nouvelle souche séquencée. Bien que le concept de pan-génome soit intuitif, les méthodes employées diffèrent selon les études (Vernikos et al., 2015). Par exemple, dans la première étude concernant *Streptococcus agalactiae*, des familles de protéines orthologues ont été définies avec un seuil de 50 % d'identité de séquence réparti sur 50 % de la longueur du peptide (Tettelin et al., 2005). Dans une autre étude de Hiller *et al.* concernant *Streptococcus pneumoniae*, ces mêmes seuils sont de 70 % (Hiller et al., 2007). Ces différents seuils rendent difficile la comparaison entre les deux études. Par exemple, certains gènes dont la séquence est variable (ex. transporteurs ou

antigènes de surface) peuvent appartenir à la partie flexible ou au core-genome suivant les critères. Des progrès restent à faire concernant la standardisation de la méthodologie, mais ces études ont quand même permis une meilleure prise de conscience de la diversité des génomes. En particulier, le nombre croissant de génomes disponibles pour des souches très proches permet d'appréhender l'évolution bactérienne, notamment le rôle et l'origine des éléments de la partie flexible du génome.

b. Le génome flexible

i. Généralités

L'idée de flexibilité du génome bactérien et d'une variabilité intra-spécifique du contenu génomique, est antérieure à l'obtention d'un grand nombre de génomes bactériens complets. En effet, des expériences d'électrophorèse en champs pulsé ont montré que différentes souches d'*E. coli* présentaient des tailles allant de 4,5 à 5,5 Kpb (Bergthorsson and Ochman, 1998). La génomique a cependant permis de mesurer plus précisément cette part variable, de connaître son organisation et d'accéder à son contenu.

Il existe une corrélation entre pan-génome fermé et niche écologique restreinte et isolée. C'est par exemple le cas de l'endosymbionte intracellulaire *Buchnera aphidicola*, pour lequel l'isolement de sa niche écologique empêche les transferts génétiques horizontaux. Cette espèce est par ailleurs dépourvue de mécanismes d'échanges génétiques et de recombinaisons. A l'inverse, les espèces vivant libres dans l'environnement naturel semblent avoir un pan-génome "ouvert", même dans le cas de la comparaison de souches isolées en même temps et au même endroit (Pena et al., 2010). Un pan-génome "ouvert" semble donc refléter un plus grand potentiel adaptatif dans des environnements ouverts et changeants.

Une caractéristique des gènes appartenant au pool flexible est qu'ils peuvent généralement conférer un avantage à la bactérie dans des conditions particulières. Ils peuvent par exemple conférer une virulence, la capacité à dégrader des xénobiotiques, ou encore l'acquisition du fer (Dobrindt et al., 2004). La partie flexible du génome confère donc des capacités adaptatives. Une deuxième caractéristique est leur regroupement en régions appelées îlots génomiques (Rodriguez-Valera and Ussery, 2012).

ii. Les îlots génomiques

Le terme d'îlot génomique est issu du concept d'îlot de pathogénicité introduit à la fin des années 1980. La définition originelle a été donnée à la fin des années 1990, alors que

seulement 12 génomes complets étaient disponibles. Les îlots génomiques (GEIs) peuvent être définis comme des fragments d'ADN, allant d'une dizaine à plus de 200 Kpb, différant entre des souches apparentées et donc supposés être ou avoir été mobiles (Juhas et al., 2009).

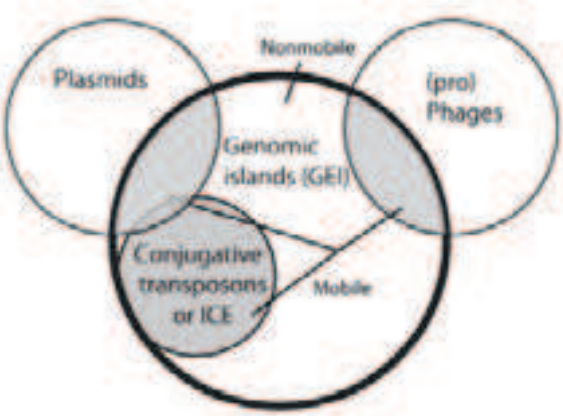


Figure 5 Illustration des éléments regroupés sous le terme d'îlot génomique. Les GEIs ne correspondent pas à une classe particulière d'éléments génétiques mobiles. D'après Juhas et al., 2009.

Les GEIs ont généralement en commun des caractéristiques indiquant leur acquisition par transfert horizontal. Juhas et al. citent six caractéristiques souvent retrouvées au niveau de ces régions : i/ statistiques nucléotidiques (e.g. %GC, CAI) différentes du reste du génome ; ii/ insertion au niveau d'ARNt ; iii/séquences répétées directes parfaites ou presque de 16-20 pb, résultant d'une insertion site-spécifique ; iv/présence d'intégrases fonctionnelles ou cryptiques et/ou facteurs apparentés à des systèmes de conjugaison plasmidiques ou des phages, impliqués dans le transfert du GEI ; v/ IS ou transposons impliqués dans la mobilité du GEI ou dans des remaniements génomiques internes vi/gènes conférant un avantage sélectif à la cellule hôte (résistance, pathogénicité, symbiose etc.).

Ainsi définis, les GEIs englobent des éléments génétiques mobiles (EGM) de diverses familles telles que les éléments intégratifs et conjugatifs (ICEs, *Integrative and Conjugative Elements*) (Figure 5), mais aussi des éléments avec une autonomie variée en terme d'excision, d'intégration, de réplication et de transfert (Bellanger et al., 2014). Ces différents éléments sont décrits dans la partie III - page 30. La présence de GEIs peut faire suite soit à la perte d'un fragment d'ADN, soit à un gain via un transfert génétique horizontal

II - Les transferts génétiques horizontaux

1 - Généralités

Les transferts génétiques horizontaux (HGTs, *Horizontal Gene Transfers*) sont des échanges de matériel génétique entre des organismes n'étant pas dans une relation de descendance directe, à l'opposé de l'hérédité dite verticale. Chez les procaryotes, les HGTs permettent l'acquisition de nouvelles fonctions et participent ainsi à la diversification et l'adaptation des populations (Wiedenbeck and Cohan, 2011). Les HGTs se font à la fois par des recombinaisons homologues entre des souches ou des espèces proches, ou part le transfert d'éléments génétiques mobiles. Il existe plusieurs mécanismes permettant des transferts horizontaux : la transformation, permettant l'acquisition d'ADN présent dans l'environnement, la conjugaison permettant l'échange d'informations génétiques entre deux cellules, et la transduction impliquant des virus. En plus de ces trois mécanismes bien connus, d'autres pistes récentes concernent l'implication des vésicules membranaires et de nanotubes.

2 - La transformation

a. Généralités

La transformation bactérienne est l'import par une cellule d'ADN libre dans l'environnement, suivi de son intégration dans le chromosome. Ce mécanisme a été découvert en 1928 par Griffith chez *Streptococcus pneumoniae* (Griffith, 1928). Contrairement aux autres mécanismes responsables des HGT (conjugaison et transduction), la transformation n'implique pas d'éléments génétiques mobiles. Toutes les espèces ne sont pas capables de transformations. Ce mécanisme a été rapporté chez plus de 80 espèces (Johnston et al., 2014), et correspond généralement a un état physiologique transitoire appelé "compétence", les bactéries étant qualifiées de "compétentes".

La transformation implique le transfert de l'ADN à travers la membrane externe (dans le cas des Gram-négatif), puis la paroi et la membrane cytoplasmique, puis son intégration dans le génome. Ce mécanisme est donc relativement complexe. L'ADN doit d'abord être internalisé, c'est à dire transféré au niveau d'un complexe protéique permettant son passage à travers la paroi.

b. Import de l'ADN

Les mécanismes impliqués dans la transformation décrits jusqu'à présent semblent être communs à toutes les bactéries compétentes connues, à l'exception près d'*Helicobacter pylori*. L'internalisation de l'ADN repose sur un ensemble de protéines conservées et principalement codées par un ensemble de gènes exprimés simultanément lors de la compétence : le régulon *com* (Johnston et al., 2014).

Le substrat de la transformation est de l'ADN sous forme double brin. Son transfert à travers l'enveloppe externe des bactéries n'est pas encore bien connu. Chez les bactéries gram-négatif, cette étape nécessite des protéines similaires à ceux impliquées dans l'assemblage du système de sécrétion de type II (SS2) et du pili de type IV (T4P), ainsi que d'un canal formé par PilQ au niveau de la membrane externe. Il semblerait également qu'un T4P soit impliqué dans ce processus à la fois chez les bactéries gram-positif et les gram-négatif. Ce pilus, appelé pilus de transformation (Tfp, *TransFormation Pilus*), se lierait à l'ADNdb extracellulaire et le transférerait au niveau du récepteur ComEA (Figure 6). Ce dernier est supposé transférer l'ADN à une protéine qui dégraderait un des deux brins avant le transfert. Ce rôle est assuré par la nucléase EndA chez *Streptococcus pneumoniae,* mais aucune protéine similaire n'a pu être identifiée chez d'autres espèces.

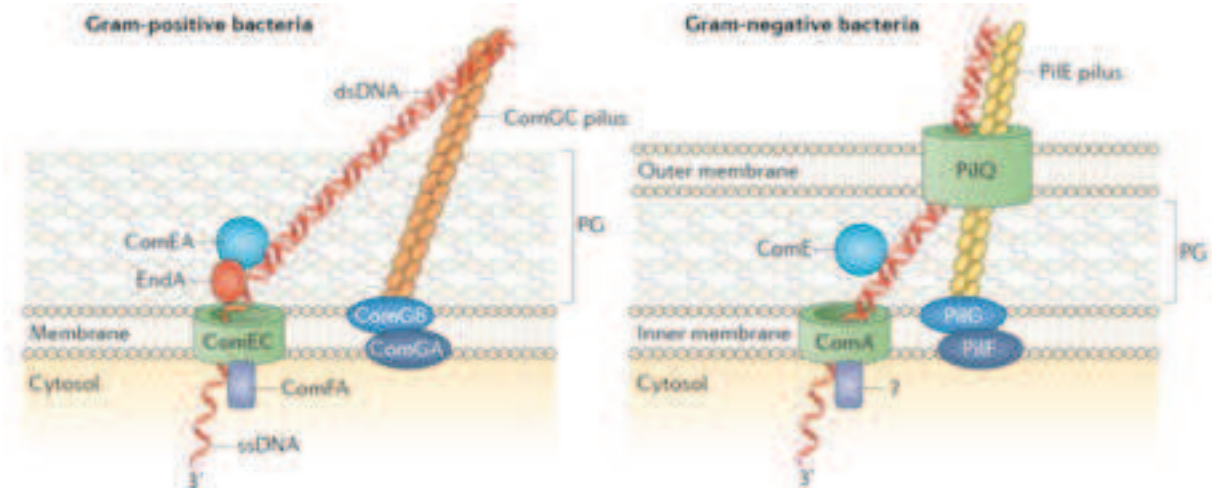


Figure 6 Schéma de l'internalisation de l'ADN lors de la transformation chez les bactéries. L'ADN sous forme double-brin est transféré par un pilus de transformation (Tfp) au niveau de la membrane cytoplasmique. Dans le cas des Gram-négatif, un canal formé par PilQ permet le transfert de l'ADN à travers la membrane externe. L'ADN est ensuite transféré au niveau de ComEA (Gram +) ou ComA (Gram -) puis transféré sous forme simple-brin dans le compartiment cytoplasmique via ComEC (Gram +) ou ComA (Gram -). D'après Jonhston et al., 2014

Le transfert de l'ADN à travers la membrane cytoplasmique a été essentiellement étudié chez *B. subtilis* où il implique trois éléments : le récepteur de l'ADN ComEA, la perméase ComEC, et la protéine de liaison à l'ATP ComFA. ComEA est lié à la membrane par sa région N-terminale et possède un domaine de liaison à l'ADN dans sa partie C-terminale pouvant interagir avec de l'ADNdb de manière aspécifique. ComEA possède également une partie flexible qui lui permettrait ainsi de transférer l'ADN au niveau du canal transmembranaire formé par ComEC.

c. Intégration de l'ADN

L'ADN étant internalisé sous forme simple-brin, il constitue de ce fait un substrat pour la recombinaison homologue. Chez *S. pneumoniae*, plusieurs études suggèrent que la protéine DprA serait une médiatrice de la recombinaison, c'est-à-dire un cofacteur aidant à l'interaction entre RecA et l'ADNsb. Le gène codant DrpA a été retrouvé dans tous les génomes de bactéries transformables disponibles, et semble donc être une caractéristique conservée des bactéries naturellement compétentes. Ce gène a également été retrouvé chez des bactéries n'étant pas décrites comme compétentes, sans que le rôle de DrpA chez ces dernières ne soit connu. Il pourrait ainsi s'agir soit de vestiges d'une compétence ancestrale, soit d'une compétence encore non détectée.

d. Le cas *Helicobacter pylori*

A la différence du mécanisme de transformation décrit ci-dessus, *H. pylori* n'utilise pas de Tfp, mais un système de sécrétion de type IV (T4SS). Ce système de sécrétion a été appelé protéine de compétence B (ComB). L'homologue de ComEC (canal transmembranaire) est cependant présent et requis pour la transformation même s'il est phylogénétiquement distant de celui décrit chez les autres bactéries. Le mécanisme d'assimilation de l'ADN extracellulaire est supposé être réalisé en deux étapes. Dans un premier temps, ComB transporte l'ADN dans le périplasme, puis ComEC l'importe dans le cytoplasme. Ce mécanisme diffère ainsi de celui décrit pour les autres bactéries qu'elles soient Gram-négatif ou Gram-positif.

3 - La conjugaison

a. Généralités

La conjugaison est un mécanisme de transfert de matériel génétique sous forme d'ADNsb entre deux cellules. Ce processus nécessite l'appariement d'une cellule donneuse avec une cellule receveuse grâce à un système de sécrétion de type IV (T4SS), puis la mobilisation (réplication parfois précédée de l'excision du génome) et le transfert du matériel génétique via ce système de sécrétion (Cabezon et al., 2015). Les gènes impliqués dans ces étapes peuvent être portés par des plasmides conjugatifs ou des ICEs (*Integrative and Conjugative Element*). Ils sont organisés en plusieurs opérons, avec une claire distinction entre ceux impliqués dans l'appariement et ceux impliqués dans le transfert et la réplication de l'ADN. Cette organisation est conservée dans la plupart des systèmes conjugatifs.

Le système de conjugaison, également appelé transferosome, peut être considéré comme une sous-famille de système de sécrétion de type IV (T4SS). Il est unique chez les bactéries du fait de sa capacité à transporter à la fois des protéines et de l'ADN (covalemment lié à des protéines). Les gènes codant ce T4SS, sont appelés gènes MPF (*Mating Pair Formation*), car ils sont impliqués dans l'appariement des cellules. Chez les protéobactéries, quatre familles de gènes MPF ont été décrits : MPF_F (plasmides conjugatifs F), MPF_I (plasmides IncI), MPF_T (plasmides Ti d'*Agrobacterium tumefacien*s), et MPF_G (associés aux ICEs). Des signatures de gènes MPF ont également été retrouvées chez les cyanobactéries, les *Firmicutes*, le genre *Bacteroides*, les *Actinobacteria* et les *Archaea*. Le processus de conjugaison chez les protéobactéries aurait donc une origine ancienne et aurait été transmis par HGT dans différents clades procaryotes.

D'un point de vue structural, l'architecture des T4SS est conservée. Quatre parties peuvent être distinguées : le pilus, le canal transmembranaire, la plateforme de la membrane interne et l'ATPase située à la base du canal. Le plus simple T4SS décrit est codé par le système MPF_T d'*A. tumefaciens* et est composé de 11 protéines nommées VirB1 à VirB11. Ces 11 protéines, avec la protéine de couplage VirD4, s'assemblent en un complexe macromoléculaire s'étendant à travers les membranes externe, interne, et le périplasme.

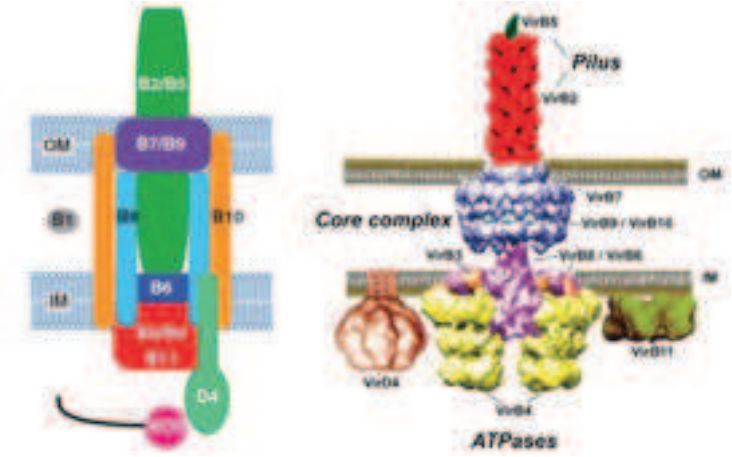


Figure 7 Schéma (gauche) et architecture (droite) du pore de conjugaison et du pilus conjugatif. D'après Guglielmini et al., 2013 et Cabezon et al., 2015.

b. Formation du pilus conjugatif

Afin d'adhérer à la cellule receveuse, la cellule donneuse synthétise un pilus au bout duquel se trouve une molécule appelée adhésine, impliquée dans les contacts intercellulaires. Le composant majeur du pilus est la protéine hydrophobe VirB2 (appelée piline), alors que l'adhésine (VirB5) est minoritaire. La synthèse de pilus n'est pas encore bien comprise mais trois étapes peuvent être distinguées : formation de la piline, extraction de la piline de la membrane interne, et élongation du pilus.

La piline est formée à partir d'un précurseur, la prépiline, contenant une séquence signal permettant son transfert à trvaers la membrane interne. La prépiline est ensuite modifiée sous l'action d'enzymes, notamment de peptidases, afin d'avoir la conformation de la piline. La piline est ensuite transférée dans le compartiment périplasmique avec l'aide des protéines VirB4 et VirB11, sans que l'action exacte de ces protéines ne soit connue. Le mécanisme d'élongation du pilus est également mal connu. Le caractère hydrophobe de la piline semble cependant jouer un rôle important pour sa polymérisation.

c. Mobilisation et transfert du matériel génétique

Les données biochimiques et génétiques suggèrent que la synthèse du pilus et le transfert d'ADN/protéines sont deux processus indépendants (Cabezon et al., 2015). Dans le cas des plasmides mobilisables et conjugatifs, l'étape de mobilisation fait intervenir un complexe protéique appelé relaxosome qui reconnait un site *nic* situé dans l'*oriT* (origine de transfert) de la molécule à transférer. Le relaxosome est constitué d'une relaxase possédant une activité

phosphodiesterase, et de protéines auxiliaires de transfert. Ces dernières se lient à la région contenant *oriT* et permettent à la relaxase d'accéder au site *nic* afin de le cliver. Le clivage se fait par une transestérification qui aboutit à une liaison covalente entre la protéine et l'ADN. La synthèse d'ADN de la cellule donneuse débute ensuite au niveau de l'extrémité 3' du site clivé. Un domaine hélicase permet la séparation des brins et la nucléoprotéine transfère l'ADN vers la souche receveuse (de la Cruz et al., 2010).

Une protéine de couplage aidée par les protéines auxiliaires se lie ensuite au relaxosome au niveau de l'*oriT* pour l'amener jusqu'au T4SS. Le mécanisme de transport à travers le canal du T4SS n'est pas connu, mais le relaxosome est supposé être dénaturé, sans quoi sa taille serait trop importante pour permettre son passage. Une fois le complexe nucléoprotéique recruté au niveau du T4SS par la protéine de couplage et transféré à VirB11, le substrat est transféré à VirB6 et VirB8. Ce processus est assuré par VirB4.

Une fois que la relaxase liée à l'ADN passe à travers le canal, VirD4 entoure l'ADN sous forme hexamérique et utilise l'énergie issue de l'hydrolyse de l'ATP pour le transférer à travers le canal. Après le transfert de l'ADN dans la cellule receveuse, la relaxase reconnaitrait le site *nic* puis accomplirait une deuxième transestérification recircularisant ainsi l'ADN transféré (de la Cruz et al., 2010).

4 - La transduction

a. Phages lytiques et lysogéniques

La transduction est un mécanisme de transfert génétique entre deux cellules par l'intermédiaire d'un bactériophage. Les bactériophages (phages) sont des virus infectant les bactéries en s'adsorbant dessus puis en leur injectant leur génome. Les virus représentent l'entité biologique la plus abondante sur Terre (Suttle, 2007). Leur forme extracellulaire, le virion, est constituée d'une enveloppe de nature protéique et parfois lipidique, la capside, contenant un génome pouvant être constitué d'ADN, d'ARN, simple ou double brin. La grande majorité des virions appartiennent à l'ordre des Caudovirales, caractérisé par une symétrie dite "tête queue" comme c'est le cas des phages lambda ou T4 d'*E. coli* par exemple. Après l'infection, deux cycles majeurs sont possibles suivant les phages : le cycle lytique et le cycle lysogénique. Les virus lytiques exploitent la machinerie cellulaire de l'hôte pour se répliquer et former de nouvelles particules, puis lysent la cellule afin de se disséminer. Dans le cas des virus dits lysogéniques, le génome viral s'intègre dans le génome hôte (prophage), et peut ainsi être transmis verticalement de génération en génération. Dans certaines conditions (ex : réponse

SOS), le génome viral s'excise du génome hôte puis entame un cycle lytique.

La transduction est causée par un défaut d'empaquetage du génome viral dans la capside lors de la formation des virions. Cette étape est réalisée par une terminase composée d'une petite sous-unité TerSΦ, qui se lie à une séquence spécifique de l'ADN viral, et d'une grande sous-unité TerL qui coupe cette séquence (Rao and Feiss, 2008). Le complexe ADN-terminase se lie ensuite à la procapside et l'ADN est transféré dans cette dernière. TerL possède une activité ATPase produisant l'énergie nécessaire au transfert de l'ADN viral.

Il existe deux cas d'empaquetage, la différence majeure résidant dans la phase finale. Dans un cas, TerL coupe à un site identique à celui du premier clivage, appelé *cos*, et la taille du génome encapsidé est toujours la même (ex : phage lambda). Dans le second cas, TerL effectue une première coupure au niveau d'un site *pac*, puis une deuxième lorsque la tête du phage est pleine (*Headfull packaging* ; ex : phage T4), et la taille du fragment encapsidé peut dépasser celle du génome viral (Penadés et al., 2015) (Figure 8).

b. La transduction générale et spécialisée

La transduction généralisée correspond à l'empaquetage d'un fragment du génome de l'hôte, et le virion est alors qualifié de particule transductrice. En théorie, n'importe quelle partie du génome peut être transduite. D'une manière générale, ce type de transduction est réalisée par des phages de type *pac*, mais des cas impliquant des phages de types *cos* ont également été décrits (Quiles-Puchalt et al., 2014). Théoriquement, la proportion de particules transductrices est faible, de l'ordre de 10^{-9} à 10^{-7} par infection. Cependant, des éléments mobiles tels que certains îlots génomiques ou plasmides exploitent ce mécanisme pour être disséminés. Dans ce cas, le nombre de particules transductrices est bien plus élevé. Par exemple, chez *Marinitoga piezophila*, le virus MPV1 encapside préférentiellement le plasmide pMPA (13,3 Kpb) à son propre génome (43,7 Kpb) (Lossouarn et al., 2015). Chez *Staphylococcus aureus*, des îlots génomiques appelées SaPIs (*S. aureus Pathogenicity Island*s) sont également connus pour leur capacité à être transférés horizontalement par phages (Maiques et al., 2007).

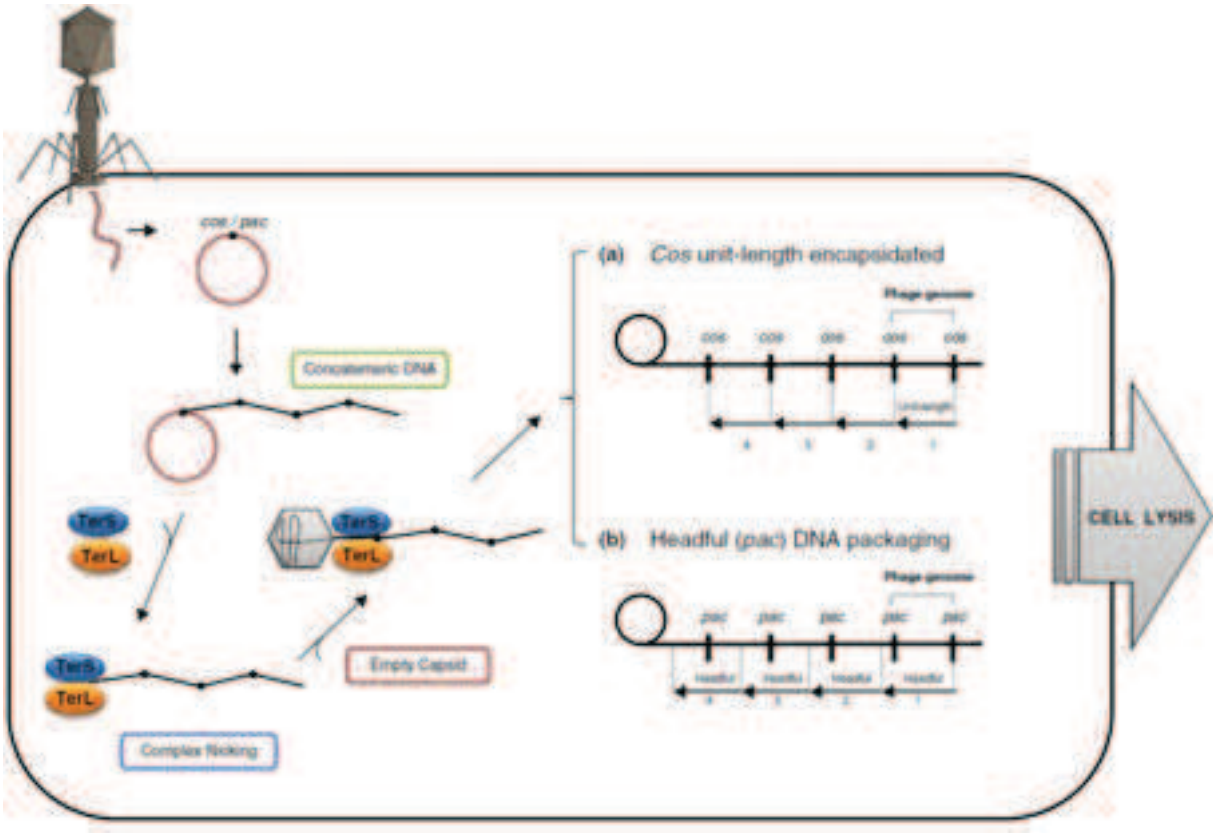


Figure 8 Schéma des deux types d'encapsidation "cos" et "pac". Après son entrée dans la bactérie, le génome du phage est répliqué en un grand nombre de copies sous forme d'un concatémère (Concatemeric DNA). Cette molécule est ensuite prise en charge par la terminase, constituée d'une petite sous-unité TerS (bleu) et d'une grande sous-unité TerL (orange). Dans le cas des phages de type "cos" (a), le terminase découpe le concatémère en fragments de longueur égales, alors que dans le cas des types "pac" (b), la première coupure est effectuée au niveau d'un site pac, puis les autres lorsque la tête de la capside est pleine (Headful). Une fois les virions formés, la cellule est lysée pour les libérer. D'après Penadès et al., 2015.

La transduction spécialisée concerne les phages lysogéniques et est causée par l'excision imprécise du prophage. Dans ce cas, seules les parties du génome hôte flaquant le site d'intégration du prophage peuvent être transduites. Le cas le mieux décrit est celui du phage λ d'*E. coli* (Rothman, 1965).

5 - Vésicules et nanotubes

En plus des mécanismes classiques décrits ci-dessus, des études récentes ont montré que des HGTs pouvaient potentiellement avoir lieu via des vésicules de membranes ou des structures tubulaires appelées "nanotubes". Les vésicules de membranes sont composées de lipides, de protéines et d'acides nucléiques. Elles sont formées par bourgeonnement de la

membrane externe chez les bactéries Gram-négatif et peuvent mesurer en moyenne de 20 à 200 nm. Elles ont été décrites pour le première fois en 1965 chez *E. coli* (Bishop and Work, 1965). Depuis, de nombreuses descriptions ont été effectuées chez les bactéries Gram-négatif (Schwechheimer and Kuehn, 2015). Cependant, bien que les bactéries Gram-positif soient dépourvues de membrane externe, des vésicules ont également été décrites chez ces dernières (Brown et al., 2014; Lee et al., 2009).

D'un point de vue fonctionnel, les vésicules ont différents rôles, et sont notamment impliquées dans les interactions avec d'autres bactéries et avec leur hôte (Kim et al., 2015). Dans certains cas, les vésicules de membranes sont supposées être impliquées dans des transferts horizontaux, ou du moins les favoriser. Il a en effet été montré que des vésicules de *Neisseria gonorrhoeae* et *E. coli* contiennent de l'ADN pouvant être transféré (Dorward et al., 1989; Kolling and Matthews, 1999). Cependant, l'implication directe des vésicules dans le transfert reste à démontrer.

Récemment, un nouveau type de structure a également été décrit comme étant impliqué dans la communication cellulaire, notamment dans les HGTs. Il s'agit de nanotubes formant des ponts intercellulaires chez *Bacillus subtilis* (Dubey and Ben-Yehuda, 2011). Il a été démontré que cette structure pouvait permettre le transfert de molécules non-héritables, mais aussi permettre le transfert d'un plasmide conjugatif. De plus, ces structures peuvent également se former de manière interspécifique, entre *B. subtilis* et *E. coli* ou *Staphylococcus aureus* par exemple.

III - Les éléments génétiques mobiles

Les éléments génétiques mobiles (EGMs) sont des fragments d'ADN contenant des gènes codant des produits permettant leur mobilité intracellulaire, ou intercellulaire (via HGTs) (Frost et al., 2005). Il existe une multitude d'EGMs différents suivant leur capacité d'intégration dans le chromosome hôte, de transfert, leur stratégie de maintien etc. Les différentes fonctions des EGM (mobilisation, réplication, transfert ...) sont codées par des gènes regroupés en cluster appelés modules (ex : module de transfert) (Toussaint and Merlin, 2002). Cette organisation modulaire est importante car la diversité des EGMs s'explique en partie par leur évolution via l'échange de ces modules par recombinaison.

1 - Les EGMs non-autonomes pour leur transfert

a. Les séquences d'insertion

i. Généralités

Les séquences d'insertion (IS, *Insertion Sequence*) sont les éléments transposables autonomes les plus simples connus. Elles mesurent de 0,7 à 2,5 Kpb, contiennent un ou deux ORFs et sont flanquées de courtes séquences répétées inversées (IR) (Siguier et al., 2006). Elles sont généralement retrouvées dans des régions inter-géniques entre des gènes orientés dans la même direction. Elles sont très répandues à la fois chez les archées et les bactéries.

Ces éléments ont été classés en une vingtaine de familles sur la base de la similarité de leur transposase et de leurs séquences IR. Les classes majeures sont définies sur la base du type de transposase (Siguier et al., 2014). Les IS "classiques" possèdent une transposase à domaine DDE, mais il en existe également à domaines DEDD et HUH, ou encore possédant une transposase à sérine (Figure 9). Les transposons (Tn) ont été différenciés des IS parce qu'ils portaient des gènes en plus de la structure de base, n'étant pas impliqués dans leur mobilité ou la régulation de leur mobilité. Cependant, il s'avère aujourd'hui que la frontière entre transposons et IS est très floue. Il existe des éléments dérivés des IS qui, bien que plus simple que les transposons, possèdent des gènes supplémentaires pouvant être par exemple impliqués dans la résistance à des antibiotiques (ex : famille *IS1595* (Siguier et al., 2009)).

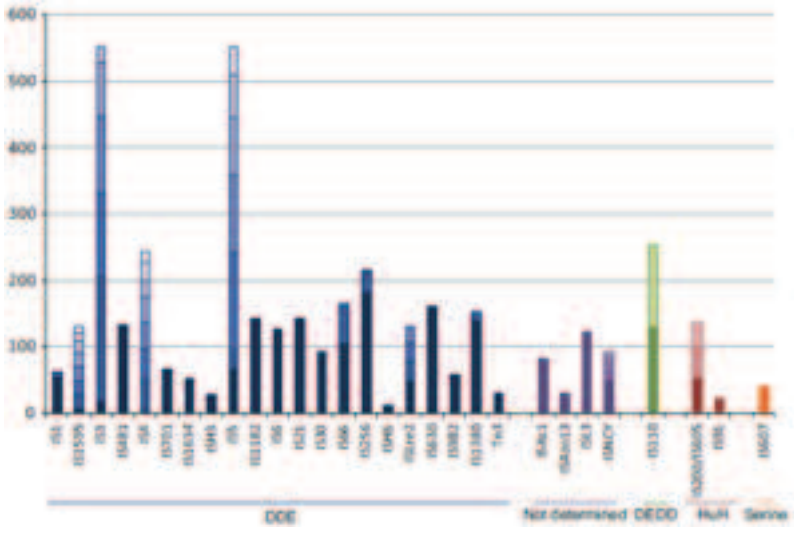


Figure 9 Nombre d'IS (ordonnée) par famille (abscisse) et type de transposase associée (en bas). D'après Siguier et al., 2014.

b. Les intégrons

Les intégrons sont des éléments relativement communs dans les génomes bactériens et considérés comme des hot-spots de variabilité génomique (Gillings, 2014). Ces éléments ont la particularité de capturer puis réarranger des ORFs dépourvus de promoteurs en opéron, s'appropriant ainsi l'expression de ces ORFs. Les intégrons sont composés d'une plateforme "stable" dans laquelle sont intégrés des cassettes de gènes. La partie stable de l'intégron est généralement composé du gène *intI* codant une recombinase à tyrosine site-spécifique, d'un site de recombinaison *attI* en amont du promoteur de *intI*, et du promoteur P_c, localisé dans *intI* ou *attI* (Figure 10).

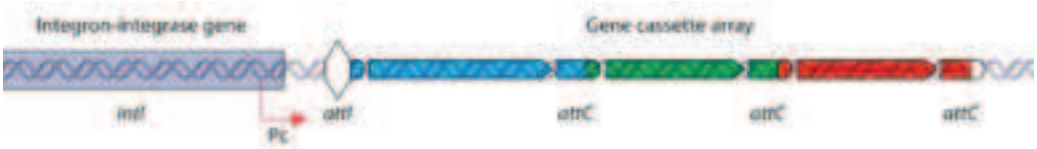


Figure 10 Schéma de l'organisation générale d'un intégron. D'après Gillings 2014.

Les cassettes de gènes sont des molécules d'ADN circulaire libres, contenant généralement un ORF dépourvu de promoteur, et un site de recombinaison *attC* dont la taille varie de 57 à 141 bp. Les cassettes sont intégrées par recombinaison site-spécifique (IntI) entre leur site *attC* et le site *attI* de l'intégron. La cassette de gènes est ainsi intégrée en aval du promoteur Pc dans l'orientation permettant en principe son expression. Les gènes suivants seront également insérés après le promoteur Pc. Plus les gènes sont éloignés du promoteur, et moins ils ont de chance d'être exprimés. Ainsi, les gènes acquis le plus récemment ont plus de chance d'être exprimés que les anciens.

Les intégrons sont considéré comme un hot-spot de diversité génomique (Hall, 2012) et ont donc un rôle important dans l'adaptation et l'évolution des bactéries. Ils permettent l'acquisition de nouvelles fonctions et participent ainsi à la diversification des populations. Ils sont par exemple à l'origine de variabilités intra-spécifiques chez *Pseudomonas stutzeri* (Holmes et al., 2003). Certains sont portés par des plasmides et peuvent ainsi être disséminés par HGTs. Ils peuvent également participer à la diversification en réponse à des stress causant des dommages à l'ADN, l'intégrase pouvant être induite lors de la réponse SOS (Guerin et al., 2009).

2 - Les EGMs autonomes pour leur transfert

a. Les plasmides conjugatifs

Les plasmides conjugatifs codent un système de conjugaison et sont donc autonomes pour leur transfert. Il est estimé qu'ils représentent environ un quart des plasmides connus (Smillie et al., 2010). Leur nombre de copies par cellule est généralement faible (~10 copies / cellule). Comme les autres EGMs, leur structure est modulaire : les gènes forment des clusters correspondant à des groupes fonctionnels (Norman et al., 2009). Certains de ces modules sont essentiels au maintien des plasmides (stabilité, réplication, transfert) et sont conservés (Figure 11), alors que d'autres sont accessoires et constituent une part variable.

La part variable des génomes plasmidiques, qui peut également être appelé "module d'adaptation", codent des traits pouvant être avantageux pour l'hôte dans certaines conditions. Les plasmides ont d'ailleurs été très étudiés à cause de leur implication dans l'émergence de résistance aux antibiotiques et aux métaux. Cette part variable peut contenir des éléments génétiques mobiles non autonomes pour leur transfert, tels que des éléments IS ou des intégrons.

Figure 11 Schéma de l'organisation modulaire d'un plasmide conjugatif. D'après Norman et al., 2009

b. Les bactériophages

Les bactériophages sont des virus, entité biologique la plus abondante sur Terre (Suttle, 2007), infectant des bactéries. Ils possèdent une forme extracellulaire, le virion, constituée d'une enveloppe généralement protéique contenant le génome. Lors de l'infection d'une bactérie, ce dernier est injecté par le phage dans le cytoplasme et deux cycles principaux peuvent se produire: lytique où de nouveaux virions sont immédiatement formés et libérés par lyse de la cellule, et lysogénique où le génome viral est intégré dans le génome de l'hôte sous la forme d'un prophage (Cf. partie II -4 - page 27).

c. Les Agents de Transfert de Gènes

Les agents de transfert de gènes (GTAs, *Gene Tranfer Agents*) sont des éléments semblables à des phages. Leur forme extracellulaire est une particule similaire aux caudovirales et contient un ADN double brin transmissible à une cellule. Ils possèdent cependant des caractéristiques particulières : l'ADN contenu dans leur capside est un fragment aléatoire du génome de leur hôte qui les a produit, et la quantité d'ADN est insuffisante pour coder la production et l'assemblage des éléments de la particule (Lang et al., 2012). Ainsi, l'ADN transmis par un GTA ne permet pas à la cellule receveuse de produire de nouveaux GTAs. Ceci les distingue des phages réalisant la transduction généralisée, où la majeure partie des virions contiennent le génome viral (Figure 12).

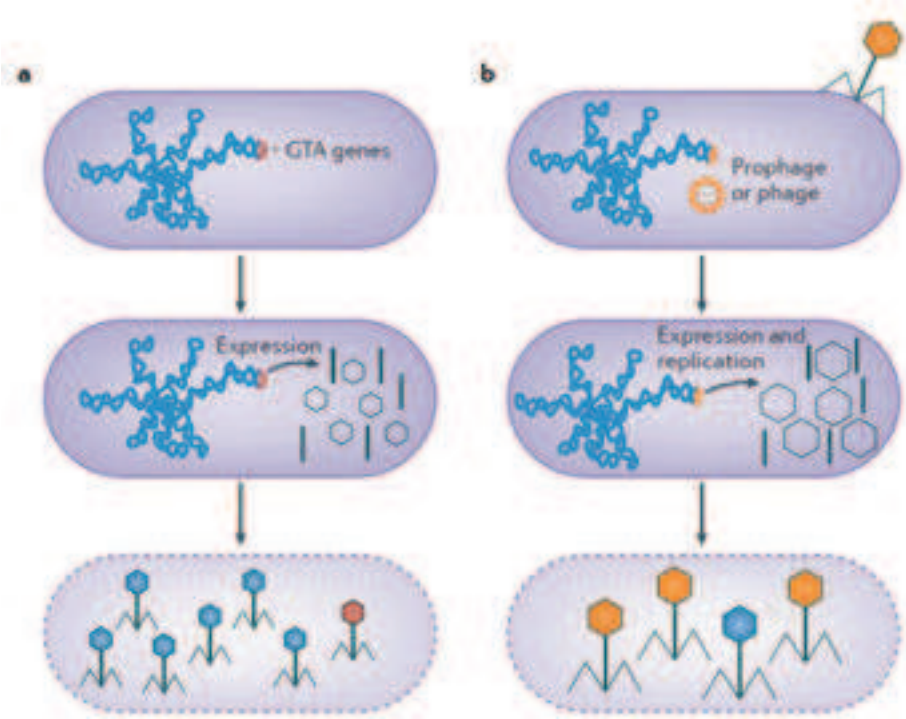


Figure 12 Schéma du cycle des GTAs (gauche) et de la transduction généralisée (droite). Dans le cas de GTAs, la majorité des particules formées contiennent des fragments du génome de l'hôte cellulaire (bleu) et quelques unes contiennent une partie de leur génome (rouge). Dans le cas de la transduction généralisée, la majorité des particules contiennent le génome viral et quelques une contiennent un fragment du génome de l'hôte cellulaire. D'après Lang et al., 2012.

Quatre GTAs différents ont actuellement été identifiés : le RcGTA chez *Rhodobacter capsulatus*, le Dd1 chez *Desulfovibrio desulfuricans*, le VSH-1 (virus of *Serpulina hyodysenteriae*) chez *Brachyspira hyodysenteriae* et le VTA (Voltae transfer Agent) chez *Methanococcus voltae*. L'organisation génétique de la partie du génome codant la capside de

ces éléments est conservée avec les génomes des phages. Ils semblent donc liés d'un point de vue évolutif. Cette similarité laisse imaginer que la liaison du GTA à la cellule hôte implique une interaction spécifique entre la queue de la particule et un récepteur membranaire.

d. Les Eléments Intégratifs et Conjugatifs (ICEs)

i. Définition

Les éléments intégratifs et conjugatifs (ICEs, *Integrative and Conjugative Elements*) sont des éléments à la fois capables de transfert par conjugaison, et d'intégration/excision dans le génome de leur hôte (Wozniak and Waldor, 2010). Cette terminologie a été proposée par Burrus et al. en 2002, afin d'unifier la nomenclature de tous les éléments possédant les caractéristiques précédemment citées (Burrus et al., 2002). Ceci a permis de regrouper les éléments anciennement appelés "transposons conjugatifs" et "plasmides intégratifs".

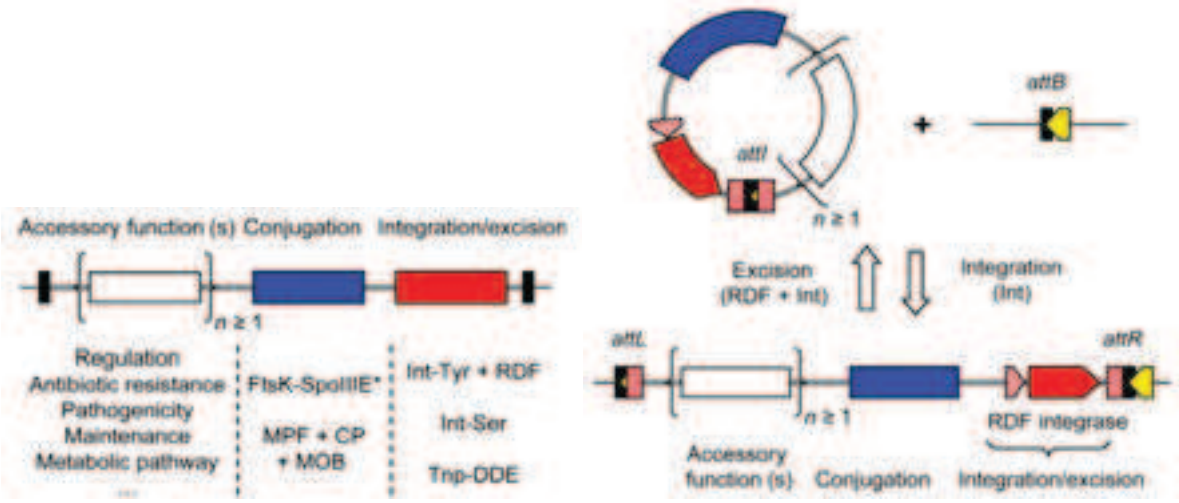


Figure 13 Organisation modulaire d'un ICE (à gauche) ; forme excisée et intégrée (à droite) d'un ICE. Noir : séquences répétées. Blanc : module de fonctions accessoires. Bleu = module de transfert par conjugaison ; FtsK-SpoIIIE = module de transfert d'ADN double-brin. MPF +CP + MOB = module de trandsfert d'ADN simple brin ; MPF = système de formation du pore de conjugaison (Mating Pair Formation system) ; CP = Protéine de Couplage (Coupling Protein); MOB = relaxase. Rouge = module d'intégration/excision. Int-Tyr = Intégrase à tyrosine ; RDF = Facteur de directionalité (Recombination Directionality Factor) ; Int-Ser = Intégrase à Sérine ; Tnp DDE = Transposase à DDE. attI, attB attL et attR = sites d'attachements. D'après Bellanger et al., 2014.

ii. Cycle de vie

Les ICEs codent leur propre intégrase, généralement une recombinase à tyrosine,

nécessaire à leur intégration site-spécifique dans le génome hôte. Il existe cependant des ICEs utilisant des recombinases à sérine ou domaine DDE (Brochet et al., 2009). L'intégration est réalisée par recombinason entre le site *attI* de l'ICE et un site *attB* du chromosome, généralement dans la région 3' d'un ARNt ou de gènes très conservés. L'intégration entraine la formation des sites *attL* et *attR* flanquant l'ICE. L'excision, qui requiert également l'intégrase, est effectuée via une recombinaison entre les sites *attR* et *attL*. La plupart du temps, le site d'intégration est laissé intact. Souvent, des facteurs de directionnalité (*RDF, Recombination Directionality Factor*) ou excisionase, interviennent pour favoriser l'excision.

Le transfert par conjugaison semble être similaire à celui des plasmides conjugatifs. Des séquences *oriT* ont été retrouvées dans certains ICEs, dont certaines prouvées expérimentalement. De plus, la plupart des ICEs contiennent des gènes homologues au T4SS, suggérant un transfert de l'ADN via ce système. Une fois transféré sous forme simple brin, le deuxième brin est synthétisé sous l'action d'une polymérase. Il existe des cas d'ICEs également capables de réplication (RICEs, *Replicating ICEs*) (Guérillot et al., 2013).

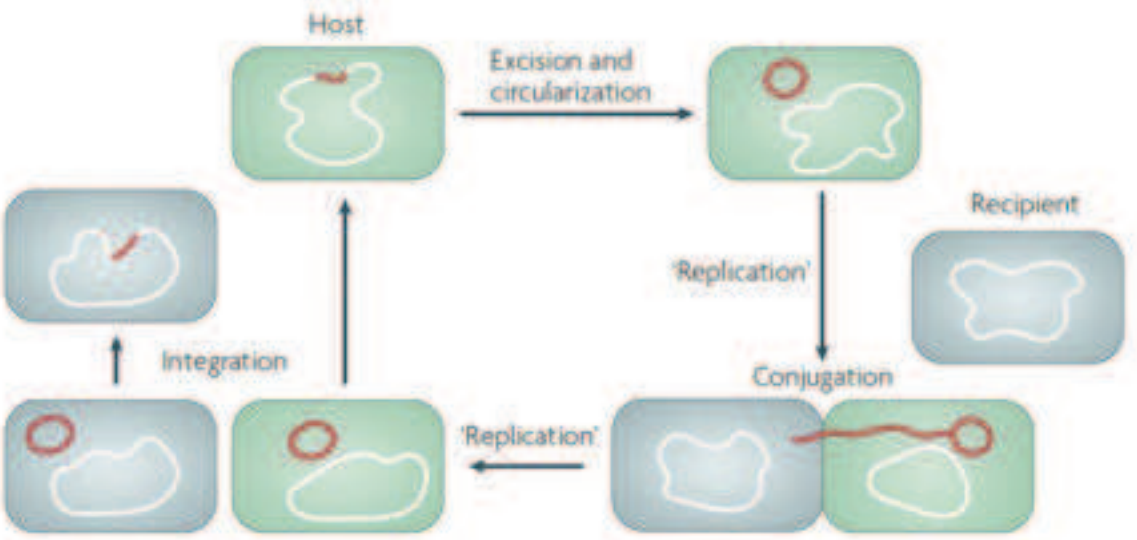


Figure 14 Schéma du cycle de vie d'un ICE. D'après Wozniak & Waldor 2010.

iii. Les éléments mobilisables intégratifs

Comme décrit dans la partie précédente, les ICEs sont une classe d'éléments diverse du fait de l'existence de différents modules de conjugaison ou de mobilisation notamment. En plus des différentes combinaisons de modules, d'autres éléments sont caractérisés par l'absence de certaines fonctions. C'est le cas des éléments mobilisables intégratifs (IME, *Integrative and Mobilizable Elements*) pour lesquels le module de transfert est en partie ou totalement absent

(Figure 15). En conséquence, ce type d'élément peut s'exciser ou s'intégrer dans le génome mais est incapable de transfert de manière autonome. Ils peuvent cependant être mobilisés en *trans* et transférés via le système de conjugaison d'un ICE ou d'un plasmide conjugatif.

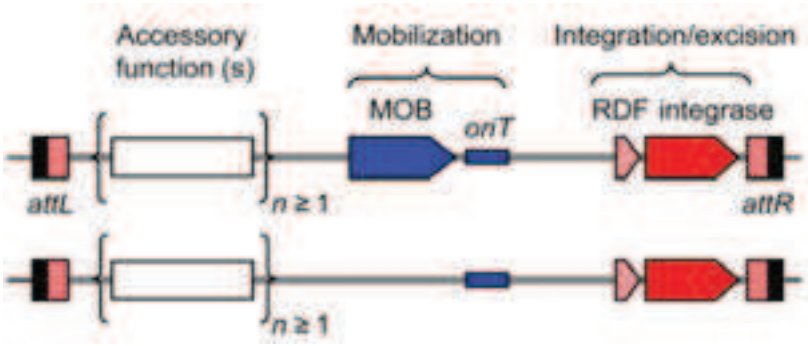


Figure 15 Schéma de l'organisation d'un IME. Noir : séquences répétées. Blanc : module de fonctions accessoires. Bleu = module de transfert par conjugaison ; MOB = relaxase. Rouge = module d'intégration/excision ; RDF = Facteur de directionalité (Recombination Directionality Factor). attL et attR =sites d'attachements. D'après Bellanger et al., 2014.

De la même manière que les ICEs, les modules d'intégration et de mobilisation sont divers. Suivant les cas, l'intégration/excision peut être réalisée par une recombinase à tyrosine, recombinase à sérine, ou recombinase à DDE, et le site d'intégration peut être très spécifique ou peu spécifique. Pour certains IMEs possédant une faible spécificité pour leur site d'intégration, plusieurs copies intégrées dans le génome peuvent être retrouvées. Leur mécanisme de mobilisation est similaire à celui des plasmides mobilisables. Ils possèdent un site *oriT* et codent la plupart du temps leur propre relaxase ainsi que d'autres protéines du relaxosome. L'existence de mécanismes de maintien de la forme circulaire n'est pas connue mais comme pour les ICEs, certains IMEs contiennent des gènes homologues à ceux impliqués dans la réplication et la répartition des plasmides. Certains IMEs codent des fonctions pouvant être avantageuses pour leur hôte, tels que des gènes de résistance aux antibiotiques.

IV - Réarrangements génomiques : les mécanismes

1 - La recombinaison et la transposition

a. Recombinaison homologue

La recombinaison homologue est un mécanisme d'échange de fragments d'ADN dont les séquences sont proches ou identiques. Ces fragments peuvent être présents sur le même chromosome ou sur deux molécules d'ADN différentes. Une séquence homologue de 20 à 100 pb minimum est nécessaire à ce type de recombinaison.

La recombinaison homologue fait partie de la réponse SOS et est impliquée dans le maintien de l'intégrité du génome en participant à la réparation des cassures double et simple-brin de l'ADN. Ces cassures de l'ADN peuvent être dues à des arrêts de la fourche de réplication suite à une collision entre l'ADN polymérase et l'ARN polymérase par exemple, ou à des agents mutagènes tels que les rayonnements ultra-violets (U.V) ou le stress oxydant. Ce type de recombinaison, associé à la présence de répétitions de grande taille, peut également causer d'importants réarrangements intragénomiques. De plus, elle peut également permettre l'intégration d'ADN étranger acquis par transfert génétique horizontal. Dans le cas de souches apparentées, cela peut entrainer des recombinaisons alléliques.

Deux mécanismes de recombinaison homologue ont été décrits : les voies RecBCD-RecA et RecFOR-RecA respectivement impliquées dans la réparation des cassures double brins et simple brins de l'ADN (Figure 16). Dans chacune de ces voies, RecA joue un rôle majeur en se liant à l'ADN simple-brin pour former un filament spiralé et permettre l'invasion d'une région double-brin homologue. Dans la voie RecA-RecBCD, RecBCD se lie à l'ADN double-brin, le déroule et le dégrade jusqu'à rencontrer un site *chi*. A partir de ce site, RecBCD génère de l'ADNsb et participe à la fixation de RecA sur l'ADN. Dans la voie RecA-RecFOR, les protéines RecF, RecO, et RecR se lient à de l'ADNsb formé par une césure, puis déplacent les protéines SSB (*Single Strand Binding*), qui stabilisent l'ADN sous forme simple-brin, afin de permettre la fixation de RecA sur l'ADNsb. RecA permet ensuite l'invasion de l'ADNdb et la recherche d'homologie afin de former une jonction de Holliday. Un complexe protéique RuvABC catalyse ensuite la migration des brins et la résolution de la jonction. Chez *E. coli*, cette résolution peut également être réalisée par l'hélicase RecG par un mécanisme restant inconnu.

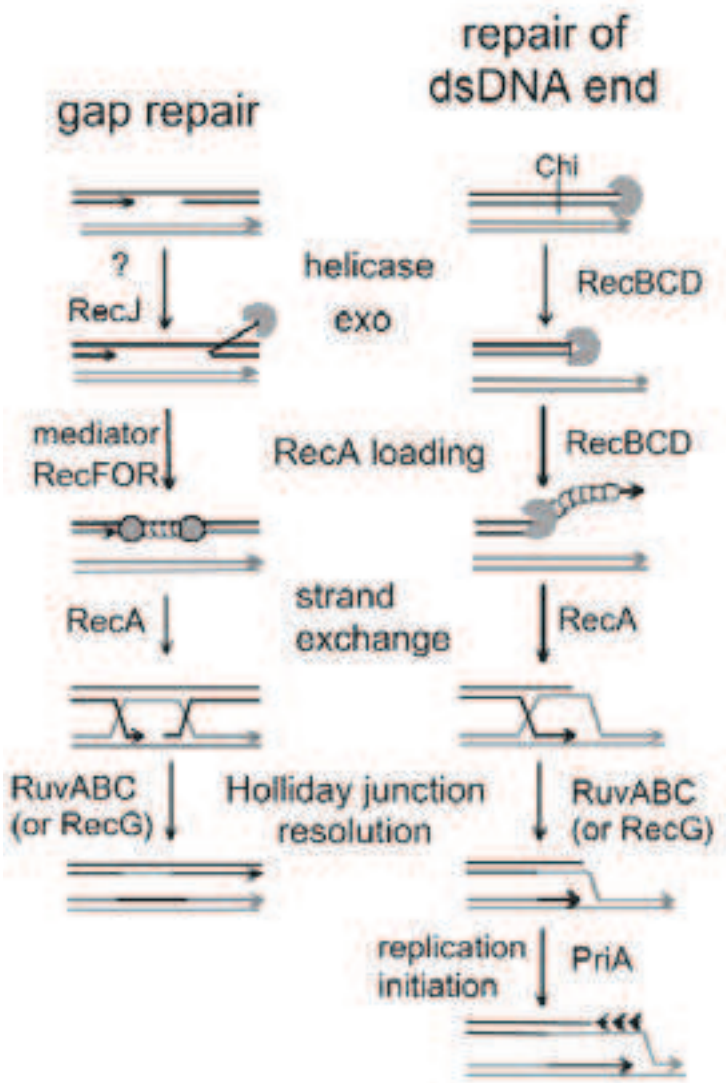


Figure 16 Schéma des voies RecBCD et RecFOR de la recombinaison homologue. D'après Rocha et al., 2005.

b. Recombinaison illégitime

Ce type de recombinaison regroupe différentes réactions se produisant généralement entre des séquences d'ADN proches physiquement mais ne partageant que peu voire aucune homologie (Ehrlich et al., 1993). Le relâchement de quelques paires de bases au niveau de sites ayant une micro-homologie favorise l'hybridation des brins d'ADN. Cette hybridation a lieu au niveau d'extrémités résultant soit de la synthèse d'ADN (cas de glissement de brins) soit de cassures (cas de jonctions d'extrémités non-homologues). Tout comme la recombinaison homologue, ce processus est RecA dépendent.

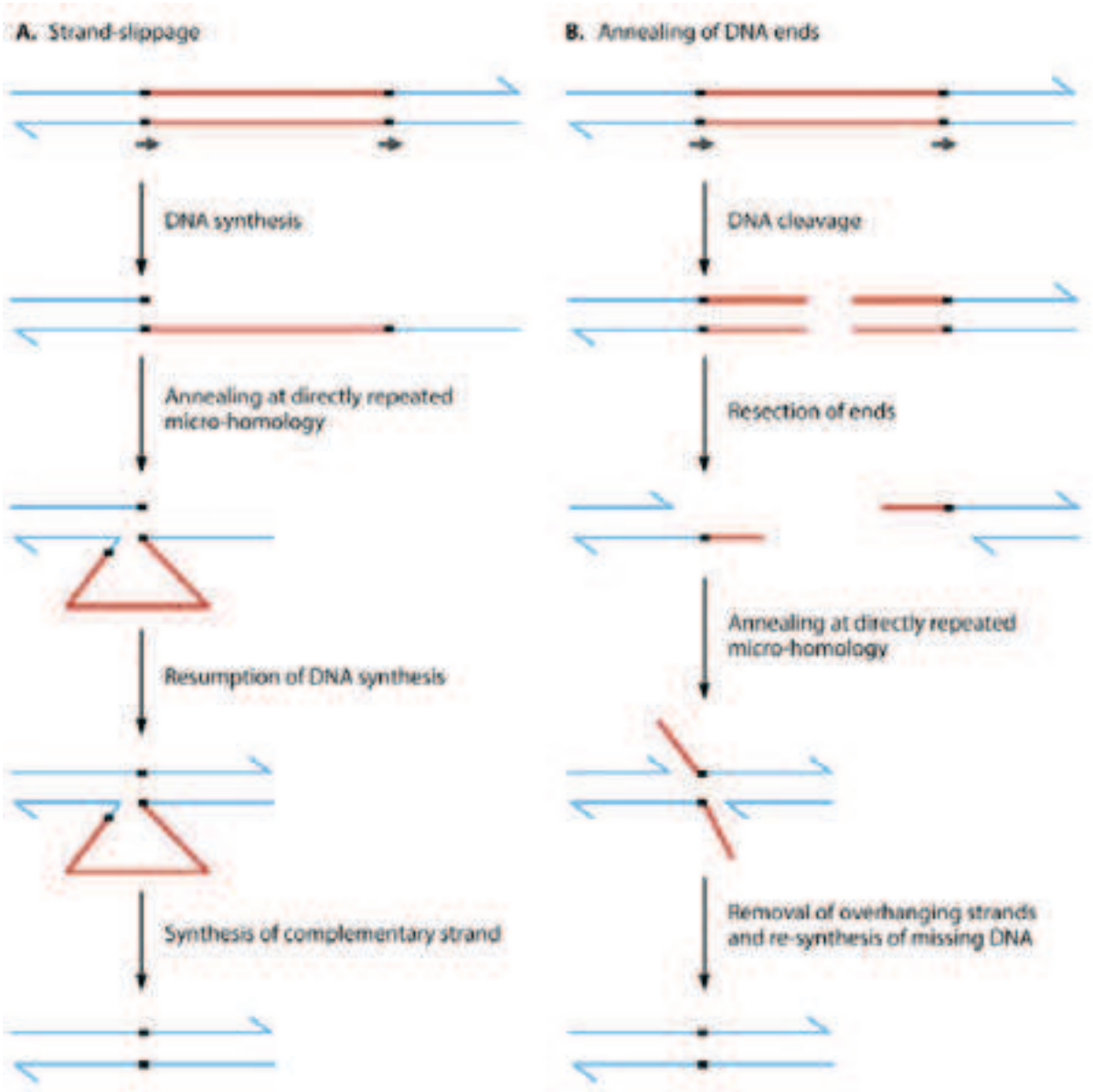


Figure 17 Schéma des deux types de recombinaisons illégitimes. A = Glissement de brin. B = Jonction d'extrémités non_homologues. Flèches noires horizontales = micro-homologies. Rouges = région d'ADN flanquées par les micro_homologies. D'après Darmon et al., 2014.

Les recombinaisons illégitimes (Figure 17) se produisant lors de la synthèse d'ADN peuvent localement causer des délétions, duplications et conversions de séquences. Cela peut par exemple se produire lors de la réplication de l'ADN (Michel and Ehrlich, 1986), mais aussi en conséquence d'irradiation UV ou à la suite de transformation avec une séquence d'ADN linéaire pour laquelle la recombinaison homologue n'est pas possible (Conley et al., 1986).

La jonction simple brin à la suite de cassures de l'ADN peut se produire après dégradation de l'ADN, entrainant une délétion locale. Des mutations au niveau de l'ADN

polymérase III ou du système de réparation des mésappariements augmentent le taux de recombinaisons illégitimes. Certaines bactéries possèdent un système de jonction d'extrémités non-homologues (*NHEJ, Non-Homologous End-Joining*) composé de deux protéines, Ku et LigD, qui agissent ensemble afin de joindre les extrémités d'ADN comportant des micro-homologies (Shuman and Glickman, 2007). Ce système revêt une importance particulière dans la réparation de l'ADN en cas d'absence d'une matrice d'ADN homologue permettant une réparation par recombinaison homologue.

c. Recombinaison site-spécifique

La recombinaison site-spécifique est un mécanisme d'échange d'ADN entre deux sites définis situés aux extrémités de la séquence d'ADN concernée. Ce processus nécessite l'intervention d'une recombinase et est qualifié de conservatif, car il n'y a ni perte ni gain d'ADN et d'énergie (Grindley et al., 2006). Il existe deux principaux types de recombinases impliquées dans la recombinaison site-spécifique : les recombinase à tyrosine et les recombinase à sérine. Les recombinase à sérine clivent les deux brins d'ADN avant le transfert (Reed and Grindley, 1981), alors que les recombinase à tyrosine clivent d'abord un brin, puis le deuxième uniquement lorsque le transfert de brin est initié (Kitts and Nash, 1987).

Ce type de recombinaison peut causer l'inversion, l'excision ou l'intégration dans le génome d'une séquence d'ADN, l'issue étant déterminée par l'orientation des deux sites. S'ils sont orientés dans le même sens, il se produira l'excision de l'élément ou bien son intégration, si ceux-ci sont situés sur deux molécules différentes. S'ils sont orientés en sens inverse, il se produira une inversion de la séquence d'ADN (Figure 18).

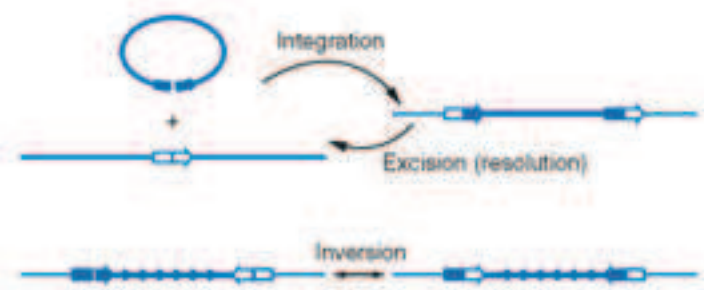


Figure 18 Schéma des différentes issues de la recombinaison site-spécifique. D'après Grindley et al., 2006.

Dans un premier temps, la recombinase se lie aux deux sites et les rapproche, formant un complexe synaptique avec les deux sites de cross-over juxtaposés. La recombinase catalyse ensuite le clivage, l'échange et la ligation des brins. Le complexe synaptique est alors défait et

les produits de la recombinaison sont libérés (Grindley et al., 2006). Les sites de recombinaison les plus simples mesurent 20-30 pb et contiennent une paire de séquences inversées qui se lient à la recombinase. Ces sites de reconnaissance contiennent le point de cassure / ligation et sont considérés comme site de cross-over.

Les systèmes d'inversions sites-spécifiques sont largement répandus chez les bactéries, que ce soit au niveau du chromosome ou d'éléments génétiques mobiles. Ces évènements se produisent généralement à une fréquence de 10^{-3} à 10^{-5} par cellule et par génération, et la taille des fragments concernés varie entre une centaine de pb et 35 Kpb. L'inversion site-spécifique est connue pour conférer un avantage aux bactéries en induisant une diversité phénotypique au sein d'une population clonale. Ces mécanismes sont appelés variation de phase, ou variation antigénique. Les mécanismes d'excision et d'intégration sont utilisés par certains éléments génétiques mobiles tels que les phages et les éléments conjugatifs et intégratifs, le mieux décrit étant le phage lambda d'*E. coli* (Cf. partie III-2). Cependant, ils peuvent également participer à des processus de différentiation. A titre d'exemple, la sporulation de *Bacillus subtilis* ou la formation d'hétérocystes (cellules spécialisées dans la fixation d'azote atmosphérique) chez les cyanobactéries filamenteuses reposent sur l'excision d'un élément par recombinaison site-spécifique. Dans les deux cas, l'élément est situé dans un gène essentiel pour la différenciation, conduisant à un arrêt d'expression.

2 - Impacts des EGMs sur les remaniements génomiques

a. Transposition

La transposition est un processus utilisé par les éléments transposables (TE, *Transposable Element*) tels que les transposons et les IS (Partie III -1 -a, page 31) afin de se déplacer d'un locus à un autre. Ces éléments ne nécessitent pas une importante homologie avec leur site cible. Le mécanisme de transposition, qui implique des coupures de l'ADN aux extrémités de l'élément, est principalement réalisé via deux mécanismes généraux. L'intermédiaire de transposition peut être soit de l'ADN, soit de l'ARN (rétrotransposons). Alors qu'une importante proportion d'éléments transposables utilise un intermédiaire ARN chez les eucaryotes, la plupart utilisent un intermédiaire ADN chez les procaryotes (Siguier et al., 2014). Deux cas de figure sont possibles. Dans un premier cas, la transposition est qualifiée de conservative, elle implique une coupure double-brin au niveau des extrémités de l'élément, puis sa transposition qui est dite "couper-coller". Dans le deuxième cas, la transposition est dite réplicative, car elle entraine la duplication de l'élément (mécanisme de "copier-coller"), et

n'implique qu'une césure simple brin aux extrémités de l'élément.

Le processus de transposition nécessite une transposase, généralement composée de plusieurs domaines. La transposase reconnait les extrémités du TE, les clive, puis forme un complexe nucléoprotéique avec le site d'intégration. Ce complexe nucléoprotéique, appelé le transposome, comprend un ou deux monomère(s) de transposase, et parfois des protéines accessoires impliquées dans l'architecture de l'ADN (ex : HU, IHF, H-NS) ou des protéines chaperons (ex : GroEL).

Une grande part des transposases sont des transposases à domaine DDE, tirant leur nom des résidus Asp(D) Asp(D) Glu(E) conservés et essentiels à la réaction de clivage des brins. Ces transposases clivent un brin d'ADN aux deux extrémités de l'élément, libérant généralement des extrémités 3'. Les extrémités libérées se lient ensuite au site cible, la réaction de transfert étant coordonnée par le transposome (Curcio and Derbyshire, 2003). Lors de la liaison au site cible, la coupure donne lieu à des extrémités sortantes dont la réparation crée des répétitions aux extrémités de l'élément, une caractéristique des transpositions médiées par des transposases DDE.

Bien que les transposases à domaine DDE réalisent toutes la même réaction, les mécanismes de transposition varient suivant les éléments. Par exemple, dans le cas de l'élément IS10, l'extrémité 3' libre envahit l'extrémité 5' ce qui forme une structure en "tête d'épingle" et permet l'excision de l'élément. Une deuxième excision des extrémités 3' permet l'obtention d'une forme linéaire qui envahit le site cible.

b. Impact des IS sur le génome hôte

L'insertion d'un élément IS peut modifier l'expression de certains gènes de l'hôte, avec des conséquences positives ou négatives suivant les gènes concernés et les conditions environnementales. L'insertion dans un gène ou dans sa séquence régulatrice peuvent conduire à son inactivation ou à son activation dans le cas d'une insertion en amont. Leur excision peut dans certains cas laisser le chromosome dans son état original, et rétablir ainsi le phénotype de départ, soit causer des mutations. Les IS peuvent également être à l'origine de remaniements génomiques. Par exemple, par recombinaison homologue d'un IS avec une autre copie du même IS.

Les GEIs contiennent souvent des éléments IS. Ces éléments peuvent entrainer des remaniements dans les îlots en formant des transposons composites ou entrainant des recombinaisons entre deux IS identiques. Ils peuvent aussi entrainer l'inactivation de certains

gènes par insertion.

Un phénomène marquant concernant l'impact des IS sur l'évolution des génomes est l'expansion. Ce phénomène correspond à l'augmentation du nombre d'éléments IS dans le génome, accompagné de remaniements génomiques ainsi que la perte et l'inactivation de certains gènes (Figure 19). Cette expansion a été observée dans des cas de bactéries ayant récemment adopté un mode vie symbiotique ou endosymbiotique. Selon certains auteurs, ce changement entrainerait un relâchement de la pression de sélection négative (purifiante) ce qui causerait une accumulation du nombre de mutations et l'augmentation du nombre d'IS (Moran and Plague, 2004). La vie endosymbiotique rend une partie des gènes des bactéries inutile à sa survie. Les éléments IS participeraient alors à la pseudogénisation et à la réduction du génome en favorisant des recombinaisons causant des délétions (Mira et al., 2001).

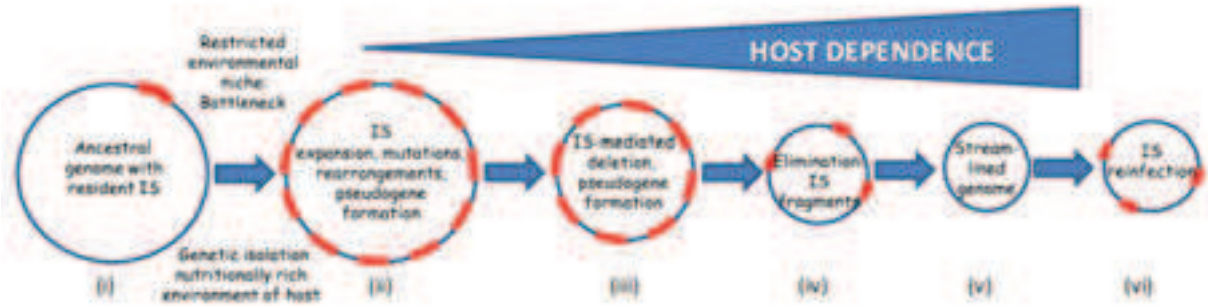


Figure 19 Schéma des étapes de l'expansion puis de l'élimination des éléments IS (rouge,) associés à une réduction de génome (bleu). D'après Siguier et al., 2014.

c. Impact des prophages sur les génomes bactériens

Les phages peuvent transférer des fragments de génome cellulaire par un processus appelé transduction. Les prophages et les séquences dérivées de prophages représentent une part non-négligeable des génomes bactériens (Bobay et al., 2013). Une conséquence de la présence de ces séquences sont des remaniements génomiques pouvant concerner des fragments de grande taille (Menouni et al., 2015). Un des cas les plus marquants est celui du pathogène *Xylella fastidiosa* où plus de dix remaniements identifiés entres différentes souches sont attribués à la présence de séquences dérivées de phages (de Mello Varani et al., 2008 ; Figure 20).

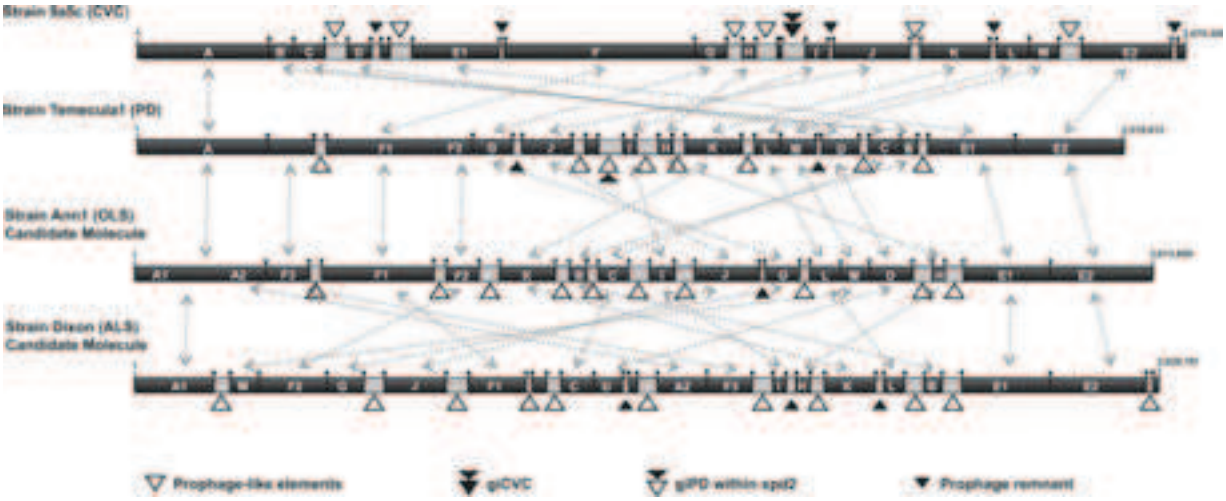


Figure 20 Séquences dérivées de prophages et remaniements génomiques chez des souches de X. fastidiosa. Flèches grises en pointillés = réarrangements génomiques entre les différentes souches. giCVC et gi PD = îlots génomiques présents dans les souches CVC et PD respectivement. D'après Mello Varani et al., 2008.

Les phages peuvent coder des recombinases qui ne sont pas retrouvées chez les bactéries. Ces recombinases sont supposées causer des remaniements génomiques par recombinaison homologues (Menouni et al., 2015). Contrairement aux remaniements évoqués ci-dessus, l'implication de ces recombinases est pour une large part hypothétique.

3 - Evolution des EGMs

a. Evolution de GEIs

i. Evolution modulaire et dégénérescence

Comme déjà évoqué précédemment pour les plasmides et les ICEs, l'échange de modules via des mécanismes de recombinaison participe à l'évolution des EGMs. Ces modules, mais également les sites de recombinaison, peuvent également être perdus, mutés, ou subir des remaniements délétères. En conséquence, certains ICEs ou IMEs peuvent perdre toute capacité de mobilité. Par exemple, dans un ICE de *Streptococcus agalactiae* (ICE*_2603_tRNALys*), le gène codant la protéine homologue à VirB4 (*orfD*), essentielle à la conjugaison, est interrompu par la présence d'un élément IS (IS*1381*) (Smyth and Robinson, 2009). Cet ICE perd ainsi sa capacité à être transféré par conjugaison. Un autre exemple concerne *Bradyrhizobium japonicum* USDA110 où une analyse génomique a suggéré que deux îlots dégénérés résultaient d'un réarrangement de grande échelle ayant coupé en deux un ICE de grande taille (Kaneko et

al., 2002). A la manière des éléments cités dans la partie précédente, certains éléments dégénérés peuvent être mobilisés par d'autres éléments fonctionnels.

ii. Accrétion en tandem et mosaïque d'éléments

Nombre de GEIs ne serait pas du à un mais à plusieurs intégrations de MGEs en tandem ou les uns dans les autres, créant ainsi des structures mosaïques. De plus, des recombinaisons peuvent ensuite entrainer des remaniements au sein de ces GEIs, ainsi que la dégénérescence de certaines parties (Bellanger et al., 2014).

Il existe une importante diversité de cas d'intégration d'éléments les uns dans les autres comme par exemple d'ICEs ou IMEs contenant d'autres éléments non-mobilisables, des prophages ou encore des ICEs (Bellanger et al., 2014). Un autre phénomène classiquement observé est l'accrétion en tandem de différents éléments. D'une manière générale, ceci est dû à l'intégration d'un élément par recombinaison de son site *attI* avec le site *attL* ou *attR* d'un autre élément déjà intégré dans le génome (Figure 21).

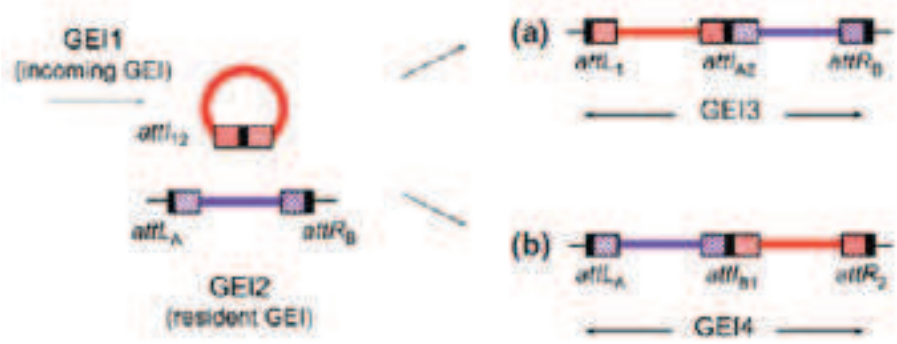


Figure 21 Schéma du phénomène d'accrétion en tandem de GEIs. D'après Bellanger et al., 2014.

Les îlots génomiques, et donc d'une manière plus générale la part variable du génome, correspond souvent à d'importantes régions flexibles. Ces régions peuvent être complexes de par l'accumulation d'évènements de gains, pertes et dégénérescences de certaines parties, ainsi que de par de nombreux remaniements génomiques.

b. Evolution des ICEs

Comme les plasmides, les ICEs ont une structure modulaire, c'est-à-dire que les gènes codant pour des processus similaires forment des clusters. Des analyses comparatives ont montré que des recombinaisons entre ICEs, phages et plasmides sont relativement fréquentes

et peuvent causer l'échange de modules. Ceci expliquerait que certains ICEs possèdent des recombinases à domaine DDE (fréquentes chez les éléments IS), mais aussi la présence de différents modules de conjugaison. Cette évolution modulaire peut aussi entrainer la création d'éléments différents de par leur autonomie de transfert et de mobilisation. Ces éléments, ICEs compris, sont regroupés sous le terme d'îlots génomiques, une classe regroupant des éléments très divers.

V - Les éléments modulant la variabilité génomique

Les remaniements génomiques et l'acquisition d'ADN étranger par des HGTs peuvent être modulés par les bactéries. Certains éléments du génome sont en effet à l'origine de mécanismes pouvant contraindre l'acquisition de matériel génétique, ou au contraire favoriser le maintien de l'ADN acquis. Trois de ces mécanismes sont présentés dans les parties qui suivent.

1 - Les systèmes de restriction-modification

a. Généralités

i. Rôle dans la défense contre l'ADN étranger

Les systèmes de restriction-modification (systèmes RM) sont considérés comme un système d'immunité inné, protégeant la cellule de l'invasion d'ADN étranger (ex. prophage). Ces éléments codent une endonucléase de restriction (REase) capable de reconnaitre et cliver l'ADN étranger, et une methyltransferase permettant de le différencier de l'ADN de l'hôte en transférant des groupements méthyle au niveau de séquences spécifiques de ce dernier (Oliveira et al., 2014). En effet, si l'ADN étranger ne possède pas la modification correcte, il est attaqué par l'endonucléase (Figure 22). La possession d'un système RM influe donc sur la capacité d'une bactérie à acquérir de l'ADN dans son génome (Vasu and Nagaraja, 2013). Les systèmes RM sont très répandus et diversifiés dans les génomes des bactéries et chez les archées.

ii. Les différents types de systèmes RM

Quatre groupes sont distingués notamment sur la base de leur composition, de leur reconnaissance de séquence, de la position de clivage et de la spécificité de substrat. Les systèmes RM de type I sont des protéines hétéro-oligomériques complexes comprenant à la fois les activités de restriction et de modification. Dans les systèmes RM de type II, les activités de

restriction et modification sont généralement codées par deux gènes différents, souvent en opéron. Les systèmes RM de type III sont des hétéromères composés des produits de deux gènes, impliqués dans la restriction et la modification. Les deux types de sous-unités sont nécessaires à la restriction alors que les sous-unités impliquées dans la modification sont autonomes. Les systèmes RM de type IV sont composés d'une ou deux endonucléases de restriction qui à la différence des autres systèmes, clivent des sites modifiés.

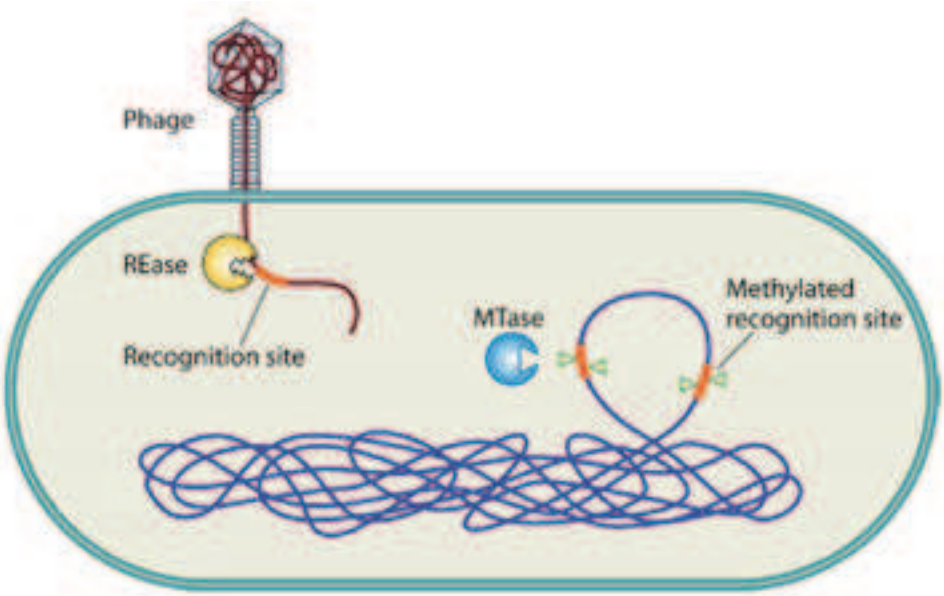


Figure 22 Schéma du rôle défensif d'un système RM contre de l'ADN phagique. D'après Vasu & Nagaraja 2013.

b. Rôle dans le maintien des EGMs

Le seul rôle défensif des systèmes RM n'explique pas la haute spécificité de séquences des endonucléases de restriction, ni la présence de plusieurs systèmes chez certaines espèces, notamment les compétentes telles que *Neisseria gonorrhoeae* (Stein et al., 1995). Ainsi, plusieurs autres rôles ont été proposés comme par exemple la nutrition, le maintien d'éléments génétiques mobiles ou encore la diversification/évolution des génomes.

Les systèmes RM peuvent dans certains cas être considérés comme des éléments égoïstes dans la mesure où ils ont tendance à se propager et à assurer avant tout leur propre maintien. Par exemple, lorsqu'ils sont présents sur un plasmide, la perte de ce plasmide par certaines bactéries lors de la division cellulaire peut entrainer leur mort à cause de l'activité de l'endonucléase plus stable que celle de la méthylase. Ces systèmes sont d'ailleurs souvent retrouvés dans des EGMs ou des parties du génome probablement acquises par HGTs (ex : les

îlots génomiques) (Furuta et al., 2010). Ils pourraient donc être impliqués dans le maintien de certaines zones flexibles du génome, comme c'est le cas des superintégrons (Szekeres et al., 2007).

Cette implication dans le maintien de certains éléments et la protection contre d'autres confèrent aux systèmes RM un rôle modulateur de l'intégrité des génomes et le maintien de l'identité des espèces (Murray, 2002). L'acquisition d'un système RM par une souche donnée peut également l'isoler génétiquement de souches proches et donner lieu à un processus de spéciation. Enfin, la dégradation d'ADN étranger par les systèmes RM est supposée favoriser la recombinaison homologue (Price and Bickle, 1986) avec le génome hôte et donc générer de la diversité .

2 - Les systèmes toxine-antitoxine

a. Définition

Les systèmes toxine-antitoxine (systèmes TA) sont de petits modules génétiques mobiles, retrouvés aussi bien dans le chromosome bactérien que dans des éléments mobiles. Des études de génomes complets ont permis de montrer que ces systèmes sont abondants et très répandus (Hayes and Van Melderen, 2011). Ils sont souvent retrouvés en multiples copies dans un génome. D'une manière générale, ils sont composés de deux gènes, le plus souvent auto-régulés, codant une toxine stable et une antitoxine instable. Actuellement, cinq types de systèmes TA ont été décrits d'après la nature et le mode d'action de l'antitoxine (Schuster and Bertram, 2013 ; Figure 23). Les toxines sont toujours de nature protéique (petites protéines < 130 acides aminés), alors que l'anti-toxine peut être soit une protéine (systèmes de type II, IV et V), soit une molécule d'ARN (systèmes de type I et III).

La perte d'un système TA par exclusion, ségrégation, recombinaison ou mutation peut mener à la mort cellulaire ou l'arrêt de la croissance, puisque la toxine est plus stable que l'antitoxine. La toxine peut affecter l'hôte de manières variées : stopper la réplication de l'ADN, la division cellulaire, la transcription, ou encore attaquer la membrane cellulaire. Les systèmes TA peuvent donc avoir une multitude d'impacts sur la physiologie, dont le rôle est plus ou moins bien compris suivant qu'ils soient présents dans le core-génome ou sur des éléments génétiques mobiles.

b. Les différentes types de systèmes TA

Dans les systèmes TA de type I, la toxine est généralement une petite protéine (< 60 aa)

hydrophobe à domaines transmembranaires. Elle est supposée déstabiliser la membrane cytoplasmique et le potentiel membranaire avec des conséquences sur la synthèse d'ATP. L'antitoxine est une molécule d'ARN qui empêche la synthèse de la toxine au niveau post-transcriptionnel par un mécanisme antisens. Dans les systèmes de type II, l'antitoxine est une protéine qui va bloquer la toxine au niveau post-traductionnel. Elle joue également le rôle de régulateur de l'opéron en réprimant la transcription de celui-ci. Les toxines du système de type III sont des endoRNases clivant les ARN cellulaires de manière aspécifique. L'antitoxine peut agir soit en bloquant la traduction de la toxine et/ou en favorisant la dégradation de son ARNm. Le système de type IV a été proposé récemment (Masuda et al., 2012). La toxine inhibe la polymérisation des protéines FtsZ et MreB impliquées dans la forme et la division de la cellule, ainsi que l'antitoxine BceA. L'antitoxine contre l'effet de la toxine en stabilisant FtsZ et MreB. Ainsi, contrairement aux autres systèmes TA, l'antitoxine n'interagit par avec la toxine mais protège ses cibles. Ce système possède donc des analogues avec les systèmes de restriction-modification (Cf. partie précédente). Enfin, dans le système TA de type V décrit chez *E. coli*, la toxine (GhoT) déstabilise la membrane, et l'antitoxine (GhoS) régule son activité au niveau post-transcriptionnel via une activité endonucléase spécifique de l'ARNm de *ghoT*.

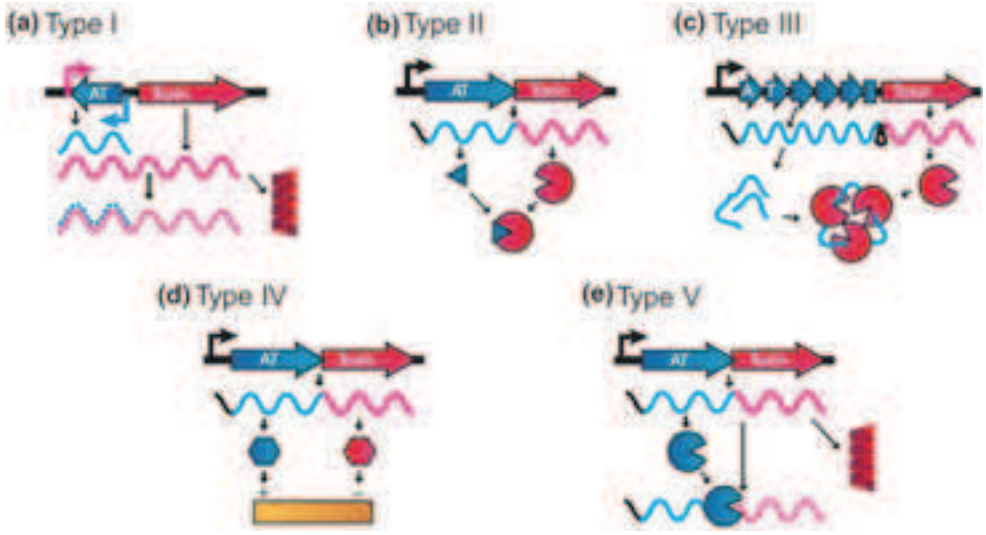


Figure 23 Schéma du fonctionnement des cinq types de systèmes toxine-antitoxine. Flèches rouges = ORF codant la toxine. Flèches bleus = ORF codant l'antitoxine. D'après Schuster & Bertram 2013.

c. Impact sur la stabilité des îlots génomiques

Lorsqu'ils sont codés par le core-génome leur rôle n'est pas toujours bien compris, mais il est maintenant bien établi qu'ils ont un rôle dans le phénotype des persisters et dans la

formation de biofilms (Wen et al., 2014). Dans certains cas, ils peuvent permettre de faire face à un système TA similaire présent sur un plasmide ou un autre EGM, et jouer ainsi un rôle anti-addictif. Lorsqu'ils sont codés par des éléments mobiles, ils participent au maintien de ces éléments dans la population en entrainant la mort ou l'arrêt de la croissance des cellules perdant l'élément. Les systèmes TA sont aussi décris comme des éléments égoïstes assurant leur propre maintien ou des vestiges d'anciens éléments mobiles.

3 - Le système CRISPR-Cas

a. Généralités

Les systèmes CRISPR-Cas (*Clustered Regularly Interspersed Short Palindromic Repeats-CRISPR associated proteins*) sont considérés comme un système d'immunité acquise des bactéries contre certaines molécules d'ADN étrangères à la cellule (Van der Oost et al., 2014). Leur présence dans un génome confère en effet une immunité adaptative et héréditaire contre des bactériophages et des plasmides. Cependant, ils sont quand même également retrouvés dans des phages, des plasmides, ou d'autres éléments mobiles, indiquant leur potentielle dissémination par HGTs. Ces éléments ont été retrouvés dans environ la moitié des génomes de bactéries séquencés et dans quasiment tous les génomes d'archées.

Trois éléments composent les systèmes CRISPR-Cas : une séquence dite "leader" de quelques centaines de pb, riche en AT et contenant un promoteur, suivie d'une série de séquences répétées palindromiques, les CRISPR, à proximité de gènes *cas*. La série de CRISPR est formée d'une succession de répétitions directes et identiques de 20-50 pb, séparée par des séquences variables, les"spacers" de 20-84 pb (Figure 24). Ces dernières peuvent être identiques à certaines séquences de plasmides ou de phages, et codent la mémoire immunitaire du système. Il existe une importante diversité de systèmes CRISPR-Cas liée à des différences au niveau des séquences répétées, des séquences des gènes *cas* et de l'organisation de l'opéron *cas*. Sur la base de ces différences, les systèmes CRISPR-Cas ont été classés en trois catégories. Les systèmes de type I possèdent la nucléase-hélicase Cas3, les systèmes de type II sont caractérisés par la nucléase Cas9 et les systèmes de type III possèdent une protéine Cas10 de fonction inconnue. Les systèmes de type I et III semblent proches phylogénétiquement alors les systèmes de type II sont distincts et ont une structure différente (Van der Oost et al., 2014).

b. Fonctionnement général

L'acquisition de la séquence "spacer" a lieu à l'extrémité de la séquence leader, mais le

mécanisme est encore mal connu. Selon un modèle proposé, la première étape consisterait en la reconnaissance puis la fragmentation de l'ADN invasif. Il semblerait que les systèmes RM puissent être impliqués dans cette étape comme chez *Streptococcus thermophilus*. Les protéines Cas reconnaîtraient ensuite dans certains fragments d'ADN produits un motif appelé PAM (*Protospacer Adjacent Motif*) (Figure 25). Le spacer est ensuite généralement intégré au niveau de la première répétition, ce qui engendre une duplication de cette répétition.

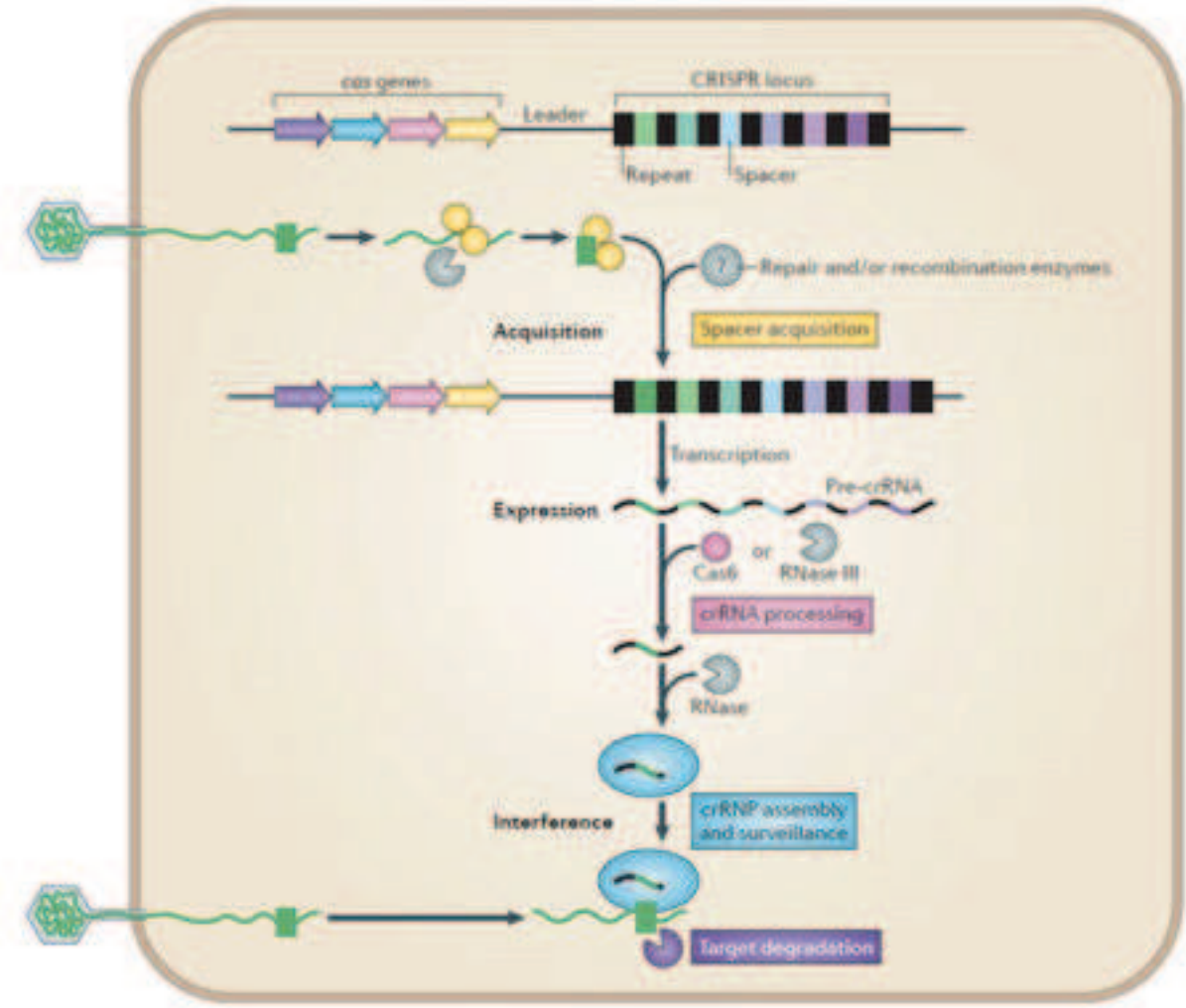


Figure 24 Schéma de la structure et du fonctionnement général du système CRISPR-Cas. Une séquence spacer est acquise après fragmentation d'une molécule d'ADN invasive. Cette nouvelle séquence spacer est insérée au niveau de la répétition à l'extrémité de la séquence leader. La série de CRISPR est transcrite en un Pre-crRNA grâce au promoteur présent dans la séquence leader. Cette molécule d'ARN est ensuite clivée par des protéines Cas ou une RNaseIII pour former un complexe nucléoprotéique (crRNP) avec d'autres protéines Cas. Ce complexe permet la reconnaissance d'une molécule d'ADN invasive lors d'une deuxième infection par un phage par exemple. Une fois détectée, la molécule est dégradée. D'après Van der Oost et al. 2014.

La détection d'une molécule d'ADN invasive, dont une partie de la séquence est gardée en mémoire dans une séquence spacer, est assurée par un complexe ribonucléoprotéique (crRNP) formé d'un transcrit de la séquence spacer en question et de protéines Cas (Figure 24). Pour former ce complexe, la série de CRISPR est tout d'abord transcrite dans son intégralité grâce au promoteur situé en amont dans la séquence leader. La molécule d'ARN obtenue, appelée Pre-crRNA, est ensuite clivée par les protéines Cas (systèmes I et III) ou une ribonucléase cellulaire (système de type II) pour former des crRNA. Ce crRNA s'associe ensuite à d'autre protéines Cas afin de former le complexe cr-RNP. La molécule d'ARN permet de guider le complexe vers les molécules d'ADN invasives possédant une séquence complémentaire, afin que ces dernières soient ensuite dégradées par des nucléases.

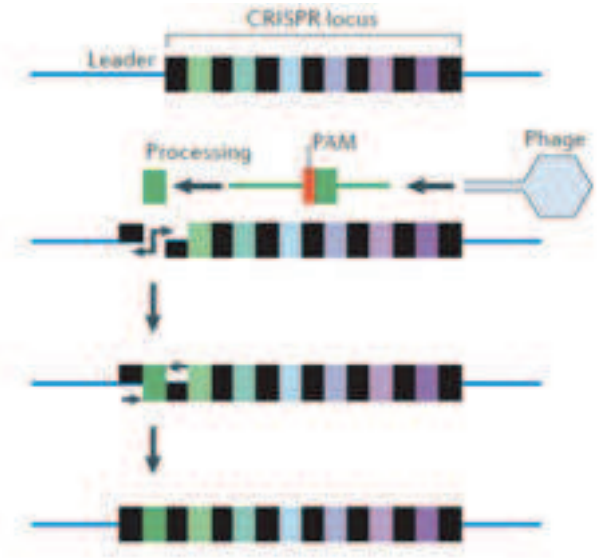


Figure 25 Schéma du modèle d'acquisition d'une nouvelle séquence spacer dans la série de CRISPR. En noir, les répétitions de la série de CRISPR ; en bleu et violet, les séquences spacers précédemment acquises ; en vert : la séquence spacer en cours d'acquisition. En amont : la séquence leader. Le spacer est inséré dans la répétition à l'extrémité de la séquence leader, engendrant une nouvelle répétition. D'après Van der Oost et al. 2014.

VI - L'influence de l'environnement et du mode de vie

En plus d'être modulée par des éléments intrinsèques aux bactéries, la variabilité génomique peut également être influencée par des facteurs environnementaux.

1 - Généralités

Les conditions environnementales sont rarement optimales, sont variables, et les bactéries sont ainsi souvent soumises à différents stress (ex : nutritionnel, thermique, acide, osmotique, oxydant). Pour faire face à ces stress, elles possèdent des réponses physiologiques adaptées telles que la réponse générale au stress (Battesti et al., 2011), ou la réponse SOS (Erill et al., 2007) dans le cas de dégâts au niveau de l'ADN. Enfin, alors que les souches utilisées en laboratoire sont généralement clonales et cultivées en culture pures en milieu liquide, les populations naturelles sont hétérogènes au sein d'une communauté complexe souvent organisées sous forme de biofilms (Costerton et al., 1995).

Les conditions environnementales et le mode de vie des bactéries peuvent avoir un impact sur leur variabilité génomique et donc leur diversification et adaptation. La réponse au stress peut induire de la variabilité génomique, que le stress soit mutagène ou pas (Foster, 2005). Le mouvement des EGMs et les HGTs peuvent également être favorisés sous certaines conditions, et les biofilms peuvent en plus constituer des conditions propices à la diversification génétique (Boles et al., 2004).

2 - La réponse au stress

Quand les bactéries sont exposées à des conditions hostiles pour leur survie, elles mettent en place des réponses au stress : réponse générale au stress, réponse au choc thermique ou encore réponse stringente. Ces réponses sont des modifications physiologiques reposant sur l'activation de facteurs sigma alternatifs qui induisent l'expression d'un ensemble de gènes particuliers (Marles-Wright and Lewis, 2007). Plusieurs mécanismes mis en place en réponse au stress peuvent conduire à un taux de mutation plus élevé chez les bactéries (Foster, 2005), sans que le stress ne cause directement des dégâts à l'ADN. Ces mécanismes pourraient avoir été sélectionnés à cause du fait qu'ils augmentent le potentiel adaptatif des bactéries, c'est-à-dire les capacités à s'adapter à une condition particulière (Tenaillon et al., 2004). La sélection agissant sur des caractères impliqués dans la capacité des organismes à s'adapter est qualifiée de "sélection de second ordre" ou "sélection indirecte".

a. Réponse générale au stress

La réponse générale au stress chez *E. coli* comprend plus de 70 gènes induits par l'activation du facteur sigma RpoS (σ^{38}). Ce dernier est activé lorsque les bactéries ont leur croissance stoppée, comme c'est le cas lors de carences nutritionnelles en phase stationnaire de croissance par exemple. Parmi les gènes induits par RpoS figure notamment celui codant l'ADN Polymérase IV (PolIV) qui produit plus d'erreurs de réplication que l'ADN polymérase III du fait de l'absence d'activité exonucléase 3'→5'. De plus, des composants du système de mismatch repair (MMR), permettant la réparation des mésappariemments produits lors de la réplication de l'ADN, sont réprimés par RpoS. Dans des conditions provoquant une diminution de la croissance cellulaire, le taux de mutation est donc susceptible d'augmenter (Foster, 2007).

b. La réponse SOS

i. Généralités

Alors que des stress d'ordres nutritionnels sont sans conséquences directes sur l'intégrité du génome, d'autres au contraire provoquent des cassures au niveau de l'ADN qui nécessitent des systèmes de réparation pour la survie des bactéries. C'est notamment le cas des rayonnements ultra-violet, de certains métaux, des quinolones ou tout facteur engendrant un stress oxydant. Le stress oxydant peut être causé par des espèces réactives de l'oxygène ou de l'azote, ou bien par des composés qui bloquent le système de protection contre ce stress, comme c'est le cas de l'arsénite. En réponse aux dommages infligés à l'ADN, les bactéries induisent un système de réparation appelé "réponse SOS". Ce mécanisme, décrit pour la première fois chez *E. coli* dans les années 1970, est aujourd'hui considéré comme universel chez les bactéries (Erill et al., 2007). Le gène codant la protéine essentielle LexA est en effet retrouvé dans quasiment tous les génomes séquencés, à quelques exceptions près. Ce mécanisme est donc à la fois ancien et conservé, témoignant de son importance. Les gènes présents dans le régulon ainsi que leur nombre sont variables suivant les bactéries. A titre d'exemple, la réponse SOS implique 33 gènes chez *B. subtilis* dont seulement 8 sont homologues chez *E. coli*. Le régulon *Pseudomonas aeruginosa* comprend quant à lui seulement 15 gènes.

ii. Le régulon SOS

La réponse SOS est un réseau global de régulations (régulon) d'une quarantaine de gènes chez *E.coli*, et régulée en amont par les protéines LexA et RecA. Les protéines LexA et RecA ont respectivement le rôle de répresseur et d'activateur. Les gènes impliqués dans la réponse SOS possèdent dans leur région promotrice une séquence palindromique de 16 pb (boîte SOS) à laquelle le répresseur LexA peut se fixer. La fixation de LexA empêche l'ARN polymérase de se fixer et donc bloque l'initiation de la transcription des gènes concernés. Ces derniers sont impliqués dans différents processus cellulaires tels que l'inhibition de la division cellulaire ou la réparation de l'ADN par excision. La protéine RecA agit à la fois comme senseur et dérépresseur de la réponse SOS. Elle se lie à de l'ADN simple brin et forme un filament, souvent suite à l'interruption de la réplication dû à des dommages de l'ADN. Une fois liée à l'ADN, RecA entre dans un état dit "activée", et provoque le clivage autocatalytique du répresseur LexA. Le clivage de LexA entraîne la levée de l'inhibition du régulon et donc de la réponse SOS. Les gènes induits vont favoriser la réparation des lésions via trois voies principales : la recombinaison homologue, la réparation par excision de nucléotides, et la synthèse translésionnelle d'ADN.

La réparation par excision de nucléotides implique le complexe UvrABC. La région du brin endommagé est excisée puis synthétisée à nouveau en utilisant le brin complémentaire intact comme matrice. La réponse SOS, bien qu'impliquée dans la réparation de l'ADN, a aussi des conséquences mutagènes. La synthèse translésionnelle est effectuée par les polymérases PolII (*polB*), PolIV (*dinB*) ou PolV (*umuCD*). Celles-ci incorporent des bases au hasard lorsqu'elles rencontrent des lésions et ne possèdent pas d'activité correctrice, ce qui entraine une augmentation du taux de mutations. La recombinaison homologue entre des régions homologues du génome entraine des réarrangements génomiques. La protéine RecX peut se lier aux filaments formés par RecA (Drees et al., 2004), ce qui faciliterait la recombinaison homologue (Cardenas et al., 2012). Des réarrangements peuvent aussi se produire lorsque la réponse SOS est induite par des HGTs, car l'ADN est souvent transféré sous forme simple brin. L'induction de la réponse SOS par les EGMs peut être une stratégie favorisant leur intégration dans le génome (Baharoglu and Mazel, 2014). La réponse SOS peut aussi induire des HGTs comme il a par exemple été démontré dans le cas du transfert d'un ICE chez *Vibrio cholerae* (Beaber et al., 2004). Le déclenchement de la réponse SOS est également connu pour réguler l'intégrase des intégrons (Guerin et al., 2009).

3 - L'implication des biofilms

a. Généralités

i. Définition

Les biofilms sont définis comme une population ou communauté microbienne fixée au sein d'une matrice sécrétée par les membres de la population / communauté. Les biofilms constituent un mode de vie très ancien (Rasmussen, 2000) et il est maintenant accepté que dans l'environnement naturel, la plupart des bactéries vivent en biofilm. La matrice est composée de polysaccharides, de protéines, d'ADN et de lipides. Cependant, la composition est variable suivant les espèces, même d'une souche à l'autre, et en fonction des conditions environnementales. De plus, cette composition complexe est difficile à déterminer expérimentalement.

ii. Développement

Le développement d'un biofilm se fait en plusieurs étapes. Dans un premier temps, les bactéries adhèrent au support, d'abord de manière réversible puis via des protéines. Ensuite, le développement du biofilm se fait via la synthèse de la matrice de substance polymérique extracellulaire (EPS, *Extracellular Polymeric Substance*) et la division de cellules et une continuelle colonisation par d'autres bactéries. Le biofilm se complexifie avec par exemple la mise en place de canaux aqueux permettant l'approvisionnement en oxygène et en nutriments. Malgré cela, les biofilms matures sont considérés comme hétérogènes, avec des parties pouvant par exemple être anoxiques et pauvres en nutriments. En conséquence, la population est hétérogène d'un point de vue phénotypique (Stewart and Franklin, 2008). Une des caractéristiques des biofilms est également l'importante densité cellulaire.

Les biofilms sont donc importants pour le maintien des bactéries dans leur environnement et de nombreuses études ont été consacrées à la compréhension de leur rôle dans la résistance au stress (e.g antibiotiques, métaux lourds), la colonisation et la persistance dans des niches écologiques pourtant peu propices à la croissance bactérienne.

iii. Biofilms et résistance

Le succès évolutif des biofilms s'explique en partie par leur implication dans la résistance face aux contraintes environnementales. Tout d'abord, la matrice constitue une

barrière physico-chimique empêchant ou ralentissant les composés toxiques. Cette matrice permet également le maintien d'une certaine homéostasie face à des conditions variables. L'hétérogénéité spatiale entraine une hétérogénéité physiologique avec notamment l'apparition de sous-populations plus résistantes appelées "persisters".

Le développement en biofilm est influencé par les conditions environnementales. Lorsque les conditions deviennent défavorables, certaines espèces ont par exemple tendance à se fixer et/ou à produire davantage de biofilm. Cependant, aucunes généralités ne peuvent être faites, les stratégies pouvant différer suivant les conditions et les espèces considérées. La mobilité peut par exemple chez d'autres organismes être favorisée afin d'échapper à des conditions adverses.

b. Biofilms et variabilité génomique

Un autre rôle des biofilms est leur implication dans la diversification et l'adaptation des bactéries. Plusieurs études ont en effet rapporté que la croissance en biofilm favorisait l'émergence de variants génétiques au sein de la population. Cette diversification ne résulte pas d'agents externes, mais est intrinsèque aux biofilms. L'hétérogénéité des conditions physico-chimiques et physiologiques ainsi que la densité cellulaire propres aux biofilms font partie des facteurs impliqués dans cette diversification. Cette dernière peut être due à un taux de mutation ou de recombinaison plus élevé ou bien à des transferts d'EGMs.

i. Mutations et remaniements

Au sein d'une population en biofilm, les conditions physico-chimiques sont hétérogènes. Une partie de la population est notamment soumise à un stress oxydant causant des dommages à l'ADN. Les cassures simple ou double brins ont pour conséquence des réparations RecA dépendantes entrainant un taux de mutations et de recombinaisons plus important et donc une diversification de la population (Allegrucci and Sauer, 2008; Boles and Singh, 2008; van der Veen and Abee, 2011). Dans certains cas, ce taux de mutation a été attribué à une répression de gènes codant les systèmes de réparation de l'ADN (Driffield et al., 2008).

ii. HGTs

Les dommages causés à l'ADN peuvent également déclencher la réponse SOS et favoriser la mobilisation et le transfert d'EGMs. Dans une population en biofilm, la dispersion des EGMs peut en plus être favorisée par la densité cellulaire. Dans le cas de la conjugaison, le

fait que le biofilm soit structuré permet un meilleur maintien du pilus (Madsen et al., 2012). Certains biofilms ont une matrice contenant beaucoup d'ADN extracellulaire. Il a été démontré que cet ADN peut avoir un rôle dans la mise en place et le maintien du biofilm. Cependant, cet ADN peut potentiellement être intégré par les bactéries via le mécanisme de transformation (Hannan et al., 2010).

Des études tendent à montrer que dans certains cas, ce sont les HGT qui favorisent le développement et le maintien des biofilms (Ghigo, 2001; Molin and Tolker-Nielsen, 2003; Petersen et al., 2005). Ils peuvent soit coder des gènes impliqués dans la formation de biofilms, soit renforcer la mise en place et le maintien d'interactions sociales au sein de la population. Ces deux mécanismes semblent ainsi très liés et sont deux phénomènes très répandus dans les populations naturelles.

CONTEXTE DU SUJET DE THESE

I - L'arsenic et les bactéries

1 - L'arsenic

L'arsenic (As) est un métalloïde de numéro atomique 33. Il appartient, comme le phosphate (P) et l'antimoine (Sb), à la famille des pnictogènes (15ème colonne de la classification périodique). Cet élément est peu abondant dans la croute terrestre (1,5 mg.kg^{-1}) mais très répandu. Il est généralement associé à des minerais de métaux tels que le plomb, le cuivre ou l'or, et peut exister sous quatre états d'oxydation différents : -III, 0, III et V (Oremland and Stolz, 2003). Il existe cependant deux formes prédominantes dans les environnements aquatiques : l'arsénite [As(III), de formule $As(OH)_3$ ou $H_2AsO_3^-$] et l'arséniate [As(V), de formule $H_2AsO_4^-$ et $HAsO_4^{2-}$].

Dans certains cas, l'arsenic trouvé dans les minéraux peut être mobilisé et atteindre des concentrations importantes dans les sédiments ou les milieux aqueux. Cette mobilisation peut résulter de processus naturels, ou d'activités humaines telles que l'utilisation de pesticides ou l'activité minière. L'arsenic étant un carcinogène, sa présence dans l'eau constitue un problème de santé public. Malgré cette contamination, certains environnements tels que les drainages miniers acides sont quand même colonisés par des microorganismes. Ces derniers possèdent des mécanismes de résistance à l'arsenic et certains l'utilisent même à des fins énergétiques (Kruger et al., 2013; Slyemi and Bonnefoy, 2012).

2 - Les mécanismes de résistance à l'arsenic chez les bactéries

a. Contrôle de l'entrée de l'As dans la cellule

Il a été montré que l'As(III) et l'As(V) ont des propriétés chimiques légèrement différentes et entrent dans les cellules par des moyens également différents. L'As(V) possède une structure analogue au phosphate, et peut donc emprunter les transporteurs phosphate pour pénétrer dans le compartiment intracellulaire. Les deux transporteurs utilisés sont Pit, qui est peu spécifique et constitue la principale porte d'entrée de l'As(V), et Pst qui est plus sélectif, laissant entrer moins d'As(V) (Figure 26). L'expression préférentielle des transporteurs Pst est donc un moyen de lutter contre l'entrée de l'As(V) dans les communautés exposées à l'As.

L'As(III) peut entrer dans la cellule par une aquaglycéroporine (GlpF) (Rosen and Tamás, 2010) chez *E. coli*, mais il existe des cas de bactéries métabolisant l'As(III) chez lesquelles aucun homologue de GlpF n'a été identifié. C'est notamment le cas de *Thiomonas* sp. 3As (Arsene-Ploetze et al., 2010), isolées dans le drainage minier acide de Carnoulès (Duquesne et al., 2008).

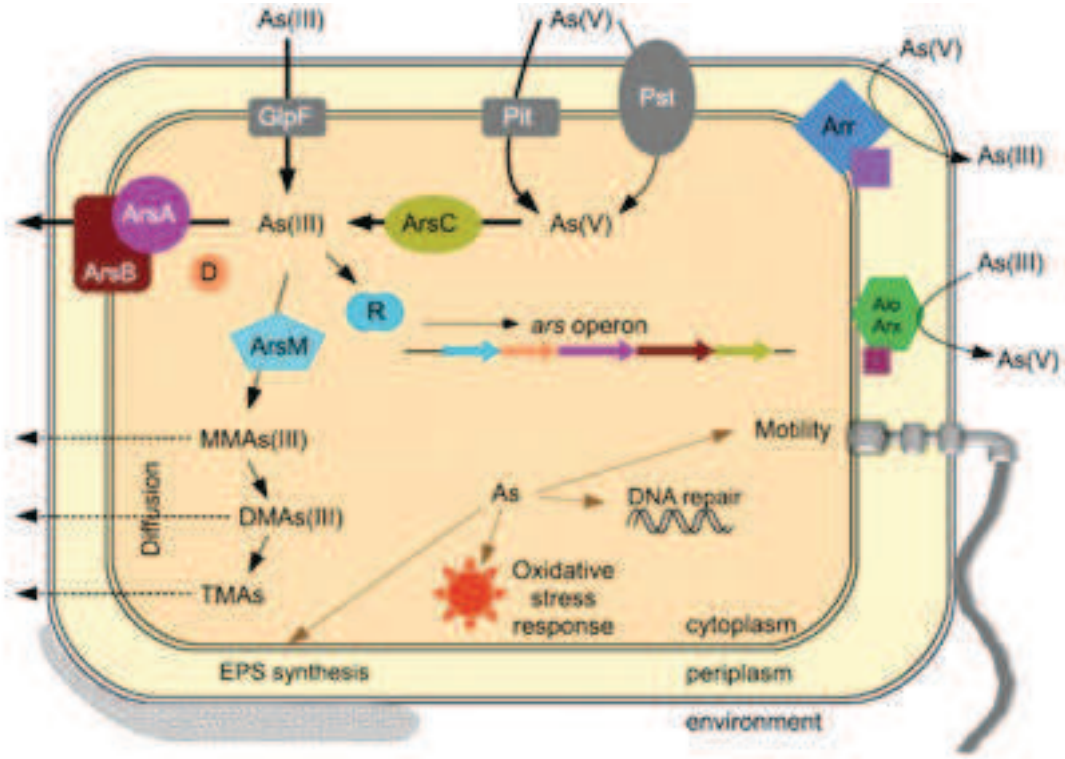


Figure 26 Schéma récapitulatif des mécanismes de résistances des bactéries à l'arsenic. Chez E. coli, l'As(III) entre dans la cellule par une aquaglycéroporine (GlpF). L'As(III) dans la cellule peut ensuite être expulsé à l'extérieur via le transporteur ArsB parfois couplé à l'ATPase ArsA, ou bien méthylé et évacué par diffusion. L'As(III)peut également être oxydé au niveau de la membrane cytoplasmique par Aio ou Arx et être utilisé comme donneur d'électrons. L'As(V) entre par les transporteurs phosphate Pit et Pst, ce dernier étant plus sélectif et laissant entrer moins d'As(V). A l'intérieur de la cellule, cet As(V) est réduit en As(III). L'As(V) peut aussi être réduit par Arr au niveau de la membrane cytoplasmique. D'après Krüger et al., 2013.

b. Mécanismes d'extrusion

Il existe deux mécanismes permettant de rejeter l'As(III) intracellulaire à l'extérieur de la cellule. Il peut être converti en forme volatile par méthylation et diffuser ensuite à travers la membrane. Ce mécanisme ne sera pas détaillé ici car il est peu connu chez les bactéries. L'As(III) peut aussi être transféré à travers la membrane par le transporteur ArsB. Ce dernier peut dans certains cas être couplé à l'ATPase ArsA qui rend l'extrusion de l'As(III) plus efficace. La protéine chaperon ArsD (Ajees et al., 2011) transfère quant à elle l'As(III) du complexe glutathion jusqu'au complexe ArsAB. Il n'existe pas de mécanisme d'extrusion de l'As(V)

connu. Celui-ci est réduit en As(III) qui sera ensuite pris en charge par ArsBA et ArsD (Figure 26). L'enzyme réduisant l'As(V) est l'arseniate reductase ArsC. Les gènes a*rs* sont organisés en un opéron dont la transcription est régulée par ArsR. Cet opéron peut être retrouvé dans le chromosome ou au niveau de plasmides, et son organisation est variable (ex : *arsRBC* dans le génome de *E. coli* et *arsRDABC* dans le plasmide R773 de *E. coli*).

c. Oxydation de l'arsénite

Chez les bactéries, l'As(III) est considéré comme généralement plus toxique que l'As(V) (Tamaki and Frankenberger, 1992). Cela est notamment lié à la biodisponibilité plus importante de l'As(III) en milieu aqueux. L'As(V) précipite plus facilement à cause de sa tendance à s'adsorber sur des minéraux tels que les hydroxydes de fer, et devient ainsi moins biodisponible. L'oxydation de l'As(III) en As(V) peut donc constituer un moyen pour les bactéries pour détoxifier leur environnement. Il existe deux enzymes capables de catalyser l'oxydation de l'As(III) en As(V) : AioBA et ArxAB. La première est largement répandue, et les gènes pourraient dans certains cas avoir été transférés par HGTs (Heinrich-Salmeron et al., 2011), la deuxième n'a été découverte que récemment (Zargar et al., 2010) et la distribution des gènes correspondants est encore mal connue.

L'opéron *aioBA* peut être associé avec les gènes accessoires *aioC* et *aioD*, codant respectivement une nitroreductase et un cytochrome C potentiellement impliqué dans un transfert d'électrons suite à l'oxydation de l'As(III). Le système à deux composants AioS/AioR, AioX (une protéine périplasmique se liant l'As(III)) et le facteur sigma RpoN sont essentiels à la régulation de l'opéron *aioBA*. Chez *Thiomonas arsenitoxydans*, la metalloprotéine AioF est impliquée dans la régulation de *aioBA* en présence d'arsenic (Moinier et al., 2014).

II - Les drainages miniers acides

Beaucoup de bactéries isolées de drainages miniers acides possèdent les systèmes de résistance décrits précédemment. Les drainages miniers acides (DMA) sont des écoulements acides chargés en métaux dus à l'exposition de sulfures (ex : sulfures de fer) à l'eau et à l'oxygène. Ils peuvent se produire de manière naturelle, mais résultent la plupart du temps d'activités anthropiques, notamment de travaux miniers. Ils représentent un problème environnemental de taille et sont ainsi devenus le sujet de nombreuses recherches. Ces dernières permettent de mieux comprendre la formation et le fonctionnement de DMA afin de mieux les prévenir, et / ou de les atténuer.

1 - Formation des DMAs

La formation d'un DMA débute par l'exposition de sulfures à l'eau ou l'air, entrainant leur oxydation et donc la libération de métaux et l'acidification du milieu. Les deux principaux oxydants sont le dioxygène et le fer ferrique Fe^{3+}. Ce processus peut être illustré avec le cas de la pyrite (FeS_2), le plus commun des sulfures metalliques.

La pyrite est d'abord oxydée par l'O_2 :

$$FeS_2 + 3{,}5O_2 + H_2O \rightarrow Fe^{2+} + 2SO_4^{2-} + 2H^+ \quad (1)$$

Le fer ferreux Fe^{2+} peut ensuite être oxydé suivant l'équation :

$$2Fe^{2+} + 0{,}5O_2 + 2H^+ \rightarrow 2Fe^{3+} + H_2O \quad (2)$$

La pyrite peut alors être oxydée par le Fe^{3+} formé :

$$FeS_2 + 14Fe^{3+} + 8H_2O \rightarrow 15Fe^{2+} + 2SO_4^{2-} + 16H^+ \quad (3)$$

Les réactions (2) et (3) permettent un maintien du processus en continu, puisque la (3) utilise le Fe^{3+} produit par la (2) et régénère ainsi du Fe^{2+} à nouveau oxydé par cette dernière. Ces deux réactions reposent sur la réaction (1) lors de laquelle est produit le Fe^{2+}. Les DMA peuvent être influencés par un nombre important de facteurs, tels que la minéralogie, le pH, la température, la concentration en oxygène ou la communauté bactérienne présente (Akcil and Koldas, 2006). En particulier, l'oxydation du Fe^{2+} dans la réaction (2) est lente à faible pH, mais peut être catalysée par certains microorganismes.

2 - Communauté microbienne associée aux DMAs

Bien que les DMA constituent un milieu avec des conditions extrêmes en termes de pH et de concentration en métaux, ils sont une niche écologique pour une communauté d'organismes. Au sein de cette communauté, des archées et des bactéries ont de par leur influence sur les paramètres physico-chimiques, un rôle important dans le fonctionnement de cet écosystème. La communauté microbienne est notamment impliquée dans l'oxydation du fer et des sulfures, favorisant ainsi la formation des DMAs.

Les bactéries ferro-oxydantes accélèrent le processus car le Fe^{3+} produit est un meilleur oxydant des sulfures que l'oxygène, mais aussi parce que la réaction d'oxydation abiotique du Fe^{2+} est lente à pH acide. *Acidithiobacillus ferrooxidans* est la bactérie la mieux caractérisée. L'oxydation du fer par celle-ci a été particulièrement étudiée. Cependant, d'autres organismes du genre *Leptospirillum*, *Gallionella*, des firmicutes tel *Sulfobacillus acidophilus* ou des archées du genre *Ferroplasma* ou *Metallosphaera* par exemple, sont aussi décrites comme

oxydant le fer au niveau des DMA.

Certains organismes sont également impliqués dans la réduction du Fe^{3+} en Fe^{2+}, tel *Sulfobacillus thermosulfidooxidans* ou *S. acidophilus* qui l'utilisent comme unique accepteur d'électrons en croissance hétérotrophe quand la quantité d'O_2 est limitée. Cependant, des conditions anoxiques ne sont pas obligatoires pour la réduction du Fe^{3+}, comme il l'a été démontré pour *A. cryptum* par exemple.

L'oxydation des sulfures est généralement réalisée par des bactéries autotrophes comme il l'a par exemple été démontré pour *Acidithiobacillus thiooxidans* ou *A. ferrooxidans* qui peut aussi se développer en conditions anoxiques avec Fe^{3+} comme accepteur d'électrons et S^0 comme donneur. Des composés soufrés avec des états d'oxydation intermédiaire formés durant l'oxydation des sulfures peuvent aussi être utilisés.

La communauté du DMA de Richmond (USA) est étudiée depuis une vingtaine d'années. La diversité de cette communauté est relativement faible que ce soit d'un point de vue taxonomique ou trophique, notamment à cause des fortes concentrations en métaux et de l'acidité. Ces caractéristiques font de cette communauté un sujet d'étude idéal pour des approches de métagénomiques, et elle est devenu un modèle d'étude (Denef et al., 2010).

3 - Le DMA de Carnoulès

a. Caractéristiques

En France, le DMA présent à Carnoulès dans le Gard est également étudié depuis plusieurs années. Le site est une ancienne mine de plomb et zinc, dont le gisement de 2,5 Mt (3,5 % de Pb et 0,8 % de Zn) était exploité à ciel ouvert. L'exploitation de ces gisements a notamment mis à jour des sulfures de fer (marcassite) et de plomb (galène) et du sulfate de baryum (baryte). En plus de ces minéraux contenant de nombreuses traces de métaux (ex : argent, cadmium, zinc, cuivre), d'autres sulfures ont été exposés en quantité moindre : des sulfures de zinc (sphalérite), et des sulfures contenant de l'arsenic (arsénopyrite) ou de l'antimoine (freibergite) (Leblanc et al., 1996). Après la fermeture de l'exploitation du gisement de Carnoulès en 1962, 0,5 Mt de déchets (3,5 % de Pb, 10 % de FeS_2 et 0,2 % d'As) ont été stockés sur la source du Reigous deerière une digue. Ce ruisseau mesure 1,5 Km, il se jette dans l'Amous qui lui-même se jette dans le Gard, ce dernier se jette dans le Rhône. Au niveau de la digue, le Reigous est caractérisé par un pH acide (2,2 – 4) et une forte concentration en As (350 $mg.L^{-1}$), en Fe (750 – 2700 $mg.L^{-1}$), et en sulfate (2000 – 7500 $mg.L^{-1}$). Les sédiments de la rivière sont également riches en As (> 22 %) (Leblanc et al., 1996). A la source, l'As est

principalement présent sous forme d'As(III) (arsénite) qui est oxydé en As(V) (arséniate) le long du Reigous. Cet As(V) s'adsorbe et co-précipite avec du Fe^{3+}. Ce processus, qui atténue la concentration d'As dans l'eau de Reigous (0,8 $mg.L^{-1}$ dans le partie la plus en aval), est en partie du à l'activité de bactéries oxydant l'As(III).

Le Reigous fait l'objet d'études à la fois minérales, chimiques et microbiologiques. Plusieurs sites sont répertoriés et nommés, allant du terril à la confluence avec l'Amous comme représenté sur le schéma ci-dessous.

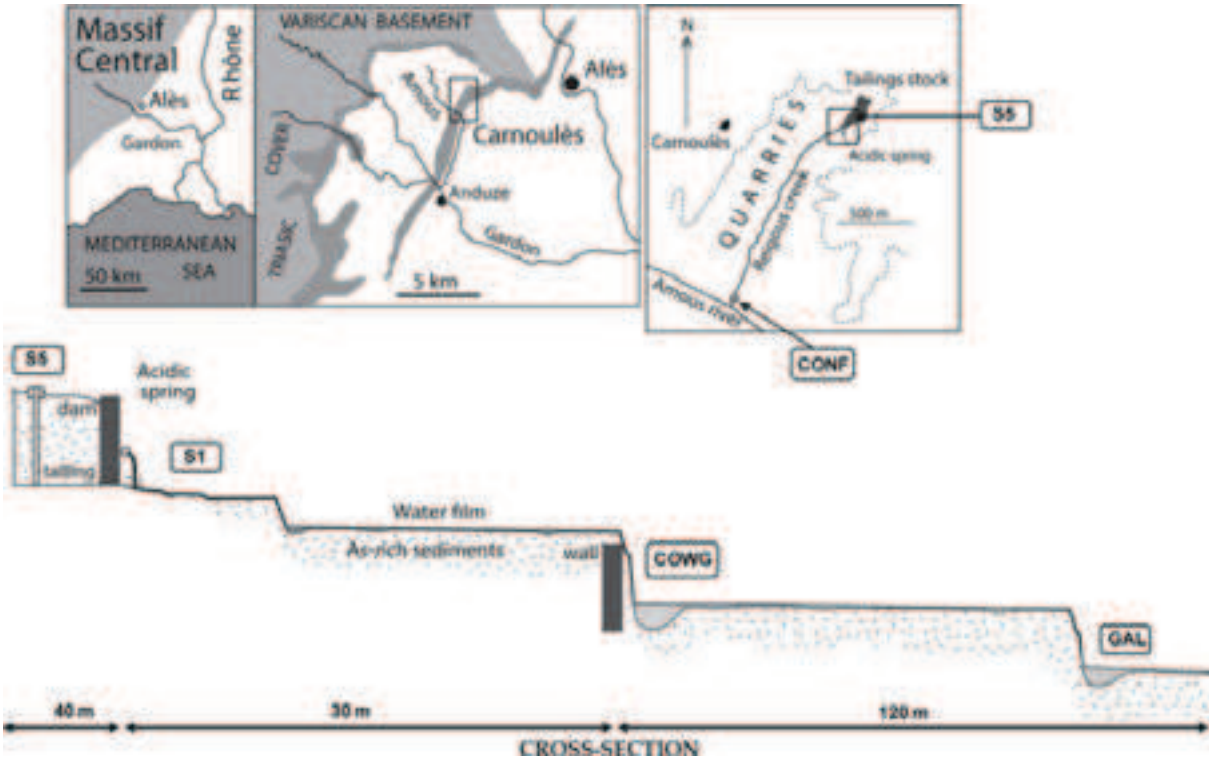


Figure 27 Plan du DMA de Carnoulès indiquant différents sites d'échantillonnages. Le site le plus en amont est un puits creusé au niveau du terril (S5), et le plus en aval est celui de la confluence avec l'Amous (CONF). Entre les deux, un site d'échantillonnage est situé au niveau de la source, un deuxième 30 m plus en aval appelé COWG et un dernier 150 m en aval de la source, le site GAL. D'après Volant et al., 2014.

b. Communauté microbienne

La communauté microbienne du DMA de Carnoulès fait l'objet d'études depuis plus d'une dizaine d'années, à la fois par des approches culturales et culture-indépendantes. Bien que les résultats obtenus diffèrent selon les études, que ce soit à cause des méthodes employées et/ou de la variabilité spatiale et temporelle, plusieurs tendances générales se dégagent.

La communauté de Carnoulès est dominée par le phylum des *Proteobacteria*, dont les *β-proteobacteria* (*Gallionellales* et *Thiobacillales* notamment), et l'ordre des

Acidithiobacillales sont généralement les mieux représentés. Les autres groupes les plus abondants sont généralement les *Firmicutes*, les *Acidobacteria*, les *Actinobacteria* et les *Nitrospirae*. Ces microorganismes sont majoritairement impliqués dans le cycle du soufre (S) et du fer (Fe). En plus des groupes majoritaires, une étude culture dépendante rapporte la détection de plusieurs genres jamais détectés auparavant à Carnoulès (Delavat et al., 2012). Ces genres sont supposés être des représentants de la biosphère rare. En 2014, une étude de Volant et al. a également mis en évidence par pyroséquençage la présence de nombreux nouveaux taxa tels que *Bacteroidetes*, *Chlorobi*, *Chloroflexi*, *Spirochaetes*, ou encore *Verrumicrobia* (Volant et al., 2014). La majorité de ces taxa représente moins de 1 % des données obtenues et est donc rare au sein de la communauté. Bien que la prise en compte de ces taxa augmente considérablement la diversité de la communauté, leur rôle écologique reste à élucider.

La diversité de la communauté ne se limite pas aux bactéries, des archées ayant également été identifiées (Bruneel et al., 2008; Volant et al., 2012). Elles appartiennent au phylum des *Euryarchaeota* (notamment des *Thermoplasmatales*) et *Thaumarchaeota* (Volant et al., 2012). La diversité estimée est faible mais comme dans le cas des bactéries, la présence d'autres taxa minoritaires n'est pas exclue. Enfin, un microorganisme eurocaryote, l'euglène *Euglena mutabilis*, est également présent sur le site. *E. mutabilis* est capable d'autotrophie et forme des tapis verdâtres à la surface de sédiments riches en précipités Fe-S, sur environ 250 mètres en aval de la source (Casiot et al., 2004).

La dynamique spatio-temporelle de la diversité microbienne a également fait l'objet de plusieurs études (Giloteaux et al., 2013; Volant et al., 2012, 2014). D'un point de vue spatial, la diversité estimée être la plus haute est celle de la source (S1), et la plus basse celle du site GAL, 150 m en aval de la source (Volant et al., 2014). La structure de la communauté diffère également suivant les sites. A titre d'exemple, les *Acidithiobacillales* sont prédominantes au niveau de l'eau souterraine du terril (S5) et de la source (S1), alors que les sites plus en aval sont généralement dominés par les *Gallionellales*. Les *β-proteobacteria* ont également tendance à être de plus en plus présentes en aval du Reigous. La présence d'arsenic, de sulfates et la température semblent être les facteurs majeurs influant la structure de la communauté (Volant et al., 2014). Des variations saisonnières sont également observées à l'instar de la communauté sulfato-réductrice (Giloteaux et al., 2013), mais aucune tendance n'a pour l'instant été démontrée.

Un dernier point concernant cette communauté est son aspect fonctionnel. Des approches métaprotéomiques (Bruneel et al., 2011) et métagénomiques (Fahy et al., 2015) ont permis de révéler des membres de la communauté active. Parmi les membres actifs ont été

retrouvés les genre *Thiomonas*, *Acidithiobacillus*, *Acidiphilum*, *Gallionella* ainsi que des bactéries affiliés aux *β- γ-* et *δ-proteobacteria*. Au niveau du site COWG (30 m en aval de la source), une étude métagénomique et métaprotéomique a permis la reconstitution des génomes de 7 microorganismes dominants (Bertin et al., 2011). Ces génomes, nommés CARN1 à CARN7, correspondent à la fois à des genres déjà décrits tels *Thiomonas* sp. (CARN2) et *At. ferrooxidans* (CARN5), à des non-cultivés tel CARN7 (proche de *Gallionella*), ou encore à des représentants d'un nouveau phylum proche de *Firmicutes* dans les cas de CARN1 et CARN4. Cette étude a montré que certains processus tels que la mobilité et la synthèse de biofilm semblaient cruciaux pour la survie dans l'AMD. Les spécificités métaboliques de ces microorganismes (hétérotrophie, autotrophie, mixotrophie) ainsi que les interactions trophiques qui en découlent semblent également être de première importance. Enfin, le protiste *Euglena mutabilis* a également un rôle important au sein de la communauté. En effet, il a été démontré par une approche métabolomique que cette euglène sécrète spécifiquement certains acides aminés sucres, polyamines ou d'autres composés tels que l'urée (Halter et al., 2012). La matière organique sécrétée par *E. mutabilis* pourrait être utilisée par les autres membres de la communauté.

III - Le genre *Thiomonas*

1 - Généralités

Le genre *Thiomonas* regroupe des bactéries gram négatif de type bacilles, appartenant à l'ordre des Burkholderiales au sein de la classe des β-proteobacteria. Ce genre a été défini en 1997 suite à une étude phylogénétique du genre *Thiobacillus* basée sur les ARN 5S et 16S (Moreira and Amils, 1997). Cette étude a en effet démontré que les espèces *Thiobacillus cuprinus* (Huber and Stetter, 1990), *Thiobacillus intermedius* (London, 1963a), *Thiobacillus perometabolis* (London and Rittenberg, 1967) et *Thiobacillus thermosulfatus* (Shooner et al., 1996), formaient un cluster phylogénétique distinct. Par la suite, *Thiobacillus delicatus* (Katayama-Fujimura et al., 1983, 1984; Mizoguchi et al., 1976) a également été transféré dans le genre *Thiomonas* en tant que *Thiomonas delicata* (Katayama et al., 2006). Les *Thiomonas* partagent avec les *Thiobacillus* la capacité à fixer le CO_2 atmosphérique et à oxyder des composés inorganiques soufrés pour leur production d'énergie. Cependant, elles ont en plus la particularité d'utiliser le carbone organique comme source d'énergie (organotrophie) et/ou de carbone (hétérotrophie) (Moreira and Amils, 1997). Ainsi, les *Thiomonas* ont un métabolisme versatile, pouvant être autotrophe, hétérotrophe, ou mixotrophe.

Les *Thiomonas* isolées jusqu'à présent sont soit issues de sources chaudes (e.g. *Tm. islandica* 6C (Vesteinsdottir et al., 2011)), soit de milieux pollués tels que les eaux usées, les boues d'épuration (e.g. *Tm. intermedia* K12 (Milde et al., 1983)), ou les drainages miniers acides (ex. *Tm. arsenitoxydans* 3As (Duquesne et al., 2008)) (
Tableau 1).

Souche	Groupe	Origine	Référence
Thiomonas arsenitoxydans 3As	I		(Duquesne et al., 2008)
Thiomonas sp. *CB1*	I		
Thiomonas sp. CB2	I	DMA, ancienne mine de plomb et de zinc (Carnoulès, Gard, France)	(Arsene-Ploetze et al., 2010)
Thiomonas sp. CB3	I		
Thiomonas sp. CB6	I		
Thiomonas sp. X19	II		(Delavat et al., 2012)
Thiomonas intermedia K12	I	Eaux usées (Hamburg, Allemagne)	(Milde et al., 1983)
Thiomonas intermedia^T	I	Rive d'un ruisseau	(London, 1963b)
Thiomonas thermosulfata ATCC 51520	I	Boues d'épurations	(Shooner et al., 1996)
Thiomonas perometabolis	I	Sol (Los Angeles, USA)	(London and Rittenberg, 1967)
Thiomonas sp. Ynys1	I	Mine de charbon (Ynysarwed, UK)	(Hallberg and Johnson, 2003)
Thiomonas sp. WJ68	I	DMA. Système de bioremédiation (Wheal Jane, UK)	(Johnson and Hallberg, 2005)
Thiomonas delicata (DSM 16361)	II	DMA, mine d'or (Cheni, France)	(Battaglia-Brunet et al., 2006)
Thiomonas delicata (DSM 17897)	II	Mine Cu-Pb-Zn-Au-Ag (Kosaka, Japan)	(Katayama-Fujimura et al., 1984)
Thiomonas delicata Hö5 (DSM 5495)	II	Mine d'uranium (Poppenreuth, Allemagne)	(Huber and Stetter, 1990)
Thiomonas islandica 6C (DSM 21436)	I	Source chaude (Greandalur, Islande)	(Vesteinsdottir et al., 2011)
Thiomonas bhubaneswarensis S10 (DSM18181)	I	Source chaude (Atri, Odisha, Inde)	(Panda et al., 2009)

Tableau 1 Quelques souches de Thiomonas et lieux d'isolement (liste non - exhaustive)

Les souches issues de drainages miniers acides ont développé des adaptations à ces milieux particuliers dont la résistance aux métaux. Dans le cas de l'arsenic, ces mécanismes comprennent à la fois l'extrusion et l'oxydation. En effet, il été démontré que certaines *Thiomonas* avaient la capacité d'oxyder l'arsénite (As(III)) en arséniate (As(V)), que ce soit en consortium (Battaglia-Brunet et al., 2002) ou en culture pure (Battaglia-Brunet et al., 2006; Bruneel et al., 2003; Duquesne et al., 2008). Chez *Tm. delicata* et *Tm. arsenitoxydans* 3As,

l'As(III) est utilisé comme donneur d'électrons dans le métabolisme énergétique (Battaglia-Brunet et al., 2006; Bryan et al., 2009; Duquesne et al., 2008). L'oxydation de l'arsenic est un mécanisme important car il contribue à la bio-atténuation naturelle de cet élément dans les DMA comme à Carnoulès. Les *Thiomonas* revêtent donc un intérêt particulier à la fois pour la compréhension du fonctionnement des écosystèmes pollués et pour leur bioremédiation.

2 - Phylogénie et évolution

Selon la phylogénie basée sur les séquences nucléotidiques de l'ARNr 16S et du gène *rpoA*, les *Thiomonas* sont réparties en deux groupes (Bryan et al., 2009). Le groupe II comprend *Thiomonas delicata* et le groupe I comprend *Thiomonas arsenitoxydans* 3As, les souches *Thiomonas* spp. CB1, CB2, CB3 et CB6, *Tm. intermedia* K12, *Tm perometabolis* et les souches *Tm.* spp. Ynys1 et WJ68. De manière intéressante, les souches 3As, CB1, CB2, CB3 et CB6 isolées du DMA de Carnoulès sont très proches et ne peuvent être distinguées que par leur gène *rpoA* (ADNr16-rpoA > 97%) ou par hybridation ADN-ADN (Arsene-Ploetze et al., 2010). Elles pourraient donc être considérées comme appartenant à la même espèce selon la définition communément acceptée.

Cependant, le séquençage de la souche 3As et l'hybridation de son génome avec les autres souches ont permis la mise en évidence de différences de contenu génomique (Arsene-Ploetze et al., 2010). Ces différences semblent correspondre à la présence ou absence de régions correspondant à des îlots génomiques. Au total, 19 îlots ont été identifiés dont un prophage putatif et certains possédant des gènes *mob*, *tra*, *trb* et *vir*, et donc potentiellement mobiles et transférables. Certains de ces îlots portent également des gènes impliqués dans la résistance au cuivre, au mercure ou à l'arsenic, indiquant que ces îlots sont probablement importants pour la survie dans ce DMA.

Les souches de *Thiomonas* spp. issues du DMA de Carnoulès ont donc probablement évolué par acquisition et perte d'îlots génomiques leur conférant des capacités adaptatives. Les différences génomiques entre ces souches sont probablement à l'origine de différences fonctionnelles. Il a notamment été observé que la souche CB2 a tendance à former des biofilms. Une étude portant sur l'implication de ces biofilms dans la résistance à l'arsénite (As(III)) a mis en évidence une surproduction de matrice en présence de cet élément (Marchal et al., 2011). En présence d'As(III), les structures observées dans le biofilm sont plus complexes et les micro-colonies contenaient plus de cellules mortes (Figure 28).

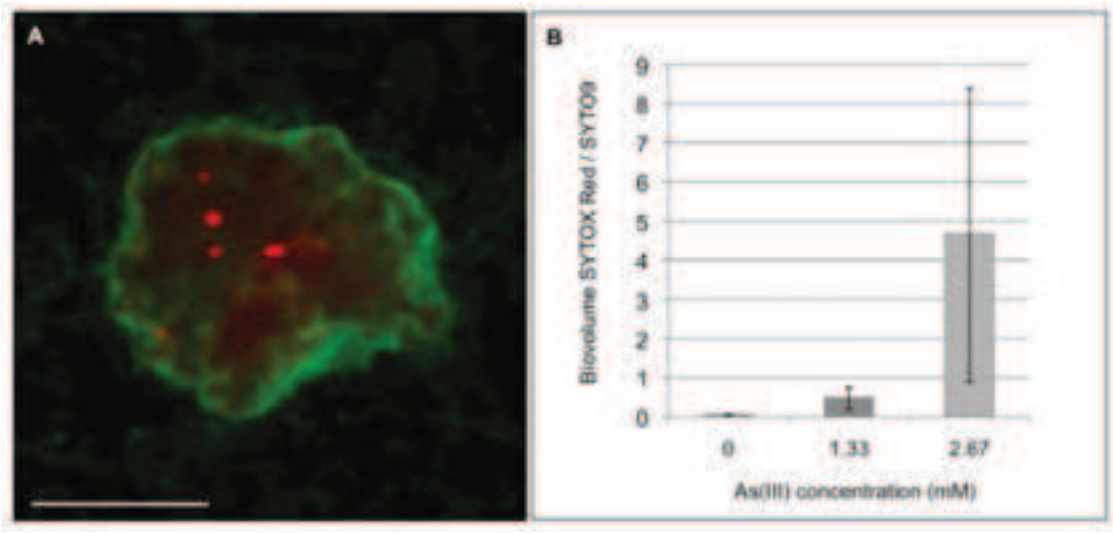


Figure 28 A-Photo d'une micro-colonie du biofilm de Thiomonas sp. CB2. Rouge = cellules mortes ; Vert = cellules vivantes. B- Ratio des biovolumes de "cellules mortes / cellules vivantes" à différentes concentrations d'As(III). D'après Marchal et al., 2011.

Enfin, au niveau des micro-colonies présentant des évènements de lyse cellulaire en présence d'As(III), une sous-population de cellules mobiles a été observée après plusieurs jours de culture. Le repiquage du biofilm sur milieux solides en présence de différentes concentrations en As(III) a mis en évidence que certaines cellules avaient acquis une résistance à cet élément supérieure à la population de départ. Ceci suggère que le développement en biofilm en présence d'As(III) pourrait favoriser l'émergence de variants plus résistants à l'As(III), et contribuerait donc à l'adaptation de *Thiomonas* sp. CB2 à cet élément.

OBJECTIFS

Cette thèse s'inscrit dans le cadre du projet ANR (Agence Nationale de la Recherche) "Thiofilm". L'objectif de ce projet est d'identifier les facteurs influençant la plasticité et l'évolution des génomes de bactéries du genre *Thiomonas*. Les travaux de thèse portent plus particulièrement sur cinq souches, 3As, CB1, CB2, CB3 et CB6, toutes isolées du DMA de Carnoulès et appartenant à la même espèce d'après la définition communément acceptée : l'espèce *Thiomonas arsenitoxydans* 3As. Il existe chez ces souches une variabilité génomique probablement liée à la présence d'îlots génomiques (GEIs). Le premier objectif de cette thèse est de caractériser ces GEIs et de mieux comprendre leur rôle dans les caractéristiques physiologiques des souches et dans leur capacité d'adaptation. Le deuxième objectif est de déterminer l'influence de l'arsénite et du développement en biofilm sur cette variabilité génomique.

La première étape a consisté à faire l'inventaire de la diversité génomique des *Thiomonas* spp. Comme seul le génome de *Tm. arsenitoxydans* 3As est disponible, celui des autres souches a donc été séquencé et l'ensemble des génomes intégrés dans la plateforme MaGe du Génoscope. Des outils de génomique comparative disponibles sur la plateforme ont ensuite permis de répertorier les différents GEIs de chaque souche. Certains GEIs ont fait l'objet d'une description plus approfondie et des approches expérimentales ont eu pour but de corréler leur présence ou absence avec des différences fonctionnelles.

Un deuxième travail a consisté à étudier l'influence de l'As(III) et du développement en biofilm sur la variabilité génomique. Cette partie a uniquement concerné la souche *Thiomonas* sp. CB2, suite à des observations préliminaires faites chez cette souche au laboratoire (Marchal et al., 2011). Il avait en effet été observé que des variants résistants à une concentration en As(III) supérieure à celle de la population de départ apparaissaient à une fréquence élevée après culture en biofilm en présence d'As(III). Dans un premier temps, la variabilité génomique a été estimée via la mesure du taux de variants ayant acquis une résistance à l'As(III) supérieure à la souche départ. Ce travail a été amorcé par Kelle Freel dans le cadre d'un travail post-doctoral. Mon travail a consisté à caractériser ces variants. Le génome de deux d'entre eux présentant des caractéristiques jugées intéressantes ont été séquencés. Ces génomes ont ensuite été analysés et comparés avec celui de la souche originelle *Tm.* sp. CB2 afin d'identifier les changements susceptibles d'expliquer les caractéristiques de ces variants.

Chapitre I - Description de la diversité génomique et fonctionnelle

Introduction - Chapitre I

Les génomes procaryotes sont généralement caractérisés par un génome dit "flexible" portant de nouvelles fonctions qui ont souvent été acquises par HGTs (Treangen and Rocha, 2011). Ces mécanismes peuvent être à l'origine d'une diversification fonctionnelle des populations naturelles, en conférant des capacités adaptatives aux populations et en leur permettant de faire face aux changements des conditions environnementales (Cohan, 2002). La comparaison des génomes de plusieurs souches issues du même environnement permet de déterminer la partie flexible de leur génome et donc de mieux comprendre leur évolution.

Plusieurs souches proches de *Thiomonas* proches phylogénétiquement ont été isolées d'un DMA à Carnoulès (Gard) : *Thiomonas arsenitoxydans* 3As qui doit son nom à sa capacité à oxyder l'arsénite (Duquesne et al., 2008; Slyemi et al., 2011), et les souches *Thiomonas* spp. CB1, CB2, CB3 et CB6 isolées par Odile Bruneel (Arsene-Ploetze et al., 2010). Il a été démontré que ces cinq souches présentaient des différences de contenu génétique probablement liées à la présence de GEIs. Ceux-ci pourraient être à l'origine de différences fonctionnelles entre ces souches.

Mes travaux présentés dans ce chapitre décrivent l'implication des GEIs dans la diversification fonctionnelle des *Thiomonas* spp. et ont également mis au jour des remaniements génomiques importants chez certaines souches. J'ai étudié la corrélation entre la présence de ces îlots et les capacités des différentes souches concernant le métabolisme de l'arsenic, la dégradation de l'urée, et la biosynthèse de biofilm, qui correspondent respectivement aux trois sous-parties du chapitre.

Partie 1 - Métabolisation de l'arsenic

I - Introduction

Le DMA de Carnoulès est caractérisé par la présence d'une importante concentration en arsénite As(III) en amont et par une diminution relativement rapide de cette concentration pendant les premiers mètres. Cette diminution de la concentration en As(III) est due en partie à son oxydation en As(V) qui s'adsorbe aux oxydes de fer puis précipite. Les sédiments de la zone où la concentration en As(III) diminue sont d'ailleurs riches en arsenic. L'oxydation de cet élément peut constituer un moyen de détoxifier l'environnement, l'As(V) étant généralement moins toxique que l'As(III). L'oxydation rapide de l'As(III) est en partie due à l'activité de bactéries arsenite-oxydantes. La souche *Thiomonas arsenitoxydans* 3As, isolée du DMA de Carnoulès, possède l'opéron *aioBA* codant l'arsénite oxydase. Une précédente étude a mis en évidence la localisation de cet opéron *aio* dans un îlot génomique (Arsene-Ploetze et al., 2010).

Les résultats présentés dans cette partie décrivent les GEIs des souches *Tm. arsenitoxydans* 3As, CB2, et *Thiomonas intermedia* K12, une espèce proche de *Tm. arsenitoxydans* 3As mais isolée d'une canalisation d'eaux usées à Hambourg. Une attention plus particulière a été portée sur des GEIs impliqués dans la métabolisation de l'arsenic, et des comparaisons ont été effectuées avec les souches CB1, CB3, CB6 ainsi que deux nouveaux isolats. En parallèle de l'approche génomique, j'ai comparé les capacités de ces souches à résister et à métaboliser l'As(III). Cette étude a fait l'objet d'une publication dans le journal PLoS ONE. J'ai contribué à cette publication de par l'analyse et la représentation des îlots génomiques impliqués dans la résistance à l'As(III), la préparation des échantillons, le traitement des résultats de la cinétique d'oxydation de l'As(III), et les comparaisons de concentration minimale inhibitrice (CMI) d'As(III) des différentes souches par rapport à *Tm. arsenitoxydans* 3As.

II-Publication

PLOS ONE

RESEARCH ARTICLE

Adaptation in Toxic Environments: Arsenic Genomic Islands in the Bacterial Genus *Thiomonas*

Kelle C. Freel[1], Martin C. Krueger[2], Julien Farasin[1], Céline Brochier-Armanet[3], Valérie Barbe[4], Jeremy Andrès[1], Pierre-Etienne Cholley[1], Marie-Agnès Dillies[5], Bernd Jagla[5], Sandrine Koechler[1], Yann Leva[6], Ghislaine Magdelenat[4], Frédéric Plewniak[1], Caroline Proux[5], Jean-Yves Coppée[5], Philippe N. Bertin[1], Hermann J. Heipieper[2], Florence Arsène-Ploetze[1]*

1 Laboratoire Génétique Moléculaire, Génomique et Microbiologie, UMR7156, CNRS-Université de Strasbourg, Département Microorganismes, Génomes, Environnement, Equipe Ecophysiologie Moléculaire des Microorganismes, Institut de Botanique, Strasbourg, France, **2** Department Environmental Biotechnology, Helmholtz Centre for Environmental Research–UFZ, Leipzig, Germany, **3** Université de Lyon, Université Lyon 1, CNRS, UMR5558, Laboratoire de Biométrie et Biologie Évolutive, Villeurbanne, France, **4** Laboratoire de Biologie Moléculaire pour l'Etude des Génomes, (LBioMEG), CEA-IG-Genoscope, Evry, France, **5** Plate-Forme Transcriptome et Epigénome, Centre d'Innovation et de Recherche Technologique—Département Génomes et Génétique, Institut Pasteur, Paris, France, **6** Université de Haute-Alsace, Biopôle–LVBE, Colmar, France

* ploetze@unistra.fr

OPEN ACCESS

Citation: Freel KC, Krueger MC, Farasin J, Brochier-Armanet C, Barbe V, Andrès J, et al. (2015) Adaptation in Toxic Environments: Arsenic Genomic Islands in the Bacterial Genus *Thiomonas*. PLoS ONE 10(9): e0139011. doi:10.1371/journal.pone.0139011

Editor: Hemant J Purohit, National Environmental Engineering Research Institute CSIR, INDIA

Received: April 28, 2015

Accepted: September 7, 2015

Published: September 30, 2015

Data Availability Statement: All relevant data are within the paper and its Supporting Information files. All genome files are available from the US DOE Joint Genome Institute database (NC_014153.1, NC_014154.1, NC_014155.1) or EMBL database (accession number(s) LK931581-LK931672, LN831666-LN831688, LN831730-LN831775, LN831689-LN831714, CTRK01000001-CTRK01000065 and CTRL01000001-CTRL01000068).

Funding: This work was supported by the Université de Strasbourg (UdS), the Centre National de la

Abstract

Acid mine drainage (AMD) is a highly toxic environment for most living organisms due to the presence of many lethal elements including arsenic (As). *Thiomonas (Tm.)* bacteria are found ubiquitously in AMD and can withstand these extreme conditions, in part because they are able to oxidize arsenite. In order to further improve our knowledge concerning the adaptive capacities of these bacteria, we sequenced and assembled the genome of six isolates derived from the Carnoulès AMD, and compared them to the genomes of *Tm*. *arsenitoxydans* 3As (isolated from the same site) and *Tm*. *intermedia* K12 (isolated from a sewage pipe). A detailed analysis of the *Tm*. sp. CB2 genome revealed various rearrangements had occurred in comparison to what was observed in 3As and K12 and over 20 genomic islands (GEIs) were found in each of these three genomes. We performed a detailed comparison of the two arsenic-related islands found in CB2, carrying the genes required for arsenite oxidation and As resistance, with those found in K12, 3As, and five other *Thiomonas* strains also isolated from Carnoulès (CB1, CB3, CB6, ACO3 and ACO7). Our results suggest that these arsenic-related islands have evolved differentially in these closely related *Thiomonas* strains, leading to divergent capacities to survive in As rich environments.

Introduction

Recent studies revealed that both horizontal gene transfer (HGT) and Genomic Islands (GEIs) may confer selective advantages to bacteria, and are essential in adaptation to extreme

Recherche Scientifique (CNRS) and the Region Alsace to JF. This study was financed by the EC2CO program (INEE, CNRS), the RARE (ANR 07-BLANC-0118) and the THIOFILM (ANR-12-ADAP-0013) projects. JF and KCF were supported by the Agence Nationale de la Recherche, ANR THIOFILM (ANR-12-ADAP-0013). FA-P was supported by the CNRS (« mise en délégation pour activité de recherche au CNRS »). CB-A is member of the Institut Universitaire de France and was supported by a Projet d'Avenir (Ancestrome, ANR-10-BINF-01-01) grant. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

environments [1–3]. GEIs consist of discrete DNA segments (ranging from 10 to 200 kbp in size) which sometimes differ in their nucleotide features (G+C content or codon usage) from the rest of the genome, and have often been found in the vicinity of tRNA or tRNA-like genes. The boundaries of these islands frequently correspond to perfect or near-perfect direct repeats (DRs). These regions often harbor functional or cryptic genes encoding integrases originated from phages or genes involved in plasmid conjugation processes. GEIs thus include elements of other kinds such as integrative and conjugative elements (ICE), conjugative transposons and cryptic or defective prophages and can result from one or several HGT events and genomic rearrangements [2]. Thus, studying bacterial colonization of diverse niches is one way to understand the forces driving bacterial evolution [4–6], which plays a particularly important role in habitats with high levels of toxic metals that provide stressful conditions. These extreme habitats can arise naturally, or as a result of human activities, although in both cases, bacteria have adapted to thrive in these challenging ecological niches. Acid mine drainage (AMD) is an excellent example of a stressful environment in which only few organisms can survive. Mining activities, involving the processing of sulfide ores, lead to the production of toxic, metal rich waste, and when mines are left exposed, weathering of the ores results in AMD [7]. Ultimately, AMD contaminates runoff and produces streams with acidic pH and high concentrations of heavy metals. Common pollutants include arsenite (As(III)) and arsenate (As(V)), two inorganic forms of arsenic (As) which are highly toxic and dramatically impact microbial community composition. The Carnoulès site in the south of France has a typical AMD profile with a pH of approximately 3 and particularly high levels of As, ranging in concentration from 50 to 350 mg/L. This AMD is therefore an excellent model site to examine extreme microbial environments and has been the focal point of a suite of studies [8–11].

Bacteria belonging to the genus *Thiomonas* are found ubiquitously at AMD sites and several strains (including CB1, CB2, CB3, CB6 and 3As), have been isolated from the Carnoulès AMD [12–16]. These facultative chemolithoautotrophic betaproteobacteria are able to use reduced inorganic sulfur compounds (RISCs) as electron donors. *Tm. delicata* (previously *Tm. arsenivorans* b6) is capable of autotrophic growth while *Tm.* spp. CB1, CB2, CB3 and CB6 survive only in the presence of organic carbon sources [13–15]. Using Comparative Genomic Hybridization (CGH), our previous work demonstrated that *Thiomonas* genomes have been shaped by the acquisition and loss of GEIs, which may have helped strains adapt to different niches within the AMD [16]. While it was possible to detect GEIs and uncover evidence of HGT using the CGH approach, forces that lead to bacterial adaptation in toxic environments are not thoroughly understood. The genome of two *Thiomonas* strains, *Thiomonas arsenitoxydans* 3As (3As) isolated from the Carnoulès AMD [16], and *Thiomonas. intermedia* K12 (K12) isolated from sewage pipes [17], (http://genome.jgi-psf.org/thiin/thiin.info.html) are available in public databases and provided the initial glimpse into the genetic capacity of these *Thiomonas* strains. In particular, *Tm. intermedia* K12 has not been well studied and, to our knowledge, its phylogenetic relationship with other characterized *Thiomonas* strains, in particular with the *Tm. arsenitoxydans* strain 3As was never analyzed. The DNA-DNA hybridization values proposed suggest that while it clearly belongs to a different species, 3As is genetically related to the *Tm. intermedia* strain, DSM 18155 [18]. The isolates designated as 3As, CB1, CB2, CB3 and CB6 are phylogenetically related and share more than 99% sequence ID according to the 16S rRNA [16]. Despite their relatedness, these isolates each have unique metabolic capacities [15,16,19], thus representing an ideal model system to analyze bacterial genomic divergence in extreme and polluted environments on a fine evolutionary scale.

In this study, we sequenced the genome of strain CB2, and optimized the assembly in order to compare this genome's architecture to that of 3As and K12. Our results revealed that the CB2 genome was subject to rearrangement and contains genomic islands involved in heavy

metal transformation and resistance. Second, we sequenced the genomes of additional five *Thiomonas* strains isolated from AMD and we compared those draft genomes with the genomes of CB2, 3As and K12. This revealed that the *Thiomonas* strains that originated from the AMD have adapted to an extremely hazardous habitat by maintaining specific GEIs that confer As resistance as well as the capacity to oxidize arsenite.

Materials and Methods

Bacterial cultivation and growth conditions

In this study, the strains *Thiomonas* spp. ACO3, ACO7 were isolated on R2A agar medium (BD Difco) after 7 days of incubation at 30°C from sediments collected at the Carnoulès AMD after being processed on a Nycodenz gradient as previously described [20]. The strains *Thiomonas* spp. CB1, CB2, CB3, CB6, and 3As (DSM 22701) were previously isolated from AMD-impacted water at the Carnoulès AMD (France) [14,16] and K12 was obtained from a corroded concrete wall of the Hamburg sewer system [17]. After isolation or recovery from cryogenic stocks (for all strains previously isolated) the *Thiomonas* spp. studied here were maintained at 30°C in modified m126 medium. The composition in one liter of diH_2O was: 4.5 g Na_2HPO_4; 1.5 g KH_2PO_4; 5.0 g $Na_2S_2O_3$; 0.3 g NH_4Cl; 0.1 g $MgSO_4.7H_20$; and 0.5 g yeast extract. The pH of the media was adjusted to 5.0 with H_2SO_4. Variations of this media included the addition of sodium arsenite (from a 500 mM stock solution), at the concentration listed in the text. *Klebsiella pneumoniae ozenae* KIIIA [21] was used in the plasmid visualization experiment, and was grown in LB medium (BD Difco) at 30°C.

Assessment of arsenite oxidation

In an initial set of experiments, 100 mL of either standard m126 or m126 supplemented with 5 g/L glucose was inoculated with CB2 from a liquid preculture to an OD_{600} ~ 0.002 and then was incubated at 30°C while horizontally shaking at 120 rpm. After 24 hours of unimpeded growth, As(III) was added from a sterile stock to a final concentration of 3 mM. Uninoculated flasks with the respective media served as abiotic controls. Samples of 1 mL were taken at daily intervals and centrifuged for 10 minutes at 16, 000 x *g*. After centrifugation, the supernatants were stored at 4°C in order to avoid precipitation of dissolved As salts observed in frozen samples. For speciation analysis, all samples were diluted 100-fold with distilled water before being subjected to inductively coupled plasma mass spectrometry (ICP-MS) or inductively coupled plasma atomic emission spectroscopy (ICP-AES) analysis. Standards of inorganic As(III) and As(V) were used for calibration throughout all experiments. In a second set of experiments, three independent 25 mL cultures of m126 supplemented with 1.33 mM of As(III) were inoculated with CB1, CB2, CB3, CB6, K12 or 3As to an OD_{600} ~ 0.002. An abiotic control (no inoculation) was also prepared. These cultures were incubated for 120 h at 30°C shaking at 120 rpm. 600 μL of the cultures were sampled at 0 h, 24 h, 48 h, 72 h, and 120 h of growth and were immediately centrifuged at 10, 000 x *g* for 10 min, after which 500 μL of the supernatant was diluted in 12 mL of ultrapure water. This dilution was then injected in an anion-exchange chromatography column (BOND ELUT JR-SAX, 500MG, Agilent Technologies) which retained the As(V), while the flow through fraction contained As(III). The bound As(V) was eluted with 0.12 M HCl solution. The two As(III) and As(V) fractions obtained were conserved at 4°C, and the As concentration was assessed by ICP-AES.

Determination of arsenite minimal inhibitory concentration (MIC)

20 mL of m126 was inoculated with K12, CB1, CB2, CB3, CB6, 3As, ACO3, or ACO7 to an OD_{600} ~ 0.002 and incubated for 40 h at 30°C, with shaking at 120 rpm. After growth, the

cultures were diluted to an OD_{600} ~ 0.1 and serial dilutions were then made in m126. Then, 2 μL of each dilution was spotted on solid m126 plates supplemented with different concentrations of arsenite (1.33 mM and from 6.67 to 16 mM), and incubated at 30°C for 10 days.

Western blot

Monoclonal antibodies raised against the AioA peptide were obtained from Perbio Science (Erembodegem, Belgium). Briefly, a peptide with the sequence CGYHAYTWDADREG GRAPHC was synthesized. This peptide corresponds to the N-terminal amino acids residues 29 to 47 of the arsenite oxidase large subunit of *Tm. arsenitoxydans* 3As. The peptide was then coupled to keyhole limpet haemocyanin (KLH). Two rabbits were injected with peptide-KLH and bled at day 90. Antibodies were partially purified on an affinity column substituted with the peptide. Excess antibodies were removed by repeated washing with PBS and then 50 mL of a stationary phase culture in m126 was prepared containing 0, 0.67, 1.33 or 2.66 mM As(III) and centrifuged at 7, 000 x *g*. The pellets were suspended in Laemmli buffer [22] and boiled for 1 min. 15 μL of these extracts were separated by SDS-PAGE using a Amersham™ ECL™ gel 4–12% (GE Healthcare). After SDS-PAGE electrophoresis, the proteins were electrotransfered to a PVDF low-fluorescent membrane (Thermo Scientific) using a TE77 Semi-dry transfer unit (Amersham Biosciences) at 45 mA for 1 h according to the manufacturer's instructions. The membranes were blocked in SuperBlock® blocking buffer (Thermo Scientific) and then washed three times in PBS2 buffer (80 mM Na2HPO4, 20 mM NaH2PO4, 100 mM NaCl, 0.01% tween 20) and incubated for 1 h with AioA antisera (1:1, 000 dilution) in PBS. After a 1 hour incubation in PBS containing the secondary antibody (1:5, 000 dilution of goat anti-rabbit IgG DylightTM 488 conjugated highly cross-adsorbed, Thermo Scientific), the membrane was washed twice with PBS in the dark and assessed using a Typhoon Scanner (GE Healthcare).

Sequencing, annotation and optical map analysis of the *Thiomonas* spp. genomes

The 3As genome was previously sequenced and described [16], while the K12 genome was sequenced by the US DOE Joint Genome Institute (NC_014153.1, NC_014154.1, NC_014155.1). The genomes analyzed for this study were integrated into the MicroScope platform [23–25]. CB1, CB2, CB3, CB6, ACO3, and ACO7 genomic DNAs were prepared using the Wizard genomic kit (Promega). The *Thiomonas* CB2 genome was sequenced *de novo* by MWG operon (http://www.operon.com/) using Roche 454 Sequencing through the use of shotgun and paired end (8 kb in size) libraries. The 425, 947 reads gave a 28x coverage. The assembly with Newbler (454 Life Science/Roche) yielded 92 contigs and ultimately 16 scaffolds. To optimize the CB2 assembly and to compare it to that of K12 and 3As, an optical map was generated using the Argus™ Whole-Genome Mapping System (www.opgen.com, Gaithersburg, Maryland, USA). The optimal restriction enzyme *Bgl*II was chosen using "Enzyme Chooser" (OpGen Inc., Gaithersburg, MD), a software program allowing for the identification of the best enzyme for analysis. The selected enzyme produced an average fragment 6–12 kb in size with no single restriction fragment larger than 80 kb across the genome. Overall, 180 single-molecule restriction maps with an average size of 214 kb each were generated. The assembly of different molecules permitted us to obtain a unique circular map (3.5 Mb). Contig fasta files were imported into MapSolver™ software and *in silico* maps were created using *Bgl*II restriction sites. The *in vitro* and *in silico* maps were compared for each of the 16 scaffolds, enabling their reorganization into 9 scaffolds. The largest scaffold spanned 3.8 Mb, which was over 98% of the CB2 genome sequence. Accession numbers for the CB2 genome (EMBL database) are from LK931581 to LK931672.

The CB1, CB3, CB6, ACO3 and ACO7 genomes were sequenced on an Illumina instrument (www.illumina.com). Genomic DNA of the strains CB1, CB3, and CB6 was fragmented and inserts approximately 8 kb in length were selected to construct mate pair indexed libraries. These libraries were loaded on a MiSeq sequencing device flowcell and sequenced on pair-ends, 150 nt length. CB1, CB3, and CB6 reads were assembled in parallel using Velvet (De Bruijn Graph) (https://www.ebi.ac.uk) and using Newbler (OLC) (www.roche.com, Basel, Switzerland) followed by SSPACE [26] to organize the contigs longer than 500 nt. The different assemblies were compared and the best were retained (performed with Newbler for CB1 and CB6 and with Velvet for CB3). The gap filling was completed for the 3 strains using GapCloser (http://soap.genomics.org.cn/soapdenovo.html) on scaffolds more than 2 kb. Genomic DNA of ACO7 and ACO3 were fragmented and inserts between 500 and 600 nt were selected to construct pair-end indexed libraries. These libraries were loaded on a MiSeq sequencing device flowcell and 300 nt length sequences were obtained for each extremity. The data were merged before assembly with Newbler and in order to reduce the number of undetermined bases, GapCloser was used. Due to the high similarities of sequence between these genomes and the reference strain 3As genome, contigs longer than 500 nt were then organized by comparison with 3As genome sequence. Accession numbers for the CB1, CB3, CB6, ACO3 and ACO7 genomes are LN831666-LN831688, LN831730-LN831775, LN831689-LN831714, CTRK01000001-CTRK01000065 and CTRL01000001-CTRL01000068, respectively (EMBL database).

The automated annotation was completed by Genoscope (Centre national de séquençage français) with the MicroScope platform [23–25]. MicroScope uses the Prokaryotic Genome DataBase (PkGDB), to analyze and compare the annotation of genomes of bacteria and Achaea. A manual annotation was completed for genes in the CB2 genomic islands RGP10, RGP19, and RGP9. The data and annotations are accessible using the MicroScope web interface, MaGe (Magnifying Genomes). Whole genome comparisons were performed using the nucmer module of the MUMmer 3.0 program [27] with default parameters. Dot plot figures were subsequently generated using mummerplot with the color parameter on. RGP finder, a module in MaGe [25] was used to initially identify and compare the genomic islands of *Thiomonas* strains. Comparative genomic analyses were performed using the MicroScope plateform tools ("Pan/core-genome" option) using the default MICFAM parameters.

Visualization of plasmids

Plasmids were visualized using the Wheatcroft's method method [28] with some modifications. *Klebsiella pneumoniae ozenae* KIIIA [21] was grown at 30°C in LB. This strain contains 3 plasmids with estimated sizes of 225 kb, 130 kb and 45 kb [29]. *Thiomonas* strains were grown at 30°C in m126 medium. For each strain, a volume of culture equivalent to 1.5 mL of cells at OD_{620nm} = 0.4 was harvested and centrifuged for 3 min at 13, 000 x *g* and 4°C. The pellet was suspended in 0.5 mL water at 4°C. The cell suspension was layered onto 1 mL of 0.3% N-lauroylsarcosine sodium salt solution and centrifuged at 13, 000 x *g* for 3 min at 4°C. The pellet was immediately resuspended in 40 μL of 40 mM Tris, 10 mM EDTA, 20% Ficoll and incubated on ice for 15 min. A 0.75% agarose gel (pulse field certified agar, Bio-Rad) was prepared with TBE buffer and 25 μL of 10% SDS containing xylene cyanole was loaded in each well. Migration was completed in TBE buffer at 100 V reversing polarities of the electrodes until the dye migrated 1 cm behind the wells. 20 μL of lysis solution was added to each cell suspension and 60 μL was loaded on the gel. The lysis solution used for *K. pneumoniae* and *T. intermedia* K12 contained 100 mM Tris, 10 mM EDTA, 0.4 mg/mL RNase and 10 mg/mL lysozyme. For CB1, CB2, CB3, CB6 and 3As, 10 mg/mL proteinase K and 4 μL/mL ß-glucanase 2 (20 mg/mL) were added to

the lysis solution and the lysozyme concentration was increased to 15 mg/mL. Migration was performed at 50 V for 30 min and then at 100 V for 5 h. The gel was stained with ethidium bromide and visualized under a UV-transilluminator.

PCR experiments

Cells were grown in m126 medium for 48h and centrifuged at 5, 000 x *g* for 15 minutes. DNA was extracted as previously described [15]. The presence of the two plasmids was verified by PCR using a Taq Core Kit (MP Biomedicals). Each reaction mixture (total volume of 25μL) contained 250 ng of DNA. Thermocycling conditions were as follows: 2 minutes at 94°C followed by 35 cycles of 40 seconds at 94°C, 45 seconds at 61°C and 4 minutes at 72°C. A final elongation step at 72°C for 7 min was added. Primers were designed as follows: 870001_280sor 5'-CAGCGATGTGCACATCGAGC-3', 870008_10220sor 5'-GCAGTAGCTCGCGGAT GATG-3', 880002_180sor 5'-CGTTCTAGGCGTACCGGTAC-3', 8800035_20340sor 5'-CAGCGCAAGTCTTCCTGGTC-3', 890001_145sor 5'-CGATCGTACTGGGTGAGCTC-3', 890016_7140sor and 5'-GAGCCTGAGCGTGAGCAATC-3' for the ascertainment of the two plasmids and 730161-830for 5'-GCCTGCTGTACGAAGCCAAC-3', 730162-1610rev 5'-CCACGATCAGGTCGACAACC-3', 780013-120for 5'-GCGGCTAAGTCTGAGGAACG-3' and 780014-850rev 5'-CCGTCGGATGTGTCGATGTG-3' for testing the circularization of the ICE19. Products of amplification were sequenced at Eurofins Genomics.

RNA sequencing

CB2 was cultivated in 20 mL of m126 liquid medium until OD_{600} = 0.2. Cells were then diluted 12.5-fold in the same medium in the absence or presence of 1.3 mM As(III) and incubated in a 6-well plate (Costar, Corning, USA) without shaking. After 48 h the planktonic cells were carefully removed and new sterile medium was added. After 72 h the planktonic cells were removed, biofilm cells were washed 2-times with NaCl 9 g/L and were manually recovered by scratching the dishes with NaCl 9 g/L. Cells were pelleted by centrifugation for 15 min at 4, 000 x *g*. RNA was extracted from these pellets as previously described [30]. For RNA-Seq, enriched mRNA was obtained from 10 μg of total RNA using the rRNA capture hybridization approach from the RiboZero kit (Epicentre, Singapore), according to the manufacturer's instructions. For high-throughput sequencing, non-directional cDNA libraries were prepared from enriched fragmented mRNA using the RNA sample preparation kit, set A (Illumina, San Diego, CA, USA). Fragments of cDNA of ± 150 bp, ligated with Illumina adapters and amplified per PCR, were purified from each library. Quality and quantity was confirmed on a Bioanalyzer (Agilent, Santa Clara, CA, USA). Sequencing of 51 bases was performed in single-end mode, using an Illumina HiSeq2000 instrument (Illumina). Reads were cleaned from the adapter sequences and from sequences of low quality using an in-house program. Only sequences with a minimum length of 25 nt were considered for further analysis.

The Bowtie program was used to align the reads to the *Thiomonas* CB2 genome. Statistical analyses were performed with R software [31] and Bioconductor packages [32]. Normalization and differential analysis were carried out according to the DESeq model and package [33]. A Benjamini and Hochberg (BH) [34] p-value adjustment was performed to take into account multiple testing and control the false positive rate to a chosen level, α set to 0.05, yielding 42 up-regulated genes, i.e. more expressed in the presence of As than in the absence. In order to compare expression levels between different genes, normalized expression values were computed as follows: counts were first normalized with DESeq in order to make reads comparable between samples; these normalized counts were then divided by gene length (in nt) and multiplied by 1, 000, in order to obtain expression levels by kilobase.

Results

Genome sequencing of *Thiomonas* strains and assembling

There were several strains used in this study that were previously isolated from the Carnoulès AMD: 3As, CB1, CB2, CB3, and CB6. Here, we also report two novel strains isolated from this AMD community: ACO3 and ACO7. Since the phylogeny of closely related strains is difficult to resolve using only the 16S rRNA gene, we combined five markers (*rpoA*, *dnaK*, *atpD*, 16S rDNA, and 23S rDNA) to clarify the relationships among these *Thiomonas* strains (Fig 1). The resulting maximum likelihood and Bayesian trees showed that 3As, CB1, CB6, ACO3, and ACO7 are very closely related (bootstrap value (BV) = 99% and posterior probability (PP) = 1.0), and cluster with CB3 and CB2, while K12 is more divergent (BV = 82% and PP = 1.0).

The genomes of CB1, CB2, CB3, CB6, ACO3, and ACO7 were sequenced. While CB2 harbors the most divergent genome compared to 3As among these strains, we attempted to obtain the most complete genome assembly of possible for this strain. The CB2 scaffold organization was determined with an optical map, and of the 16 scaffolds, of which organization of twelve was possible. Five scaffolds could not be mapped precisely to the genome: two were too small (5 and 2.4 kb) to be assembled according to the optical map, and the other three (20.9, 10.8 and 7.9 kb) likely correspond to plasmid DNA as is indicated by the presence of *mob* and *par* genes which are involved in plasmid mobilization, replication, and stabilization [35–38]. PCR experiments were performed showing that the 7.9 Kb scaffold correspond to one small plasmid and

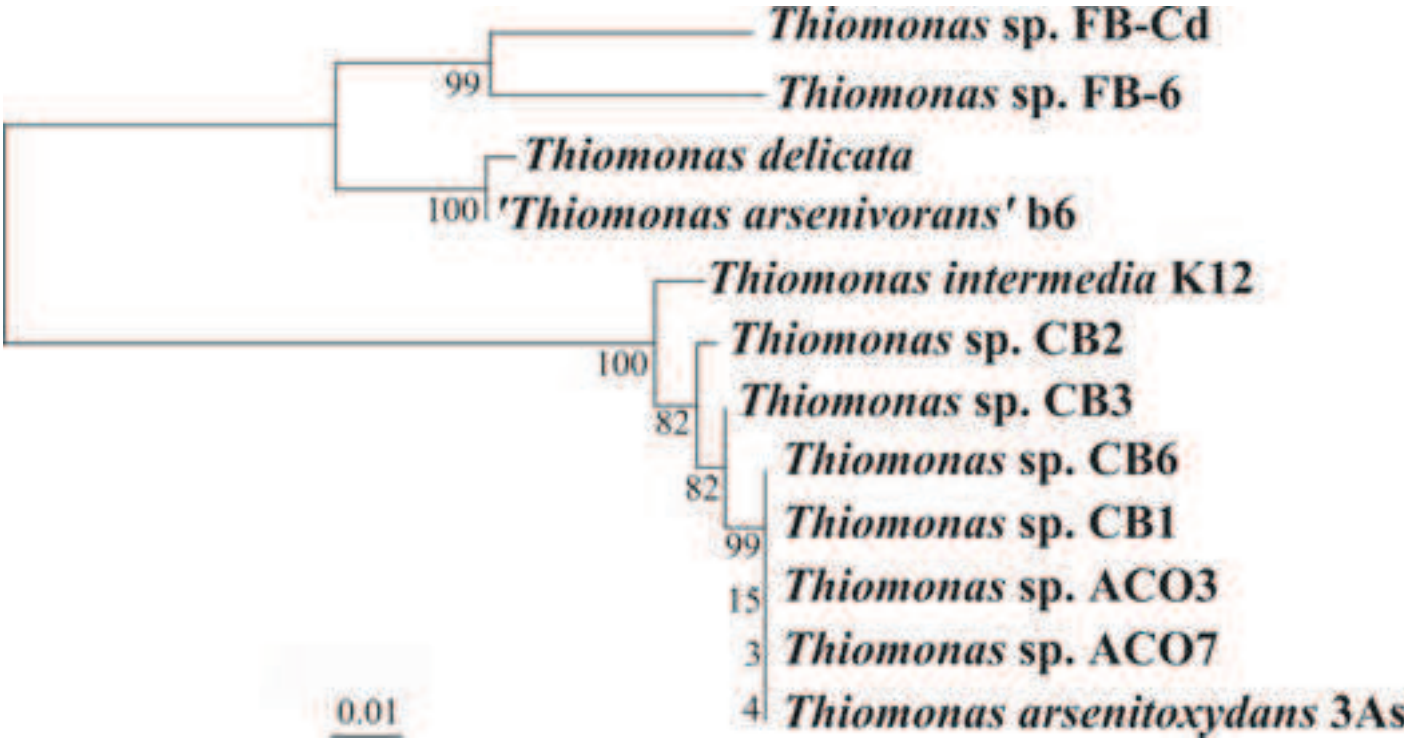


Fig 1. Maximum likelihood tree of the *Thiomonas* strains included in this study. The tree is based on the concatenated alignment of *rpoA*, *dnaK*, *atpD*, 16S rDNA, and 23S rDNA genes (8628 nucleotide positions). The sequences were aligned with the MUSCLE and trimmed with GBLOCKS programs implemented in SEAVIEW software version 4.5.4 [46] and then concatenated. The tree was inferred with PhyML version 3.1 [47], using the GTR+Γ4 model as suggested by the proposed model tool implemented in TREEFINDER v2011 [48], using the NNI+SPR option for topology search. Values at branch nodes correspond to bootstrap values (100 replicates) calculated with PhyML and to posterior probabilities calculated with MrBayes version 3.2.2 [49], with the GTR + G4 model. The scale bar represents the average number of substitutions per site.

doi:10.1371/journal.pone.0139011.g001

the two other scaffolds (20.9 and 10.8 Kb) correspond to one plasmid of 31.7 Kb (data not shown). Moreover, electrophoresis of CB2 gDNA detected an extra chromosomal element less than 40 kb in length (S1 Fig), which further supports the presence of the 31.7 Kb plasmid. Similarly, one plasmid was detected in CB1 (approximately 40 kb), and two were detected in CB3 (one of approximately 40 kb and one less than 40 kb) however, no plasmid was detected in CB6 (S1 Fig). These observations are in agreement with genomic data since, in the case of CB1 and CB3, the presence of regions similar to the 3As plasmid pTHI (46.8 kb) [16] suggested that plasmids similar to pTHI were present in CB1 and CB3. These observations revealed that *Thiomonas* strains harbor various plasmids of different sizes with different sets of genes.

General features including the GC-contents (63.7%– 63.9%) are similar among the eight strains, while the genome size and gene content varied from 3.4 Mb (K12) to 4.3 Mb (CB3) (S1 Table). The CB3 genome was larger, due in part to multiple duplications (77 tandem duplications including five large duplications (>10 genes)) and had a higher percentage of repetitive sequences compared to the other genomes (S1 Table). Finally, a comparative genomic analysis of CB2, 3As, and K12 revealed that the pan genome of these three strains is comprised of 11, 016 genes, while the core-genome encompassed 7, 381 genes (80% of amino acid identity, 80% alignment coverage), representing 66.8%, 65.0%, and 69.4% of the 3As, CB2, and K12 proteomes, respectively. This analysis also highlighted that 285 genes are present in CB2 and 3As (isolates from the Carnoulès AMD,) but not in K12 (https://www.genoscope.cns.fr/agc/microscope/home/index.php). These genes encode enzymes involved in glutamate, biotin, and cobalamin biosynthesis, toxic metal resistance, nitrite and nitrate reduction, sulfate assimilation, transcriptional regulation, permeases, methyl-accepting chemotaxis proteins, transporters, as well as proteins of unknown function (https://www.genoscope.cns.fr/agc/microscope/home/index.php) representing 7.5% (CB2) to 7.8% (3As) of the AMD genomes, and possibly required for survival in AMD.

Comparison of the genome structure of *Thiomonas* strains CB2, 3As, and K12

Previous analysis based on the use of CGH revealed dramatic variation in the genomic content of eight *Thiomonas* strains (CB1, CB2, CB3, CB6, 3As, *Tm. arsenivorans*, Ynys1 and *Tm. perometabolis*) despite their phylogenetic proximity. Indeed, the 16S rDNA/*rpoA*-based phylogeny of the CB1, CB2, CB3, CB6, and 3As isolates (which share a 97% nucleotide identity), as well as DNA-DNA hybridization experiments, suggested that while they are indeed 5 different strains, they are of the same species. However, subtle physiological differences were observed among the isolates, and only 74.7% of the genome of strain 3As was sufficiently conserved to allow hybridization using the CGH approach [16]. Nevertheless, this did not allow for the comparison of genomic synteny or to detect if the genomes shared similar architectures. Here, due to the high quality of sequence assembly, the architecture of the 3As, K12 and CB2 genomes could be compared and the genomic islands identified. Dot plots were constructed for pairwise genomic comparisons among the three isolates, revealing striking similarities between 3As and K12 compared to the CB2 genome (Fig 2A). The 3As and K12 genomes are highly syntenic and similar in gene content (Fig 2A), although 3As harbors several insertions (indicated by breaks in the line) and one inversion compared to K12. In contrast, genome organization between CB2 and 3As is not well conserved due to a variety of rearrangements and inversions (Fig 2A). Since the 3As and K12 strains are phylogenetically distant but harbor similar genomic organization (Figs 1 and 2A), we hypothesize that these genomes represent the ancestral state, while that of CB2 is more derived (Fig 2B). Accordingly, three major genomic rearrangements and inversions could have occurred in the CB2 lineage after its divergence from the two others

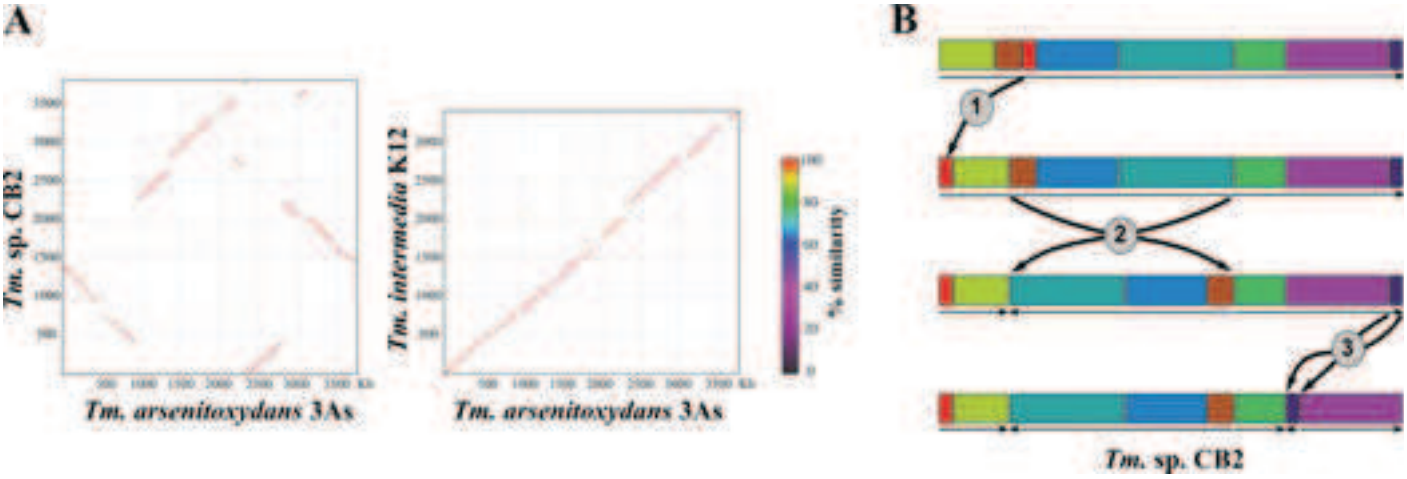


Fig 2. Comparison among the CB2, K12, and 3As genomes. (A) Dot plot, CB2 vs 3As and 3As vs K12. The genomes of 3As and K12 are well conserved while the CB2 genome appears to have undergone chromosomal rearrangements. (B) Scheme demonstrating the key differences between the *Thiomonas* ancestor and CB2 genome. Above, "1", "2" and "3" highlight a translocation of 0.1 Mb, an inversion of 1.8 Mb and a translocation and inversion of 0.13 Mb, respectively.

doi:10.1371/journal.pone.0139011.g002

strains: (i) the translocation of a 0.1 Mb region; (ii) the inversion of a 1.8 Mb region, and (iii) the translocation and inversion of a 0.13 Mb region (Fig 2B).

The CB2 genome is larger than that of 3As and K12 due to the presence of sizable regions that may correspond to GEIs. Careful analyses were completed on the CB2, 3As and K12 genomes to define GEIs (also referred to as Regions of Genomic Plasticity or RGPs by "RGP finder"), a module in the MaGe platform (https://www.genoscope.cns.fr/agc/microscope/home/index.php). According to the criteria used, the 3As genome harbored 30 GEIs, among which 19 were identified previously using the CGH approach [16]. Twelve GEIs were specific to 3As and 5 were shared with K12. We identified 24 GEIs in the K12 genome, 11 of which were specific to this strain. We then carefully analyzed the 20 GEIs identified in CB2, which were $>$ 5 kb, and accounted for more than 27.8% (1, 065 CDS) of the genome (S2 and S3 Tables). The island identified as RGP19 was composed of at least two distinct islands, labeled RGP19 and RGP20 (S3 Table), since these genes were found in a different scaffold. Remarkably, genes involved in plasmid conjugation (*tra* or *trb* genes) were found on five of these GEIs, suggesting that these sequences were perhaps originally acquired by plasmid conjugation and integration. Four CB2 GEIs, i.e. RGP5 (29.8 kb, 26 genes), RGP8 (35.7 kb, 32 genes), RGP12 (36.3 kb, 33 genes) and RGP17 (9.9 Kb, 12 genes) contain genes with homologs in distantly related bacteria (https://www.genoscope.cns.fr/agc/microscope/home/index.php) but not in 3As and K12, or in other *Thiomonas* genomes (S3 Table). In RGP12, of the 33 genes 4 genes were found duplicated in CB2 in another GEI and involved in urea transport. Overall, these data suggests that CB2 acquired the corresponding genes via horizontal gene transfer. The 3As and K12 genomes shared five and four distinct GEIs with CB2, respectively, while only two islands (RGP1 and 13) were found in all three genomes. Interestingly, genes found in RGP7 and a large part of RGP19 of CB2 were present in all other AMD strains (CB1, CB3, CB6, 3As, ACO3, and ACO7) but not in K12, suggesting that these two islands were most likely important for survival in AMD. The RGP7 contains almost exclusively proteins of unknown function, while RGP19 carries genes involved in As resistance and transformation (S3 Table) and was further analyzed in detail.

Detailed analysis of the *ars* and *aio* genes of CB2, involved in As metabolism

Microbial mechanisms to mitigate the toxic effects of As include the expression of *ars* genes, which encode an arsenite (As(III)) efflux pump (*arsB*) and an arsenate (As(V)) reductase (*arsC*) and allow bacteria to reduce arsenate to arsenite in order to ultimately extrude it from the cell [35]. These genes are usually present as a three-gene operon, *arsRBC*, with *arsR* encoding a transcriptional regulator, or an extended five-gene operon *arsRDABC*, with *arsA* encoding an ATPase involved in As(III) extrusion, and *arsD* encoding an arsenite chaperone that transfers arsenite from glutathione-bound complexes to the ArsA subunit of the ArsAB complex. The *arsA* and *arsD* genes supposedly confer higher levels of resistance [35]. Another mechanism of As resistance involves the *aio* operon that encodes proteins involved in detoxifying arsenite by oxidizing As(III) to arsenate As(V) or utilizing As(III) as an electron donor [35].

The CB2, K12, and 3As *Thiomonas* genomes harbor an *arsRICB* operon, sharing more than 97% of similarity at the nucleotide level, that is not located in a GEI but in a very conserved region of these three genomes. Both CB2 and 3As have an additional *arsRDACB* operon with a shared 91% identity at the nucleotide level, located in a GEI in both genomes. In CB2, this operon is found on RGP19, directly upstream from an *aio* locus, while in 3As, it is located on The ThGEI-O described previously [16]. In CB2, two versions of the *aio* operon, sharing less than 90% identity at the nucleotide level, are located in two distinct GEIs (referred to from now on as "arsenic genomic islands") (Fig 3). One copy is encoded in RGP19 of CB2, while the second is located in RGP10. In contrast, a single *aio* operon is found in 3As [16] and in K12 (this work). The CB2 AioA and AioB proteins encoded by the operon located on RGP19 both share 100% identity with the proteins encoded by 3As, and 84% and 88% identity with those of K12, respectively (Table 1). However, the AioA and AioB encoded by the operon on RGP10 share 82% and 89% identity with the 3As proteins and 87% and 98% identity with those of K12 (Fig 3 and Table 1). These observations revealed than one copy of the *aio* locus and associated genes of CB2 (located on RGP19), were more related to those in 3As, whereas the second copy of the *aio* and associated genes in CB2 (located on RGP10), were more related to those found in the unique *aio* locus of K12 (Fig 3, Table 1 and S2 Fig). To test if the genes are functional in CB2, we assessed the extent of As (III) oxidation, and found that more than 60% of As(III) was oxidized to As(V) within 48 h, followed by a slow reduction of As(V) during the stationary growth phase (Fig 4). In other *Thiomonas* isolates, where arsenite oxidation is involved in energy production, this activity was reduced in the presence of a preferred substrate. In our work, the addition of glucose to m126 caused a severe inhibition of As(III) oxidation in CB2 (Fig 4), suggesting that CB2 may use either arsenite or glucose as an electron donor and the inhibition of the arsenite oxidation could be due to a catabolite repression. Finally, we investigated if both copies of the *aio* genes were expressed in two growth conditions. First, RNAseq was performed on cells grown in conditions favoring biofilm formation (without shaking) and revealed that the expression of the *ars* and the *aio* genes (two copies of each operon), and also other genes from these two islands, are induced by As(III) (S4 Table). Second, a Western blot was performed to determine if both copies were synthetized during aerobic growth. In shaken liquid cultures, we observed that indeed, the two *aioA* genes were expressed only in the presence of As(III) (S3 Fig), as was found previously in 3As [18]. Altogether, these experiments demonstrated that both CB2 *aioAB* operons are functional.

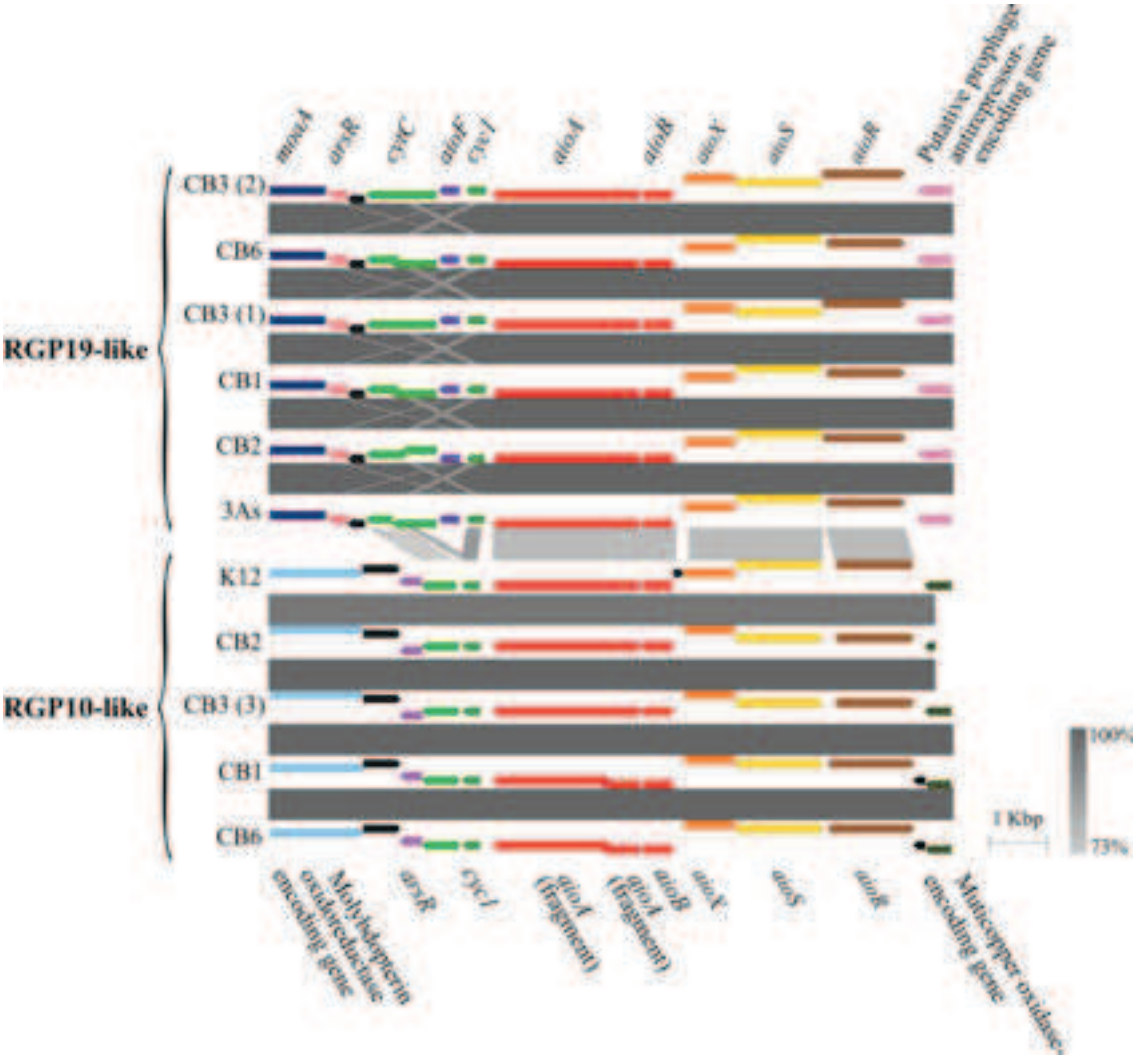


Fig 3. Synteny of the *aio* loci of the *Thiomonas* spp. isolates. The top portion represents the RGP19-like region present in each genomes and the bottom portion represents the RGP10-like region. The % of nucleotide identity is expressed along a grey scale. The genes involved in the arsenic oxidation are localized on the *aio* operon (red). The *aioX* (orange), *aioR* (brown) and *aioS* (yellow) code for proteins involved in regulation and expression of *aioBA*. The cytochrome genes (green) are likely involved in electron transfer. The *moaA* gene (dark blue) is involved in the biosynthesis of molybdopterin, an essential co-factor for the catalytic activity of the protein. Black arrow represents CDS with no known function. Figures were generated with Easyfig [50].

doi:10.1371/journal.pone.0139011.g003

Table 1. Percent of identity between AioA and AioB of CB2, 3As and K12. Whereas both 3As and K12 encoded a single copy of AioA and AioB, CB2 encoded two copies of each protein, one in RGP19 and another in RGP10. AioA and AioB of CB2 encoded in the RGP19 shared more identity with AioA and AioB of 3As (99.9% and 100% respectively). However, AioA and AioB encoded on the RGP10 share more identity with AioA and AioB of K12 (86.9% and 98.3% respectively).

	AioA				AioB			
	3As	**K12**	**CB2-1**	**CB2-2**	**3As**	**K12**	**CB2-1**	**CB2-2**
3As	-	83.7	99.9	82.2	-	87.7	100	88.8
K12	83.7	-	83.8	86.9	87.7	-	87.7	98.3
CB2-1	99.9	83.8	-	82.4	100	87.7	-	88.8
CB2-2	82.2	86.9	82.4	-	88.8	98.3	88.8	-

doi:10.1371/journal.pone.0139011.t001

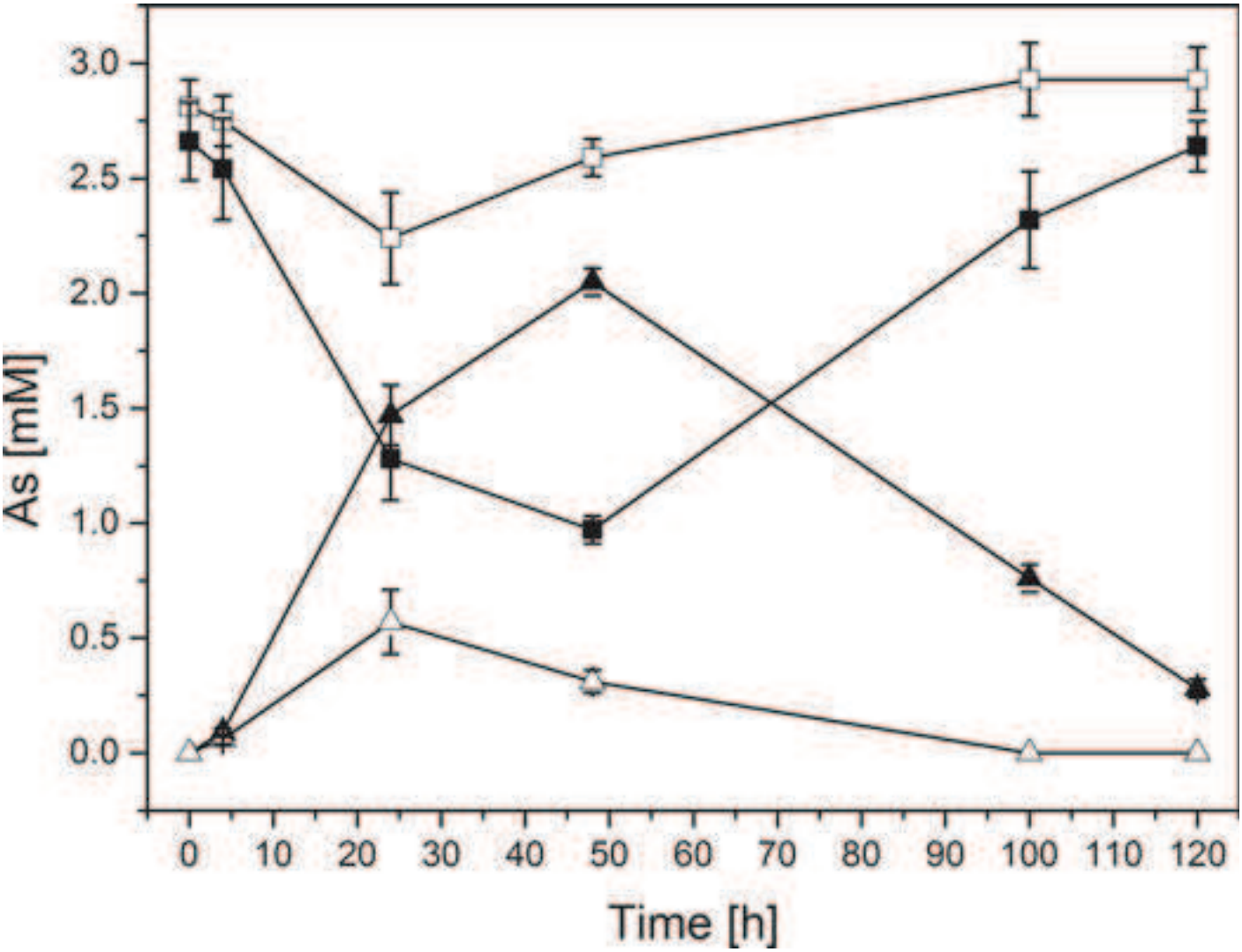


Fig 4. CB2 capacity to oxidize arsenite. Concentrations of As(III) (squares) and As(V) (triangles) are shown for CB2 cells grown in m126 medium in the presence of 2.66 mM of As(III) (full symbols) and m126 medium in the presence of 2.66 mM of As(III) and supplemented with glucose (hollow symbols). Error bars indicate standard deviations of triplicate cultures. No As(III) oxidation was observed in abiotic controls (data not shown).

doi:10.1371/journal.pone.0139011.g004

Comparison of the two CB2 arsenic genomic islands with those found in other *Thiomonas* strains

RGP19 is 338.7 kb long and contains genes implicated in resistance to various heavy metals (As, mercury, copper, zinc, and cadmium). Genes involved in sulfate assimilation (*cys* genes) and biotin synthesis (*bio* genes), were also present in this genomic island. In fact, many integrases are also located on RGP19, as well as genes involved in conjugation, including phage integrase and relaxase. 49 bp nearly perfect direct repeats are located at the left extremity of RGP19 and between gene THICB2v3-780013 and -780014, and the region between these two repeats contains *tra* and *trb* genes involved in conjugation (Fig 5). The excision and circularization of this part have been confirmed by PCR in CB2 (Fig 5C and data not shown). All these observations emphasized the fact that this region has been acquired after insertion and is or, at

Fig 5. Schematic diagram of ICE19 in CB2. This figure shows one part of the arsenic island of CB2 (RGP19) that has the characteristics of an ICE (called ICE19) (A) Alignment of ICE19 of *Thiomonas* sp. CB2 with its cognate portion found in the strain 3As (ThGEI-O). The arsenic island RGP19 of CB2 is localized in a different genomic region that ThGEI-O in 3As. The ThGEI-O of 3As (upper portion of the figure) is localized in a gene encoding a 4Fe-4S ferredoxin which is intact in CB2. Direct repeat sequences are indicated. (B) Sequence comparison of the almost perfect direct repeat flanking the ICE19. *attL*: left DR; *attR*: right DR. (C) Schematic representation of the integrated and circular form of the ICE19 of CB2, which were detected by PCR. The sequence *attI* is identical to *attR* and *attB* identical to *attL*. The % of nucleotide identity is expressed along a grey scale. Figures were generated with Easyfig [50].

doi:10.1371/journal.pone.0139011.g005

some point, was mobile. Consequently, this part of RGP19 (from 3533461 to 3796598, 185,5 Kb) has the characteristics of an integrative and conjugative element (ICE) [36], and has been designated ICE19 and the repeats designed *attL* and *attR* (Fig 5). We observed that the arsenic island is not syntenic in 3As as compared to CB2 and only part of this ICE19 is found in 3As (in the genomic island previously named ThGEI-O [16], Fig 5). In particular, the *tra* and *trb* genes were not found in the ThGEI-O and therefore this island has not the characteristics of an ICE. In 3As, ThGEI-O is located in the middle of a gene encoding a 4Fe-4S ferredoxin, whereas in CB2, ICE19 is found near the *fbp* gene encoding fructose-1,6-bisphosphatase and the Thr tRNA. The presence of direct repeats (DR) at the both sides of the ThGEI-O (RGP19-like island) in 3As was previously described [16] and indicates a probable insertion by site-specific recombination in the gene encoding a 4Fe-4S ferredoxin. Interestingly, similar repeats are also found in the ICE19 of CB2, although one is slightly different, perhaps due to punctual mutations occurred after the insertion or rearrangement of genes in this region (Fig 5). These observations suggest that the two islands, ThGEI-O in 3As and RGP19 in CB2, may have similar origin but evolved differently in the two strains.

RGP10 is shorter (91.2 kb) than RGP19 and contains genes encoding transporters as well as those involved in resistance to cobalt or acriflavine. In addition, RGP10 encodes a hydrogenase similar to that of a gastropod endosymbiont Gammaproteobacteria, which possibly acquired the relevant genes by HGT [37]. A region containing genes similar to those in RGP10 was detected in K12, however, it is shorter in K12 in comparison to RGP10 (S2 Fig). In CB2, an additional region was identified that is not harbored by K12, which includes genes encoding transposases, integrases, and proteins of unknown function. All in all, these observations suggest that the two CB2 arsenic islands RGP19 and RGP10 are different, at least in part, from those of 3As and K12, respectively.

To further analyze these two arsenic genomic islands, we searched for the genes involved in As metabolism in the CB1, CB3, CB6, ACO3, and ACO7 draft genomes. Interestingly, multiple copies of the *aio/ars* genes were in some strains (Fig 3). In fact, the ACO3 and ACO7 genomes are very similar to the 3As genome and encode only one copy of the *aio* genes (https://www.genoscope.cns.fr/agc/microscope/home/index.php). CB1 and CB6 strains harbour one copy of the *aio* genes, found in RGP10, with a frameshift in the *aioA* sequence, and one copy of the *aio/ars* genes is found in CB2 RGP19 (Fig 3 and S2 Fig). A Western blot revealed that only one copy is indeed expressed in CB1, CB6, ACO3, and ACO7 strains with a similar apparent molecular weight when grown in the presence of arsenite (S3C Fig). Interestingly, the genome of the CB3 strain contains five copies of the *ars* genes as well as four copies of the *aio* loci, which were clearly located on different contigs. The *aio* genes in CB2 RGP19 were identified in 3 copies in CB3 (Fig 3); however, one *aio* cluster was only partially intact (only one *aioA* fragment and no *aioB*) and found on a short scaffold. Interestingly, at least two copies of *aioA* were expressed in the presence of arsenite in CB3 and the proteins migrated similarly to the two CB2 AioA proteins (S3C Fig). The synteny around the *aio* genes was analysed in more detail in these *Thiomonas* strains (Fig 3 and S2 Fig). It has to be mentioned that this synteny analysis was limited by the scaffolds limits in each of the draft genomes. Indeed, the synteny of the entire CB2 RGP was occasionally not clarified in other genomes since the scaffold organization could not be deciphered. Nevertheless, this analysis revealed that in all strains, the synteny of the *aio*/*ars* genes was conserved (Fig 3), whereas the synteny of the other genes from these islands was not (S2 Fig). These observations further support the hypothesis that these two arsenic islands have evolved differently in these *Thiomonas* strains through multiple genomic rearrangements or HGT.

Finally, we investigated if the number of *aio* and *ars* genes influenced resistance to As(III) and its oxidation efficiency in *Thiomonas* strains. The MIC for As(III) was previously determined in liquid culture, however, in this study we tested the As(III) resistance on solid media with a narrow range of As(III) concentrations. We observed that the CB2 and CB3 strains are slightly more resistant than K12, ACO3, and ACO7 while the strains 3As, CB1, and CB6 demonstrated the lowest levels of As(III) resistance (S5 Table). We also determined if the kinetics of As(III) oxidation to As(V) were similar in all these strains by measuring the As(III) and As (V) concentrations over 120 h (S4 Fig). The As(III) oxidation activity of K12 was lower compared to the strains isolated from the AMD strains (CB1, CB2, CB3, CB6 and 3As). No dramatic difference in the As(III) oxidation was observed between the AMD strains suggesting that despite several copies of *aio* genes, CB2 and CB3 did not oxidize As(III) more rapidly than the other strains, this activity was even slightly slower in CB2. In contrast, once the As(III) was completely oxidized, As(V) was then reduced in 3As, CB1, and CB6. Such As(V) reduction activity was not observed in CB2 or CB3 in 1.3 mM As(III). These results suggest that arsenate reduction is either less efficient or that As(III) oxidation activity is maintained for a longer time period in these two strains. Altogether, these data revealed that in relation to the genes involved in As metabolism, the As(III) resistance and oxidation capacities are different between these closely related strains.

Discussion

With the growing number of genome sequences available, it is now possible to utilize comparative tools to understand adaptations that microorganisms have acquired in diverse environments. Here, we compared the CB2 genome to that in other *Thiomonas* strains, including two complete (3As and K12) and five draft (ACO3, ACO7, CB1, CB3 and CB6) genomes. These isolates, as well as CB2 and 3As, were isolated from an AMD rich in toxic metals ([14,16] and

this work), while K12 was originally obtained from a sewer in Hamburg, an environment not known for high As concentrations [17]. The genetic relatedness of K12 with other *Thiomonas* strains has never been analyzed previously. Here, a detailed phylogenetic analysis was completed that demonstrated that 3As, CB1, CB6, ACO3, and ACO7 are very closely related and clade with CB3 and CB2, while the K12 strain clearly diverges from this group.

CB1, CB6, 3As, ACO3, and ACO7 expressed only one copy of the *aio* genes, while CB2 and CB3 have two or more copies which are expressed, each likely playing a role in As tolerance and metabolism. To date, the Carnoulès AMD, which is highly contaminated with As(III), is the only location that has yielded isolates with multiple copies of the *aio* operon. Expressing two *aio* operons may allow CB2 and CB3 strains to maintain arsenite oxidation activity for a longer time period, as suggested by the results we've obtained in this study, ultimately allowing these strains to be slightly more resistant to As(III) than the others assessed.

Due to a robust assembly, the comparison among K12, CB2, and 3As allowed us to define essential GEIs, which might be important for survival in AMD. In CB2, we described 20 GEIs, while 30 were found in 3As ([16] and this work). Genes found in only two GEIs were shared between all the isolates from AMD strains but not in K12, revealing that there is not a large pool of AMD-specific genes in *Thiomonas* strains from the Carnoulès AMD. In fact, our data demonstrated that the *Thiomonas* genomes are variable, due to the presence, expansion (possibly via gene duplication and/or HGT), or absence of several GEIs. In addition, several *Thiomonas* strains have at least one plasmid. Remarkably, the size and content of the plasmids found in the strains analyzed here are variable and the functions they encoded were different and may be specific to some strains. Thus, a large portion of the genome in these strains may confer different ecological capacities and functions, acting as a flexible region influencing gene content diversity and suggesting that they adapt differently to their niches. Finally, the K12 genome (3, 467 genes) is smaller than the others (which range from 3, 628 to 4, 508 genes) possibly indicating its native environment requires fewer genes for survival. Additionally, we found a larger number of CDS with unknown functions in the AMD strains, in particular in RGP7, which is found in all the strains isolated from AMD. It may be possible that these genes confer specific capacities not yet characterized, which could allow for survival in AMD.

Our comparative analyses revealed that the CB2 chromosome has a unique architecture in comparison to 3As and K12. Surprisingly, the 3As genome structure was more similar to K12, possibly indicating that the CB2 genome underwent specific genomic events, exemplified by three major inversions and rearrangements compared to 3As. Furthermore, some CB2 GEIs appear to contain multiple islands, most likely acquired after the divergence of this strain. Some of these GEIs, in particular the ICE19 possibly allowing for survival in the harsh AMD environment may have been acquired by HGT [38,39]. In particular, within biofilm structures, HGT is more likely to shuffle genes among organisms [40,41]. Indeed CB2 forms biofilms, which are stimulated in the presence of arsenic [30], which may have played a role in the acquisition of exogenous DNA. Previous work demonstrated genetic exchange is more common between closely related organisms [42], thus *Thiomonas* sp. could potentially transfer the portions of the GEIs that confer a selective survival advantage to other *Thiomonas* strains. Our data and a recent analysis of GEIs involved in As metabolism [43] suggest that both vertical inheritance and HGT favor the spread of such *aio/ars* genes. Thus, rearrangement, HGT and gene loss events probably impacted CB2 genome content after strain divergence, leading to the formation of similar but not identical composite GEIs. In particular, the arsenic islands may have been modified by HGT or rearrangement processes, since it appears that the gene content of these islands is similar but not identical in *Thiomonas* strains.

The analysis of the *Thiomonas* genomes is a step towards understanding environmental pressures on genome evolution. The high frequency of modifications within these sequences

possibly depicts the first phase in the creation of novel species and the extinction of others [5,44,45]. Readily shuffling DNA allows microbial community members to sample the environment for potentially beneficial genes, which has implications in both community diversity and strain-specific lineage survival [39,41]. Our results clearly demonstrate the importance of genomic islands in *Thiomonas* adaptation. These islands allow organisms to inhabit diverse environments by providing genomic plasticity and the rapid acquisition of functional traits. In the case of *Thiomonas*, the environment appears to have played an important role in driving the acquisition of genomic islands, and is possibly the source of selective pressures leading to rearrangements and HGT. Therefore, in order to obtain a detailed understanding of bacterial adaptation, speciation, and evolution, as well as to delineate lineages, a holistic view taking into account bacterial species as well as traditional metrics needs to be adopted.

Supporting Information

S1 Fig. Plasmid detection and visualization in CB1, CB2, CB3, CB6, 3As, and K12 *Thiomonas* strains in agarose gels after ethidium bromide staining. K3: the 3 plasmids from *Klebsiella pneumoniae* with estimated sizes of 225 kb, 130 kb and 45 kb. One extra chromosomal element less than 40 kb was detected in CB2, CB3, and K12, while an additional one of approximately 40 kb was detected in 3As, CB1, CB3, and K12.
(TIF)

S2 Fig. Comparison of the *Thiomonas* genes similar to those found in RGP19 and RGP10 in CB2 genome. The synteny between the 3As, K12, and CB2 genomes and CB1, CB3, and CB6 scaffolds is represented. The % of nucleotide identity is expressed on a grey scale. Genes: red = *aio* genes; purple = *ars* genes; pink: genes involved in metals resistance (Cu, Hg, Cd. . .); yellow: genes encoding transposases, integrases; brown: mobile genetic elements associated genes (including *mob*, *tra*, and *trb)*; orange and green: genes conserved around *aioBA* (see Fig 3); black: unknown function and /or no homology known; blue: others genes. Figures were generated with Easyfig (Sullivan *et al*., 2011). *Thiomonas sp*. CB2 has two *aioBA* operons localized on the RGP19 and RGP10. The strains 3As and K12 have one copy localized in a region of conserved synteny with the RGP19 (RGP19-like) and the RGP10 (RGP10-like) respectively. *Thiomonas* sp. CB3 have two *aioBA* localized in a RGP19-like region and another found in a RGP10-like region. Both CB1 and CB6 have two *aioBA* copies in a RGP19-like and RGP10-like region respectively, but with a frameshift in *aioA* in the latter region.
(TIF)

S3 Fig. Western blots to detect *aioBA* operon expression in *Thiomonas* strains. Expression of two AioA proteins in the presence of arsenic is indicated with arrows. (A) Results from a planktonic culture of CB2 grown without or with arsenic. The two AioA are induced in the presence of As(III). (B) Planktonic cultures of CB2 grown at a range of concentration of As (III). The two AioA are expressed from 0.67 mM to 2.66 mM of As(III). (C) Results from planktonic cultures of CB1, CB2, CB3, CB6 and 3As grown with 1.33 mM of As(III). Two AioA are expressed for CB2 and CB3 and only one of similar weight for CB1, CB6, ACO3 and ACO7.
(TIF)

S4 Fig. Capacity of *Thiomonas* strains to oxidize arsenite, As(III), to arsenate, As(V). Concentrations of As(III) (squares) and As(V) (triangles) are shown for cells grown in m126 with an initial concentration of 1.33 mM As(III). On the left of the figure the concentrations measured at t_{0h}, t_{+24h}, t_{+48h}, t_{+72h} and t_{+120h} are represented, while on the right the concentrations measured at t_{0h}, t_{+6h}, t_{+12h} and t_{+24h} are shown. Error bars indicate standard deviation of data

obtained from triplicate cultures. The As(III) oxidation activity of K12 was lower than the strains from the AMD. After 24 h, and once all As(III) was oxidized, 3As, CB1, and CB6 reduced the As(V), which was not observed for CB2 or CB3.
(TIF)

S1 Table. General characteristics of the *Thiomonas* genomes. Calculation of repetitive regions did not include undetermined ('N') bases. The genome size and gene content range from 3.4 Mb and 4, 367 CDS (K12) to 4.3 Mb and 4, 508 CDS (CB3). *Thiomonas* sp. CB3 also has a higher proportion of repetitive regions (19.21%). The others general features are similar between the strains.
(DOCX)

S2 Table. Regions of genomic plasticity (RGP) identified after comparison of CB2 with 3As and K12. The RGP were identified with the *RGP finder* module in MaGe, which identify synteny breaks and search for HGT features (tRNA hotspots, genes involved in mobility), and for compositional bias (AlienHunter (Vernikos and Parkhill, 2006), SIGI-HMM (Waack *et al*., 2006), and GC deviation computation).
(DOCX)

S3 Table. Details of the RGP found in CB2 genome.
(XLSX)

S4 Table. RNAseq results for genes that are significantly more expressed in the presence than in the absence of arsenite. The *ars* and the *aio* genes (two copies of each operon), as well as other genes from these two islands (RGP10 and RGP19) are induced by As(III).
(XLSX)

S5 Table. Minimal inhibitory concentration (MIC) of As(III) for *Thiomonas* strains on solid m126. *Thiomonas* spp. CB2 and CB3 are slightly more resistant than the others strains. *Thiomonas* spp. 3As, CB1 and CB6 demonstrate lower levels of resistance to As(III).
(DOCX)

Acknowledgments

The authors thank Odile Bruneel, Violaine Bonnefoy and Sabine Heinhorst for providing the CB1, CB2, CB3, and CB6 strains, the 3As strain and the K12 strain, respectively, Yan Ploetze for the figures and Jürgen Mattusch, and Karine Jezequel for the ICP-MS and ICP-AES experiments. We thank the PRABI (Pôle Rhône-Alpes de Bioinformatique) for providing computing facilities.

Author Contributions

Conceived and designed the experiments: KF MCK JF VB FP J-YC PNB HJH, FA-P. Performed the experiments: KF MCK JF JA SK YL GM CP. Analyzed the data: KF MCK JF CB-A VB P-EC M-AD BJ FP J-YC PNB HJH FA-P. Contributed reagents/materials/analysis tools: CB-A VB YL J-YC. Wrote the paper: KF MCK JF CB-A PNB HJH FA-P.

References

1. Coleman ML, Sullivan MB, Martiny AC, Steglich C, Barry K, Delong EF, et al. Genomic islands and the ecology and evolution of *Prochlorococcus*. Science. 2006; 311: 1768–1770. doi: 10.1126/science.1122050 PMID: 16556843

2. Juhas M, van der Meer JR, Gaillard M, Harding RM, Hood DW, Crook DW. Genomic islands: tools of bacterial horizontal gene transfer and evolution. FEMS Microbiol Rev. 2009; 33: 376–393. doi: 10.1111/j.1574-6976.2008.00136.x PMID: 19178566
3. Penn K, Jenkins C, Nett M, Udwary DW, Gontang EA, McGlinchey RP, et al. Genomic islands link secondary metabolism to functional adaptation in marine *Actinobacteria*. ISME J. 2009; 3: 1193–1203. doi: 10.1038/ismej.2009.58 PMID: 19474814
4. Cohan FM. What are bacterial species? Annu Rev Microbiol. 2002; 56: 457–487. doi: 10.1146/annurev.micro.56.012302.160634 PMID: 12142474
5. Cohan FM, Perry EB. A systematics for discovering the fundamental units of bacterial diversity. Curr Biol CB. 2007; 17: R373–386. doi: 10.1016/j.cub.2007.03.032 PMID: 17502094
6. Koeppel A, Perry EB, Sikorski J, Krizanc D, Warner A, Ward DM, et al. Identifying the fundamental units of bacterial diversity: a paradigm shift to incorporate ecology into bacterial systematics. Proc Natl Acad Sci U S A. 2008; 105: 2504–2509. doi: 10.1073/pnas.0712205105 PMID: 18272490
7. Cheng H, Hu Y, Luo J, Xu B, Zhao J. Geochemical processes controlling fate and transport of arsenic in acid mine drainage (AMD) and natural systems. J Hazard Mater. 2009; 165: 13–26. doi: 10.1016/j.jhazmat.2008.10.070 PMID: 19070955
8. Casiot C, Morin G, Juillot F, Bruneel O, Personné JC, Leblanc M, et al. Bacterial immobilization and oxidation of arsenic in acid mine drainage (Carnoulès creek, France). Water Res. 2003; 37: 2929–2936. doi: 10.1016/S0043-1354(03)00080-0 PMID: 12767295
9. Bruneel, Duran R, Casiot C, Elbaz-Poulichet F, Personné J-C. Diversity of microorganisms in Fe-As-rich acid mine drainage waters of Carnoulès, France. Appl Environ Microbiol. 2006; 72: 551–556. doi: 10.1128/AEM.72.1.551–556.2006 PMID: 16391091
10. Volant A, Desoeuvre A, Casiot C, Lauga B, Delpoux S, Morin G, et al. Archaeal diversity: temporal variation in the arsenic-rich creek sediments of Carnoulès Mine, France. Extrem Life Extreme Cond. 2012; 16: 645–657. doi: 10.1007/s00792-012-0466-8
11. Volant A, Bruneel O, Desoeuvre A, Héry M, Casiot C, Bru N, et al. Diversity and spatiotemporal dynamics of bacterial communities: physicochemical and other drivers along an acid mine drainage. FEMS Microbiol Ecol. 2014; 90: 247–263. doi: 10.1111/1574-6941.12394 PMID: 25070063
12. Bruneel, Personné J-C, Casiot C, Leblanc M, Elbaz-Poulichet F, Mahler BJ, et al. Mediation of arsenic oxidation by *Thiomonas* sp. in acid-mine drainage (Carnoulès, France). J Appl Microbiol. 2003; 95: 492–499. PMID: 12911697
13. Battaglia-Brunet F, Joulian C, Garrido F, Dictor M-C, Morin D, Coupland K, et al. Oxidation of arsenite by *Thiomonas* strains and characterization of *Thiomonas arsenivorans* sp. nov. Antonie Van Leeuwenhoek. 2006; 89: 99–108. doi: 10.1007/s10482-005-9013-2 PMID: 16341463
14. Duquesne K, Lieutaud A, Ratouchniak J, Muller D, Lett M- C, Bonnefoy V. Arsenite oxidation by a chemoautotrophic moderately acidophilic *Thiomonas* sp.: from the strain isolation to the gene study. Environ Microbiol. 2008; 10: 228–237. doi: 10.1111/j.1462-2920.2007.01447.x PMID: 17894815
15. Bryan, Marchal M, Battaglia-Brunet F, Kugler V, Lemaitre-Guillier C, Lièvremont D, et al. Carbon and arsenic metabolism in *Thiomonas* strains: differences revealed diverse adaptation processes. BMC Microbiol. 2009; 9: 127. doi: 10.1186/1471-2180-9-127 PMID: 19549320
16. Arsène-Ploetze F, Koechler S, Marchal M, Coppée J-Y, Chandler M, Bonnefoy V, et al. Structure, function, and evolution of the *Thiomonas* spp. genome. PLoS Genet. 2010; 6: e1000859. doi: 10.1371/journal.pgen.1000859 PMID: 20195515
17. Milde K, Sand W, Wolff W, Bock E. *Thiobacilli* of the corroded concrete walls of the Hamburg sewer system. J Gen Microbiol. 1983; 129: 1327–1333.
18. Slyemi D, Moinier D, Brochier-Armanet C, Bonnefoy V, Johnson DB. Characteristics of a phylogenetically ambiguous, arsenic-oxidizing *Thiomonas* sp., *Thiomonas arsenitoxydans* strain 3As(T) sp. nov. Arch Microbiol. 2011; 193: 439–449. doi: 10.1007/s00203-011-0684-y PMID: 21409355
19. Moreira D, Amils R. Phylogeny of *Thiobacillus cuprinus* and other mixotrophic thiobacilli: proposal for *Thiomonas* gen. nov. Int J Syst Bacteriol. 1997; 47: 522–528. PMID: 9103643
20. Bertin, Heinrich-Salmeron A, Pelletier E, Goulhen-Chollet F, Arsène-Ploetze F, Gallien S, et al. Metabolic diversity among main microorganisms inside an arsenic-rich ecosystem revealed by meta- and proteo-genomics. ISME J. 2011; 5: 1735–47. doi: 10.1038/ismej.2011.51 PMID: 21562598
21. Rinkel M, Hubert JC, Roux B, Lett MC. Identification of a new transposon Tn5403 in a *Klebsiella pneumoniae* strain isolated from a polluted aquatic environment. Curr Microbiol. 1994; 29: 249–254. PMID: 7765419
22. Laemmli UK. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature. 1970; 227: 680–685. PMID: 5432063

23. Vallenet D, Labarre L, Rouy Z, Barbe V, Bocs S, Cruveiller S, et al. MaGe: a microbial genome annotation system supported by synteny results. Nucleic Acids Res. 2006; 34: 53–65. doi: 10.1093/nar/gkj406 PMID: 16407324
24. Vallenet D, Engelen S, Mornico D, Cruveiller S, Fleury L, Lajus A, et al. MicroScope: a platform for microbial genome annotation and comparative genomics. Database J Biol Databases Curation. 2009; 2009: bap021. doi: 10.1093/database/bap021
25. Vallenet D, Belda E, Calteau A, Cruveiller S, Engelen S, Lajus A, et al. MicroScope—an integrated microbial resource for the curation and comparative analysis of genomic and metabolic data. Nucleic Acids Res. 2013; 41: D636–647. doi: 10.1093/nar/gks1194 PMID: 23193269
26. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. Scaffolding pre-assembled contigs using SSPACE. Bioinformatics. 2011; 27: 578–579. doi: 10.1093/bioinformatics/btq683 PMID: 21149342
27. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, et al. Versatile and open software for comparing large genomes. Genome Biol. 2004; 5: R12. doi: 10.1186/gb-2004-5-2-r12 PMID: 14759262
28. Wheatcroft R. Changes in the *Rhizobium meliloti* Genome and the Ability to Detect Supercoiled Plasmids During Bacteroid Development. Mol Plant Microbe Interact. 1990; 3: 9. doi: 10.1094/MPMI-3-009
29. Albiger B, Hubert JC, Lett MC. Identification of the plasmid-mobilization potential of the strain *Klebsiella pneumoniae ozenae* KIIIA isolated from a polluted aquatic environment. Plasmid. 1999; 41: 30–39. doi: 10.1006/plas.1998.1372 PMID: 9887304
30. Marchal M, Briandet R, Halter D, Koechler S, Dubow MS, Lett M-C, et al. Subinhibitory arsenite concentrations lead to population dispersal in *Thiomonas* sp. PloS One. 2011; 6: e23181. doi: 10.1371/journal.pone.0023181 PMID: 21876737
31. R Development Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: the R Foundation for Statistical Computing.; 2011. Available: http://www.R-project.org/.
32. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, et al. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. 2004; 5: R80. doi: 10.1186/gb-2004-5-10-r80 PMID: 15461798
33. Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol. 2010; 11: R106. doi: 10.1186/gb-2010-11-10-r106 PMID: 20979621
34. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J R Stat Soc Ser B Methodol. 1995; 57: 289–300.
35. Kruger MC, Bertin PN, Heipieper HJ, Arsène-Ploetze F. Bacterial metabolism of environmental arsenic—mechanisms and biotechnological applications. Appl Microbiol Biotechnol. 2013; 97: 3827–3841. doi: 10.1007/s00253-013-4838-5 PMID: 23546422
36. Bellanger X, Payot S, Leblond-Bourget N, Guédon G. Conjugative and mobilizable genomic islands in bacteria: evolution and diversity. FEMS Microbiol Rev. 2014; 38: 720–760. doi: 10.1111/1574-6976.12058 PMID: 24372381
37. Nakagawa S, Shimamura S, Takaki Y, Suzuki Y, Murakami S-I, Watanabe T, et al. Allying with armored snails: the complete genome of gammaproteobacterial endosymbiont. ISME J. 2014; 8: 40–51. doi: 10.1038/ismej.2013.131 PMID: 23924784
38. Denef VJ, Mueller RS, Banfield JF. AMD biofilms: using model communities to study microbial evolution and ecological complexity in nature. ISME J. 2010; 4: 599–610. doi: 10.1038/ismej.2009.158 PMID: 20164865
39. Aminov RI. Horizontal gene exchange in environmental microbiota. Front Microbiol. 2011; 2: 158. doi: 10.3389/fmicb.2011.00158 PMID: 21845185
40. Harrison JJ, Ceri H, Turner RJ. Multimetal resistance and tolerance in microbial biofilms. Nat Rev Microbiol. 2007; 5: 928–938. doi: 10.1038/nrmicro1774 PMID: 17940533
41. Madsen JS, Burmølle M, Hansen LH, Sørensen SJ. The interconnection between biofilm formation and horizontal gene transfer. FEMS Immunol Med Microbiol. 2012; 65: 183–195. doi: 10.1111/j.1574-695X.2012.00960.x PMID: 22444301
42. Ochman H, Lawrence JG, Groisman EA. Lateral gene transfer and the nature of bacterial innovation. Nature. 2000; 405: 299–304. doi: 10.1038/35012500 PMID: 10830951
43. Li H, Li M, Huang Y, Rensing C, Wang G. In silico analysis of bacterial arsenic islands reveals remarkable synteny and functional relatedness between arsenate and phosphate. Front Microbiol. 2013; 4: 347. doi: 10.3389/fmicb.2013.00347 PMID: 24312089
44. Cohan FM. Bacterial species and speciation. Syst Biol. 2001; 50: 513–524. PMID: 12116650
45. Koeppel AF, Wu M. Species matter: the role of competition in the assembly of congeneric bacteria. ISME J. 2014; 8: 531–540. doi: 10.1038/ismej.2013.180 PMID: 24132075

46. Gouy M, Guindon S, Gascuel O. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol Biol Evol. 2010; 27: 221–224. doi: 10.1093/molbev/msp259 PMID: 19854763

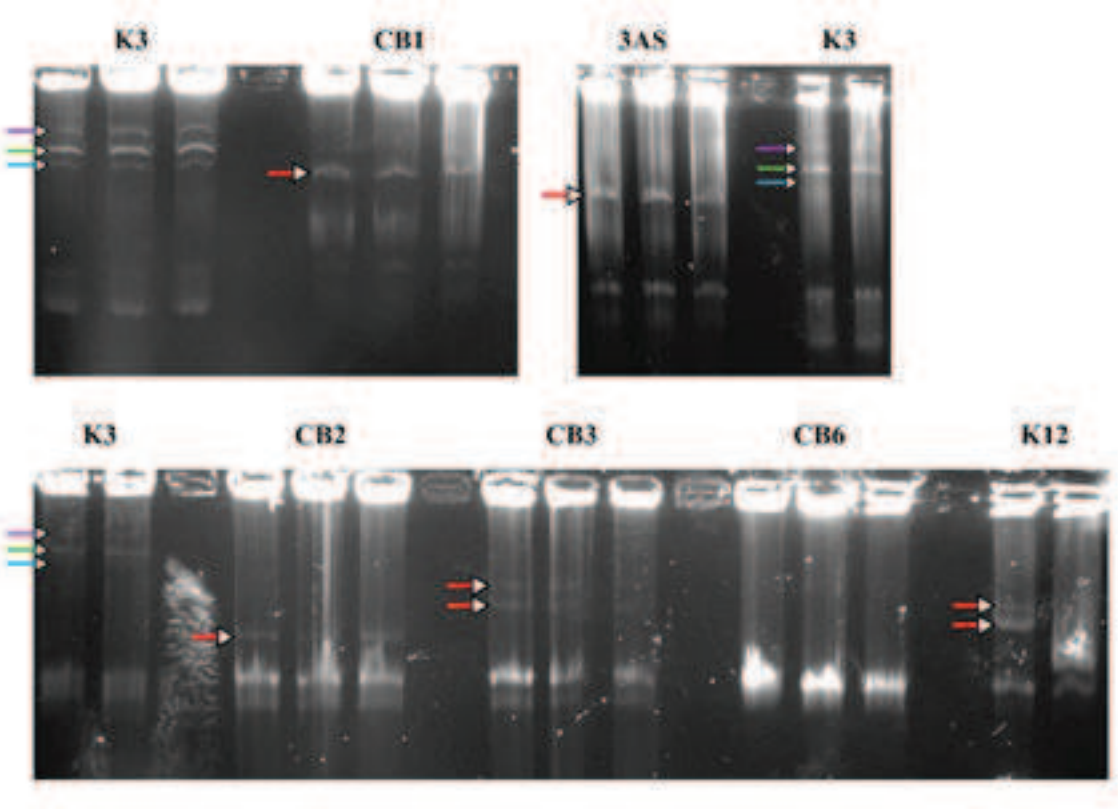
47. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol. 2010; 59: 307–321. doi: 10.1093/sysbio/syq010 PMID: 20525638

48. Jobb G, von Haeseler A, Strimmer K. TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. BMC Evol Biol. 2004; 4: 18. doi: 10.1186/1471-2148-4-18 PMID: 15222900

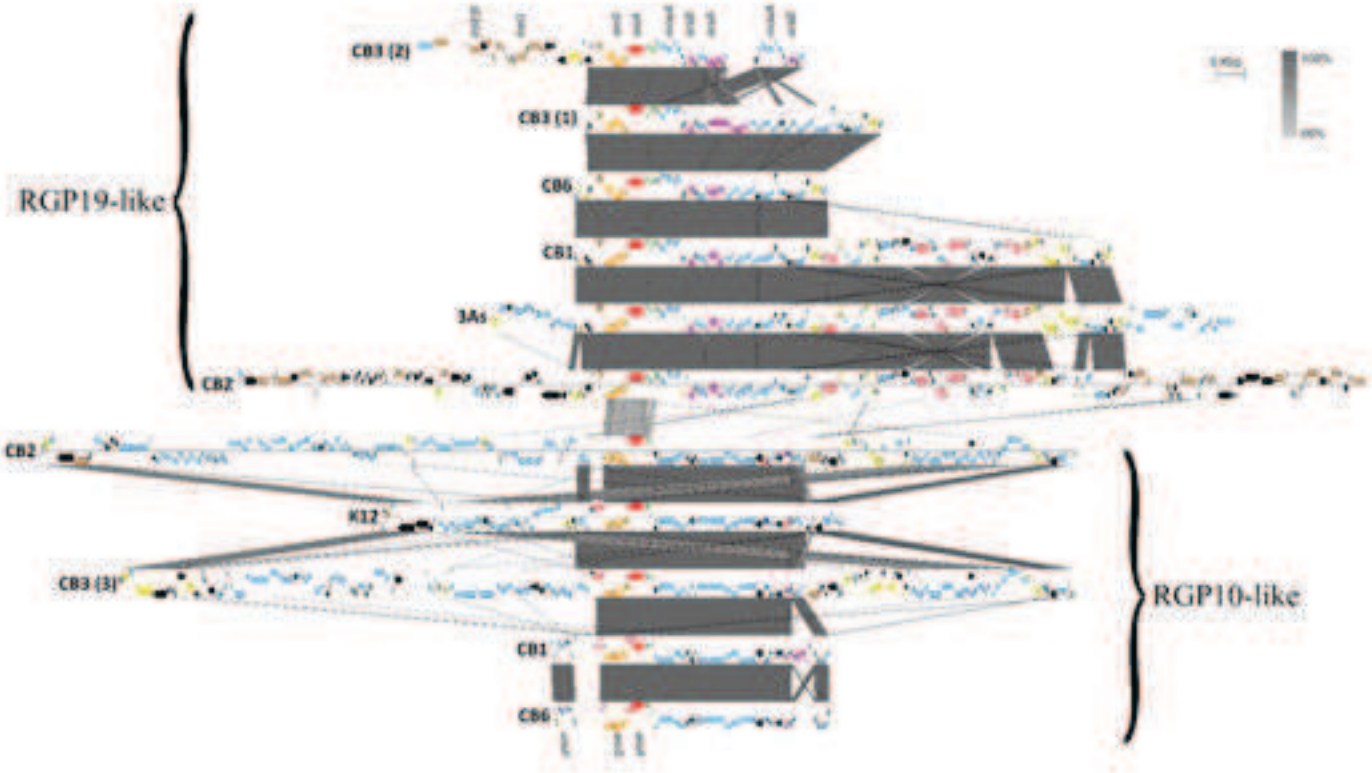
49. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol. 2012; 61: 539–542. doi: 10.1093/sysbio/sys029 PMID: 22357727

50. Sullivan MJ, Petty NK, Beatson SA. Easyfig: a genome comparison visualizer. Bioinforma Oxf Engl. 2011; 27: 1009–1010. doi: 10.1093/bioinformatics/btr039

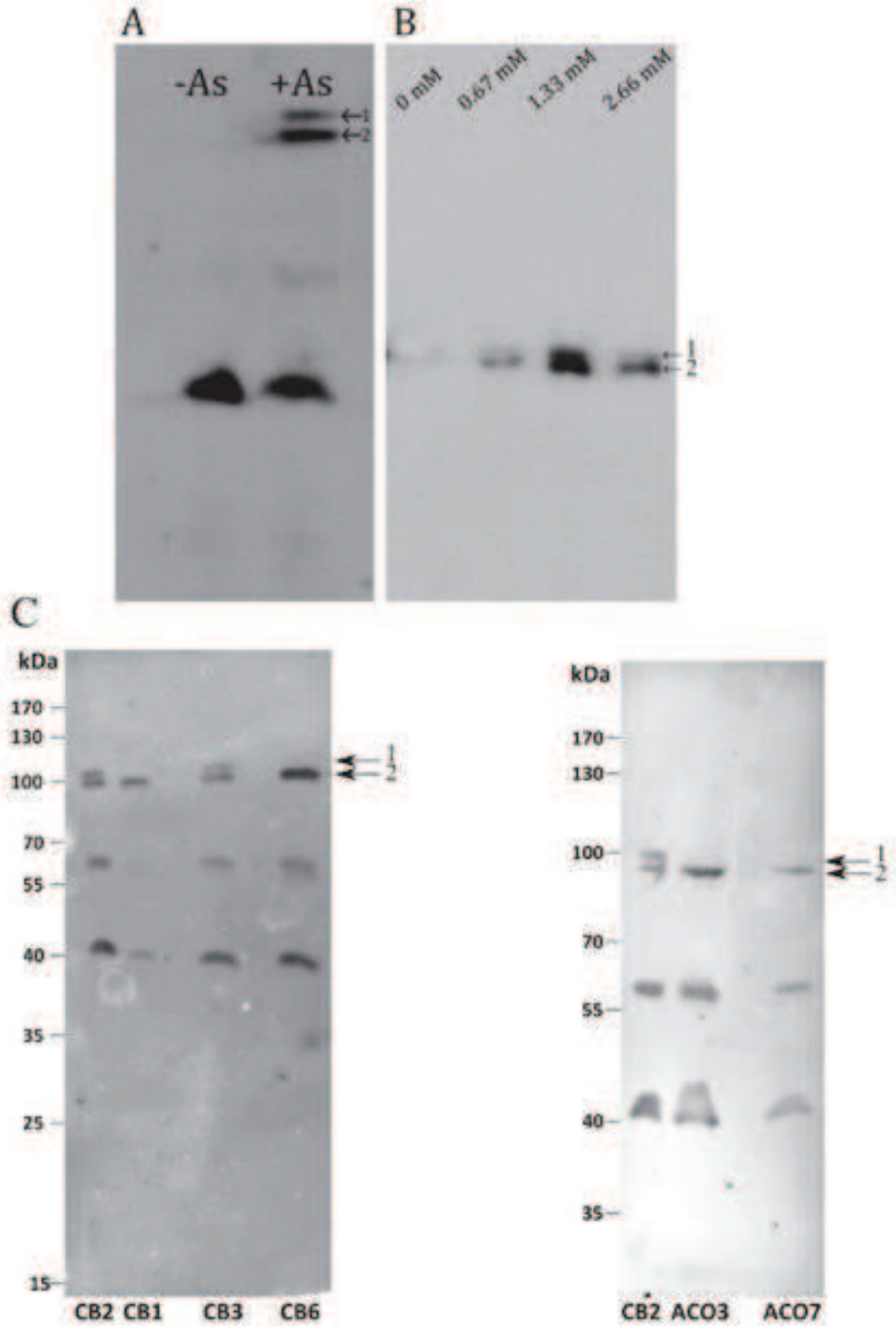
Note : Le "Supplementary table 3" est présenté en Annexe 1.



Supplementary figure 1

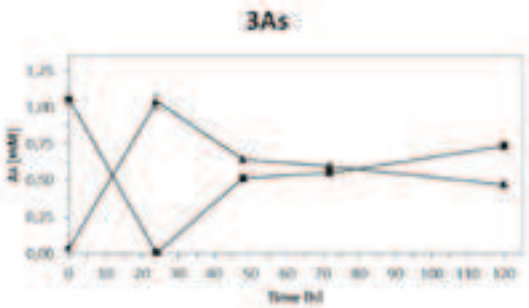


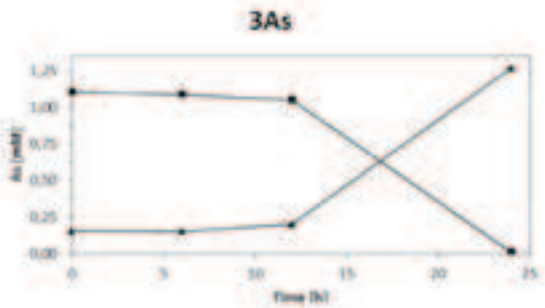
Supplementary figure 2

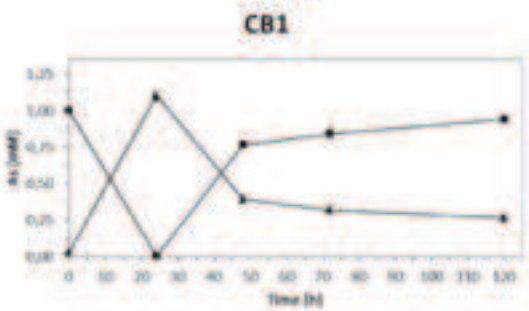


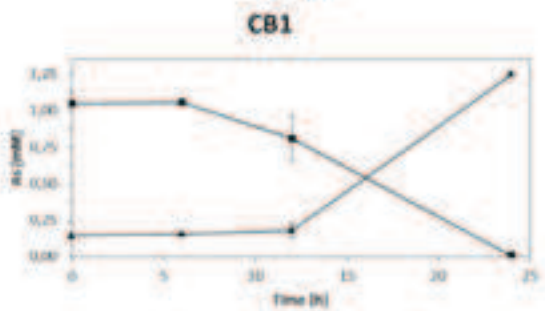
Supplementary figure 3

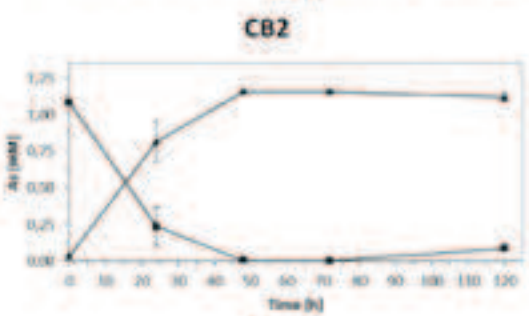
Supplementary figure 4 (page suivante)

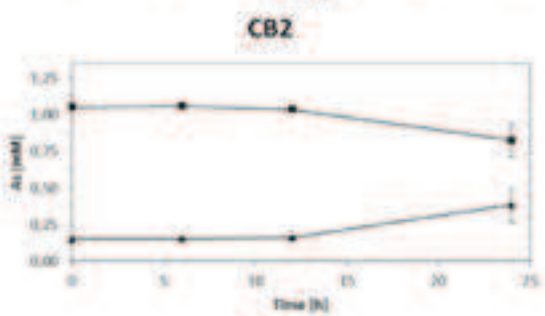

3As

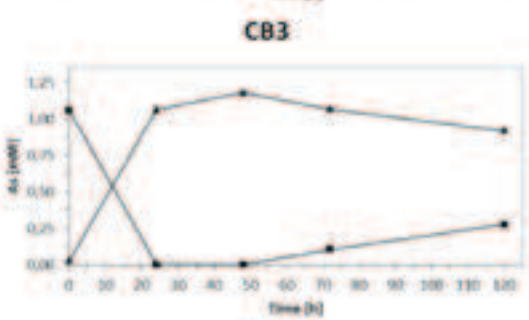

3As

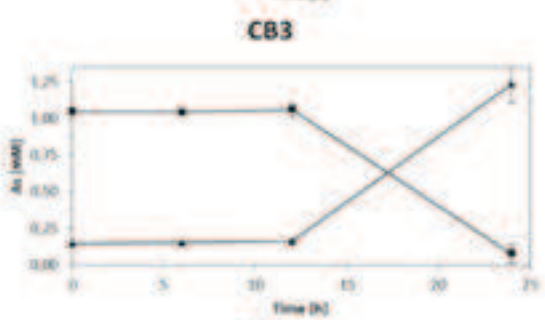

CB1

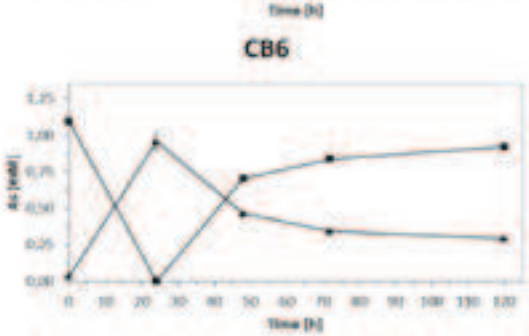

CB1

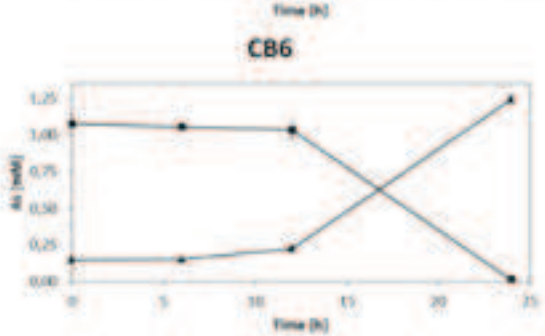

CB2

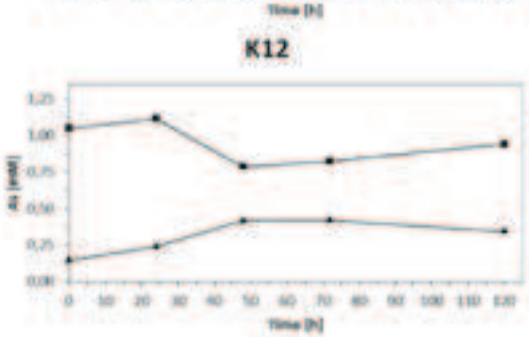

CB2

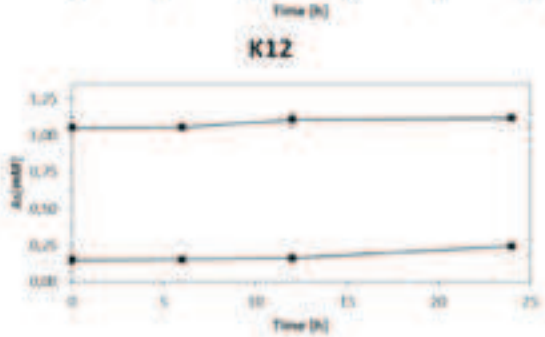

CB3


CB3


CB6


CB6


K12


K12

Genome characteristics	CB2	3As	CB1	CB3	CB6	ACO3	ACO7	K12
Size (bp)	3872510	3738778	3880244	4323494	3861602	3673425	3730618	3396378
GC-content (%)	63.67	63.8	63.8	63.93	63.83	63.8	63.78	63.88
Predicted CDS	3820	3631	3927	4508	3880	3628	3680	3467
Genome coding (%)	91.23	90.19	91.75	91.64	92.04	92.95	9181	90.11
Average CDS size (bp)	917.51	956.15	918.52	887.02	920.4	929.49	929.18	927.59
Number of tRNA	45	43	44	43	44	42	42	45
ORF of unknown function (%)	1192 (31.20)	939 (25.86)	1238 (31,53)	1514 (33.58)	1212 (31.24)	1097 (30.24)	1130 (3071)	218 (6.29%)
Repetitive regions (%)	6.31	7.26	7.3	19.21	7.52	3.11	5.04	4.86

Supplementary table 1 ***General characteristics of the Thiomonas genomes.*** *Calculation of repetitive regions did not include undetermined ('N') bases. The genome size and gene content range from 3.4 Mb and 4, 367 CDS (K12) to 4.3 Mb and 4, 508 CDS (CB3). Thiomonas sp. CB3 also has a higher proportion of repetitive regions (19.21%). The others general features are similar between the strains.*

RGP	Begin	End	Length	Feature Score	Feature	THIOM_3As	THIIK_K12	Specificity Score
RGP1	409	12446	12038	4	tRNA-int-IVOM-Specific_Region	71	71	142
RGP2	607500	625666	18167	1	Specific_Region	83	96	179
RGP3	750094	824454	74361	4	int-SIGI-IVOM-Specific_Region	88	70	158
RGP4	837543	916150	78608	3	tRNA-IVOM-Specific_Region	73	29	102
RGP5	1014685	1044464	29780	3	SIGI-IVOM-Specific_Region	93	93	186
RGP6	1111788	1130387	18600	2	IVOM-Specific_Region	82	71	153
RGP7	1188698	1206631	17934	0	none	13	100	113
RGP8	1393011	1428722	35712	4	int-SIGI-IVOM-Specific_Region	100	100	200
RGP9	1538398	1622571	84174	2	misc_RNA-IVOM	3	94	97
RGP10	1626740	1717910	91171	4	int-SIGI-IVOM-Specific_Region	72	27	99
RGP11	2196265	2202854	6590	1	Specific_Region	100	25	125
RGP12	2233184	2269464	36281	5	tRNA-int-SIGI-IVOM-Specific_Region	82	94	176
RGP13	2628929	2697519	68591	4	tRNA-int-IVOM-Specific_Region	52	77	129
RGP14	2794075	2816184	22110	3	tRNA-IVOM-Specific_Region	68	100	168
RGP15	2865731	2874024	8294	3	tRNA-int-IVOM	100	50	150
RGP16	2959834	2965781	5948	0	none	100	20	120
RGP17	3304256	3314146	9891	5	tRNA-int-misc_RNA-IVOM-Specific_Region	83	100	183
RGP18	3327601	3382919	55319	3	int-IVOM-Specific_Region	75	70	145
RGP19*	3533461	3872209	338749	6	tRNA-int-mob-SIGI-IVOM-Specific_Region	50	69	119

* this island possibly corresponds to at least two distinct GEIs

Supplementary table 2 ***Regions of genomic plasticity (RGP) identified after comparison of CB2 with 3As and K12.*** *The RGP were identified with the RGP finder module in MaGe, which identify synteny breaks and search for HGT features (tRNA hotspots, genes involed in mobility), and for compositional bias (AlienHunter (Vernikos and Parkhill, 2006), SIGI-HMM (Waack et al., 2006), and GC deviation computation).*

gene number	RGP number		Annotation	normalized counts, without AsIII [a]	normalized counts, with AsIII [a]	Fold change (FC) [b]	log2(FC)	raw p-value [c]	adjusted p-value [d]
THICB2_v3_390037		operonic ?	Permease	21	248	11,83	3,56	7,06e-26	1,89e-23
THICB2_v3_390036			conserved exported protein of unknown function	23	200	8,78	3,13	1,13e-19	2,32e-17
THICB2_v3_390035			Thiol-disulfide isomerase and thioredoxins family protein	30	220	7,42	2,89	1,14e-18	2,23e-16
THICB2_v3_390033		operonic ?	Arsenical resistance protein ArsH	39	166	4,28	2,10	1,57e-10	2,08e-08
THICB2_v3_390032			putative Carboxymuconolactone decarboxylase	671	1496	2,23	1,16	1,36e-06	1,65e-04
THICB2_v3_390031			conserved protein of unknown function	1827	3293	1,80	0,85	2,09e-04	2,20e-02
THICB2_v3_390024		operonic ?	OsmC family protein	6	74	13,05	3,71	5,88e-14	8,94e-12
THICB2_v3_390023			Outer membrane efflux protein	15	109	7,41	2,89	3,49e-13	5,11e-11
THICB2_v3_390022	RGP10		Efflux transporter, RND family, MFP subunit	9	115	12,16	3,60	1,10e-17	1,96e-15
THICB2_v3_390021			Acriflavin resistance protein/ efflux pomp	82	315	3,85	1,94	1,69e-11	2,39e-09
THICB2_v3_390020			FAD-dependent pyridine nucleotide-disulfide oxidoreductase	112	344	3,08	1,62	4,65e-09	5,97e-07
THICB2_v3_390019			Lipid A 3-O-deacylase-related protein	24	71	3,00	1,59	8,18e-05	9,08e-03
THICB2_v3_390010		*aio* operon	Transcriptional regulator, ArsR family	92	407	4,44	2,15	1,90e-14	3,00e-12
THICB2_v3_390009			Nitroreductase-like	95	1079	11,34	3,50	2,10e-37	1,08e-34
THICB2_v3_390008			Cytochrome c-552	109	1232	11,30	3,50	8,85e-38	5,19e-35
THICB2_v3_390007			Arsenite oxidase subunit AioA	336	5624	16,73	4,06	1,32e-53	2,72e-50
THICB2_v3_390006			Arsenite oxidase subunit AioB	32	969	30,68	4,94	2,49e-59	1,02e-55
THICB2_v3_840004	RGP20		protein of unknown function	54	126	2,32	1,22	3,70e-04	3,71e-02
THICB2_v3_220018	RGP3		conserved protein of unknown function	69	149	2,18	1,12	4,31e-04	4,21e-02
THICB2_v3_760066		*ars* operon	Arsenic operon regulator ArsR	108	734	6,79	2,76	9,20e-24	1,99e-21
THICB2_v3_760067			Arsenate reductase ArsC2a	142	1058	7,47	2,90	2,06e-27	6,50e-25
THICB2_v3_760068	RGP19		Arsenical resistance operon trans-acting repressor ArsD	8	404	47,70	5,58	1,91e-51	2,61e-48
THICB2_v3_760069			Arsenical pump-driving ATPase ArsA	317	1982	6,25	2,64	1,13e-25	2,73e-23
THICB2_v3_760070			Arsenate reductase ArsC2b	23	330	14,29	3,84	8,67e-32	3,95e-29

THICB2_v3_760071			putative FAD-dependent pyridine nucleotide-disulphide oxidoreductase	149	1209	8,11	3,02	4,76e-30	1,95e-27
THICB2_v3_760072			conserved protein of unknown function	45	463	10,37	3,37	6,34e-30	2,37e-27
THICB2_v3_760073			arsenite/antimonite transporter ArsB	88	455	5,17	2,37	1,71e-17	2,93e-15
THICB2_v3_760074			protein of unknown function	6	34	5,67	2,50	2,57e-05	3,01e-03
THICB2_v3_760075			Transposase of ISThsp18, IS1182 family (fragment)	60	492	8,18	3,03	7,37e-26	1,89e-23
THICB2_v3_760082		*aio* operon	putative nitroreductase	51	120	2,35	1,23	2,39e-04	2,45e-02
THICB2_v3_760083			putative Cytochrome c, class ID (fragment)	16	86	5,22	2,38	6,94e-09	8,63e-07
THICB2_v3_760084			putative Bacterial regulatory protein, ArsR family	9	97	11,01	3,46	3,11e-15	5,10e-13
THICB2_v3_760085			Cytochrome c-551	39	425	10,79	3,43	2,49e-29	8,50e-27
THICB2_v3_760086			Arsenite oxidase subunit AioA	939	13571	14,45	3,85	2,65e-50	2,72e-47
THICB2_v3_760087			Arsenite oxidase subunit AioB	188	2693	14,29	3,84	3,74e-47	3,07e-44
THICB2_v3_640015	no	*ars* operon	putative transcriptional regulator, ArsR family	46	399	8,62	3,11	1,02e-24	2,31e-22
THICB2_v3_640014	no		Glyoxalase/bleomycin resistance protein/dioxygenase	106	554	5,22	2,38	7,01e-18	1,31e-15
THICB2_v3_640013	no		arsenate reductase ArsC	171	1230	7,19	2,84	5,71e-27	1,67e-24
THICB2_v3_640012	no		arsenite/antimonite transporter ArsB	126	424	3,38	1,76	1,04e-10	1,42e-08

[a] normalized counts per biological condition; [b] FC fold change of expression, calculated on the basis of mean normalized counts as with As(III)/withoutAs(III); [c] raw p-value from the statistical test; [d] adjusted p-value on which the cut-off α is applied

Supplementary table 4

Thiomonas strains	K12	3As	CB1	CB2	CB3	CB6	ACO3	ACO7
As(III) MIC (mM)	9.3	8	8	12	12	8	9.3	9.3

Supplementary table 5

III - Conclusion

Bien que l'ensemble des souches étudiées soient très proches phylogénétiquement, la résolution de l'arbre construit montre que les souches *Thiomonas* spp. CB2, CB3 et *Thiomonas intermedia* K12 divergent des autres, bien que *Tm.* spp. CB2 et CB3 aient été isolées du même DMA que *Tm. arsenitoxydans* 3As. De manière plutôt surprenante, la conservation de la synténie ne dépend pas de la divergence phylogénétique. En effet, elle n'est pas conservée entre *Tm.* sp. CB2 et *Tm. arsenitoxydans* 3As alors qu'elle l'est entre *Tm. arsenitoxydans* 3As et *Tm. intermedia* K12, plus éloignée phylogénétiquement de *Tm. arsenitoxydans* 3As que ne l'est *Tm.* sp. CB2. Le génome de la souche *Tm.* sp. CB2 a donc probablement subi plusieurs remaniements importants de son génome. De plus, la présence de nombreux îlots génomiques (GEIs) montre une importante partie flexible des génomes. Cette diversité est probablement due à des HGTs, et concerne notamment l'oxydation de l'As(III).

J'ai montré que l'opéron *aioBA* a été retrouvé chez toutes les souches au niveau de GEIs. Les souches *Tm.* spp. CB2 et CB3, contrairement aux autres, possèdent plusieurs copies exprimées de cet opéron. Dans le cas de *Tm.* sp. CB2, deux copies divergentes sont présentes sur deux GEIs dont un est similaire à *Tm. arsenitoxydans* 3As (RGP19) et l'autre à *Tm. intermedia* K12 (RGP10). Le RGP19 possède les caractéristiques d'un ICE et le RGP10 code une hydrogénase similaire à celle retrouvée chez une γ-protéobactérie. Ceci suggère que ces deux copies de l'opéron *aio* ont probablement été acquises par HGT. De plus, les RGP19 et RGP10 de CB2 ont évolué différemment par réarrangements génomiques.

La présence de plusieurs copies de l'opéron *aio* est corrélée avec une CMI plus élevée pour les souches *Tm.* spp. CB2 et CB3, ainsi qu'une cinétique de l'oxydation de l'As(III) différente. Ces deux GEIs semblent donc bien avoir un impact sur l'écologie des *Thiomonas* spp. Les souches *Tm.* spp. CB2 et CB3 pourraient être adaptées à des micro-niches légèrement différentes au sein du DMA. De plus, dans le cas où ces GEIs sont mobiles, ils constituent un potentiel adaptatif face aux changements des conditions environnementales (augmentation de la concentration en As). Une étude de protéomique ciblée, pour laquelle j'ai participé à la préparation des échantillons, a montré que l'arsénite oxydase était bien exprimée sur le site, tout le long du Reigous et à différents temps (**Annexe 2**).

Partie 2 - Dégradation de l'urée

I - Introduction

La résistance peut impliquer des interactions directes avec les composés toxiques, comme c'est le cas avec l'oxydation à l'As(III). Cependant, les *Thiomonas* spp. dans leur milieu naturel font partie d'une communauté où elles sont intégrées dans un réseau d'interactions complexes entre les différents membres. Certaines de ces interactions pourraient permettre une meilleure résistance aux composés toxiques et constituer une adaptation. Nous avons voulu étudier d'autres fonctions en lien avec les conditions du DMA, qui permettraient une adaptation à différentes micro-niches au sein de ce milieu et seraient codées sur des GEIs.

Une étude metagénomique réalisée sur le DMA de Carnoulès (site COWG) a souligné l'existence de potentielles interactions entre sept membres dominants (CARN1 à CARN7) de la communauté (Bertin et al., 2011). Il a été démontré que de l'urée est produite et excrétée par un eucaryote photosynthétique sur ce site, l'euglène *Euglena mutabilis* (Halter et al., 2012) mais aussi par les bactéries CARN1 et CARN4. Sa dégradation pourrait être catalysée par des *Thiomonas* comme CARN2 ou *Tm.* sp. CB2. La dégradation de l'urée peut être impliquée dans la résistance à l'acidité, comme c'est le cas chez *Helicobacter pylori* (Stingl et al., 2002), et pourrait donc conférer un avantage dans le DMA de Carnoulès.

Les travaux présentés dans la publication qui suit avaient pour but de comparer les capacités à dégrader l'urée chez les souches *Thiomonas* spp. 3As, CB1, CB2, CB3 et CB6, puis de comprendre quel avantage pouvait leur conférer cette capacité dans le DMA de Carnoulès. Comme dans la partie précédente, *Tm. intermedia* K12 a également été incluse dans cette étude. Dans ce travail j'ai comparé les régions du génome impliquées dans la dégradation de l'urée. J'ai également comparé l'activité uréase des différentes souches et mesuré la viabilité cellulaire des souches cultivées en milieu m126 avec et sans addition d'urée par microcopie à fluorescence (marquage "live-dead"). Enfin, j'ai regardé si l'urée avait un effet sur l'activité arsénite oxydase de *Tm.* sp. CB2. Le reste de l'étude réalisée dans l'eau de l'AMD a été effectué par les co-auteurs de la publication.

II-Publication

frontiers in Microbiology

ORIGINAL RESEARCH
published: 28 September 2015
doi: 10.3389/fmicb.2015.00993

OPEN ACCESS

Edited by:
Philippe M. Oger, Centre National de la Recherche Scientifique and Ecole Normale Supérieure de Lyon, France

Reviewed by:
Ronald Oremland, United States Geological Survey, USA
William P. Inskeep, Montana State University, USA

***Correspondence:**
Florence Arsène-Ploetze, UMR7156, Université de Strasbourg - Centre National de la Recherche Scientifique, Institut de Botanique, 28 Rue Goethe, 67083 Strasbourg, France
ploetze@unistra.fr

†Present Address:
David Halter, Santé de la Vigne et Qualité du Vin, Institut National de la Recherche Agronomique UMR 1131, Colmar, France

Specialty section:
This article was submitted to Extreme Microbiology, a section of the journal Frontiers in Microbiology

Received: *15 May 2015*
Accepted: *07 September 2015*
Published: *28 September 2015*

Citation:
Farasin J, Andres J, Casiot C, Barbe V, Faerber J, Halter D, Heintz D, Koechler S, Lièvremont D, Lugan R, Marchal M, Plewniak F, Seby F, Bertin PN and Arsène-Ploetze F (2015) Thiomonas sp. CB2 is able to degrade urea and promote toxic metal precipitation in acid mine drainage waters supplemented with urea. Front. Microbiol. 6:993. doi: 10.3389/fmicb.2015.00993

Thiomonas sp. CB2 is able to degrade urea and promote toxic metal precipitation in acid mine drainage waters supplemented with urea

Julien Farasin[1], Jérémy Andres[1], Corinne Casiot[2], Valérie Barbe[3], Jacques Faerber[4], David Halter[1†], Dimitri Heintz[5], Sandrine Koechler[1], Didier Lièvremont[1], Raphael Lugan[5], Marie Marchal[1], Frédéric Plewniak[1], Fabienne Seby[6], Philippe N. Bertin[1] and Florence Arsène-Ploetze[1*]

[1] Laboratoire Génétique Moléculaire, Génomique et Microbiologie, UMR7156, Université de Strasbourg - Centre National de la Recherche Scientifique, Institut de Botanique, Strasbourg, France, [2] Laboratoire Hydrosciences Montpellier, UMR 5569, Centre National de la Recherche Scientifique-UM I, UM II, IRD, Université Montpellier 2, CCMSE, Montpellier, France, [3] Laboratoire de Biologie Moléculaire Pour l'Etude des Génomes, CEA-IG-Genoscope, Evry, France, [4] Institut de Physique et de Chimie des Matériaux de Strasbourg, Université de Strasbourg, CNRS UMR 7504, Strasbourg, France, [5] Plateforme Métabolomique, UPR2357, Centre National de la Recherche Scientifique, Institut de Biologie Moléculaire des Plantes, Institut de Botanique, Strasbourg, France, [6] Hélioparc Pau Pyrénées, Pau, France

The acid mine drainage (AMD) in Carnoulès (France) is characterized by the presence of toxic metals such as arsenic. Several bacterial strains belonging to the *Thiomonas* genus, which were isolated from this AMD, are able to withstand these conditions. Their genomes carry several genomic islands (GEIs), which are known to be potentially advantageous in some particular ecological niches. This study focused on the role of the "urea island" present in the *Thiomonas* CB2 strain, which carry the genes involved in urea degradation processes. First, genomic comparisons showed that the genome of *Thiomonas* sp. CB2, which is able to degrade urea, contains a urea genomic island which is incomplete in the genome of other strains showing no urease activity. The urease activity of *Thiomonas* sp. CB2 enabled this bacterium to maintain a neutral pH in cell cultures *in vitro* and prevented the occurrence of cell death during the growth of the bacterium in a chemically defined medium. In AMD water supplemented with urea, the degradation of urea promotes iron, aluminum and arsenic precipitation. Our data show that *ureC* was expressed *in situ*, which suggests that the ability to degrade urea may be expressed in some *Thiomonas* strains in AMD, and that this urease activity may contribute to their survival in contaminated environments.

Keywords: aluminum, arsenic, adaptation, metabolomics, acid stress

Introduction

Although acid mine drainage (AMD) environments are highly toxic to most living organisms due to the acidic conditions and the presence of elements such as arsenic, a stable microbial community composed of bacteria, archaea, and protists has been known for several years to inhabit the AMD-impacted Reigous creek near Carnoulès (France; Bruneel et al., 2003, 2006, 2011; Duquesne et al., 2003, 2008; Bryan et al., 2009; Halter et al., 2011; Slyemi et al., 2011; Volant et al., 2012). Various *Thiomonas* bacteria isolated from this river were previously characterized, and it has been suggested that by oxidizing arsenite, *Thiomonas* strains may promote the sorption of arsenic by iron oxides and their coprecipitation, resulting in a natural process of arsenic attenuation (Bruneel et al., 2003; Morin et al., 2003; Duquesne et al., 2008; Bryan et al., 2009; Arsène-Ploetze et al., 2010; Egal et al., 2010; Slyemi et al., 2011, 2013). The survival of these *Thiomonas* bacteria probably involves several processes. In particular, metabolic interactions between *Thiomonas* and other microorganisms may be essential to the survival and development of the microbial communities inhabiting the AMD-impacted Carnoulès waters. Interactions of this kind are known to contribute importantly to the survival and development of microbial communities in toxic environments (Jones et al., 2012; Johnson and Amarasekare, 2013). These interactions may be crucial to the functioning of ecosystems and the recycling of chemical elements in toxic environments such as AMDs where the growth conditions are particularly stringent (Wilmes and Bond, 2009). In addition, some of the genes involved in these interactions or in the general adaptation of strains in a particular niche are often found to exist in genomic islands (GEIs). GEIs consist of discrete DNA segments (ranging from 10 to 200 kbp in size), that are present in one species, but are not present in several other related species. These genomic regions sometimes differ in their nucleotide features (G + C content or codon usage) from the rest of the genome, and have often been found to exist in the vicinity of tRNA or tRNA-like genes. The boundaries of these islands frequently correspond to perfect or near-perfect direct repeats (DRs), usually resulting of site-specific integration. These regions often harbor functional or cryptic genes encoding integrases originated from phages or genes involved in plasmid conjugation processes. GEIs, which include elements of other kinds such as integrative and conjugative elements (ICE), conjugative transposons and cryptic or defective prophages, can result from one or several HGT events and genomic rearrangements (Juhas et al., 2009; Bellanger et al., 2014). They can be detected *in silico* by performing genomic comparisons between closely related strains. Studies on these GEIs shed interesting light on the bacterial survival strategies at work in inhospitable environments such as AMD waters.

The pH of the acidic AMD waters in Carnoulès sometimes drops to values of <3. One of the defense mechanisms whereby bacteria resist acidity is based on the degradation of urea: urease catalyzes the hydrolysis of urea, giving ammonia and carbonic acid. In aqueous solutions, the released carbonic acid and the two molecules of ammonia are in equilibrium with their deprotonated and protonated forms, respectively. This reaction increases the pH (Mobley, 2001). A defense mechanism of this kind has been previously described in the pathogen *Helicobacter pylori*, which resists the acidic conditions pertaining in the human stomach (Athmann et al., 2000; Stingl et al., 2002). In sediments collected at the AMD-impacted creek in Carnoulès, urea was detected in the interstitial water (Bertin et al., 2011; Halter et al., 2012). Urea is one of the organic substances produced and excreted by *Euglena mutabilis*, a photosynthesizing bio-indicator of AMD (Brake et al., 2001a,b; Brake and Hasiotis, 2010; Bertin et al., 2011; Halter et al., 2012). Other hitherto uncultivated bacteria such as "*Candidatus* Fodinabacter communificans" (Carn1 and Carn4) may also produce urea in the Carnoulès AMD, since genes involved in the production of urea were detected in their genomes (Bertin et al., 2011). Urea produced by organisms of this kind may then be used by other microorganisms to withstand the acidity. Interestingly, studies on metagenomic sequences have shown that at least one *Thiomonas*-like strain present in this community harbors genes possibly involved in urea degradation processes (Arsène-Ploetze et al., 2010; Bertin et al., 2011).

In this study, we searched for genes involved in urea degradation in the genomes of several *Thiomonas* isolates. Chemical and microscopic tests were first conducted under laboratory conditions in order to determine whether these *Thiomonas* strains are able to degrade urea and if so, whether their urease activity may contribute to acid tolerance and survival. Genes involved in urea degradation processes were found to be present in one GEI in *Thiomonas* sp. CB2. This GEI was incomplete in the other strains tested. The urease activity was found to confer greater viability on *Thiomonas* sp. CB2 in synthetic medium, and to promote metal precipitation in AMD water supplemented with urea. Interestingly, the *Thiomonas* sp. CB2 *ureC* gene was expressed *in situ,* which suggests that urea degradation may favor the survival of *Thiomonas* strains in the Carnoulès AMD.

Materials and Methods

Bacterial Growth Conditions

Thiomonas spp. CB2, CB1, CB3, CB6, and *Thiomonas arsenitoxydans* 3As were isolated from AMD-impacted water collected at the Reigous creek in Carnoulès (France; Duquesne et al., 2008; Arsène-Ploetze et al., 2010), and *Thiomonas intermedia* K12 was isolated from a corroded concrete wall in the Hamburg sewer system (Milde et al., 1983). These strains were grown in m126 medium: (Na_2HPO_4 4.5 g; KH_2PO_4 1.5 g; $Na_2S_2O_3$ 5.0 g; NH_4Cl 0.3; $MgSO_4.7H_2O$ 0.1 g; and yeast extract 0.5 g. The pH of the medium was adjusted to 5.0 using H_2SO_4). In addition, 4 $g.L^{-1}$ sodium arsenite (from a 500-mM stock solution) were added to the medium when required. The growth potential of *Thiomonas* cells was tested in AMD-impacted Reigous creek water collected on 24 January 2012, which was filtered twice using 0.22 μm filters. Since no urea was detected using a metabolomic approach (see below) in the AMD-impacted water sampled from the creek, urea was added to the AMD-impacted water as specified below. A final urea concentration of 0.02 M (from a 1.67 M stock solution) was adopted because

urease activity was detected in the m126 medium when the cells had been previously incubated at this concentration but not at lower concentrations (see the Results Section). Before being incubated in AMD-impacted water, *Thiomonas* strains were grown in the m126 liquid medium. After their growth, the cells were centrifuged for 15 min at 4500 × *g*, washed in 9 g.L^{-1} NaCl and resuspended in filtered AMD-impacted water at the initial Optical Density (OD_{600nm}) specified in the figure legends.

Viability Assessment

Thiomonas strains were grown in 20 mL of m126 medium at an initial OD_{600nm} of 0.002. Urea from a 1.67 M sterile stock solution was added to obtain a final concentration of 0.02 M. $O.D_{600nm}$ and pH were measured after 2, 3, 5, and 9 days. 500 µL of each cell culture (3As and K12) and 1 mL (CB2) were centrifuged separately at 10,000 × *g* for 10 min and pellets were recovered in 500 µL of sterile NaCl solution (9 g.L^{-1}). Viability staining was carried out using a Live/Dead BacLight Bacterial Viability Kit (Invitrogen, Cergy Pontoise, France): 0.5 µL of both SYTO9 green (5 mM) and iodide propidium iodide (1 g.L^{-1}) were mixed with 500 µL of cells and incubated for 20 min in the dark. 2.4 µL were used to observe viable or dead cells using a fluorescence microscope (Zeiss Axio-Observer used with a X63 oil immersion lens). Wavelengths of 488 nm and 532 nm were used for SYTO9 and propidium iodide excitation purposes, respectively. Numbers of viable and dead cells were calculated using NIH ImageJ analysis software (http://rsbweb.nih.gov/ij/). These tests were performed in triplicate, except on the 9th day in the case of strains 3As and K12, which were tested in duplicate. When tested in the AMD-impacted water, the mortality rate of *Thiomonas* strains was measured using cell plating methods: cell cultures were diluted in 9 g.L^{-1} NaCl solution and plated onto m126. After 10 days of incubation at 30°C, the colonies were counted.

Urease Activity

The ability of each *Thiomonas* strain to degrade urea was tested with cells grown in liquid m126 media with and without urea. Cells were centrifuged for 15 min at 4500 × *g*, washed and resuspended in PBS buffer (8 g.L^{-1} NaCl, 0.2 g.L^{-1} KCl, 1.44 g.L^{-1} Na_2HPO_4, and 0.24 g.L^{-1} KH_2PO_4) with the pH adjusted to 5.1, and the volume was adjusted in order to obtain the same number of cells, based on the OD at 600nm. One volume of cell suspension (5 × 10^8cells) was mixed with 1 volume of BCP reagent (1 g.L^{-1} of Bromo-Cresol Purple with or without 0.83 M urea). In BCP containing urea, the presence of a violet color indicated an increase in the pH due to the urea degradation activity, resulting in the formation of ammonia and carbonic acid. The effects of urea degradation activity on metal precipitation in the AMD-impacted water were assessed by adding 10 U of purified urease (from *Canavalia gladiata* (sword beans), Merck) to 5 mL samples of sterile AMD-impacted water in the presence or absence of 1 g.L^{-1} urea. As a negative control, urease was inactivated before being added to samples by heating it for 5 min at 95°C. Samples were then incubated at 30°C for 24 h and images of the precipitates were taken after a 10-min centrifugation step at 3750 × *g*. Precipitate formation was tested in three independent replicates in AMD-impacted water supplemented or not with urea and active urease.

Chemical and Metabolomic Analysis

The AMD-impacted water was filtered at the time of sampling and again at the laboratory and stored at 4°C during the experiments, all of which were completed within 1 year. Due to the characteristics of the creek water (**Table 1**), Fe(II) precipitated slowly even in non-inoculated water. Abiotic controls were therefore systematically performed and Fe(II) oxidation/precipitation was expressed as the difference between the Fe(II) concentrations measured in the soluble fraction of the non-inoculated and inoculated samples. In addition, the characteristics of the water were determined in 10-month old 0.22 µm filtered AMD-impacted water and compared with those observed at the time of sampling. No significant decrease in the concentrations of elements was detected, except for lead, arsenic and iron, the levels of which decreased by 50, 40, and 10%, respectively, after 10 months and cobalt and barium, the levels of which dropped to below the detection limit in both cases.

Concentrations of aluminum, calcium, manganese, iron, nickel, strontium, cadmium, thallium, cobalt, barium, lead, iron, total arsenic, and redox species arsenite (As(III)) and arsenate (As(V)) were determined by performing ICP-MS and HPLC-ICP-MS as previously described (Casiot et al., 2011) or inductively coupled plasma-atomic emission spectroscopy (ICP-AES). As(III) and As(V) species were also observed during metabolomic analysis using previously described GC-MS profiling methods (Lugan et al., 2010). Briefly, samples were vacuum-dried prior to derivatization with methoxyamine hydrochloride (20 mg.mL^{-1} in pyridine) and N,O-Bis(trimethylsilyl) trifluoroacetamide. A mixture of n-alkanes was introduced as retention index standards. The GC-MS system consisted of a 7683 injector, a 6890 gas chromatograph

TABLE 1 | Chemical parameters of the creek water at the time of sampling.

Parameters	Values[a]
pH	3.68
SO_4^{2-} (mg.L^{-1})	2.991
Fe (mg.L^{-1})	891.50
As (mg.L^{-1})	115.20
Al (mg.L^{-1})	36.95
Ca (mg.L^{-1})	266.80
Mn (mg.L^{-1})	7.41
Zn (mg.L^{-1})	18.53
Co (µg.L^{-1})	249
Ni (µg.L^{-1})	418
Cu (µg.L^{-1})	60.9
Sr (µg.L^{-1})	197
Cd (µg.L^{-1})	85.4
Sb (µg.L^{-1})	<QL
Ba (µg.L^{-1})	6.4
Tl (µg.L^{-1})	287
Pb (µg.L^{-1})	890
U (µg.L^{-1})	20.5

[a]*Measurements taken 30 m downstream from the source ("COWG") (Egal et al., 2010). DL, detection limit; QL, quantification limit.*

and a 5973 mass spectrometer (Agilent Technologies, Santa clara, USA). Samples were injected twice, in the splitless and split modes (with a split ratio of 50), on an Agilent J&W HP5-MS capillary column with helium as the carrier gas. As(III) and As(V) were identified by comparing the mass spectra obtained with those available in databases (NIST 2008). Total Fe and Fe(II) concentrations were measured using ICP-AES or ICP-MS and the previously described phenanthroline method (Fadrus and Malý, 1975), respectively.

FE-SEM and Energy-Dispersive X-ray Spectroscopy (EDXS, Thermo-Noran Vantage) were performed in order to assess the content of precipitates. In each condition, one pinhead of precipitate was treated with 1 mL hydrogen peroxide in order to eliminate the organic matter, using plastic tubes to prevent metal adsorption. Samples were then centrifuged at 3000 × *g* for 15 min and rinsed with Milli-Q water before performing an additional centrifugation step. Precipitates were then suspended in 100% ethanol. A droplet of the ultra-sonicated suspension was placed on a plasma-treated piece of silicon or carbon wafer providing a smooth, conductive support. After being dried at room temperature, the samples were carbon coated. Standard conditions were used at 12 keV primary beam energy. Since the arsenic L lines at 1.282 keV overlapped with the magnesium K lines at 1.254 keV, control measurements were also performed at 15 keV to check the presence/absence of the arsenic K lines at 10.543 keV. Depending on whether a silicon or carbon wafer was used, the spectra consistently showed a silicon K line at 1.735 keV or a carbon K line at 0.26 keV, respectively, the intensity of which depended on the thickness of the precipitate analyzed. Acquisitions were performed either in the fixed spot mode or by scanning a selected area, and at least three analyses were performed per sample. Due to the topography and the chemically inhomogeneous characteristics of the samples, it was not possible to calculate absolute concentrations in all the samples.

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

Genomic Comparisons

The *Thiomonas arsenitoxydans* 3As genome was previously sequenced and described (Arsène-Ploetze et al., 2010), while the K12 genome was sequenced by the US DOE Joint Genome Institute (NC_014153.1, NC_014154.1, NC_014155.1). The CB2, CB1, CB3, CB6 genomes analyzed in this study were obtained previously (Freel et al., in press) and integrated into the MicroScope platform (Vallenet et al., 2006, 2009, 2013) for analysis. The genome accession numbers of CB2, CB1, CB3 and CB6 (EMBL database) are LK931581-LK931672, LN831666-LN831688, LN831730-LN831775, LN831689-LN831714, respectively. These genomes were compared using the *RGP finder* tool provided with the MaGe platform (http://www.genoscope.cns.fr/agc/microscope/home/) in order to identify GEIs. The *RGP finder* interface can be used to search for potentially horizontally transferred genes (HGT) present in genomic regions (Regions of Genomic Plasticity). The *RGP Finder* method starts by identifying synteny breaks between a query genome and other closely related genomes. It then searches for HGT features (tRNA hotspots, mobility genes) and for the presence of any compositional bias (AlienHunter (Vernikos and Parkhill, 2006), SIGI-HMM (Waack et al., 2006), and GC deviation computation) in the query genome (https://www.genoscope.cns.fr/agc/microscope/compgenomics/genomicIsland.php?act=logout).

Total RNA Extraction from Sediment Originating from the Reigous Creek and Reverse Transcription

The total microbial community was recovered after Nycodenz gradient density separation as previously described (Bertin et al., 2011) from sediments sampled in May 2007 and stored at −80°C. RNA extraction and RT-PCR were performed as described previously (Bertin et al., 2011), with some slight modifications. Cells recovered from sediments were centrifuged at 10,000 × *g* for 15 min at 4°C. Pellets were resuspended in 400 µL of suspension solution [25 mM Tris-HCl (pH 7.6)-10 mM EDTA + 20% glucose (vol/vol)] and transferred into microtubes containing 0.5 g glass beads (0.1 mm in diameter) and 0.5 µL acidic phenol (pH 4.5). Cells were disrupted mechanically with a Retsch's Mixer Mill apparatus (three 30-s cycles of homogenization at maximum speed with 1 min intervals on ice). Microtubes were centrifuged at 13,000 × *g* for 5 min at 4°C. The aqueous phase was placed in a new tube with 1 mL TRIzol reagent (Life Technologies). Microtubes were incubated for 5 min at room temperature. 100 µL of chloroform was added and homogenized by pipetting. The tubes were centrifuged at 13,000 × *g* for 5 min at 4°C. The aqueous phase was then recovered and treated with 200 µL chloroform before being centrifuged at 13,000 × *g* for 5 min at 4°C. The aqueous phase was recovered and purified using the RNeasy Plus Mini kit (Qiagen). RNA integrity was checked by electrophoregram using a BioAnalyser (Agilent) and quantified by measuring A_{260} and A_{280} with a Nanodrop device. Reverse transcription was performed using the SuperScript® III One-Step RT-PCR System with Platinum® Taq DNA Polymerase (Life Technologies). Each reaction (total volume: 12.5 µL) contained 55 µg total RNA. Thermocycling conditions were as follows: 30 min at 50°C followed by 3.5 min at 94°C, followed by 40 cycles of 15 s at 94°C, 30 s at 55°C and 1 min at 68°C. A final 5-min elongation step at 68°C was added. Primers were designed using three genes: CARN2_aioA_for 5′-CCTGCCATTTCTGCATCG-3′ and CARN2_aioA_rev 5′-GCATTCGGAGTTGTACGC-3′ in the case of the *aioA* gene of the *Thiomonas*-like organisms previously detected at this site (CARN2_0821; Bertin et al., 2011); THICB2v2_ureC_for2 5′-CGAAGGCATGATCCTCAC-3′ and

THICB2v2_ureC_rev2 5′-CTCGTCGATGGTGTTGAC-3′ in the case of the *ureC* gene (THICB2V3_370058); THICB 2v2_360024_for2 5′-GGGCTGTACACCAACTTC-3′ and THIC B2v2_360024_rev2 5′-GCCATCTTCAAGCTGCAG-3′ to amplify the transcript encoding the putative allophanate hydrolase gene (THICB2V3_360023). Negative controls were performed by omitting the reverse transcription step with each pair of primers.

Results and Discussion

Genes Involved in Urea Transport and Degradation are Present in One Genomic Island in Some *Thiomonas* Strains

The genomes of several closely related *Thiomonas* strains, seven originating from the same AMD (CB1, CB2, CB3, CB6, 3As, ACO3, ACO7) (Freel et al., in press) and one isolated from a sewage pipeline (*Tm. intermedia* K12), have been recently sequenced. A comparative analysis of these genomes using *RGP finder* (see Materials and Methods) showed the presence of 19 GEIs in the *Tm.* sp. CB2 genome (Freel et al., in press). One GEI 88.2 Kb in size (95 CDS) detected using this comparative genomic approach, which is present in the vicinity of another GEI (RGP10; Freel et al., in press) involved in arsenic resistance, was found to have the following characteristics: it is flanked by a miscRNA gene, contains several transposases and shows a compositional bias indicating possible acquisition by HGT (see Supplementary Table 1). Interestingly, this genomic island carries genes involved in urea degradation. In this study, we therefore refer to this GEI as the "urea island." Two distinct enzymes, urease and urea amidolyase, are known to degrade urea into ammonia and carbonic acid. In bacteria, the urea amidolyase activity results from two separate enzymes, a urea carboxylase and an allophanate hydrolase (Kanamori et al., 2004). The "urea island" in *Tm.* sp. CB2 carries genes encoding these enzymes (*ure* genes encoding urease and two genes which are essential to the urea carboxylase (UCA)-allophanate hydrolase pathway) and *urt* genes involved in urea import (**Figure 1A**). A second copy of the *urtABCDE* genes was present at another locus in the CB2 genome (from 2246372 to 2251947). This duplication was not detected in the genomes of the other strains tested.

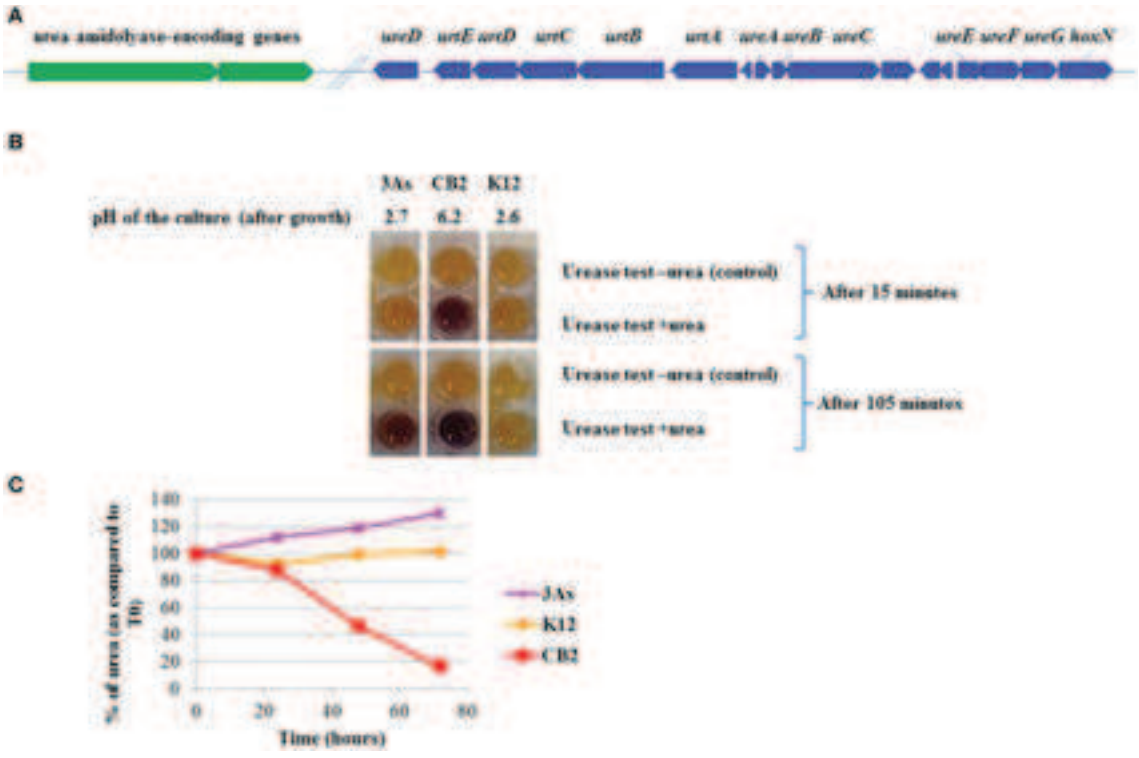


FIGURE 1 | Genes involved in urea degradation in CB2 and urea degradation activity in *Thiomonas* strains. (A) Genes involved in urea import (*urt*) and degradation (*ure,* in blue or genes encoding the urea amidolyase involved in the UCA-allophanate hydrolase pathway, in green) were detected in a genomic island in the CB2 genome. **(B)** After growth of cell cultures in m126 supplemented with urea (1 g.L^{-1}), urea degradation activity was detected using Bromo-cresol purple as described in Materials and Methods. The purple color indicates urea degradation. In the negative control where no urea was added to the BCP reagent, the orange color stands for no activity observed (as indicated by "urease test−urea"). Urea degradation affected the pH of the media when cells were incubated in the synthetic medium m126 supplemented with urea. **(C)** Urea was detected in the the supernatant of CB2, 3As, or K12 cell cultures in synthetic medium m126 supplemented with 1 g.L^{-1} urea, at T_0 and after 1, 2, or 3 days using mass spectrometry (UPLC-MS/MS). The data are expressed as the percentages of urea (area under the peak) measured at T_0.

The urea island also carries genes involved in metal, sulfur, lipid and nitrogen metabolism, including 27 genes involved in nitrate transport and reduction in particular. It also contains genes encoding ABC transporters and several hydrolases that contribute to the degradation of various compounds such as cyanate and creatinine (Supplementary Table 1).

The synteny of the genes detected in and around the urea island in CB2 was compared with that of the corresponding genes in CB1, 3As, CB3, CB6, and K12 (**Figures 2**, **3**), and these analyses showed that some blocks of genes were highly conserved, although possibly translocated or inverted, whereas others were lacking in some *Thiomonas* strains. In particular, genes involved in urea degradation and transport processes were found to be present in 4 genomes from *Thiomonas* strains (CB2, CB1, CB6, and 3As) but not in CB3 or K12 (**Figure 4**), while the genes detected in the vicinity of the *ure* genes in CB2 are present in both of these strains (**Figure 2**). These results suggest that these urea islands have evolved differentially in these closely related *Thiomonas* strains. GEIs may endow bacterial strains with particular abilities (Juhas et al., 2009). Since urea degradation is thought to be involved in acid tolerance in some bacteria such as *H. pylori* (Stingl et al., 2002), and since *Thiomonas* strains were isolated from acidic AMDs, the contribution of these genes to acid tolerance was further investigated.

Urea Degradation Activity Prevents Acidification of the Medium and Cell Death during the Growth of *Thiomonas* sp. CB2

First, the urea degradation activity of *Thiomonas* sp. CB2 was tested with cells previously grown in culture media supplemented with various urea concentrations (Supplementary Figure 1). An optimized colorimetric test (see Materials and Methods) showed the occurrence of urea degradation in the case of cells grown with 0.02 M urea but not in those pre-grown with lower urea concentrations (Supplementary Figure 1). These findings suggest that the expression of the corresponding genes is induced in the presence of urea. This urea degradation activity could not be quantified but was reproduced at least five times and confirmed by the results of metabolomic studies (**Figures 1B,C**). No activity was detected with either *Thiomonas intermedia* K12 or *Thiomonas* sp. CB3 (**Figures 1**, **4**; Supplementary Figure 1), in agreement with the genomic analyses, which showed that these two strains carry no genes involved in urea degradation (**Figures 2**, **4**). The fact that strains 3As, CB1, and CB6 showed very low levels of activity (**Figures 1**, **4**; and data not shown) suggests that under the conditions tested here, the genes involved in urea degradation processes are not strongly expressed in these strains or that the corresponding enzymes are not fully active. Close analysis of the genes detected in the urea islands of these

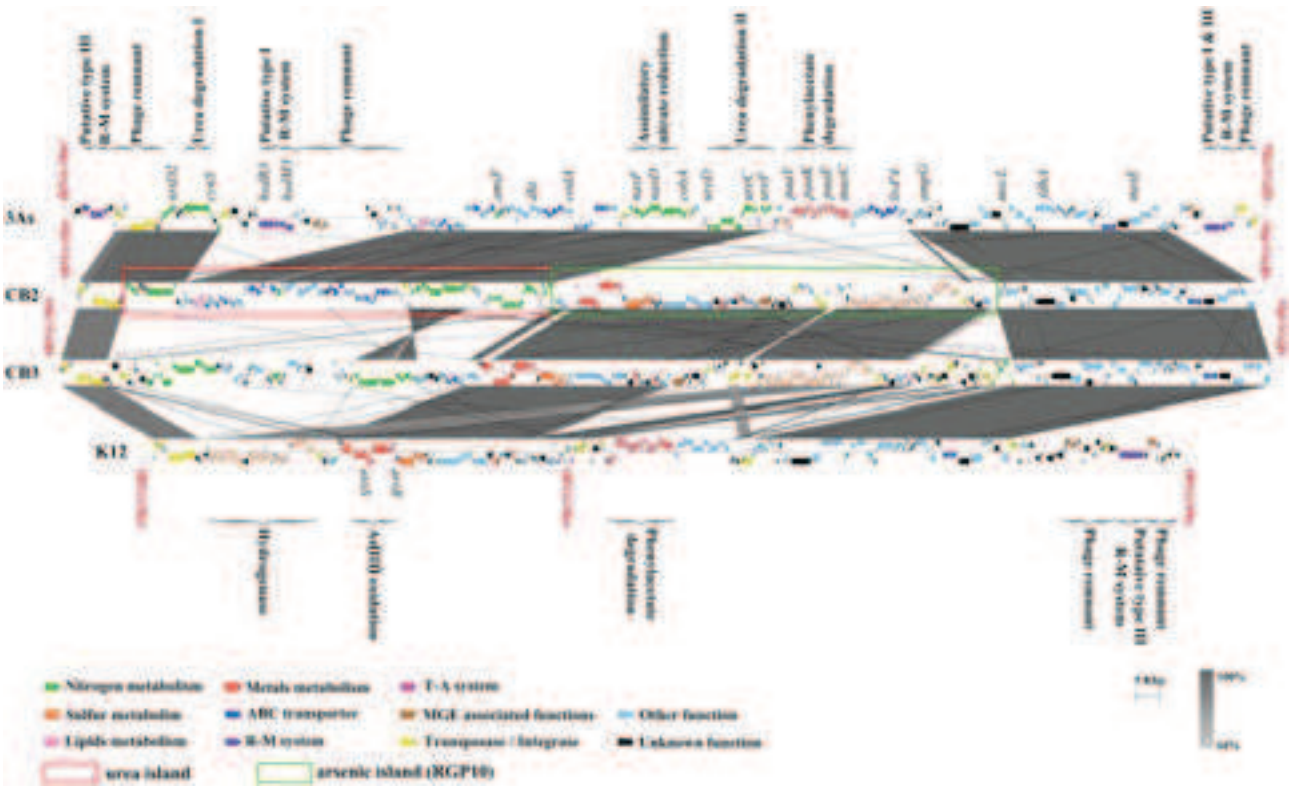


FIGURE 2 | Comparison between the *Thiomonas* genomic regions carrying urea islands. The large genomic region comprising the urea island and another island, RGP10 in CB2 and the synteny between 3As, K12, CB3, and CB2 genomes are shown. The percentage of nucleotide identity is expressed in shades of gray (see the gray scale). Figures were generated with Easyfig (Sullivan et al., 2011).

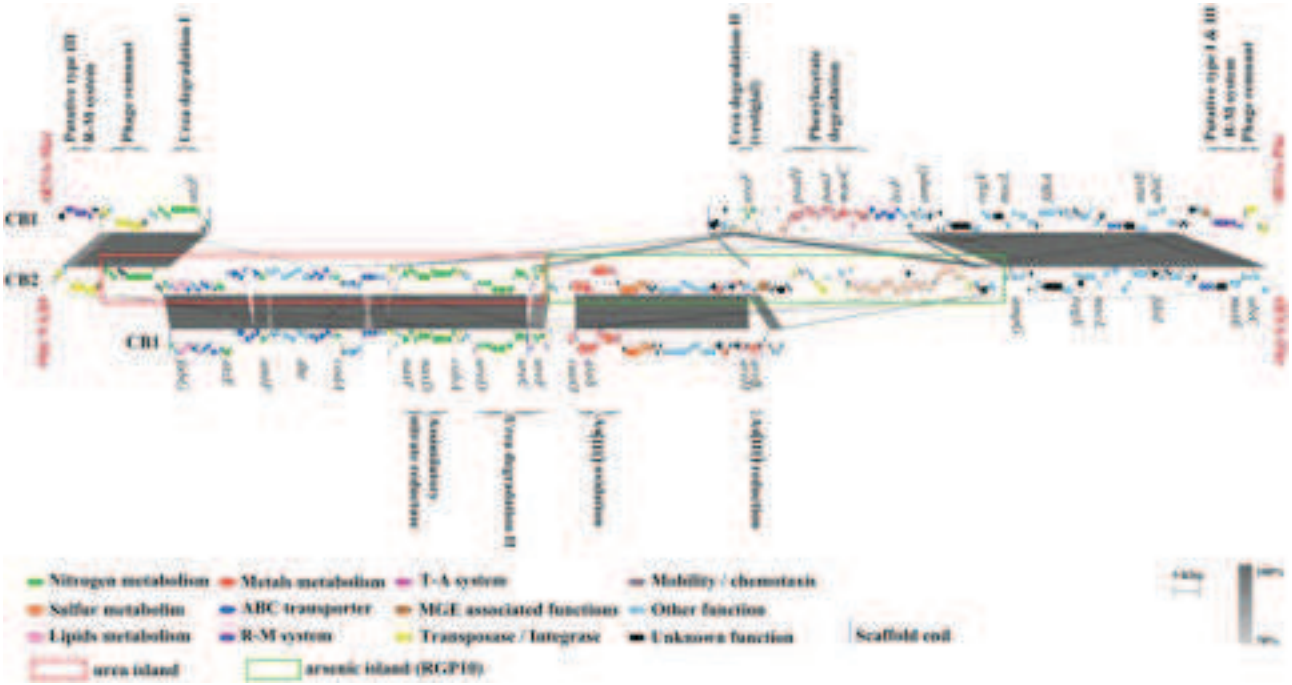


FIGURE 3 | Comparison between the *Thiomonas* genomic regions carrying a urea island. The large genomic region comprising the urea island and another island, RGP10 in CB2 and the synteny between CB1 and CB2 genomes are shown. The percentage of nucleotide identity is expressed in shades of gray (see the gray scale). Figures were generated with Easyfig (Sullivan et al., 2011).

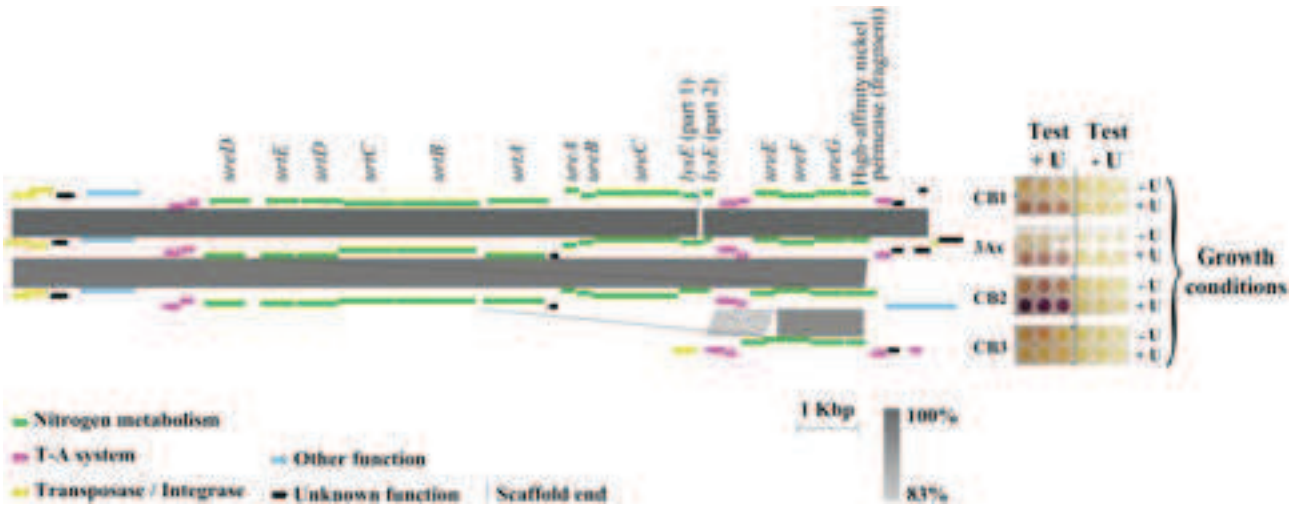


FIGURE 4 | Comparison between the genes involved in urea degradation and urease activity in the *Thiomonas* strains investigated. Only part of the urea island carrying the *ure* and *urt* genes involved in urea degradation is shown. The figure shows the synteny between CB1, CB2, CB3, and 3As. The percentage of nucleotide identity is expressed in shades of gray (see the gray scale). Figures were generated with Easyfig (Sullivan et al., 2011). The urea degradation activity is shown on the right: cells were grown in the absence (−U) and presence (+U) of urea. After growth, cells were incubated in the presence of urea and Bromo-cresol purple for 30 min, as described in Materials and methods (as indicated by "test + urea"). As a negative control, the test was performed in the absence of urea in the reaction mixture (as indicated by "test − urea"). In the case of the test performed in the presence of urea, the purple color indicates urea degradation.

Thiomonas genomes showed that two genes differed in CB2 in comparison with 3As, CB1, and CB6: (i) one gene encoding a putative transporter belonging to the *lysE* family was fragmented in 3As; (ii) a gene encoding a high-affinity nickel permease was shorter in CB1, CB6, and 3As than in CB2 (**Figure 4**, and data not shown). Since nickel plays an important role in urease activity

(Stingl et al., 2002), this may explain why the degradation activity was lower in these strains than in CB2. This possible explanation for the difference in urea degradation activities will have to be tested when appropriate genetic tools have been developed for studying these bacteria. All in all, the present findings suggest that several *Thiomonas* strains contain a genomic island harboring genes which are either directly or indirectly involved in the degradation of urea, whereas other strains lack these features, which results in differences between strains in terms of their ability to degrade urea.

The effects of urea degradation on acid tolerance were further tested in three strains, CB2, 3As, and K12. During the growth of these three strains in m126 in the absence of urea, the pH of the culture medium dropped to approximately 3.0 or less within 24 h, and stabilized after 50 h (**Figure 5**). The viability of the cells decreased rapidly, probably due to this pH decrease. In the presence of urea, the pH of the CB2 growth medium remained at a value of approximately 6 because of the production of ammonia and carbonic acid during the urea degradation process, whereas in both K12 and 3As, the pH dropped to 3 even in the presence of urea in the growth medium (**Figure 5**). CB2 cells were less viable when grown without than with urea (**Figure 5**). The viability of 3As and K12 was similar whether cells were grown in the presence or absence of urea, as well as being similar to that of CB2 cells grown without any urea (**Figure 5**). These data show that the urea degradation activity of CB2 in synthetic media contributed to preventing the acidification of the medium during cell growth and enabled CB2 to survive more efficiently than 3As and K12 under the conditions tested here. The acid resistance conferred on CB2 by urease activity may result from a similar mechanism to that found to occur in the human pathogen *H. pylori*, which resists acid in the human stomach by degrading urea into carbonic acid and ammonia, the deprotonated and protonated forms of which are in a state of equilibrium (Athmann et al., 2000). It was then proposed to test whether this activity played a role when this bacterium was incubated under AMD conditions.

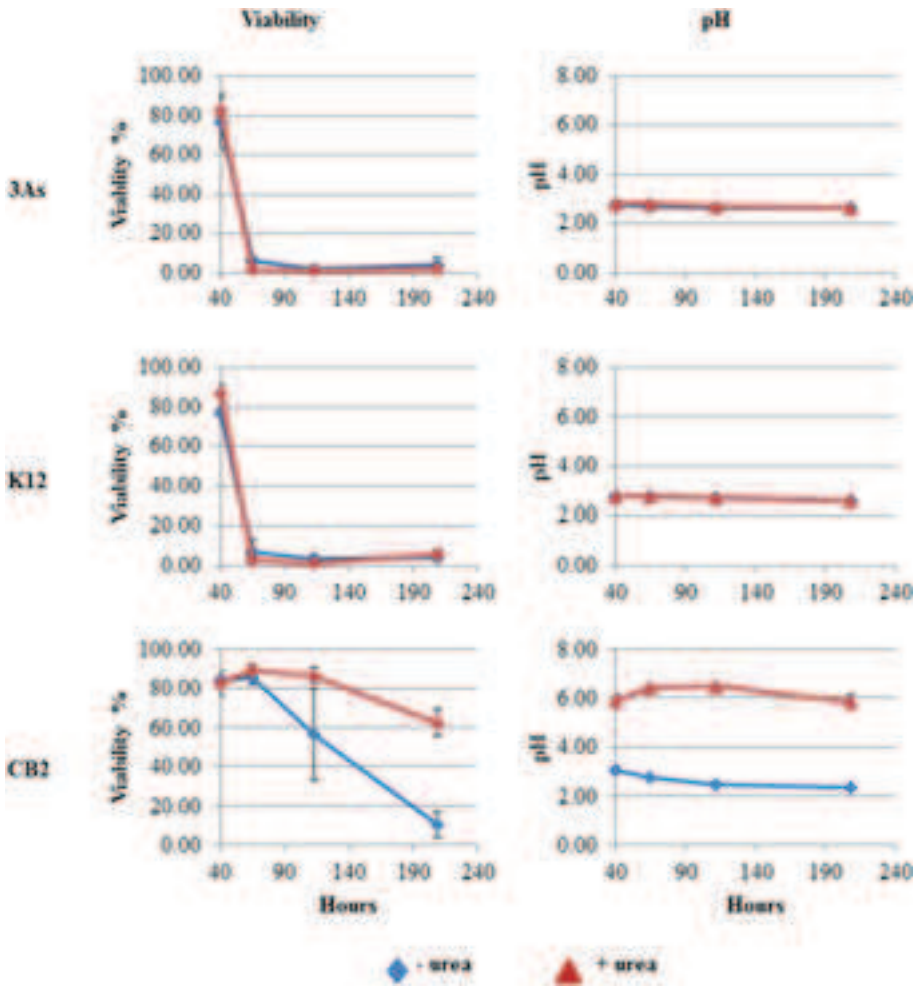


FIGURE 5 | Effect of urea degradation on cell viability. Viability and pH were measured during the growth of the *Thiomonas* strains in m126 medium in the absence and presence of urea. Error bars indicate standard deviations based on triplicate cultures.

The Degradation of Urea by *Thiomonas* sp. CB2 Promotes the Precipitation of the Toxic Metals in AMD-impacted Water Supplemented with Urea

The effects of the CB2 urea degradation activity were therefore tested in contaminated water from the AMD-impacted Carnoulès creek under laboratory conditions (**Table 1**). No urea was detected in the AMD-impacted water used in these experiments (data not shown). CB2, 3As, and K12 could therefore be incubated in this AMD-impacted water, whether or not it was supplemented with urea. When no urea was added to the AMD-impacted water, the growth of all three strains, as shown by an increase in the OD_{600nm}, resulted in a decrease in the pH (**Figures 6A,C**). However, when CB2 was incubated in the AMD-impacted water with urea, the pH increased within 48 h (**Figure 6C**), reflecting the occurrence of urea degradation activity. The number of CFU.mL^{-1} decreased with time, which suggests that the *Thiomonas* cell survival ability decreases under these conditions (**Figure 6B**). Contrary to what was observed in m126 medium, the viability of the CB2 cells did not increase in the AMD water in the presence of urea, and was even slightly impaired. The urea degradation process therefore did not improve the rate of CB2 survival under these conditions tested. The acid resistance mechanism involving urea degradation activity may not have been sufficiently strong to improve the viability of *Thiomonas* sp. CB2 in the AMD-impacted water, the pH of which was lower than that of the synthetic medium.

An orange precipitate accumulated during the experiments with CB2 in AMD-impacted water in the presence of urea, which was not observed in samples where no urea degradation activity was possible (**Figures 7A,B**). To test whether this precipitate was correlated with urease activity, purified urease and urea were added to sterile AMD-impacted water. A substantial orange precipitate was observed when both urease and urea were added, whereas no precipitate occurred when only urea was added, and less precipitate when urease was heat-inactivated (**Figure 7D**). The presence of this orange material suggested that Fe was precipitated when either urease or bacteria with urea degradation activity were present in the AMD water supplemented with urea. A covariance analysis was performed with MATLAB R2014a (using the *aoctool* and *multcompare* commands) on the quantity of soluble Fe(II) measured in the dissolved phase vs. time. This analysis showed that the rate of Fe(II) oxidation and subsequent Fe(III) precipitation was significantly higher ($p < 0.01$) in the case of CB2 grown with urea than in that of the other cell cultures tested (CB2 without urea and 3As or K12 with and without urea) (**Figure 7C**). CB2 is not able to oxidize ferrous iron in synthetic medium, (data not shown), which is consistent with the finding that its genome lack any detectable genes normally associated with Fe(II) oxidation. This is also true in the case of other *Thiomonas* strains (Slyemi et al., 2011). The iron precipitation observed here in AMD-impacted water in the presence of urea was therefore probably due to abiotic oxidation. In view of the

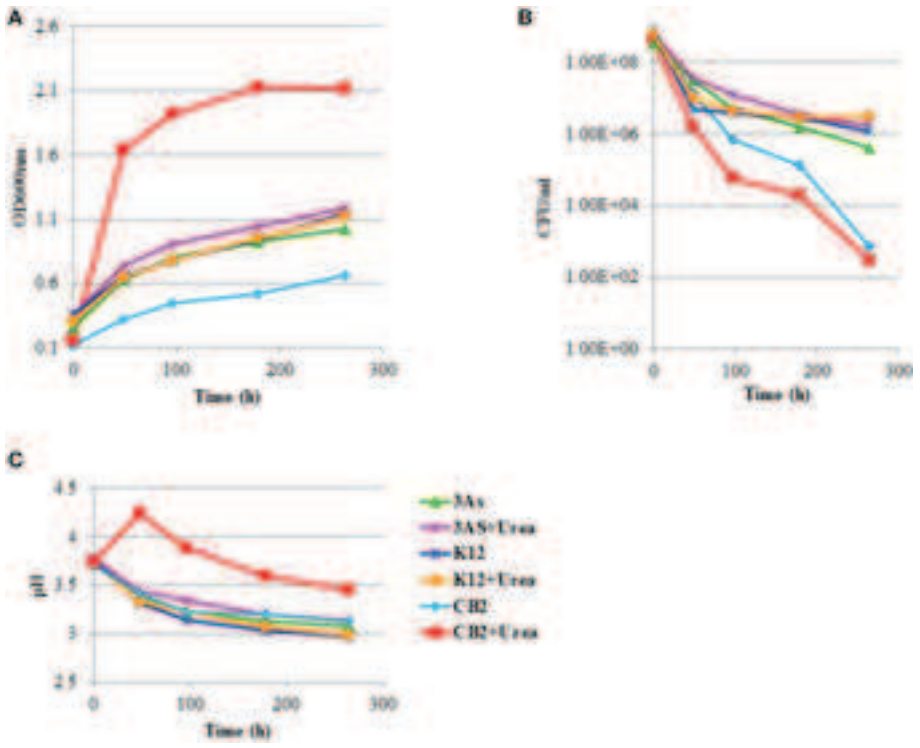


FIGURE 6 | Effects of incubating *Thiomonas* strains in AMD-impacted water. (A) OD_{600nm}, **(B)** Viability expressed in CFU/mL, and **(C)** pH was measured in AMD-impacted water incubated with *Thiomonas* strains in the presence and absence of 1 g.L^{-1} urea. This experiment was representative of three independent experiments.

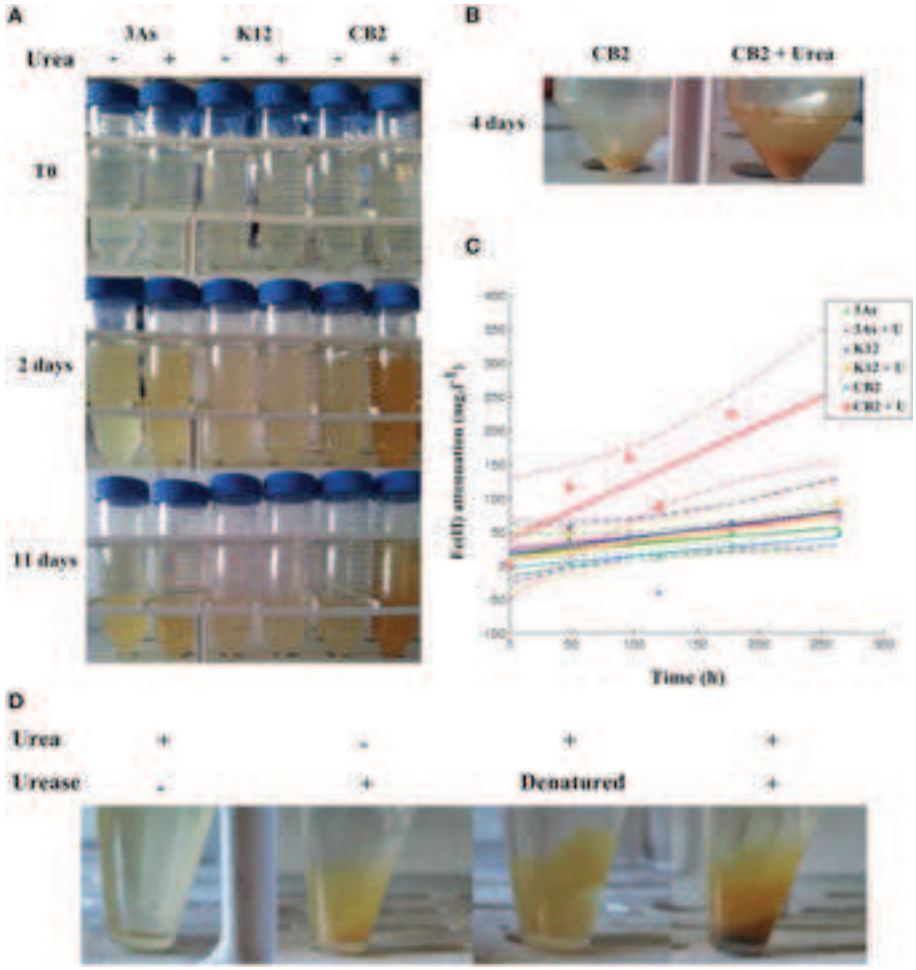


FIGURE 7 | *Thiomonas* urea degradation activity promotes Fe(II) oxidation/precipitation in AMD-impacted water supplemented with urea. K12, 3As, and CB2 were incubated at an initial OD_{600nm} of 0.2 – 0.3 in AMD-impacted water in the absence and presence of 1 g.L^{-1} urea. **(A)** The medium supplemented with urea and CB2 acquired an orange color with time, which did not occur with 3As or K12. **(B)** Orange precipitate detected after centrifuging 4-day cell cultures of CB2 in AMD-impacted water supplemented or not with 1 g.L^{-1} urea. **(C)** Kinetics of Fe(II) oxidation/precipitation. Fe(II) oxidation/precipitation is expressed as the difference between the Fe(II) concentrations measured in the soluble fractions of the non-inoculated and inoculated samples. The dotted line gives the 5% confidence interval of each of the regression lines, computed with the MATLAB fitlm command. **(D)** Effects of urease activity on the formation of orange precipitate in the AMD-impacted water. AMD-impacted water was supplemented with 10 U of purified urease in the presence and absence of 1 g.L^{-1} urea. As an additional control, urease was heat-inactivated for 5 min at 95°C.

TABLE 2 | Metal content of the Precipitates and Supernatant obtained from 3As, CB2, and K12 cell cultures in AMD-impacted water.

Metal[a]	Supernatants						Precipitates					
	CB2	Urea + CB2	3AS	Urea + 3AS	K12	Urea + K12	CB2	Urea + CB2	3AS	Urea + 3AS	K12	Urea + K12
Al	45.7	35.6	45.3	42.9	43.6	43.6	69.6	1.106	98.9	80.2	87.5	103
As	80.9	17.1	83.0	80.4	59.4	58.5	2.355	6.899	3660	2676	4.797	6.134
Cd	0.066	0.067	0.070	0.066	0.067	0.067	0.025	0.07	0.020	0.021	0.039	0.051
Co	0.24	0.23	0.24	0.23	0.23	0.23	0.27	0.25	0.30	0.24	0.24	0.29
Cu	0.072	0.07	0.074	0.073	0.074	0.073	0.22	0.75	0.76	0.59	0.66	0.58
Fe	882	714	874	824	847	851	5.961	20.144	11478	8488	8.189	10.224
Ni	0.47	0.46	0.46	0.46	0.47	0.45	0.64	0.73	0.81	0.64	0.62	0.67
Pb	0.2	0.075	0.22	0.20	0.065	0.066	10.3	23.9	12.6	10.4	28.4	35.3
Zn	20.5	20.2	20.8	19.5	19.9	20.1	21.2	25	27.6	21.5	21.8	26.9

[a]*Analysis of precipitates and supernatants was performed by ICP-MS after incubating CB2, 3As, and K12 in AMD-impacted water for 11 days in the presence and absence of 1 g.L^{-1} urea. Metal concentrations are given in mg.kg^{-1}.*

dependence of the Fe(II) oxidation rate on the pH in natural waters (Sigg et al., 2006), the iron oxidation observed here may have been at least partly due to the increase in the pH observed under our experimental conditions as the result of urea degradation activity.

The pH increase observed here may have induced the co-precipitation of other metals together with Fe under these experimental conditions. Using inductively coupled plasma-mass spectrometry (ICP-MS, **Table 2**) and X-ray microanalysis (**Figure 8**), we established that in samples where urea degradation had occurred (samples inoculated with CB2 in the presence of urea), the soluble iron, arsenic, and aluminum concentrations were lower than in the other samples tested. In addition, high amounts of these elements were detected in the precipitates obtained from these samples (**Table 2**; **Figure 8**). Fe(III) precipitation probabaly causes the co-precipitation of arsenic, as previously found to occur in AMD-impacted waters (Casiot et al., 2003; Morin et al., 2003; Duquesne et al., 2008). Previous studies have shown that *Thiomonas* strains express arsenite oxidases in AMD-impacted creek waters (Bertin et al., 2011). As(III) may be oxidized into As(V), which is less soluble and precipitates with Fe(III) more efficiently than As(III) under these conditions (Casiot et al., 2003; Morin et al., 2003; Cheng et al., 2009; Maillot et al., 2013). *Thiomonas* sp. CB2 oxidizes As(III) in the presence of organic compounds (Bryan et al., 2009), and we observed in the present study that this bacterium is also able to oxidize As(III) in the presence of urea in m126 medium (**Figure 9**). Therefore, to test whether CB2 is able to oxidize arsenite in AMD-impacted water, the concentrations of As(III) and As(V) present in the soluble fraction were measured using ICP-AES or GC-MS. The concentrations of these two forms of arsenic decreased under all the conditions tested (**Figure 10**). No As(V) accumulation was observed in the soluble phase because As(V) is more efficiently adsorbed by Fe precipitates than As(III) in AMD-impacted water (Morin et al., 2003; Cheng et al., 2009; Maillot et al., 2013). These explanations are consistent with results obtained in previous

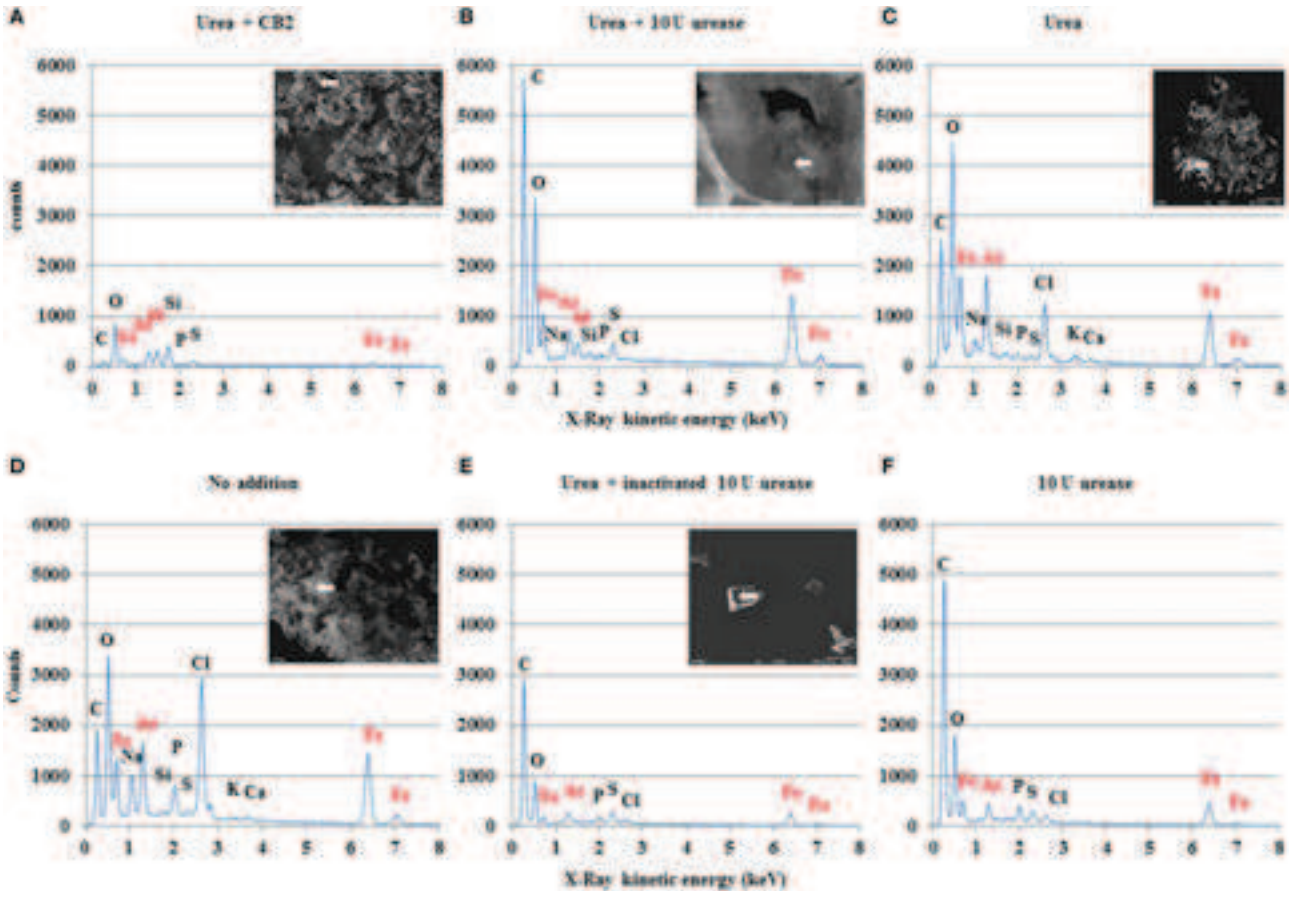


FIGURE 8 | Analysis of precipitates' composition. SEM photographs and microanalytical X-ray graphs of precipitates. These were obtained with AMD-impacted water supplemented **(A)** with 1 g.L^{-1} urea after incubating CB2 for 4 days and **(B)** after incubating 10 U of urease in AMD-impacted water supplemented with 1 g.L^{-1} urea for 1 day, respectively. White arrows on the photographs show spots on which the X-ray microanalyses were conducted to obtain the graphs presented below. Since the magnesium X-ray K lines interfered with the arsenic L lines around 1250–1300 eV, the presence of As was confirmed by checking its K lines at higher energy levels. Silicon in the second graph and carbon in the first and third ones originate from the supports used in these studies. Pictures and graphs are representative of at least three points analyzed. These graphs show that aluminum is present in the precipitate only under the conditions where urea degradation activity was possible. Controls were also performed with **(C)** urea added alone, **(D)** no urea added **(E)** with inactivated urease, and **(F)** with urease but no urea.

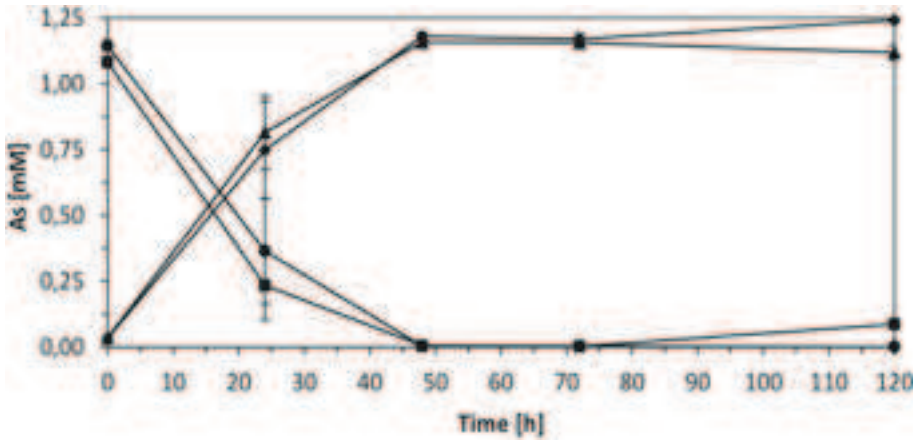


FIGURE 9 | Ability of *Thiomonas* strains to oxidize arsenite in m126. Cells were grown in m126 medium supplemented with 1.33 mM arsenite, in the absence (squares and triangles) and presence of 1 g.L^{-1} urea (diamonds and circles) and concentrations of As(III) (squares and circles) and As(V) (triangles and diamonds) were measured in culture supernatants. Error bars give the standard deviations based on triplicate cultures.

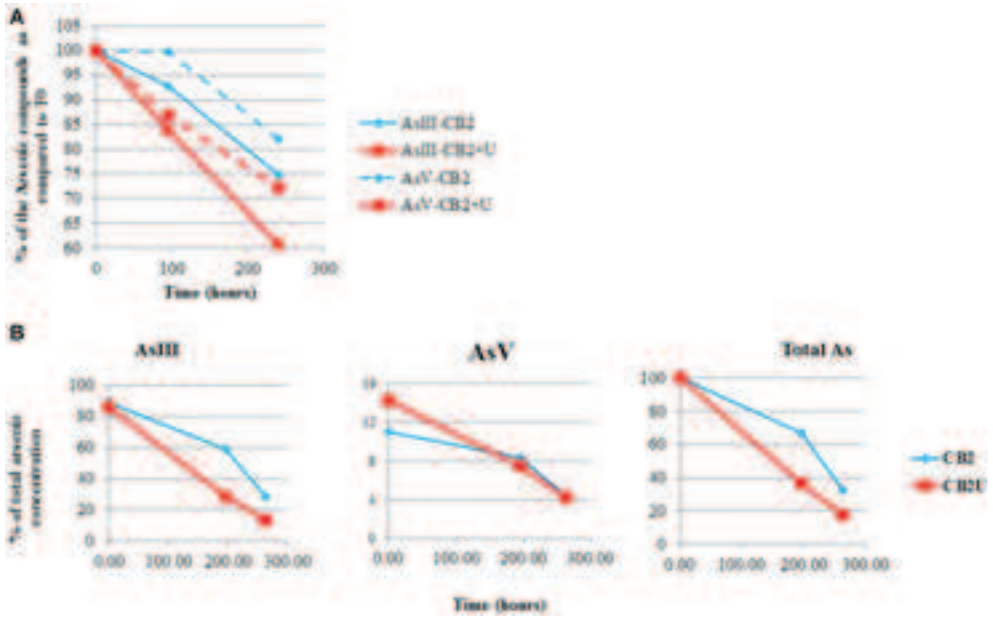


FIGURE 10 | Arsenic concentration in AMD-impacted water inoculated with CB2 in the presence and absence of urea. As(III) and As(V) were measured by performing **(A)** GC-MS or **(B)** ICP-AES in duplicate. Arsenic concentrations are expressed as a percent of the total arsenic concentration measured at T_0.

studies on the geochemical processes underlying iron and arsenic solubility in AMD (Cheng et al., 2009; Klein et al., 2014). All in all, the present findings indicate that As(III) may be oxidized into As(V) and precipitated with iron more readily when urease activity is possible.

Another noteworthy finding made in this study was that bacterial urea degradation activity induces aluminum precipitation. It has been established that in aluminum sulfate solutions and homogeneous solutions of aluminum acetate, sodium sulfate and urea, the reaction between water and the ammonia formed by urease activity yields hydroxide ions, which react in turn with aluminum sulfate, generating particles of aluminum basic sulfate (Simpson et al., 1998; Kara and Sahin, 2000). The AMD-impacted water used in this study certainly has a high sulfate content (**Table 1**). Alternatively, aluminum may precipitate in the form of aluminum oxyhydroxides, as suggested by our X-ray microanalyses (**Figure 8**). The pH value reached in our samples after urea degradation was slightly higher than 4.2 (**Figure 6**). Previous authors have indeed reported that aluminum hydroxide or microcrystalline gibbsite flocs are formed at pH values ranging from 4.2 to 4.9 in AMD and form aggregates at pH levels above 5 (Furrer et al., 2002).

The present data suggest that the bacterial degradation of urea may accelerate iron, arsenic, and aluminum precipitation in AMD in the presence of urea. Since urea was detected in the interstitial water of sediments from the AMD-impacted creek, it may be available to bacteria inhabiting this AMD to use for degradation purposes (Halter et al., 2012). However, the effects of urea degradation on metal precipitation were observed in this study under laboratory conditions using a higher urea concentration than that previously measured *in situ* (Halter et al., 2012). In addition, some of the *Thiomonas* strains inhabiting AMD waters are able to degrade urea whereas others are not, which suggests that urea degradation activity is not crucial to the survival of *Thiomonas* in AMD environments. In order to determine whether these biological activities are relevant *in situ*, the levels of expression of the *Thiomonas* genes encoding the arsenite oxidase and urease were therefore examined by performing RT-PCR with RNA extracted from the Reigous creek sediment community, as previously (Bertin et al., 2011). These *Thiomonas* cells extracted from the Reigous creek sediment were found to express the two genes *aioA* and *ureC* (**Figure 11**) responsible for the two activities promoting toxic metal precipitation, i.e., arsenite oxidation and urea degradation, respectively. These findings suggest that some of the organisms in the Carnoulès AMD community are able to produce large enough amounts of urea to promote the expression of *ure* genes *in situ*. CB2 may benefit indirectly from microbial activities in microscale environmental niches where the urea concentrations may be higher than those previously measured in the macroscale environment. This bacterium is known to be able to form biofilms, for instance (Marchal et al., 2011): other organisms surrounding *Thiomonas* cells in biofilms may provide metabolites such as urea or make the environmental conditions more suitable for CB2 survival and growth. These organisms may in turn benefit from the urea degradation and arsenite oxidation activities performed by *Thiomonas*, since both activities accelerate the precipitation of toxic metals. The urea degradation process promoted by *Thiomonas* may therefore contribute to enhancing the survival or the fitness of other microorganisms in the surrounding biofilm community, as previously suggested to occur in the case of other activities and other multi-species biofilms exposed to toxic metals (Koechler et al., 2015). Previous studies have established that the protist *E. mutabilis* present in the Carnoulès AMD is a primary producer excreting organic compounds which may be consumed by bacterial species (Bertin et al., 2011; Halter et al., 2012). Among the organic substances produced by this protist, urea was found to be excreted in synthetic medium (Halter et al., 2012). Other bacteria belonging to this community, namely "*Candidatus* Fodinabacter communificans" (Carn1 and Carn4), also carry genes involved in urea production (Bertin et al., 2011). If one of these organisms produces urea, this compound may then be degraded by *Thiomonas*. Since several organisms may be responsible of urea production, some of which have not yet been grown and studied *in vitro*, a new experimental approach will be required to identify the source(s) of urea and the relevance of its degradation to the survival of *Thiomonas* in AMDs. Generally speaking, the results presented here confirm the complexity of these processes, which probably involve interactions both between *Thiomonas* and the urea producers (*E. mutabilis* or other less well documented bacteria such as "*Candidatus* Fodinabacter communificans") and with other mechanisms (such as arsenite oxidation). In conclusion, it emerges that several closely related *Thiomonas* strains co-exist in the Carnoulès AMD waters, which have different abilities in this toxic environment. Recent studies have shown that several closely-related bacteria co-existing in AMD biofilms express different proteins and therefore play distinct ecological roles (Denef et al., 2010). Further studies on this fine-scale heterogeneity and the interactions occurring between *Thiomonas* strains and other members of the AMD community will now be required in order to understand more clearly how these communities function and survive in these highly toxic ecosystems.

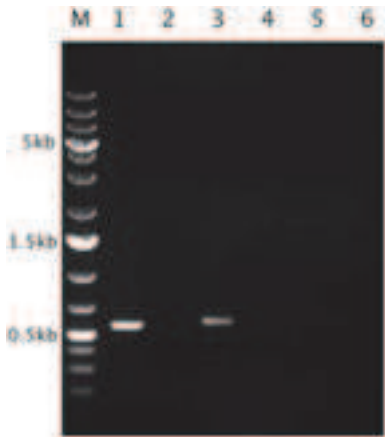


FIGURE 11 | *In situ* expression of the *Thiomonas* arsenite oxidase and urease-encoding genes. Agarose gel analysis of transcripts corresponding to *aioA*, *ureC,* and the gene encoding the allophanate hydrolase amplified by performing RT-PCR on the RNA extracted from the bacterial community inhabiting the Carnoulès AMD. Lane M: GeneRuler^{TM} 1kb DNA Ladder Plus (Fermentas). Lane 2, 4, and 6: negative controls (with each gene, the same reaction was performed but without any reverse transcriptase). Lane 1: RT-PCR product in the case of *aioA* (555 bp). Lane 3: RT-PCR product in that of *ureC* (570 bp). Lane 5: RT-PCR product in that of the allophanate hydrolase gene (515 bp). The amplification products were sequenced and the sequences predicted were obtained (i. e., those of *aioA* and *ureC*).

Author Contributions

JF, JA, DaH, MM, DH, RL, CC, PB, and FA designed the research project; JF, JA, DaH, CC, DH, SK, RL, MM, FS, and FA performed the research; VB, FP, and JaF contributed new reagents/analytical tools; JF, JA, DaH, CC, RL, DH, MM, FP, PB, and FA analyzed the data; and JF, PB, FP, and FA wrote the paper.

Acknowledgments

The authors thank Odile Bruneel for providing the CB1, CB2, CB3, and CB6 strains, Violaine Bonnefoy for providing the 3As strain and Sabine Heinhorst for providing the K12 strain. This research was supported by grants from the Université

de Strasbourg (UdS), the Centre National de la Recherche Scientifique (CNRS), and the Region Alsace to JF. This study was financed by the EC_2CO program (INEE, CNRS), the "Observatoire de Recherche Méditerranéen de l'Environnement-Observatoire des Sciences de l'Univers" (OSU OREME), the RARE (ANR 07-BLANC-0118) and the THIOFILM (ANR-12-ADAP-0013) projects. MM and JF were supported by the Agence Nationale de la Recherche (ANR COBIAS) project PRECODD 2007 and ANR THIOFILM (ANR-12-ADAP-0013). FA was supported by the CNRS ("Mise en Délégation Pour Activité de Recherche au CNRS") and JA and DaH were supported by a grant from the French Ministry of Education and Research. The authors thank Odile Bruneel for providing the *Thiomonas* strains and Yan Ploetze for his help with the figures.

Supplementary Material

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fmicb.2015.00993

References

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. doi: 10.1371/journal.pgen.1000859

Athmann, C., Zeng, N., Kang, T., Marcus, E. A., Scott, D. R., Rektorschek, M., et al. (2000). Local pH elevation mediated by the intrabacterial urease of *Helicobacter pylori* cocultured with gastric cells. *J. Clin. Invest.* 106, 339–347. doi: 10.1172/JCI9351

Bellanger, X., Payot, S., Leblond-Bourget, N., and Guédon, G. (2014). Conjugative and mobilizable genomic islands in bacteria: evolution and diversity. *FEMS Microbiol. Rev.* 38, 720–760. doi: 10.1111/1574-6976.12058

Bertin, P. N., Heinrich-Salmeron, A., Pelletier, E., Goulhen-Chollet, F., Arsène-Ploetze, F., Gallien, S., et al. (2011). Metabolic diversity among main microorganisms inside an arsenic-rich ecosystem revealed by meta- and proteo-genomics. *ISME J.* 5, 1735–1747. doi: 10.1038/ismej.2011.51

Brake, S. S., Dannelly, H. K., and Conners, K. A. (2001a). 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. doi: 10.1007/s002540000181

Brake, S. S., Dannelly, H. K., Connors, K. A., and 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. *Appl. Geochem.* 16, 1641–1652. doi: 10.1016/S0883-2927(01)00033-6

Brake, S. S., and Hasiotis, S. T. (2010). Eukaryote-dominated biofilms and their significance in acidic environments. *Geomicrobiol. J.* 27, 534–558. doi: 10.1080/01490451003702966

Bruneel, O., Personné, J.-C., Casiot, C., Leblanc, M., Elbaz-Poulichet, F., Mahler, B. J., et al. (2003). Mediation of arsenic oxidation by *Thiomonas* sp. in acid-mine drainage (Carnoulès, France). *J. Appl. Microbiol.* 95, 492–499. doi: 10.1046/j.1365-2672.2003.02004.x

Bruneel, O., Duran, R., Casiot, C., Elbaz-Poulichet, F., and Personné, J.-C. (2006). Diversity of microorganisms in Fe-As-rich acid mine drainage waters of Carnoulès, France. *Appl. Environ. Microbiol.* 72, 551–556. doi: 10.1128/AEM.72.1.551-556.2006

Bruneel, O., Volant, A., Gallien, S., Chaumande, B., Casiot, C., Carapito, C., et al. (2011). Characterization of the active bacterial community involved in natural attenuation processes in arsenic-rich creek sediments. *Microb. Ecol.* 61, 793–810. doi: 10.1007/s00248-011-9808-9

Bryan, C. G., Marchal, M., Battaglia-Brunet, F., Kugler, V., Lemaitre-Guillier, C., Lièvremont, D., et al. (2009). Carbon and arsenic metabolism in *Thiomonas* strains: differences revealed diverse adaptation processes. *BMC Microbiol.* 9:127. doi: 10.1186/1471-2180-9-127

Casiot, C., Egal, M., Bruneel, O., Verma, N., Parmentier, M., and Elbaz-Poulichet, F. (2011). Predominance of aqueous Tl(I) species in the river system downstream from the abandoned Carnoulès mine (Southern France). *Environ. Sci. Technol.* 45, 2056–2064. doi: 10.1021/es102064r

Casiot, C., Morin, G., Juillot, F., Bruneel, O., Personné, J. C., Leblanc, M., et al. (2003). Bacterial immobilization and oxidation of arsenic in acid mine drainage (Carnoulès creek, France). *Water Res.* 37, 2929–2936. doi: 10.1016/S0043-1354(03)00080-0

Cheng, H., Hu, Y., Luo, J., Xu, B., and Zhao, J. (2009). Geochemical processes controlling fate and transport of arsenic in acid mine drainage (AMD) and natural systems. *J. Hazard. Mater.* 165, 13–26. doi: 10.1016/j.jhazmat.2008.10.070

Denef, V. J., Mueller, R. S., and Banfield, J. F. (2010). AMD biofilms: using model communities to study microbial evolution and ecological complexity in nature. *ISME J.* 4, 599–610. doi: 10.1038/ismej.2009.158

Duquesne, K., Lebrun, S., Casiot, C., Bruneel, O., Personné, J.-C., Leblanc, M., et al. (2003). Immobilization of arsenite and ferric iron by *Acidithiobacillus ferrooxidans* and its relevance to acid mine drainage. *Appl. Environ. Microbiol.* 69, 6165–6173. doi: 10.1128/AEM.69.10.6165-6173.2003

Duquesne, K., Lieutaud, A., Ratouchniak, J., Muller, D., Lett, M.-C., and Bonnefoy, V. (2008). Arsenite oxidation by a chemoautotrophic moderately acidophilic *Thiomonas* sp.: from the strain isolation to the gene study. *Environ. Microbiol.* 10, 228–237. doi: 10.1111/j.1462-2920.2007.01447.x

Egal, M., Casiot, C., Morin, G., Elbaz-Poulichet, F., Cordier, M. A., and Bruneel, O. (2010). An updated insight into the natural attenuation of as concentrations in Reigous Creek (southern France). *Appl. Geochem.* 25, 1949–1957. doi: 10.1016/j.apgeochem.2010.10.012

Fadrus, H., and Malý, J. (1975). Suppression of iron(III) interference in the determination of iron(II) in water by the 1,10-phenanthroline method. *Analyst* 100, 549–554. doi: 10.1039/an9750000549

Freel, K., Krueger, M., Farasin, J., Brochier-Armanet, C., Barbe, V., Andres, J., et al. (in press). Adaptation in toxic environments: arsenic genomic islands in the bacterial genus *Thiomonas*. *PLoS ONE*.

Furrer, G., Phillips, B. L., Ulrich, K.-U., Pöthig, R., and Casey, W. H. (2002). The origin of aluminum flocs in polluted streams. *Science* 297, 2245–2247. doi: 10.1126/science.1076505

Halter, D., Goulhen-Chollet, F., Gallien, S., Casiot, C., Hamelin, J., Gilard, F., et al. (2012). *In situ* proteo-metabolomics reveals metabolite secretion by the acid mine drainage bio-indicator, *Euglena mutabilis*. *ISME J.* 6, 1391–1402. doi: 10.1038/ismej.2011.198

Halter, D., Casiot, C., Heipieper, H. J., Plewniak, F., Marchal, M., Simon, S., et al. (2011). Surface properties and intracellular speciation revealed an original adaptive mechanism to arsenic in the acid mine drainage bio-indicator *Euglena mutabilis*. *Appl. Microbiol. Biotechnol.* 93, 1735–1744. doi: 10.1007/s00253-011-3493-y

Johnson, C. A., and Amarasekare, P. (2013). Competition for benefits can promote the persistence of mutualistic interactions. *J. Theor. Biol.* 328, 54–64. doi: 10.1016/j.jtbi.2013.03.016

Jones, E. I., Bronstein, J. L., and Ferrière, R. (2012). The fundamental role of competition in the ecology and evolution of mutualisms. *Ann. N.Y. Acad. Sci.* 1256, 66–88. doi: 10.1111/j.1749-6632.2012.06552.x

Juhas, M., van der Meer, J. R., Gaillard, M., Harding, R. M., Hood, D. W., and Crook, D. W. (2009). Genomic islands: tools of bacterial horizontal gene transfer and evolution. *FEMS Microbiol. Rev.* 33, 376–393. doi: 10.1111/j.1574-6976.2008.00136.x

Kanamori, T., Kanou, N., Atomi, H., and Imanaka, T. (2004). Enzymatic characterization of a prokaryotic urea carboxylase. *J. Bacteriol.* 186, 2532–2539. doi: 10.1128/JB.186.9.2532-2539.2004

Kara, F., and Sahin, G. (2000). Hydrated aluminium sulfate precipitation by enzyme-catalysed urea decomposition. *J. Eur. Ceram. Soc.* 20, 689–694. doi: 10.1016/S0955-2219(99)00202-2

Klein, R., Tischler, J. S., Mühling, M., and Schlömann, M. (2014). Bioremediation of mine water. *Adv. Biochem. Eng. Biotechnol.* 144, 109–172. doi: 10.1007/10_2013_265

Koechler, S., Farasin, J., Cleiss-Arnold, J., and Arsène-Ploetze, F. (2015). Toxic metal resistance in biofilms: diversity of microbial responses and their evolution. *Res. Microbiol.* doi: 10.1016/j.resmic.2015.03.008. [Epub ahead of print].

Lugan, R., Niogret, M.-F., Leport, L., Guégan, J.-P., Larher, F. R., Savouré, A., et al. (2010). Metabolome and water homeostasis analysis of *Thellungiella salsuginea* suggests that dehydration tolerance is a key response to osmotic stress in this halophyte. *Plant J. Cell Mol. Biol.* 64, 215–229. doi: 10.1111/j.1365-313X.2010.04323.x

Maillot, F., Morin, G., Juillot, F., Bruneel, O., Casiot, C., Ona-Nguema, G., et al. (2013). Structure and reactivity of As(III)- and As(V)-rich schwertmannites and amorphous ferric arsenate sulfate from the Carnoules acid mine drainage, France: comparison with biotic and abiotic model compounds and implications for As remediation. *Geochim. Cosmochim. Acta* 104, 310–329. doi: 10.1016/j.gca.2012.11.016

Marchal, M., Briandet, R., Halter, D., Koechler, S., Dubow, M. S., Lett, M.-C., et al. (2011). Subinhibitory arsenite concentrations lead to population dispersal in *Thiomonas* sp. *PLoS ONE* 6:e23181. doi: 10.1371/journal.pone.0023181

Milde, K., Sand, W., Wolff, W., and Bock, E. (1983). *Thiobacilli* of the corroded concrete walls of the Hamburg sewer system. *J. Gen. Microbiol.* 129, 1327–1333. doi: 10.1099/00221287-129-5-1327

Mobley, H. L. T. (2001). "Urease," in *Helicobacter Pylori: Physiology and Genetics*, eds H. L. Mobley, G. L. Mendz, and S. L. Hazell (Washington, DC: ASM Press). Available online at: http://www.ncbi.nlm.nih.gov/books/NBK2417/ (Accessed July 1, 2015).

Morin, G., Juillot, F., Casiot, C., Bruneel, O., Personné, J.-C., Elbaz-Poulichet, F., et al. (2003). Bacterial formation of tooeleite and mixed arsenic(III) or arsenic(V)-iron(III) gels in the Carnoulès acid mine drainage, France. A XANES, XRD, and SEM study. *Environ. Sci. Technol.* 37, 1705–1712. doi: 10.1021/es025688p

Sigg, L., Black, F., Buffle, J., Cao, J., Cleven, R., Davison, W., et al. (2006). Comparison of analytical techniques for dynamic trace metal speciation in natural freshwaters. *Environ. Sci. Technol.* 40, 1934–1941. doi: 10.1021/es051245k

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

Slyemi, D., Moinier, D., Brochier-Armanet, C., Bonnefoy, V., and Johnson, D. B. (2011). Characteristics of a phylogenetically ambiguous, arsenic-oxidizing *Thiomonas* sp., *Thiomonas arsenitoxydans* strain 3As(T) sp. nov. *Arch. Microbiol.* 193, 439–449. doi: 10.1007/s00203-011-0684-y

Slyemi, D., Moinier, D., Talla, E., and Bonnefoy, V. (2013). Organization and regulation of the arsenite oxidase operon of the moderately acidophilic and facultative chemoautotrophic *Thiomonas arsenitoxydans*. *Extremophiles* 17, 911–920. doi: 10.1007/s00792-013-0573-1

Stingl, K., Altendorf, K., and Bakker, E. P. (2002). Acid survival of *Helicobacter pylori*: how does urease activity trigger cytoplasmic pH homeostasis? *Trends Microbiol.* 10, 70–74. doi: 10.1016/s0966-842x(01)02287-9

Sullivan, M. J., Petty, N. K., and Beatson, S. A. (2011). Easyfig: a genome comparison visualizer. *Bioinformatics* 27, 1009–1010. doi: 10.1093/bioinformatics/btr039

Vallenet, D., Belda, E., Calteau, A., Cruveiller, S., Engelen, S., Lajus, A., et al. (2013). MicroScope–an integrated microbial resource for the curation and comparative analysis of genomic and metabolic data. *Nucleic Acids Res.* 41, D636–D647. doi: 10.1093/nar/gks1194

Vallenet, D., Engelen, S., Mornico, D., Cruveiller, S., Fleury, L., Lajus, A., et al. (2009). MicroScope: a platform for microbial genome annotation and comparative genomics. *Database (Oxford)* 2009:bap021. doi: 10.1093/database/bap021

Vallenet, D., Labarre, L., Rouy, Z., Barbe, V., Bocs, S., Cruveiller, S., et al. (2006). MaGe: a microbial genome annotation system supported by synteny results. *Nucleic Acids Res.* 34, 53–65. doi: 10.1093/nar/gkj406

Vernikos, G. S., and Parkhill, J. (2006). Interpolated variable order motifs for identification of horizontally acquired DNA: revisiting the *Salmonella* pathogenicity islands. *Bioinformatics* 22, 2196–2203. doi: 10.1093/bioinformatics/btl369

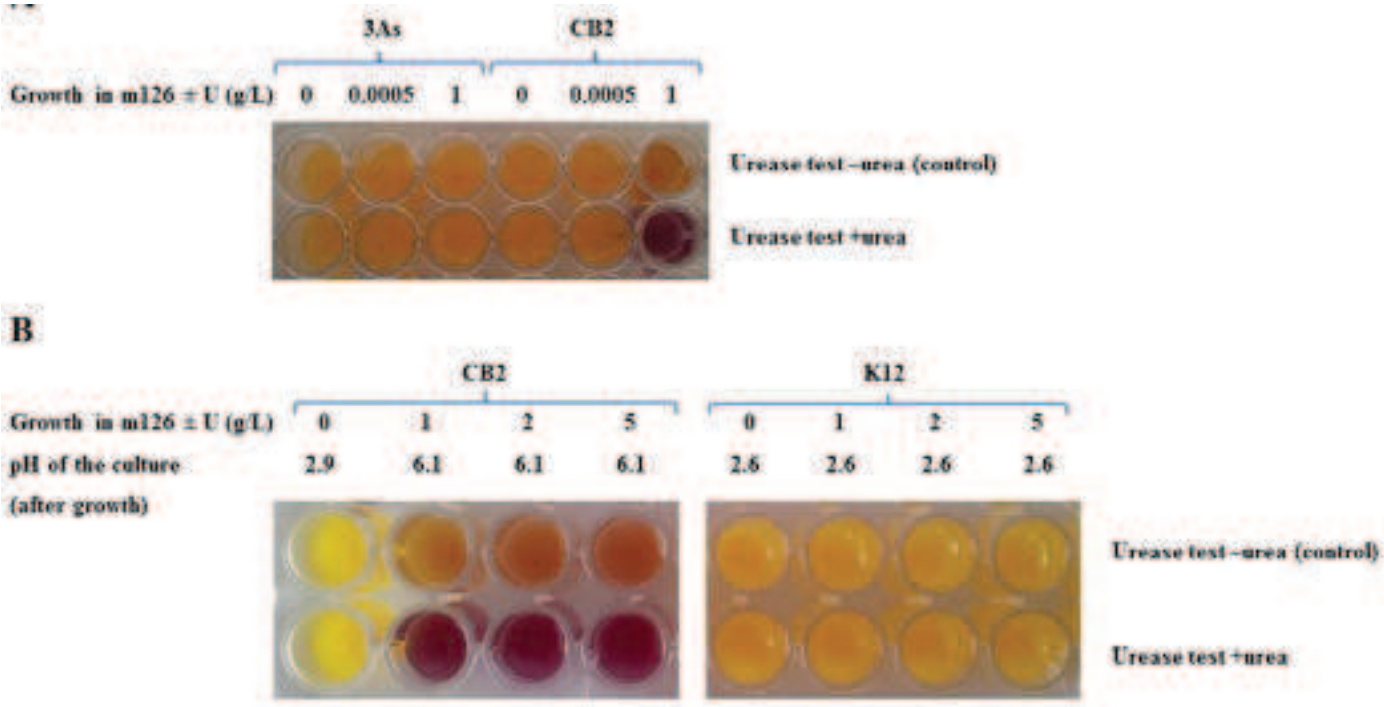
Volant, A., Desoeuvre, A., Casiot, C., Lauga, B., Delpoux, S., Morin, G., et al. (2012). Archaeal diversity: temporal variation in the arsenic-rich creek sediments of Carnoulès Mine, France. *Extremophiles* 16, 645–657. doi: 10.1007/s00792-012-0466-8

Waack, S., Keller, O., Asper, R., Brodag, T., Damm, C., Fricke, W. F., et al. (2006). Score-based prediction of genomic islands in prokaryotic genomes using hidden Markov models. *BMC Bioinformatics* 7:142. doi: 10.1186/1471-2105-7-142

Wilmes, P., and Bond, P. L. (2009). Microbial community proteomics: elucidating the catalysts and metabolic mechanisms that drive the Earth's biogeochemical cycles. *Curr. Opin. Microbiol.* 12, 310–317. doi: 10.1016/j.mib.2009.03.004

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Note : Le "Supplementary table 1" est présenté en Annexe 3.



Supplementary figure 1 **. Urea degradation activity in Thiomonas strains.** *After growth of cell cultures in m126 supplemented without or with 0.0005, 1, 2 or 5 g.L^{-1} urea, urea degradation activity was detected using Bromo-cresol purple as described in Material and Methods after incubation at room temperature for 15 min. The purple color indicates urea degradation. In the negative control where no urea was added to the BCP reagent, the orange color stands for no activity observed (as indicated by "urease test – urea"). Urea degradation affected the pH of the media when cells were incubated in the synthetic medium m126 supplemented with urea.*

III - Conclusion

Les variations génomiques liées à des îlots sont à l'origine de différences fonctionnelles pouvant conférer de nouvelles capacités à certaines souches. Les gènes codant l'uréase chez CB2 sont situés sur un îlot génomique (îlot urée) mesurant 88,2 Kpb et accolé au RGP10 (contenant un opéron *aio*) évoqué dans la première partie du chapitre. Le RGP10 mesurant 91,2 Kpb, ces deux îlots constituent une région flexible avoisinant les 200 Kpb et ayant les caractéristiques d'un IME. Les remaniements génomiques au niveau de l'îlot urée de CB2, et des régions équivalentes des autres souches, sont probablement à l'origine des différences concernant leur activité uréase. La souche CB2 est la seule à posséder une activité uréase élevée, celle-ci étant absente chez CB3 et *Tm. intermedia* K12 et faible chez 3As, CB1 et CB6.

L'activité uréase de CB2 lui permet de maintenir le pH à ~6 et confère à cette souche une meilleure survie en milieu m126 supplémenté en urée. Elle entraîne également une augmentation du pH dans l'eau du DMA de Carnoulès supplémentée en urée et cette augmentation du pH accélère la précipitation du fer, de l'aluminium et de l'arsenic. La viabilité de CB2 dans l'eau du DMA n'est par contre pas augmentée par l'ajout d'urée. La seule précipitation des métaux et l'augmentation du pH ne confèrent donc visiblement pas d'avantages à cette souche dans les conditions testées en laboratoire. Les souches 3As, CB1 et CB6 ne possèdent qu'une légère activité uréase et CB3 n'en possède même aucune. La dégradation de l'urée n'est donc pas indispensable dans le DMA de Carnoulès. Ceci n'exclut pas que sur le site, dans un contexte d'interactions complexes avec d'autres organismes, la dégradation soit utile à la survie de *Tm.* sp. CB2 et d'autres souches de *Thiomonas*. Comme dans le cas des multiples copies de l'opéron *aioBA*, elle pourrait également conférer un avantage dans une micro-niche particulière.

Partie 3 - Biosynthèse de biofilm et mobilité

I - Introduction

La formation de biofilms et la mobilité sont deux processus très liés. Certaines voies de régulation sont communes et permettent aux bactéries de basculer d'un mode de vie fixé en biofilm à un mode de vie mobile planctonique. De plus, les flagelles responsables de la nage, sont souvent impliqués dans les premières étapes d'adhésion des biofilms. Une fois matures, les biofilms peuvent être très complexes tant d'un point de vue de leur architecture que de la population. Des sous-populations mobiles peuvent émerger dans le biofilm et se déplacer au sein de celui-ci ou bien se disséminer, et des cellules mobiles peuvent également envahir les biofilms (Houry et al., 2012).

La formation de biofilm et la mobilité peuvent également être considérées comme deux stratégies de survie différentes face à un stress. Une fois le danger détecté par les cellules, celles-ci peuvent soit utiliser la mobilité pour fuir vers de meilleures conditions, soit se fixer et former un biofilm pour se protéger. La formation de biofilm peut notamment constituer une réponse face aux métaux toxiques (Harrison et al., 2007).

Au cours de cette thèse, j'ai participé à la rédaction d'une revue portant sur l'impact des métaux toxiques sur la formation de biofilm et l'implication des biofilms dans la résistance aux métaux. Cette revue constitue la suite de cette introduction.

Institut Pasteur
Research in Microbiology xx (2015) 1−10
www.elsevier.com/locate/resmic

Toxic metal resistance in biofilms: diversity of microbial responses and their evolution

Sandrine Koechler, Julien Farasin, Jessica Cleiss-Arnold, Florence Arsène-Ploetze*

Laboratoire Génétique Moléculaire, Génomique et Microbiologie, UMR7156, CNRS-Université de Strasbourg, Département Microorganismes, Génomes, Environnement, Equipe Ecophysiologie Moléculaire des Microorganismes, Institut de Botanique, 28 Rue Goethe, 67083, Strasbourg, France

Received 29 January 2015; accepted 25 March 2015
Available online ■ ■ ■

Abstract

Since biofilms are an important issue in the fields of medicine and health, several recent microbiological studies have focused on their formation and their contribution to toxic compound resistance mechanisms. In this review, we describe how metals impact biofilm formation and resistance, and how biofilms can help cells resist toxic metals. First, the organic matrix acts as a barrier isolating the cells from many environmental stresses. Secondly, the metabolism of the cells changes, and a slowly-growing or non-growing sub-population of cells known as persisters emerges. Thirdly, in the case of multispecies biofilms, metabolic interactions are developed, allowing cells to be more persistent or to have greater capacity to survive than a single species biofilm. Finally, we discuss how the high density of the cells may promote horizontal gene transfer processes, resulting in the acquisition of new features. All these crucial mechanisms enable microorganisms to survive and colonize toxic environments, and probably accelerate ongoing evolutionary processes.

Keywords: Adaptation; Metal tolerance; Sequestration; Acid mine drainage; Persisters; Variants

1. Introduction

In many environments, populations of microorganisms concentrate at the solid/liquid or liquid/gas interfaces, resulting in formation of periphytons (biofilms forming on submerged solid surfaces), stromatolites (mushroom- and tower-shaped biofilms in quiescent waters) or aggregates of cells not attached to a solid surface (named flocs) or streamers (filamentous biofilms in flowing water), which have similar characteristics [1,2]. All these types of biofilms consist mainly of cells embedded in an organic matrix composed of hydrated biopolymers such as proteins (including enzymes), exopolysaccharides, glycolipids and extracellular DNA [3]. Considerable interest has focused during the last few years on the process of formation of surface-adherent communities and their dispersion, partly because of the advent of microscopic methods such as confocal microscopy and the development of new cell culture systems (which were recently reviewed in [1,3−6]). Briefly, planktonic bacteria first adhere to a biotic or abiotic surface; this first step often requires, but is not restricted to, the presence of flagella or pili (Fig. 1). Once they have become firmly attached, the cells migrate together using a process of twitching motility or recruit planktonic cells and grow to form microcolonies. Sessile bacteria excrete extracellular polymeric substances (EPS) forming a heterogeneous matrix interspersed with open water channels. The specific components of the matrix determine the architecture and its stability, and influence the functional efficiency of the biofilm, in particular, its adhesive properties, molecular exchange processes, matrix water content, charge and sorption properties [3]. The composition of the EPS depends on the microorganisms present and the environmental conditions, including the temperature and availability of nutrients. Cell lysis may be

* Corresponding author.
E-mail addresses: sandrine.koechler@unistra.fr (S. Koechler), farasin@unistra.fr (J. Farasin), cleiss.j@unistra.fr (J. Cleiss-Arnold), ploetze@unistra.fr (F. Arsène-Ploetze).

http://dx.doi.org/10.1016/j.resmic.2015.03.008
0923-2508/

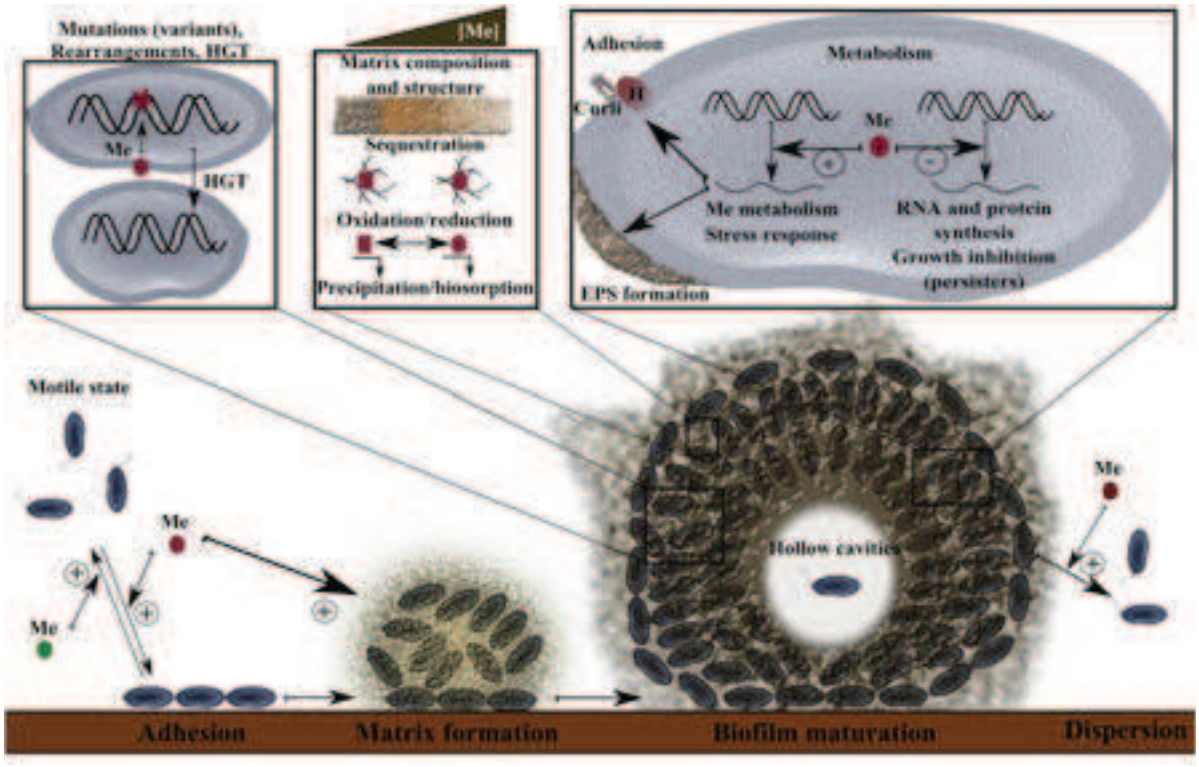


Fig. 1. Overview of biofilm formation and factors that contribute to toxic metal resistance or tolerance. The reduced sensitivity to toxic metals in biofilms is due to multiple factors, including changes in biofilm formation efficiency, matrix structure, metabolism and gene expression, as well as the onset of mutations. (Me: metal; H: hemagglutinin).

responsible for the formation of hollow cavities in a mature biofilm (Fig. 1). Other cells can be passively or actively (via a process of motility-driven dispersal) detached from the sessile community and disperse either in the fluid (via their swimming motility) or on the surface (by swarming on moist surfaces or via a process of gliding or twitching motility) and colonize other sites (Fig. 1). A link between the regulation of biofilm formation and swarming, swimming or twitching motility processes has been recently described and reviewed [1,7]. Biofilm formation is a coordinated developmental process based on cell-to-cell communications which involves several regulation processes such as quorum sensing (QS), two-component systems- or c-di-GMP-mediated regulation [2,7,8].

There exist several lines of evidence suggesting that biofilm formation protects cells from a wide range of environmental challenges [9–11]. In the field of microbial ecology, the role of biofilms in cell resistance to toxic compounds such as metal ions is of particular interest [2,10]. Several metal ions are toxic since they block many crucial cellular processes in all the domains of life [12,13]. In eukaryotic and prokaryotic cells, the toxicity of metals depends on whether or not they are essential metals, their speciation and their concentration [12,13]. Essential redox-active metals such as iron (Fe), copper (Cu), chromium (Cr) and cobalt (Co), which undergo cycling reactions involved in electron transfer processes, are toxic when they are present in excess, whereas non-essential metals such as arsenic (As), mercury (Hg), lead (Pb) and cadmium (Cd) are toxic even at very low concentrations. The mechanisms involved in the toxicity of metals have been recently reviewed [12,13]. Briefly, metals can directly or indirectly produce reactive radicals via Fenton and Haber-Weiss reactions, resulting in the formation of reactive oxygen and nitrogen species (ROS or RNS). This mechanism was observed in the case of Fe, As, Cr, tellurium (Te), Co, Hg, silver (Ag) and Cu. ROS, such as the hydroxyl radical ($^{\bullet}$OH) in particular, react with DNA, resulting in damaged bases or strand breaks, lipid peroxidation (leading indirectly to damaged bases) or protein modifications. In addition, metal toxicity may be linked to glutathione metabolism, which plays an important antioxidant role in cells and induces the chelation of several metals. The presence of excess As, Ag, Cd, Co, Cr or Te induces glutathione depletion, affecting the cellular redox balance. Metals interfere with protein activity by binding to thiols or other functional groups, catalyzing the oxidation of amino acid side chains, displacing essential metals ions in metalloproteins and interfering with folding processes and promoting protein aggregation [12–14]. Some metals can also inhibit cellular activity because they present a structural homology with enzyme substrates. This occurs, for example, with pentavalent arsenic (arsenate), which substitutes for phosphate used in many cellular processes including oxidative phosphorylation. Lastly, in microorganisms, some metals such as Ag and aluminium (Al) impair membrane function by interfering with nutrient uptake processes and disrupting the electron transport chain and/or the proton gradient force

Please cite this article in press as: Koechler S, et al., Toxic metal resistance in biofilms: diversity of microbial responses and their evolution, Research in Microbiology (2015), http://dx.doi.org/10.1016/j.resmic.2015.03.008

[12,13]. Despite such toxic effects, many microorganisms are able to colonize contaminated ecosystems because they develop resistance mechanisms, which can affect metal speciation, cause extrusion of metal out of the cell, efficiently repair macromolecules, circumvent metabolic pathways or sequester the metal within the structure of a biofilm [10,15].

In this review, we propose to report on the latest findings on how biofilms contribute to the adaptive bacterial mechanisms enabling cells to survive in extreme environments containing toxic metals. Data on the link between biofilm formation or dispersion and microbial toxic metal resistance processes will be presented. Biofilms are generally formed by communities containing several species [16], and recent findings on how multispecies biofilms promote resistance to toxic metals will also be presented.

2. Effect of metals on the formation and development of biofilms

Most studies on the role of bacterial biofilms in resistance to toxic metals were performed on preformed biofilms challenged with metals since, in most cases, bacterial biofilm formation was not possible at toxic metal concentrations above the planktonic minimum inhibitory concentration (MIC) [17,18]. In such studies, it was shown that once initiated, the biofilm provides a slight advantage in the presence of toxic metals, as the occurrence of cell death is generally postponed. However, and contrary to what is observed in the presence of antiobiotics, planktonic and biofilm cells are in most cases almost equally susceptible to killing by metal oxyanions and cations [17,19–21], i.e. their minimum biofilm-eradication concentration (MBEC) in biofilm is not higher than the minimum bactericidal concentration (MBC) observed under planktonic growth conditions. Growth in a bacterial biofilm therefore does not necessarily increase the resistance of the cells to toxic metals, but rather, endows them with greater time-dependent tolerance [19,20,22,23]. In some bacteria, cells in biofilm are even more susceptible to oxyanions than planktonic cells, as observed for tellurite in the case of *Staphylococcus aureus* ATCC 29213, where the MBC is higher than the MBEC [24]. Nevertheless, in few studies, it was possible to test the bacterial biofilm formation at inhibitory concentrations. For example, formation of biofilm was found to be promoted in *Rhizobium alamii* YAS34 by Cd exposure at inhibitory concentrations [25], and in the marine methanogen archaeon *Methanosarcina acetivorans* C2A when pre-adapted cells were exposed to Cd at concentrations which are very toxic for control cells [26]. *Candida tropicalis* 99916 yeast biofilms were more tolerant to toxic metals than planktonic cells, and the death of the cells in the biofilms was not time-dependent [27]. Thus, the effects of metals on biofilm depend on the bacterial strain, the metal and the growth environment (the composition of the media, carbon source and buffer conditions) [18,21,22]. Recent studies that more specifically reveal how metals can affect the various steps in the formation and development of biofilms will be described below.

2.1. Attachment and adherence

One protective strategy adopted by several bacteria when they are exposed to toxic metals consists in promoting the formation of a biofilm. Bacterial attachment is the first step in the development of a biofilm, as well as playing a key role in this process. Metals often have an impact on the adhesion of cells to the surface and/or cell-to-cell aggregation processes, which probably promote the formation of biofilms and enhance their resistance. However, the efficiency of the process of attachment and/or aggregation of cells seems to depend on the metal involved and its concentration. For example, increasing concentrations of Cd, even at inhibitory concentrations, induce the attachment of cells, and thus, biofilm formation in *R. alamii* YAS34 [25]. Several studies have also been performed on these processes in the case of *Xylella fastidiosa*. Increasing the Cu concentration (from 0 to 200 μM, <MIC) enhanced biofilm formation of the *X. fastidiosa* strain Temecula, as shown by crystal violet detection assays on microtiter plates, although higher concentrations (>200 μM) reduced the amount of biofilm formed. On the other hand, in studies using microfluid chambers, the addition of Cu (50 μM) promoted attachment of this bacterium and accelerated cell aggregation, while higher concentrations (400 μM, <MIC) inhibited aggregation and prevented formation of biofilm [28]. Subinhibitory manganese (Mn) and Zn concentrations affect cell-to-cell attachment in *X. fastidiosa* strain Temecula, but Mn increases the process of biofilm formation in this bacterium, whereas zinc impairs this process, probably by decreasing the surface-cell attachment process [28,29]. In another bacterium, *X. fastidiosa* 9a5c, the responses of the biofilm to Cu and tetracycline were compared. In this case, the biofilm was found to be less susceptible than planktonic cells to Cu. The biofilms incubated in the presence of inhibitory concentrations of tetracycline seemed to be thinner than those formed in the presence of inhibitory concentrations of Cu. In line with these results, it was shown that the gene and protein expressions of hemagglutinins, which are known to be involved in cell-to-cell adhesion processes in *X. fastidiosa*, were greater in response to Cu than to tetracycline [30]. Cu may therefore speed up the settling rate and the process of biofilm formation in the case of both *X. fastidiosa* 9a5c and Temecula strains. Some *Escherichia coli* K-12 strains formed a biofilm in response to subinhibitory nickel (Ni) concentrations. In this biofilm, cell growth was found to be less affected by metal exposure than under planktonic conditions [31]. In addition, the adherence properties and the process of biofilm formation were increased and transcription of the adhesive curli encoding genes was induced in response to Ni [31]. All in all, these findings suggest that bacterial cells exposed to metals generally respond by promoting adhesion processes, and hence, formation and maintenance of biofilms (Fig. 1). However, some exceptions to this strategy have been observed, and in other cases, the switch between the planktonic and biofilm ways of life was prevented in the presence of toxic metals. Indeed, biofilm formation may compete with flagellar motility in bacteria, and in some cases, the motility

Please cite this article in press as: Koechler S, et al., Toxic metal resistance in biofilms: diversity of microbial responses and their evolution, Research in Microbiology (2015), http://dx.doi.org/10.1016/j.resmic.2015.03.008

strategy is favored rather than biofilm formation in the presence of toxic metals (Fig. 1). In *Herminiimonas arsenicoxydans* ULPAs1 and *Rhizobium* sp. NT26, for example, As(III) induces swimming motility, delays biofilm initiation [32] and reduces the amount of biofilm formed [33]. Bacteria may therefore have to choose between the sessile and planktonic lifestyles in order to either embed the cells in a protective matrix or escape harsh conditions and reach more favorable niches (Fig. 1).

2.2. Changes in biofilm structure and maturation

In addition to impacting cell adhesion, heavy metal exposure can lead to structural changes in the biofilm matrix (Fig. 1). The formation of a stiffer or more compact biofilm that is more resistant to erosion by shear forces has been observed, for example, at concentrations higher than the MIC for Fe^{2+} and higher than the MBC for Al^{3+}, Cu^{2+}, Fe^{3+}, Zn^{2+} in *Bacillus subtilis* B1, and at subinhibitory concentration of Zn^{2+} in *X. fastidiosa* strain Temecula [29,34]. The amount of extracellular matrix increases considerably in *Thiomonas* sp. CB2 at subinhibitory concentrations of As(III), thus probably improving the integrity of the biofilm [35]. It has been established that the complex 3-D development of the *Tm.* sp. CB2 biofilm contributes to its physiological heterogeneity and to the development of a subpopulation of cells that are remobilized in a rapid process of dispersal [35] (Fig. 1). In this case, the exopolysaccharides contribute to enhancing the complex 3-D architecture of the biofilm matrix in response to As(III) exposure. In the case of *R. alamii* YAS34, the biofilm is more condensed in the presence of subinhibitory concentrations of Cd than in the absence of this metal [25]. Metal-specific changes in structure and hydrodynamic properties have also been observed in *Shewanella oneidensis* MR-1 biofilm exposed to either chromate Cr(VI) or uranyl U(VI), resulting in specific spatial stratifications [36]. In yeast strains, metals affect the biofilm structure and influence the 3-D structure in a similar way via the cell differentiation process from yeast to hyphal cells in *C. tropicalis* 99916 and *Candida albicans* 3153A. Subinhibitory concentrations of Zn^{2+}, Co^{2+}, Cu^{2+}, Ag^{+}, Cd^{2+}, Hg^{2+}, AsO_2^- and SeO_3^{2-} abolish hyphal formation in both strains, whereas CrO_4^{2-} induces a mycelial structure in *C. tropicalis* 99916 [37]. All these findings suggest that exposure to metals results in a spatial restructuring of the biofilm, which may help the microorganisms to acquire greater tolerance to toxic metals. It has been hypothesized that these structural changes may affect the metals' bioavailability, cellular nutrient acquisition processes and the metabolic activities described in the next section.

3. Role of biofilms in toxic metal resistance

Once a biofilm has been established, it will impact the cells either via a passive process and/or by influencing the pattern of gene expression, resulting in the specific resistance or tolerance mechanisms described below.

3.1. Sequestration and immobilization within the biofilm matrix

It has been clearly established that the extracellular polymeric matrix acts as a barrier to toxic metals, in which they can be sequestered, immobilized, mineralized and/or precipitated [1,10,38] (Fig. 1). Some toxic metals bind to functional groups of EPS matrix components (such as the carboxyl, phosphoryl, hydroxyl and amino groups) [38]. The combination of positive and negative charges in the matrix restricts diffusion of metal ions because the number of electrostatic binding sites available for biosorption of metals is 20- to 30-fold greater in extracellular polymers than in planktonic cells [39]. This biosorption of toxic metals results in their sequestration in the extracellular matrix and may therefore decrease their bioavailability, retard their diffusion and thus reduce cell exposure and increase bacterial resistance [22,34,40]. This process may protect the bacteria until the matrix becomes saturated. Indeed, in agreement with this idea, a gradient in the distribution of living cells has been observed in *Pseudomonas aeruginosa* PAO1 biofilm exposed to subinhibitory concentrations of Cu. In this case, death of the cells occurred preferentially on the outer surface and periphery of the biofilm, which is more exposed to Cu, whereas the number of living cells was greater in the bottom layers of the biofilm where cells are probably not affected by the metal [41]. Increasing the amounts of EPS or changing the carbohydrate composition of biofilms exposed to metal ions in comparison with non-exposed cells would probably improve the metal trapping process [29,42,43]. This process has been mainly described in the case of cations, since EPS are often anionic [22,27]. It has been reported that in biofilm matrices of bacterium *E. coli* JM109 and yeast *C. tropicalis* 99916, metal oxyanions have lower binding affinity than cations and may therefore diffuse more efficiently [17,27]. In *Pseudomonas putida* ATCC 33015, sugars are present in great amounts in the biofilm matrix exposed to chromium, which probably facilitates cation binding processes. In addition, large amounts of extracellular DNA (eDNA) were found to bind to Cr(III) and contribute to metal biostabilization [43]. In *X. fastidiosa* 9a5c, the amount of exopolysaccharides able to form a complex with Cu changes depending on the growth medium [44]. Likewise, *Pseudomonas pseudoalcaligenes* KF707 biofilms grown in minimal medium were able to counteract metals better than those obtained in rich medium ($MBEC_{minimal\ media} > MBEC_{rich\ media}$) [21]; whereas *E. coli* JM109 was less susceptible to metals in rich than in minimal medium ($MBEC_{rich\ media} > MBEC_{minimal\ media}$) [22]. These findings show that both the environmental conditions and the quantity of EPS influence the biosorption behavior and the sequestration efficiency of biofilms.

Mechanisms involving biofilms that influence metal speciation or precipitation processes have been described. Changes in the oxidative state may favor metal precipitation and render metals less toxic, which enables the cells to tolerate higher metal concentrations than planktonic cells (Fig. 1). Several studies have established that inorganic compounds

Please cite this article in press as: Koechler S, et al., Toxic metal resistance in biofilms: diversity of microbial responses and their evolution, Research in Microbiology (2015), http://dx.doi.org/10.1016/j.resmic.2015.03.008

such as sulfides and phosphates promote metal depletion from a solution by concentrating them, which results in the formation of insoluble precipitates in the matrix [10,38,45]. These precipitates can be seen to form finely grained minerals when observed microscopically, as in the case of electro-dense cadmium and sulfur grains trapped throughout the biofilm matrix of *M. acetivorans* C2A [26]. Similar selenium−sulfur granules were observed in the biofilm matrix of a *Desulfomicrobium norvegicum*-like strain (Dundee isolate 1). In this case, it was established that the bacterium produced sulfide, which promoted abiotic reduction of selenite at subinhibitory concentrations [46]. On the other hand, the immobilization and modification of metal in the biofilm matrix may be linked to an active extracellular enzymatic reduction process possibly occurring in the presence of a suitable electron donor. For example, the biofilm of *Geobacter sulfurreducens* PCA showed enhanced capacity to immobilize and reduce U(VI) to the less mobile form U(IV) in comparison with planktonic cells. This biofilm contains proteins involved in electron transfer, i.e. outer membrane c-type cytochromes, and electrically conductive pili, which immobilize and reduce uranium in the top stratum of the biofilm. This process is largely dependent on the size of the metalloid-exposed surface of the biofilm [47]. Another biomineralization process has been described in *Cupriavidus metallidurans* CH34, which is able to form gold (Au) nanoparticles in biofilms through an active mechanism of reduction and precipitation of toxic Au(III)-complexes [48]. All these processes are tightly regulated and require a switch in the pattern of gene expression as well as changes in the general metabolism of the cells, as described in the following section.

3.2. Biofilm-specific gene induction and metabolic changes in response to toxic metals

Changes in gene expression and in metabolism are thought to be involved in biofilm-mediated metal tolerance. Three main changes have been described to date. First, it has been reported in two studies that cells under metal exposure favored the synthesis of EPS, which meant that they would be protected in the biofilm from exposure to toxic metals during their growth (Fig. 1, see also part 3.1). A shift in the exopolysaccharide-related metabolism was observed in *Pseudomonas fluorescens* ATCC 13525 biofilm metabolome exposed to a subinhibitory concentration of Cu ions using both GC-MS and H_NMR approaches, which resulted in an increase in the matrix production, whereas metabolism corresponding to an oxidative stress response was favored in planktonic cells exposed to Cu [49]. Likewise, the expression of genes involved in exopolysaccharide synthesis increased in *X. fastidiosa* strain Temecula cells exposed to subinhibitory concentrations of Zn in the biofilm as compared to non-exposed cells [29]. Secondly, one metabolic pathway can sometimes be favored versus another during biofilm formation. In *Desulfovibrio desulfuricans* strain ND132, for example, the process of Hg methylation was faster in the biofilm than in unattached bacteria. In addition, the process of methylation in planktonic cultures was found to be independent of the acetyl-CoA pathway, whereas this pathway seems to play an important role in the methylation processes occurring within the biofilm [50]. In addition, *P. pseudoalcaligenes* KF707 biofilm and planktonic cells adopt different strategies for counteracting toxic metals: the cells present in biofilms are resistant to selenite and tellurite, but contrary to what is observed with planktonic cells, this resistance is not linked to the reduction of these metals [21]. Thirdly, a microarray approach was used to investigate the response to Cu of the plant-pathogen *X. fastidiosa* 9a5c biofilm and compare it with the response to tetracycline. In these experiments performed with subinhibitory and inhibitory Cu concentrations, a dose-dependent response mechanism specific to this metal was observed. The number of regulated genes involved in similar biological processes increases with increasing Cu concentrations. The genes induced in this way included phage-related genes, pathogenic, virulence and adaptation genes, whereas the genes repressed included genes involved in RNA and protein synthesis [30]. Thus, phage-related activity and biofilm formation seem to represent a specific response to Cu in this *X. fastidiosa* strain, whereas the shutdown of DNA replication and protein translation, as well as the repression of movement and the induction of toxin-antitoxin systems, seem to comprise a common response to both Cu and tetracyclsine in *X. fastidiosa* biofilm (Fig. 1). These findings suggest that bacterial cells may reduce their metabolism and enter a resistant physiological state when they are exposed to toxic compounds [30]. A dormant state was previously described in studies on the responses of other bacteria to antibiotics, and these cells were termed persister cells.

3.3. Persister cells formation

A mechanism commonly associated with antibiotic resistance is the formation of persister cells, a subpopulation of cells that neither grow nor die and survive prolonged exposure to toxic compounds [51−53]. Only a few examples of persister cells able to survive exposure to metals in biofilms have been described, including *P. aeruginosa* ATCC 27853 exposed to metal cations at concentrations higher than the MBC [20], and *E. coli* strain JM109 exposed to oxyanions at concentrations slightly higher than the MIC [17]. This dormant state differs from the acquisition of resistance, since the MBC of these cells does not increase, and it was suggested that these persister cells may contribute to the time-dependent tolerance described in section 2 [17,20]. In these studies, it was observed that these cells were able to grow again after metal removal and show normal susceptibility to it [17,20]. Similarly, *X. fastidiosa* strain Temecula may be able to enter a viable but non-cultivable state at subinhibitory concentration of Zn, at which growth is slow [29]. In *X. fastidiosa* 9a5c, the presence of persister cells (accounting for 0.05% of the population) was observed in response to inhibitory concentrations of both tetracycline and Cu [30] and the number of these persister cells increased when the population was pretreated with a subinhibitory concentration of Cu. The expression of toxin−antitoxin related genes, known to be associated with the rate

Please cite this article in press as: Koechler S, et al., Toxic metal resistance in biofilms: diversity of microbial responses and their evolution, Research in Microbiology (2015), http://dx.doi.org/10.1016/j.resmic.2015.03.008

of occurrence of persister cells in response to antibiotics in *E. coli* [52,53], was induced in *X. fastidiosa* 9a5c biofilm exposed to Cu. All in all, these findings show that, as in the case of antibiotics, some bacteria are able to enter a dormant state in the presence of toxic metals, awaiting better growth conditions.

4. Metals and multispecies biofilms

The existence of biodiversity is crucial to any ecosystem, and this is even more so under fluctuating environmental conditions, where it may serve to maintain the functional integrity of ecosystems. Several studies demonstrate that the presence of toxic metals can have either negative [54] or positive effects on functional relationships at work in microbial multispecies biofilms [55]. On the other hand, the presence of several species in biofilms may favor the functioning of the communities and their resistance to toxic metals. For example, it was demonstrated that in the presence of Cd [56], a single-species biofilm accumulated only 25–30% of the amount of Cd accumulated by a multispecies biofilm. It was suggested that the low amounts of Cd accumulated by the single-species biofilm may be linked to its poor capacity to grow and produce extracellular Cd-binding components, since a higher amount of both proteins and EPS accumulated in the multispecies biofilm than in the single-species biofilm [57]. In another study, a microbial community showed a two-to threefold higher Cu ion uptake capacity than that of single-species cultures [58]. Likewise, the mixed cyanobacterial and bacterial biofilms removed Cr(VI) from the aqueous phase more effectively than each strain separately [59]. In a third study, experimental bioreactors inoculated with a single resistant strain and fed with sterile chlorine/alkali wastewater collapsed when the influent Hg concentrations increased, whereas stable performances were achieved using a mixture of strains [60]. This study reveals that stable, efficient bioreactor performances are facilitated even under variable influent conditions by the presence of a mixed Hg-resistant community.

Biomineralization seems to be an important process for metal detoxification in natural biofilms. A central role has indeed been demonstrated for *C. metallidurans* CH34, a bacterium that dominates the communities of biofilms on Au grains in Australian mines, in the detoxification of Au complexes [48,61] (See also part 3.1). Recently, multispecies biofilms grown from an oil sand tailings pond in Northern Alberta, Canada were found to be extremely resistant to several metals (up to 20 mg L^{-1} of Pb, 16 mg L^{-1} of Zn, 1000 mg L^{-1} of Sr and 3.2 mg L^{-1} of Ni). In this community, a process of Cu, Ag, Pb and Sr mineralization was observed on the surface of the biofilms where a high level of bacterial species complexity was maintained. Interestingly, little change in the microbial communities exposed to silver, lead or strontium was observed in comparison with stress-free multispecies biofilm [62]. *Pseudomonas*, *Ensifer*, *Azospirillum* and species belonging to the *Rhizobiaceae* family were the main members of this community. One species, *Rhodococcus erythropolis*, was isolated from this community and found to be less tolerant to lead and strontium than the other isolates, and it did not biomineralize these metals. In contrast to what is observed in the presence of Pb and Sr, the viability of *R. erythropolis* biofilm did not decrease in the presence of Cu as it did in a multispecies biofilm, which suggests that this species is fairly Cu-resistant [62]. These observations indicate that its resistance depended on the metals but also on the presence of the other species. It should be noted that the structure and diversity of bacterial communities of this kind do not depend on toxic metal concentrations alone, but also, as mentioned above in the case of single-species biofilms, on the temperature [63], the pH [64], the organic content [65] and interactions between these factors.

The development and persistence of biofilms are also affected not only by their environment, but also by the variety of species present. Some recent studies have been performed with a view to understanding how toxic metals may influence the diversity of the biofilm community. Chodak et al. [66] reported that a high heavy metal content in forest soil had a negative influence on bacterial activity and diversity, suggesting that community members do not necessarily all benefit equally from belonging to a multispecies biofilm. Although a loss of biodiversity within the biofilm has been observed in some studies, positive effects of heavy metal exposure on biodiversity have been frequently reported. High levels of biodiversity were detected in rocks and mines, where the typical number of OTUs amounted to almost 100 [67–70]. In one study on a rock biofilm from a former gold and arsenic mine in Zloty Stok (Poland), a high level of microbial diversity (more than 190 OTUs) dominated by Rhizobiales (*α-Proteobacteria*) (60 mg kg^{-1} of dry mass) was observed [71]. Several organisms potentially involved in transformation of arsenic, iron-oxidizing/reducing microorganisms, methanotrophs and many microorganisms involved in the nitrogen cycle were detected in the biofilm originating from this mine. The authors presented an "arsenium commensalism" model, in which they speculated that the majority of the bacteria in the biofilm community may gain energy from sources other than arsenic compounds, but oxidize arsenic and produce the less toxic oxidized form, arsenate. The reduction of arsenate is then catalyzed by *Shewanella* and used to promote its own growth [71].

More generally, microbial diversity yields a reservoir of strains with complementary ecological niches resulting in high levels of resistance to toxic metal conditions. It seems that the driving forces responsible for bacterial community development in toxic metal environments are self-organization and cooperation rather than competition between individual microorganisms. Multispecies biofilms probably constitute a favorable bacterial habitat under toxic metal stress conditions, providing protection and the possibility of engaging in mutually beneficial cooperative interactions. Finally, species diversity within a biofilm constitutes an important means of dealing with stress and variable conditions by producing an "insurance effect" [72]. The "insurance" hypothesis is based on the assumption that different species or strains will respond differently to environmental conditions: some bacteria may tolerate various kinds of abiotic perturbations more than others and their growth rates may be less severely affected [73–75].

Please cite this article in press as: Koechler S, et al., Toxic metal resistance in biofilms: diversity of microbial responses and their evolution, Research in Microbiology (2015), http://dx.doi.org/10.1016/j.resmic.2015.03.008

Altogether, the very small changes in the microbial communities observed in response to As, Hg, Zn, Pb, Cr or Cd challenge in all the studies described above indicate that the abundance of more resistant species is not enhanced by the presence of these compounds. They tend, rather, to indirectly promote growth of multispecies biofilms because these metals induce resistance and/or tolerance upon its members when they are exposed to toxic metal stress. This finding may be partly attributable to differences in the genetic potential of this set of microorganisms, which allow the coexisting members of the community to withstand and sometimes take advantage of increasing environmental stresses such as toxic metals.

Since more than a decade, several studies have been performed on the biofilm communities found in acid mine drainages (AMDs) at the Richmond Mine (California), which contains extremely high As, Cd, Cu and Zn concentrations [76]. It was demonstrated that a biofilm formed by several microorganisms (bacteria, archaea, protist and fungi) at the air-solution interface stimulates pyrite dissolution to form an acidic and metal-rich solution. At various stages of this biofilm development, metabolic roles have been assigned to different microorganisms or even populations of closely-related bacteria or archaea [76]. In particular, it was demonstrated that several closely-related *Leptospirillum* co-exist and play crucial roles, expressing different proteins and therefore having distinct ecological roles. More generally, fine-scale heterogeneity was observed in these biofilms and unexpected high diversity was recently demonstrated in these ecosystems [77–79]. Environmental genomics performed on these biofilms have led to studies of microbial evolution at the community level and demonstrate that such bacteria and archaea have evolved over a relatively short time scale (a few years) by mutation, migration and genome rearrangement, but also horizontal gene transfer (HGT) involving plasmids or phages [76,78–80]. Such mechanisms probably enable maintaining the high diversity and heterogeneity observed. A comparative study on several metagenomics data suggest that microorganisms found in extreme environments evolve faster, and the frequency of HGT is higher than in those found in non-extreme environment [81]. Interestingly, an enrichment in genes involved in recombination, replication and DNA repair was observed in the genomes of microorganisms found in AMDs, and in particular in the biofilms of the Richmond mine [81], and genomic islands carrying metal resistance genes have been detected in several bacteria capable of forming biofilms which may have been acquired by HGT processes [33,82–84]. All these data suggest that toxic metals may favor mutation occurrence, HGT and thus accelerate evolution of such microorganisms. To date, only one study was performed in *Tm.* sp. CB2 biofilm to test such a hypothesis. In this bacterium, in the presence of As(III), dead cells and cell lysis events were observed, as well as rapid dispersal of part of the population [35]. After the dispersal, 20% of the cells were more resistant, i.e. the MIC was higher than that of the planktonic "mother" cells. These cells were defined as variants, since this resistance is conserved after several generations of growth in the absence of As(III) (Farasin, unpublished data). Such resistance acquisition mechanism may increase bacterial capacity to colonize new niches after cell dispersal (Fig. 1). The occurrence of variants in biofilm metal resistance and the mechanisms involved should be studied in the future to understand how metals can increase biofilm diversity and accelerate the evolution processes.

5. Conclusion

The fact that bacterial biofilms are often more tolerant or resistant to toxic metals than their planktonic counterparts is due to complex processes involving many factors (Fig. 1). This resistance makes biofilms an attractive biotechnological tool, since they may increase the efficiency of the current toxic metal remediation processes. For example, since toxic metal sequestration or immobilization processes by EPS are energy-independent, EPS and biofilms could be used for heavy metal removal or in situ heavy metal immobilization purposes [45,85,86]. As established in the literature, multispecies biofilms are probably more efficient and more stable than single-species biofilms, and they promise to have useful applications in the future. Generally speaking, understanding how multispecies biofilms form, resist pollutants such as toxic metals and remove them is of crucial importance not only from the bioremediation point of view, but also in medical processes and water distribution systems, where biofilms are undesirable because they are liable to interfere with decontamination processes. Finally, only a few works have been performed to study how metals influence bacterial evolution in biofilms. Various processes such as HGT are known to impact species evolution in multispecies biofilms. Less information is available thus far on the ability of plasmids, transposons, bacteriophages and eDNA to mediate HGT in bacterial cells growing in biofilms under toxic metal stress. Nevertheless, HGT may have a strong impact on the spreading of resistance genes and may therefore strongly contribute to overall resistance to toxic metals in nature. In the future, it will be important to study the mechanisms involved in HGT in biofilms exposed to toxic metal.

Conflict of interest

The authors declare no conflict of interests.

Acknowledgments

This research was supported by the Université de Strasbourg (UdS) and the Centre National de la Recherche Scientifique (CNRS). This study was financed by the THIOFILM project. JF and JCA were supported by the Agence Nationale de la Recherche (ANR-12-ADAP-0013) and the Region Alsace, and the Université de Strasbourg, respectively.

References

[1] Hall-Stoodley L, Costerton JW, Stoodley P. Bacterial biofilms: from the natural environment to infectious diseases. Nat Rev Microbiol 2004;2:95–108. http://dx.doi.org/10.1038/nrmicro821.

[2] Vega LM, Alvarez PJ, McLean RJC. Bacterial signaling ecology and potential applications during aquatic biofilm construction. Microb Ecol 2014;68:24–34. http://dx.doi.org/10.1007/s00248-013-0321-1.
[3] Flemming H-C, Wingender J. The biofilm matrix. Nat Rev Microbiol 2010;8:623–33. http://dx.doi.org/10.1038/nrmicro2415.
[4] Bridier A, Tischenko E, Dubois-Brissonnet F, Herry J-M, Thomas V, Daddi-Oubekka S, et al. Deciphering biofilm structure and reactivity by multiscale time-resolved fluorescence analysis. Adv Exp Med Biol 2011;715:333–49. http://dx.doi.org/10.1007/978-94-007-0940-9_21.
[5] Karimi A, Karig D, Kumar A, Ardekani AM. Interplay of physical mechanisms and biofilm processes: review of microfluidic methods. Lab Chip 2014. http://dx.doi.org/10.1039/c4lc01095g.
[6] Coenye T, Nelis HJ. In vitro and in vivo model systems to study microbial biofilm formation. J Microbiol Methods 2010;83:89–105. http://dx.doi.org/10.1016/j.mimet.2010.08.018.
[7] Verstraeten N, Braeken K, Debkumari B, Fauvart M, Fransaer J, Vermant J, et al. Living on a surface: swarming and biofilm formation. Trends Microbiol 2008;16:496–506. http://dx.doi.org/10.1016/j.tim.2008.07.004.
[8] Guttenplan SB, Kearns DB. Regulation of flagellar motility during biofilm formation. FEMS Microbiol Rev 2013;37:849–71. http://dx.doi.org/10.1111/1574-6976.12018.
[9] Van Acker H, Van Dijck P, Coenye T. Molecular mechanisms of antimicrobial tolerance and resistance in bacterial and fungal biofilms. Trends Microbiol 2014;22:326–33. http://dx.doi.org/10.1016/j.tim.2014.02.001.
[10] Harrison JJ, Ceri H, Turner RJ. Multimetal resistance and tolerance in microbial biofilms. Nat Rev Microbiol 2007;5:928–38. http://dx.doi.org/10.1038/nrmicro1774.
[11] Coenye T. Response of sessile cells to stress: from changes in gene expression to phenotypic adaptation. FEMS Immunol Med Microbiol 2010;59:239–52. http://dx.doi.org/10.1111/j.1574-695X.2010.00682.x.
[12] Jomova K, Valko M. Advances in metal-induced oxidative stress and human disease. Toxicology 2011;283:65–87. http://dx.doi.org/10.1016/j.tox.2011.03.001.
[13] Lemire JA, Harrison JJ, Turner RJ. Antimicrobial activity of metals: mechanisms, molecular targets and applications. Nat Rev Micro 2013;11:371–84. http://dx.doi.org/10.1038/nrmicro3028.
[14] Tamás MJ, Sharma SK, Ibstedt S, Jacobson T, Christen P. Heavy metals and metalloids as a cause for protein misfolding and aggregation. Biomolecules 2014;4:252–67. http://dx.doi.org/10.3390/biom4010252.
[15] Silver S, Phung LT. A bacterial view of the periodic table: genes and proteins for toxic inorganic ions. J Ind Microbiol Biotechnol 2005;32:587–605. http://dx.doi.org/10.1007/s10295-005-0019-6.
[16] Yang L, Liu Y, Wu H, Hóiby N, Molin S, Song Z. Current understanding of multi-species biofilms. Int J Oral Sci 2011;3:74–81. http://dx.doi.org/10.4248/IJOS11027.
[17] Harrison JJ, Ceri H, Roper NJ, Badry EA, Sproule KM, Turner RJ. Persister cells mediate tolerance to metal oxyanions in *Escherichia coli*. Microbiology 2005;151:3181–95. http://dx.doi.org/10.1099/mic.0.27794-0 (Reading, Engl).
[18] Booth SC, George IFS, Zannoni D, Cappelletti M, Duggan GE, Ceri H, et al. Effect of aluminium and copper on biofilm development of *Pseudomonas pseudoalcaligenes* KF707 and *P. fluorescens* as a function of different media compositions. Metallomics 2013;5:723–35. http://dx.doi.org/10.1039/c3mt20240b.
[19] Harrison JJ, Ceri H, Stremick CA, Turner RJ. Biofilm susceptibility to metal toxicity. Environ Microbiol 2004;6:1220–7. http://dx.doi.org/10.1111/j.1462-2920.2004.00656.x.
[20] Harrison JJ, Turner RJ, Ceri H. Persister cells, the biofilm matrix and tolerance to metal cations in biofilm and planktonic *Pseudomonas aeruginosa*. Environ Microbiol 2005;7:981–94. http://dx.doi.org/10.1111/j.1462-2920.2005.00777.x.
[21] Tremaroli V, Fedi S, Turner RJ, Ceri H, Zannoni D. *Pseudomonas pseudoalcaligenes* KF707 upon biofilm formation on a polystyrene surface acquire a strong antibiotic resistance with minor changes in their tolerance to metal cations and metalloid oxyanions. Arch Microbiol 2008;190:29–39. http://dx.doi.org/10.1007/s00203-008-0360-z.
[22] Harrison JJ, Turner RJ, Ceri H. High-throughput metal susceptibility testing of microbial biofilms. BMC Microbiol 2005;5:53. http://dx.doi.org/10.1186/1471-2180-5-53.
[23] Tremaroli V, Vacchi Suzzi C, Fedi S, Ceri H, Zannoni D, Turner RJ. Tolerance of *Pseudomonas pseudoalcaligenes* KF707 to metals, polychlorobiphenyls and chlorobenzoates: effects on chemotaxis-, biofilm- and planktonic-grown cells. FEMS Microbiol Ecol 2010;74:291–301. http://dx.doi.org/10.1111/j.1574-6941.2010.00965.x.
[24] Harrison JJ, Ceri H, Stremick C, Turner RJ. Differences in biofilm and planktonic cell mediated reduction of metalloid oxyanions. FEMS Microbiol Lett 2004;235:357–62. http://dx.doi.org/10.1016/j.femsle.2004.05.005.
[25] Schue M, Fekete A, Ortet P, Brutesco C, Heulin T, Schmitt-Kopplin P, et al. Modulation of metabolism and switching to biofilm prevail over exopolysaccharide production in the response of *Rhizobium alamii* to cadmium. PLoS One 2011;6:e26771. http://dx.doi.org/10.1371/journal.pone.0026771.
[26] Lira-Silva E, Santiago-Martínez MG, García-Contreras R, Zepeda-Rodríguez A, Marín-Hernández A, Moreno-Sánchez R, et al. Cd2+ resistance mechanisms in *Methanosarcina acetivorans* involve the increase in the coenzyme M content and induction of biofilm synthesis. Environ Microbiol Rep 2013;5:799–808. http://dx.doi.org/10.1111/1758-2229.12080.
[27] Harrison JJ, Rabiei M, Turner RJ, Badry EA, Sproule KM, Ceri H. Metal resistance in *Candida* biofilms. FEMS Microbiol Ecol 2006;55:479–91. http://dx.doi.org/10.1111/j.1574-6941.2005.00045.x.
[28] Cobine PA, Cruz LF, Navarrete F, Duncan D, Tygart M, De La Fuente L. *Xylella fastidiosa* differentially accumulates mineral elements in biofilm and planktonic cells. PLoS One 2013;8:e54936. http://dx.doi.org/10.1371/journal.pone.0054936.
[29] Navarrete F, De La Fuente L. Response of *Xylella fastidiosa* to zinc: decreased culturability, increased exopolysaccharide production, and formation of resilient biofilms under flow conditions. Appl Environ Microbiol 2014;80:1097–107. http://dx.doi.org/10.1128/AEM.02998-13.
[30] Muranaka LS, Takita MA, Olivato JC, Kishi LT, de Souza AA. Global expression profile of biofilm resistance to antimicrobial compounds in the plant-pathogenic bacterium *Xylella fastidiosa* reveals evidence of persister cells. J Bacteriol 2012;194:4561–9. http://dx.doi.org/10.1128/JB.00436-12.
[31] Perrin C, Briandet R, Jubelin G, Lejeune P, Mandrand-Berthelot M-A, Rodrigue A, et al. Nickel promotes biofilm formation by *Escherichia coli* K-12 strains that produce curli. Appl Environ Microbiol 2009;75:1723–33. http://dx.doi.org/10.1128/AEM.02171-08.
[32] Marchal M, Briandet R, Koechler S, Kammerer B, Bertin PN. Effect of arsenite on swimming motility delays surface colonization in *Herminiimonas arsenicoxydans*. Microbiology 2010;156:2336–42. http://dx.doi.org/10.1099/mic.0.039313-0 (Reading, Engl).
[33] Andres J, Arsène-Ploetze F, Barbe V, Brochier-Armanet C, Cleiss-Arnold J, Coppée J-Y, et al. Life in an arsenic-containing gold mine: genome and physiology of the autotrophic arsenite-oxidizing bacterium *Rhizobium* sp. NT-26. Genome Biol Evol 2013;5:934–53. http://dx.doi.org/10.1093/gbe/evt061.
[34] Grumbein S, Opitz M, Lieleg O. Selected metal ions protect *Bacillus subtilis* biofilms from erosion. Metallomics 2014;6:1441–50. http://dx.doi.org/10.1039/c4mt00049h.
[35] Marchal M, Briandet R, Halter D, Koechler S, Dubow MS, Lett M-C, et al. Subinhibitory arsenite concentrations lead to population dispersal in *Thiomonas* sp. PLoS One 2011;6:e23181. http://dx.doi.org/10.1371/journal.pone.0023181.
[36] Cao B, Majors PD, Ahmed B, Renslow RS, Silvia CP, Shi L, et al. Biofilm shows spatially stratified metabolic responses to contaminant exposure. Environ Microbiol 2012;14:2901–10. http://dx.doi.org/10.1111/j.1462-2920.2012.02850.x.
[37] Harrison JJ, Ceri H, Yerly J, Rabiei M, Hu Y, Martinuzzi R, et al. Metal ions may suppress or enhance cellular differentiation in *Candida albicans* and *Candida tropicalis* biofilms. Appl Environ Microbiol 2007;73:4940–9. http://dx.doi.org/10.1128/AEM.02711-06.
[38] Hullebusch ED van, Zandvoort MH, Lens PNL. Metal immobilisation by biofilms: mechanisms and analytical tools. Reviews Environ Sci Bio Technol 2003;2:9–33. http://dx.doi.org/10.1023/B: RESB.0000022995.48330.55.

Please cite this article in press as: Koechler S, et al., Toxic metal resistance in biofilms: diversity of microbial responses and their evolution, Research in Microbiology (2015), http://dx.doi.org/10.1016/j.resmic.2015.03.008

[39] Liu H, Fang HHP. Characterization of electrostatic binding sites of extracellular polymers by linear programming analysis of titration data. Biotechnol Bioeng 2002;80:806–11. http://dx.doi.org/10.1002/bit.10432.

[40] Choi O, Yu C-P, Esteban Fernández G, Hu Z. Interactions of nanosilver with *Escherichia coli* cells in planktonic and biofilm cultures. Water Res 2010;44:6095–103. http://dx.doi.org/10.1016/j.watres.2010.06.069.

[41] Teitzel GM, Parsek MR. Heavy metal resistance of biofilm and planktonic Pseudomonas aeruginosa. Appl Environ Microbiol 2003;69:2313–20.

[42] Jang A, Kim SM, Kim SY, Lee SG, Kim IS. Effect of heavy metals (Cu, Pb, and Ni) on the compositions of EPS in biofilms. Water Sci Technol 2001;43:41–8.

[43] Priester JH, Olson SG, Webb SM, Neu MP, Hersman LE, Holden PA. Enhanced exopolymer production and chromium stabilization in *Pseudomonas putida* unsaturated biofilms. Appl Environ Microbiol 2006;72:1988–96. http://dx.doi.org/10.1128/AEM.72.3.1988-1996.2006.

[44] Rodrigues CM, Takita MA, Coletta-Filho HD, Olivato JC, Caserta R, Machado MA, et al. Copper resistance of biofilm cells of the plant pathogen *Xylella fastidiosa*. Appl Microbiol Biotechnol 2008;77:1145–57. http://dx.doi.org/10.1007/s00253-007-1232-1.

[45] Pal A, Paul AK. Microbial extracellular polymeric substances: central elements in heavy metal bioremediation. Indian J Microbiol 2008;48:49–64. http://dx.doi.org/10.1007/s12088-008-0006-5.

[46] Hockin SL, Gadd GM. Linked redox precipitation of sulfur and selenium under anaerobic conditions by sulfate-reducing bacterial biofilms. Appl Environ Microbiol 2003;69:7063–72.

[47] Cologgi DL, Speers AM, Bullard BA, Kelly SD, Reguera G. Enhanced uranium immobilization and reduction by *Geobacter sulfurreducens* biofilms. Appl Environ Microbiol 2014;80:6638–46. http://dx.doi.org/10.1128/AEM.02289-14.

[48] Reith F, Etschmann B, Grosse C, Moors H, Benotmane MA, Monsieurs P, et al. Mechanisms of gold biomineralization in the bacterium *Cupriavidus metallidurans*. PNAS 2009;106:17757–62. http://dx.doi.org/10.1073/pnas.0904583106.

[49] Booth SC, Workentine ML, Wen J, Shaykhutdinov R, Vogel HJ, Ceri H, et al. Differences in metabolism between the biofilm and planktonic response to metal stress. J Proteome Res 2011;10:3190–9. http://dx.doi.org/10.1021/pr2002353.

[50] Lin TY, Kampalath RA, Lin C-C, Zhang M, Chavarria K, Lacson J, et al. Investigation of mercury methylation pathways in biofilm versus planktonic cultures of *Desulfovibrio desulfuricans*. Environ Sci Technol 2013;47:5695–702. http://dx.doi.org/10.1021/es400079n.

[51] Keren I, Kaldalu N, Spoering A, Wang Y, Lewis K. Persister cells and tolerance to antimicrobials. FEMS Microbiol Lett 2004;230:13–8.

[52] Lewis K. Persister cells, dormancy and infectious disease. Nat Rev Microbiol 2007;5:48–56. http://dx.doi.org/10.1038/nrmicro1557.

[53] Lewis K. Persister cells. Annu Rev Microbiol 2010;64:357–72. http://dx.doi.org/10.1146/annurev.micro.112408.134306.

[54] Steudel B, Hector A, Friedl T, Löfke C, Lorenz M, Wesche M, et al. Biodiversity effects on ecosystem functioning change along environmental stress gradients. Ecol Lett 2012;15:1397–405. http://dx.doi.org/10.1111/j.1461-0248.2012.01863.x.

[55] Lee KWK, Periasamy S, Mukherjee M, Xie C, Kjelleberg S, Rice SA. Biofilm development and enhanced stress resistance of a model, mixed-species community biofilm. ISME J 2014;8:894–907. http://dx.doi.org/10.1038/ismej.2013.194.

[56] Li J-T, Duan H-N, Li S-P, Kuang J-L, Zeng Y, Shu W-S. Cadmium pollution triggers a positive biodiversity–productivity relationship: evidence from a laboratory microcosm experiment. J Appl Ecol 2010;47:890–8. http://dx.doi.org/10.1111/j.1365-2664.2010.01818.x.

[57] White C, Gadd GM. Accumulation and effects of cadmium on sulfate-reducing bacterial biofilms. Microbiology 1998;144:1407–15. http://dx.doi.org/10.1099/00221287-144-5-1407.

[58] Carpio IEM, Machado-Santelli G, Sakata SK, Ferreira Filho SS, Rodrigues DF. Copper removal using a heavy-metal resistant microbial consortium in a fixed-bed reactor. Water Res 2014;62:156–66. http://dx.doi.org/10.1016/j.watres.2014.05.043.

[59] Sabri AN, Sabri S, Sultan S. Metallurgical microscopy of bacterial biofilm Cr+6 conversion into Cr+3 by *Acienetobacter calcoaceticus*, *Staphylococcus aureus* and *Oscillatoria* strain. In: Méndez-Vilas A, editor. Curr. Microsc. Contrib. Adv. Sci. Technol., vol. 1; 2012. p. 749–55.

[60] Von Canstein H, Kelly S, Li Y, Wagner-Döbler I. Species diversity improves the efficiency of mercury-reducing biofilms under changing environmental conditions. Appl Environ Microbiol 2002;68:2829–37.

[61] Fairbrother L, Etschmann B, Brugger J, Shapter J, Southam G, Reith F. Biomineralization of gold in biofilms of *Cupriavidus metallidurans*. Environ Sci Technol 2013;47:2628–35. http://dx.doi.org/10.1021/es302381d.

[62] Golby S, Ceri H, Marques LLR, Turner RJ. Mixed-species biofilms cultured from an oil sand tailings pond can biomineralize metals. Microb Ecol 2014;68:70–80. http://dx.doi.org/10.1007/s00248-013-0331-z.

[63] Streten-Joyce C, Manning J, Gibb KS, Neilan BA, Parry DL. The chemical composition and bacteria communities in acid and metalliferous drainage from the wet-dry tropics are dependent on season. Sci Total Environ 2013;443:65–79. http://dx.doi.org/10.1016/j.scitotenv.2012.10.024.

[64] Belnap CP, Pan C, Denef VJ, Samatova NF, Hettich RL, Banfield JF. Quantitative proteomic analyses of the response of acidophilic microbial communities to different pH conditions. ISME J 2011;5:1152–61. http://dx.doi.org/10.1038/ismej.2010.200.

[65] Bouskill NJ, Barker-Finkel J, Galloway TS, Handy RD, Ford TE. Temporal bacterial diversity associated with metal-contaminated river sediments. Ecotoxicology 2010;19:317–28. http://dx.doi.org/10.1007/s10646-009-0414-2.

[66] Chodak M, Gołębiewski M, Morawska-Płoskonka J, Kuduk K, Niklińska M. Diversity of microorganisms from forest soils differently polluted with heavy metals. Appl Soil Ecol 2013;64:7–14. http://dx.doi.org/10.1016/j.apsoil.2012.11.004.

[67] Labrenz M, Banfield JF. Sulfate-reducing bacteria-dominated biofilms that precipitate ZnS in a subsurface circumneutral-pH mine drainage system. Microb Ecol 2004;47:205–17. http://dx.doi.org/10.1007/s00248-003-1025-8.

[68] Mendez MO, Neilson JW, Maier RM. Characterization of a bacterial community in an abandoned semiarid lead-zinc mine tailing site. Appl Environ Microbiol 2008;74:3899–907. http://dx.doi.org/10.1128/AEM.02883-07.

[69] Santelli CM, Edgcomb VP, Bach W, Edwards KJ. The diversity and abundance of bacteria inhabiting seafloor lavas positively correlate with rock alteration. Environ Microbiol 2009;11:86–98. http://dx.doi.org/10.1111/j.1462-2920.2008.01743.x.

[70] Pasić L, Kovce B, Sket B, Herzog-Velikonja B. Diversity of microbial communities colonizing the walls of a Karstic cave in Slovenia. FEMS Microbiol Ecol 2010;71:50–60. http://dx.doi.org/10.1111/j.1574-6941.2009.00789.x.

[71] Tomczyk-Żak K, Kaczanowski S, Drewniak Ł, Dmoch Ł, Sklodowska A, Zielenkiewicz U. Bacteria diversity and arsenic mobilization in rock biofilm from an ancient gold and arsenic mine. Sci Total Environ 2013;461–462:330–40. http://dx.doi.org/10.1016/j.scitotenv.2013.04.087.

[72] Boles BR, Thoendel M, Singh PK. Self-generated diversity produces "insurance effects" in biofilm communities. Proc Natl Acad Sci U S A 2004;101:16630–5. http://dx.doi.org/10.1073/pnas.0407460101.

[73] Yachi S, Loreau M. Biodiversity and ecosystem productivity in a fluctuating environment: the insurance hypothesis. Proc Natl Acad Sci U S A 1999;96:1463–8.

[74] Loreau M, de Mazancourt C. Biodiversity and ecosystem stability: a synthesis of underlying mechanisms. Ecol Lett 2013;16(Suppl. 1):106–15. http://dx.doi.org/10.1111/ele.12073.

[75] Awasthi A, Singh M, Soni SK, Singh R, Kalra A. Biodiversity acts as insurance of productivity of bacterial communities under abiotic perturbations. ISME J 2014. http://dx.doi.org/10.1038/ismej.2014.91.

[76] Denef VJ, Mueller RS, Banfield JF. AMD biofilms: using model communities to study microbial evolution and ecological complexity in nature. ISME J 2010;4:599–610. http://dx.doi.org/10.1038/ismej.2009.158.

[77] Aliaga Goltsman DS, Comolli LR, Thomas BC, Banfield JF. Community transcriptomics reveals unexpected high microbial diversity in

Please cite this article in press as: Koechler S, et al., Toxic metal resistance in biofilms: diversity of microbial responses and their evolution, Research in Microbiology (2015), http://dx.doi.org/10.1016/j.resmic.2015.03.008

acidophilic biofilm communities. ISME J 2015;9:1014–23. http://dx.doi.org/10.1038/ismej.2014.200.

[78] Yelton AP, Comolli LR, Justice NB, Castelle C, Denef VJ, Thomas BC, et al. Comparative genomics in acid mine drainage biofilm communities reveals metabolic and structural differentiation of co-occurring archaea. BMC Genomics 2013;14:485. http://dx.doi.org/10.1186/1471-2164-14-485.

[79] Simmons SL, Dibartolo G, Denef VJ, Goltsman DSA, Thelen MP, Banfield JF. Population genomic analysis of strain variation in *Leptospirillum* group II bacteria involved in acid mine drainage formation. PLoS Biol 2008;6:e177. http://dx.doi.org/10.1371/journal.pbio.0060177.

[80] Denef VJ, Banfield JF. In situ evolutionary rate measurements show ecological success of recently emerged bacterial hybrids. Science 2012;336:462–6. http://dx.doi.org/10.1126/science.1218389.

[81] Li S-J, Hua Z-S, Huang L-N, Li J, Shi S-H, Chen L-X, et al. Microbial communities evolve faster in extreme environments. Sci Rep 2014;4:6205. http://dx.doi.org/10.1038/srep06205.

[82] Navarro CA, von Bernath D, Jerez CA. Heavy metal resistance strategies of acidophilic bacteria and their acquisition: importance for biomining and bioremediation. Biol Res 2013;46:363–71. http://dx.doi.org/10.4067/S0716-97602013000400008.

[83] Muller D, Médigue C, Koechler S, Barbe V, Barakat M, Talla E, et al. A tale of two oxidation states: bacterial colonization of arsenic-rich environments. PLoS Genet 2007;3:e53. http://dx.doi.org/10.1371/journal.pgen.0030053.

[84] Arsène-Ploetze F, Koechler S, Marchal M, Coppée J-Y, Chandler M, Bonnefoy V, et al. Structure, function, and evolution of the *Thiomonas* spp. genome. PLoS Genet 2010;6:e1000859. http://dx.doi.org/10.1371/journal.pgen.1000859.

[85] Singh R, Paul D, Jain RK. Biofilms: implications in bioremediation. Trends Microbiol 2006;14:389–97. http://dx.doi.org/10.1016/j.tim.2006.07.001.

[86] Marsili E, Beyenal H, Di Palma L, Merli C, Dohnalkova A, Amonette JE, et al. Uranium immobilization by sulfate-reducing biofilms grown on hematite, dolomite, and calcite. Environ Sci Technol 2007;41:8349–54.

Please cite this article in press as: Koechler S, et al., Toxic metal resistance in biofilms: diversity of microbial responses and their evolution, Research in Microbiology (2015), http://dx.doi.org/10.1016/j.resmic.2015.03.008

II - Matériels et méthodes

1 - Souches et conditions de cultures

Thiomonas arsenitoxydans 3As et les souches *Thiomonas* spp. CB1, CB2, CB3 et CB6 ont été isolées du drainage minier acide de Carnoulès (Gard) (Arsene-Ploetze et al., 2010; Duquesne et al., 2008). Elles ont été cultivées en milieu m126 dont la composition pour 1 litre est la suivante : Extrait de levure : 0,5 g ; Thiosulfate de sodium [$Na_2S_2O_3$] : 5 g ; Phosphate de potassium monobasique [KH_2PO_4] : 1,5 g ; Phosphate de sodium dihydrate [Na_2HPO_4-$2H_2O$] : 4,5 g ; Sulfate de magnésium heptahydrate [$MgSO_4$-$7H_2O$] : 0,1 g ; Chlorure d'ammonium [$(NH_4)Cl$] : 0,3 g. Le pH était ensuite ajusté à 5 avec du H_2SO_4. L'arsénite (As(III)) a ensuite été ajouté aux concentrations désirées à partir d'une solution mère stérile de $NaAsO_2$ (Prolabo Fontenay-sous-Bois, France) à 667 mM.

Les cultures en biofilm ont été réalisées dans des plaques en polystyrène 12 puits (Nunc) pour les expériences de coloration au crystal violet, et 6 puits pour les expériences de RNAseq. Pour les observations en microscopie confocale, les biofilms ont été cultivés dans des plaques 12 puits à fond en verre (InVitroScientific). Les plaques 12 puits ont été inoculées avec 1 mL de culture dilué dans du m126 à $D.O_{600nm}$ = 0,002 par puits, supplémenté ou non en As(III). Les plaques 6 puits ont été inoculées de la même manière mais avec 2,75 mL/puits. Ces cultures ont été incubées à 30°C et le milieu a été renouvelé au bout de 24 h.

2 - Quantification de biofilm au crystal violet

Les colorations au crystal violet ont été effectuées sur des biofilms de 72 h en triplicat (inoculation à partir de trois cultures indépendantes) avec trois puits pour chaque condition (0 mM, 2,67 mM et 5,33 mM d'As(III)). Au bout de 72 h d'incubation, le milieu a été retiré des puits et les biofilms ont été délicatement rincés avec 1 mL d'eau physiologiques (NaCl = 9 $g.L^{-1}$). Les plaques ont ensuite été séchées par incubation à 30°C pendant 20 min environ, puis 1 mL de crystal violet à 0,1 % a été ajouté dans chaque puits. Les plaques ont alors été à nouveau incubées à 30°C pendant 20 min puis le crystal violet a été retiré et l'excès enlevé par un rinçage avec 1 mL d'eau physiologique. 1 mL/puits d'éthanol à 95 % a été ajouté pour re-suspendre le crystal violet fixé dans les biofilms puis des dilutions ¼ ont été réalisées dans des plaques 96 puits (transfert de 50 µL des puits de la plaques 12 puits dans 150 µL d'éthanol 95 % dans la plaque 96 puits). La $D.O_{595nm}$ de la plaque 96 puits a ensuite été mesurée avec un lecteur de

plaque (Synergy HT ; BioTek).

3 - Test de mobilité en gélose molle

Trois gouttes de 1 µl d'une culture diluées à $D.O_{600nm}$ = 0,002 ont été déposées sur une boîte de Pétri contenant 20 mL de m126 semi-solide (0,3% d'agar) supplémenté ou non avec 2,67 mM ou 5,33 mM d'As(III). Les boîtes ont ensuite été incubées à 30°C pendant 7 jours. En cas de mobilité des souches, un halo était visible autour du dépôt.

4 - Etude des biofilms en microscopie confocale

Les biofilms ont été cultivés dans des plaques 12 puits à fond en verre de la même manière que décrit plus haut pour les colorations en crystal violet. Deux triplicats biologiques (trois cultures indépendantes) ont été réalisés : un pour les observations à 24 h, l'autre pour les observations à 72 h (avec renouvellement du milieu à 24 h). Les biofilms ont marqués avec trois colorants différents : 2,5 µM de SYTO9 green (485/498) pour révéler les cellules viables ; 2,5 µM de iodure de propidium (305-538/617) pour révéler les cellules mortes ; 100 $\mu g.mL^{-1}$ de tetramethylrhodamine conjuguée à de la concanavaline A (tetramethylrhodamine-ConA) (555/580) pour révéler la matrice. Après ajout des colorants, les plaques ont été incubées à l'obscurité pendant 30 min.

L'observation et l'acquisition des images ont ensuite été réalisées à l'aide d'un microscope confocal (Leica TCS SP8, objectif 63 X à immersion). Deux z-stacks (série d'images sur un plan vertical) avec un intervalle de 1 µm ont été acquis pour chaque puits. Les longueurs d'ondes d'excitation étaient de 488 nm (SYTO9 green et iodure de propidium) et 561 nm (tetramethylrhodamine-ConA). Les plages de longueurs d'ondes émises étaient de 495 nm à 547 nm pour le SYTO9 green, de 657 nm à 757 nm pour l'iodure de propidium et de 567 nm à 598 nm pour la tetramethylrhodamine-ConA.

Les images ont ensuite été traitées pour obtenir une représentation en trois dimensions avec le logiciel IMARIS (Bitplane, Zürich, Suisse). Les biovolumes de cellules (SYTO9 green) et de matrice (tetramethylrhodamine-ConA) ont été quantifiés à partir des séries d'images avec le logiciel ICY (http://icy.bioimageanalysis.org).

5 - Recherche d'îlots génomiques impliqués dans la synthèse de biofilm et la mobilité

Afin de rechercher les origines génétiques potentielles des différences observées concernant la biosynthèse de biofilm et la mobilité, les génomes des cinq souches (3As, CB1, CB2, CB3 et CB6) ont été comparés. Pour cela l'outil *RGP Finder* de la plateforme MaGe (https://www.genoscope.cns.fr/agc/microscope/compgenomics/genomicIsland.php?&wwwpkgdb=efd0cd3f90e8c722e87b7593695c23f4) a été utilisé afin de repérer les ruptures de synténie entre les souches. Ensuite, le contenu des régions identifiées comme étant variable a été analysé manuellement afin rechercher la présence de gènes potentiellement impliqués dans la synthèse de biofilm ou la mobilité.

6 - Transcriptomique : RNAseq différentiel

Afin de rechercher des gènes potentiellement impliqués dans la formation de biofilm chez *Tm.* sp. CB2, notamment la réponse à l'As(III), une expérience de transcriptomique par RNAseq a été réalisée.

a. Cultures

Les biofilms ont été cultivés dans des plaques 6 puits sans As(III) ou avec 5,33 mM d'As(III) pendant 24 h, 48 h et 72 h, avec renouvellement du milieu à 24 h. Au total, 8 plaques ont été utilisées pour chaque condition et l'expérience a été réalisée en triplicat (3 cultures indépendantes). Après l'incubation à 30°C, le milieu a été délicatement retiré des puits et remplacé par 1,5 mL d'eau physiologique à 4°C. Les biofilms ont alors été décollés du fond des puits à l'aide de grattoirs à cellules (Biologix) afin de mettre en suspension les cellules dans l'eau physiologique. Cette étape a été réalisée sur glace pour éviter la dégradation des ARN. Les suspensions ont ensuite été centrifugées à 8000 *x g* pendant 15 min à 4°C et après élimination des surnageants, les culots cellulaires ont été stockés à -80°C.

b. Extractions d'ARN et séquençage

Les extractions d'ARN ont été réalisées par Sandrine Koechler. Après décongélation à température ambiante, 60 µL d'EDTA (0,5 M ; pH 7,6) à 4°C ont été ajoutés dans chaque tube. Les cellules ont ensuite été reprises dans 400 µL de solution de re-suspension (20 % glucose ; 25 mM Tris ; 10 mM (v/v) EDTA) afin de lyser les cellules par choc osmotique. Les suspensions ont ensuite été transférées dans des tubes contenant 2 mL de phénol, 0,5 g de microbilles et

500 µL de phénol acide. Les tubes ont été agités à l'aide d'un broyeur (FastPrep) durant deux cycles de 30 secondes, puis centrifugés à 16 000 *x g* pendant 5 minutes à 4°C. Les phases aqueuses ont alors été transférées dans de nouveaux tubes puis 1 mL de trizol a été ajouté. La solution obtenue a été homogénéisée à l'aide d'une pipette jusqu'à que l'aspect devienne translucide, et 200 µL de chloroforme / alcool isoamylique ont ensuite été ajoutés. Le mélange obtenu a de nouveau été homogénéisé à la pipette, puis après avoir reposé 3 min à température ambiante, les tubes ont été centrifugés à 16 000 *x g* pendant 5 min à 4°C. La phase aqueuse a alors été transférée dans de nouveaux tubes et 200 µL de chloroforme / alcool isoamylique ont à nouveau été ajoutés. Le contenu des tubes a été homogénéisé à l'aide d'une pipette puis les tubes ont été centrifugés à 16 000 x g pendant 5 min à 4 °C. Les surnageants ont été transférés dans de nouveaux tubes et les ARN contenus dans cette phase ont été purifiés sur colonne (Kit Qiagen RNeasy Plus Mini Kit) puis envoyés à séquencer au Génoscope.

III - Résultats et discussion

1 - Quantification de biofilms au crystal violet et test de mobilité

La quantification de biofilms au crystal violet a été réalisée afin d'avoir une vue globale des capacités de chacune des souches à synthétiser des biofilms, que ce soit les unes par rapport aux autres ou en réponse à la présence d'As(III). Cette quantification a été réalisée sur des biofilms de 72 h cultivés dans des plaques de 12 puits en polystyrène.

Deux groupes peuvent être distingués : d'un côté 3As, CB1 et CB6 qui produisent relativement peu de biofilms, et de l'autre CB2 et CB3 qui en produisent beaucoup plus (Figure 29). En présence de 5,33 mM d'As(III), la D.O est plus élevée pour CB3 alors qu'elle l'est moins pour CB2 et CB6 (Figure 29). Ceci suggère que la production de biofilms semble être différemment régulée par la présence d'As(III) suivant les souches considérées. Aucun effet de l'As(III) n'a pu être observé chez 3As et CB1, peut-être à cause du manque de résolution de cette méthode.

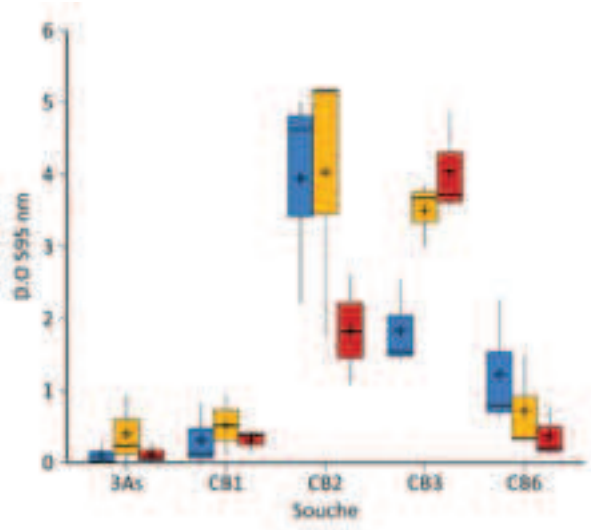


Figure 29 D.O 595 nm mesurée après coloration des biofilms en crystal violet. Bleu = 0 mM d'As(III) ; Jaune = 2,67 mM d'As(III) ; Rouge = 5,33 mM d'As(III). La croix "+" indique la moyenne.

La production de biofilms et la mobilité flagellaire sont des processus liés au niveau de leur régulation et de par l'implication des flagelles dans l'adhésion lors des étapes précoces de développement des biofilms. Dans le but de savoir si les différences de production de biofilms révélées par marquage au crystal violet sont corrélées avec des différences concernant la mobilité, des tests de mobilité en gélose molle (0,03 % d'agar) ont été réalisés en présence ou

en absence d'As(III). Au bout de 7 jours d'incubation, un halo de mobilité était observable pour les souches 3As, CB1 et CB6 en présence d'As(III), indiquant l'induction d'une mobilité plus importante par cet élément chez ces trois souches.

Les souches *Tm.* spp. 3As, CB1 et CB6 sont les trois pour lesquelles la production de biofilms s'était révélée être la plus faible par marquage au crystal violet (Figure 29). Ces résultats montrent donc une corrélation inverse entre la mobilité en présence d'As(III) et la production de biofilms. Ceci peut être interprété comme une stratégie différente adoptée par les souches : *Tm.* spp. 3As, CB1 et CB6 qui synthétisent peu de biofilms, deviennent mobiles afin de fuir les conditions adverses, alors que *Tm.* spp. CB2 et CB3 favoriseraient la formation de biofilms dans ces conditions. Cependant, à forte concentration, *Tm.* sp. CB2 formerait moins de biofilm que la souche *Tm.* sp. CB3.

La production de biofilms est un processus complexe, comprenant l'adhésion des cellules, la biosynthèse de la matrice, la maturation du biofilm et la dispersion des cellules. La coloration au crystal violet ne permettant pas de différencier les cellules de la matrice, il est impossible de savoir si l'augmentation de D.O est due à une meilleure adhésion cellulaire ou à l'augmentation de la biosynthèse de matrice par exemple. Afin de pallier ce problème, des analyses en microscopie confocale ont donc été réalisées.

2 - Etudes des biofilms en microscopie confocale

Afin de mieux comprendre les différences de biosynthèse de biofilms entre les différentes souches et en réponse à l'As(III), une analyse en microscopie confocale a été réalisée sur des biofilms après 24 h et 72 h de croissance. Pour cette analyse, les souches CB2 et CB3 non mobiles et produisant beaucoup de biofilms ainsi que la souche CB1 mobile et produisant peu de biofilms ont été sélectionnées. Les biovolumes de matrice extracellulaire et les cellules viables ont été marqués avec de la tetramethylrhodamine couplée à de la concanavaline A et du SYTO9 green respectivement. Les biovolumes de cellules mortes n'ont pas pu être calculés car le marquage à l'iodure de propidium produisait trop de bruit de fond.

A 24 h en absence d'As(III) (Figure 30), le biofilm de *Tm.* sp. CB3 parait être plus structuré que celui des deux autres souches. Les cellules forment des agrégats assimilables à des micro-colonies et de nombreux agrégats de matrice relativement denses sont visibles. De manière surprenante, comparé aux résultats obtenus par coloration au crystal violet, le biofilm de *Tm.* sp. CB1 est plus développé que celui de *Tm.* sp. CB2. Au bout de 72 h toujours en absence d'As(III) (Figure 31), un important volume de cellules est visible chez *Tm.* sp. CB2

mais sans organisation particulière comparé à *Tm.* sp CB3. Le volume de matrice a diminué chez les trois souches. Cette observation est également surprenante étant donné que les biofilms sont à maturité. Des structures plus complexes seraient attendues, et il est impossible ici de déterminer si ces observations sont liées à un processus naturel de dissémination, ou bien à la fragilité des structures polysaccharidiques qui seraient éliminées lors du renouvellement du milieu à 24h. En présence d'As(III) à 24 h (Figure 30), aucune différence très marquée n'est visible au niveau des cellules. Les agrégats de matrice paraissent par contre un peu plus nombreux et plus denses chez *Tm.* spp. CB1 et CB2. Ces deux souches semblent donc répondre à l'As(III) dès 24 h en formant un biofilm un peu plus structuré. En présence d'As(III) à 72 h (Figure 31), le biofilm de *Tm.* sp. CB1 est marqué par un nombre important de cellules mortes. Cette souche parait donc plus sensible à l'As(III) que les autres souches. Le biofilm de *Tm.* sp. CB2 contient moins de cellules qu'à 24 h mais possède plus de structures polysaccharidiques, exactement à l'inverse de *Tm.* sp. CB3. Ces deux souches répondent donc différemment à la présence d'As(III) comme il l'a été mis en évidence précédemment par coloration au crystal violet.

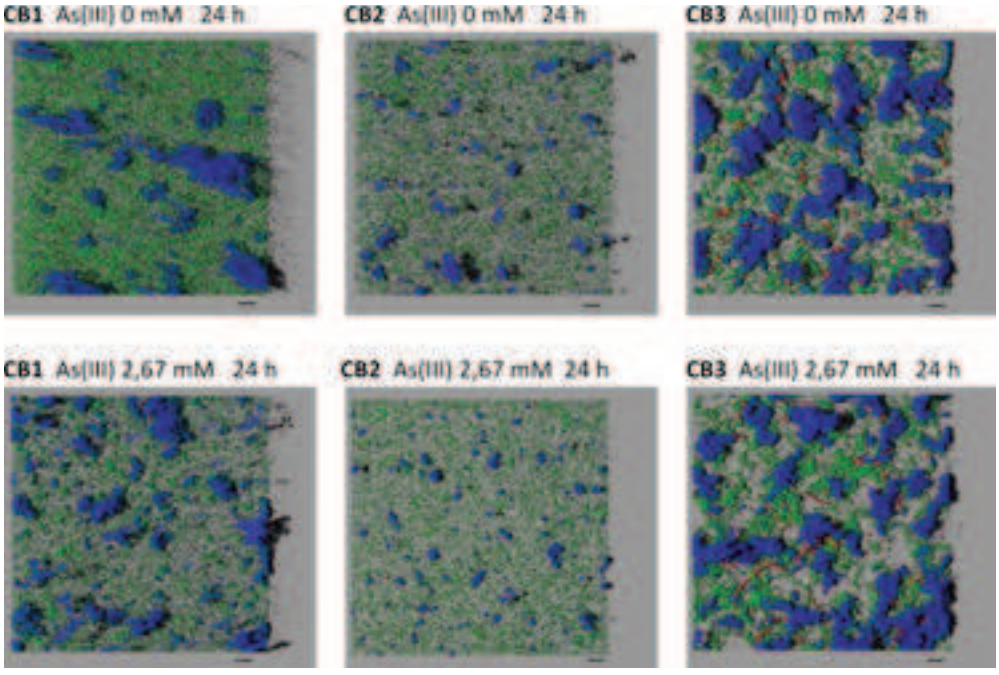


Figure 30 Reconstitution 3D d'observations réalisées au bout de 24 h des biofilms de Thiomonas spp. CB1, CB2 et CB3 cultivées en absence ou en présence de 2,67 mM d'As(III). Bleu = matrice polysaccharidique ; Vert = Cellules vivantes ; Rouge = Cellules mortes. Echelle = 50 µm. La souche CB1 produit du biofilm en quantité comparable aux deux autres souches. Le biofilm de CB3 est plus développé que CB1 et CB2 à 24h.

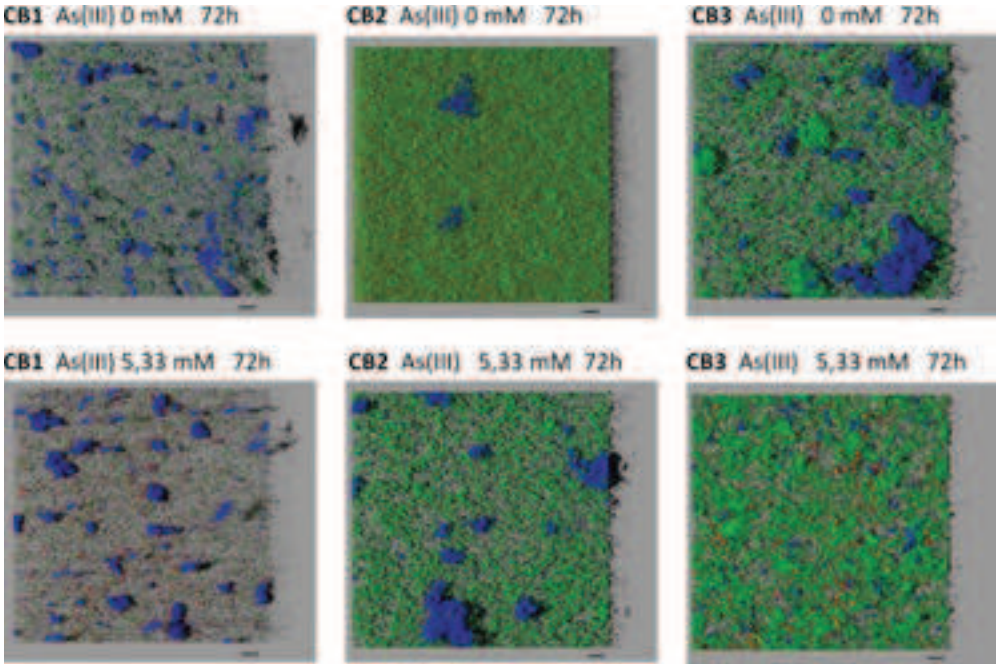


Figure 31 Reconstitution 3D d'observations réalisées au bout de 72 h des biofilms de Thiomonas spp. CB1 CB2 et CB3 cultivées en présence de 0 mM et 5,33 mM d'As(III). Bleu = matrice polysaccharidique ; Vert = Cellules vivantes ; Rouge = Cellules mortes. Echelle = 50 µm. L'effet de l'As(III) semble différent chez CB2 et CB3 : augmentation du volume de la matrice et diminution du volume de cellules pour CB2, inversement pour CB3.

D'après les biovolumes mesurés à 24 h en absence d'As(III) (Figure 32), la souche *Tm.* sp. CB3 produit une quantité de matrice plus élevée que les deux autres, y compris ramenée au volume de cellules. Le volume de matrice rapporté à celui de cellules est également élevé chez *Tm.* sp. CB1. Au bout de 72 h, toujours sans As(III) (Figure 33), la production de matrice rapportée au volume de cellules est plus importante chez *Tm.* spp. CB1 et CB3 que chez *Tm.* sp. CB2. Contrairement à ce qui pouvait être pensé à partir des colorations en crystal violet, la souche *Tm.* sp. CB2 produit donc moins de biofilm que *Tm.* sp. CB1.

En présence d'As(III) à 24 h (Figure 32), le volume cellulaire chez *Tm.* sp. CB3 est inférieur comparé à l'absence d'As(III). Ceci pourrait être lié soit à une adhésion plus faible des cellules soit à une croissance plus faible en présence d'As(III). Aucune autre différence significative n'est visible comparée à l'absence d'As(III) au bout de 24 h. En présence d'As(III) au bout de 72 h (Figure 33), le volume de cellules de *Tm.* sp. CB3 est plus élevé qu'en absence de cet élément. Cependant, le volume de matrice produit rapporté au volume de cellules est plus bas en présence d'As(III). Les cellules de cette souche ont donc clairement tendance à adhérer pour se développer en biofilm en présence d'As(III), mais produisent individuellement moins de matrice. La souche *Tm*. sp. CB2 répond à la présence d'As(III) de manière inverse à *Tm.* sp.

CB3, puisque le nombre de cellules fixées diminue mais qu'elles forment individuellement plus de matrice. Une augmentation de la production de matrice chez cette souche avait déjà été mise en évidence chez *Tm.* sp. CB2 (Marchal et al., 2011). Le volume de matrice rapporté aux cellules augmente également chez *Tm.* sp. CB1.

Cette analyse en microscopie confocale a donc permis de déterminer de manière plus précise les capacités de formation de biofilms des différentes souches. Ceci a notamment permis de constater que la souche *Tm.* sp. CB1 produit une quantité de biofilms comparable à *Tm.* sp. CB2 et de mieux comprendre les différences entre *Tm.* sp. CB2 et *Tm.* sp. CB3 en réponse à l'As(III).

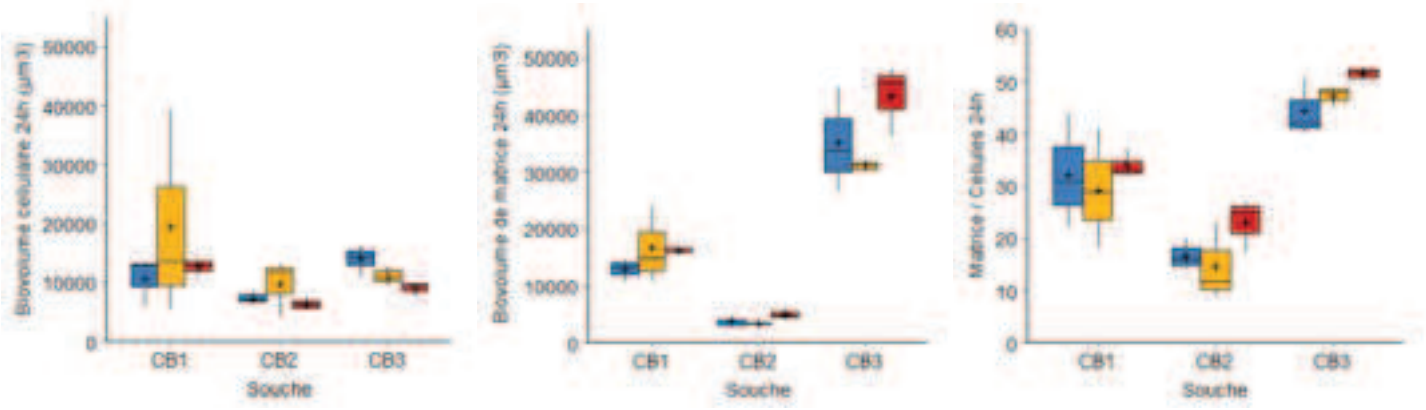
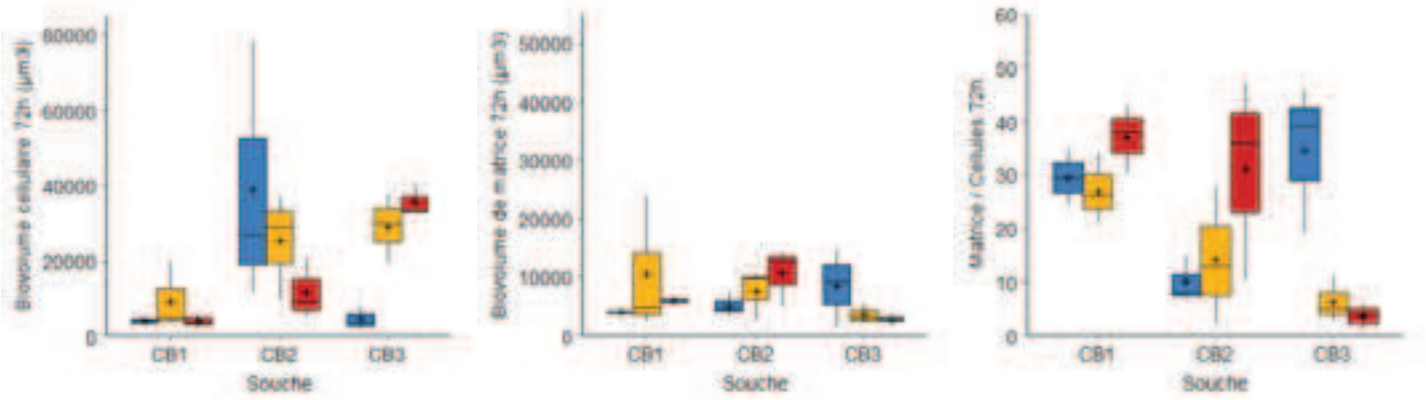


Figure 32 Volume de cellules (gauche) et de matrice (milieu), et rapport des volumes "matrice/cellules" (droite) des Thiomonas spp. CB1, CB2 et CB3 à 24h suivant différentes concentration d'As(III). Bleu = 0 mM ; Jaune = 2,67 mM ; Rouge = 5,33 mM. La croix "+" indique la moyenne.



.

Figure 33 Volume de cellules (gauche) et de matrice (milieu), et rapport des volumes "matrice/cellules" (droite) des Thiomonas spp. CB1, CB2 et CB3 à 72 h suivant différentes concentration d'As(III) à 72 h. Bleu = 0 mM ; Jaune = 2,67 mM ; Rouge = 5,33 mM. La croix "+" indique la moyenne.

3 - Ilots impliqués dans la synthèse de biofilm et la mobilité

Afin de rechercher des différences génomiques potentiellement impliquées dans les différences observées, les îlots génomiques des souches ont été analysés. Nous avons recherché si des GEIs étaient présents spécifiquement chez *Tm. arsenitoxydans* 3As et *Tm.* spp. CB1 et CB6 d'une part, et chez *Tm.* spp. CB2 et CB3 d'autre part. La comparaison des génomes des cinq souches a permis l'identification de plusieurs régions variables (RGPs, *Region of Genomic Plasticity*) dans chaque souche : 30 RGPs chez 3As, 32 chez CB1, 20 chez CB2, 39 chez CB3 et 31 chez CB6. Un RGP commun à 3As, CB1 et CB6 (RGP28, RGP23 et RGP20 respectivement) et absent de CB2 et CB3 contient des gènes *eps*, connus pour être impliqués dans la synthèse de matrice. Cet RGP a été appelé "îlot eps" (Figure 34).

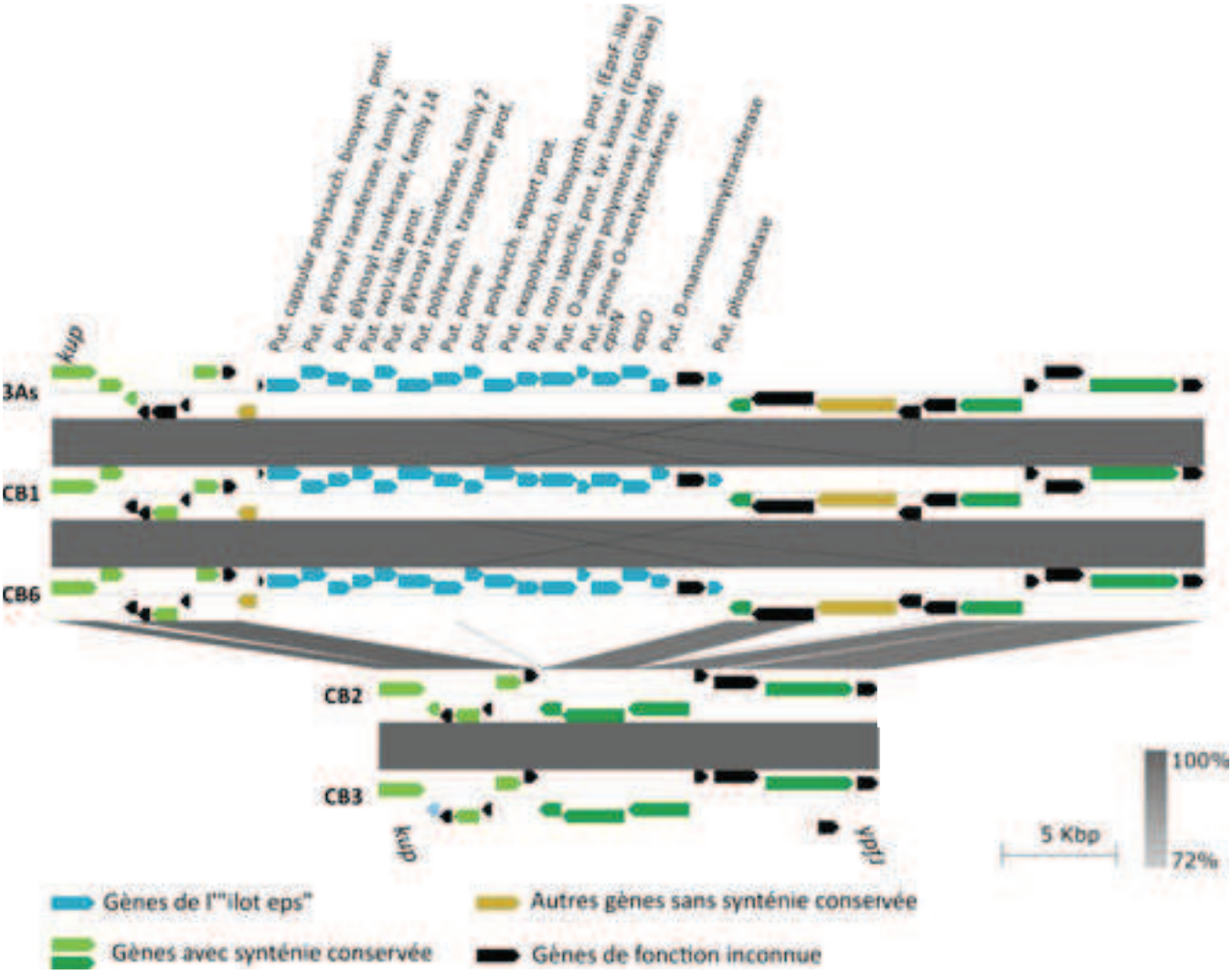


Figure 34 Synténie au niveau de l'îlot "eps". L'échelle grise indique le % d'identité nucléotidique (BlastN).

Deux autres RGPs d'intérêt sont présents chez *Tm.* spp. CB2 et CB3 mais absentes des trois autres souches. Un contient des gènes *rfb* (nommé "îlots rfb"), décrits comme étant impliqués dans la biosynthèse du dTDP-rhamnose, un précurseur des exopolysaccharides présents dans les biofilms (Barreto et al., 2005) (Figure 35). L'autre contient les gènes *cheBDR* et un gène codant une diguanylate cyclase/phosphodiesterase. Cet îlot pourrait être responsable de différence concernant la régulation de la mobilité et de la synthèse de biofilms.

L'îlot "eps" ne contient pas de gènes impliqués dans son éventuelle mobilité. Il pourrait soit avoir été acquis par un ancêtre commun aux souches 3As, CB1 et CB6, ou bien avoir été perdu par un ancêtre commun à CB2 et CB3. C'est également le cas de l'îlot "che" : gain par un ancêtre à CB2 et CB3 ou perte par un ancêtre commun à 3As, CB1 et CB6. De manière intéressante, l'îlot "rfb" de CB3 est par contre flanqué par deux transposases, et pourrait donc être toujours mobile.

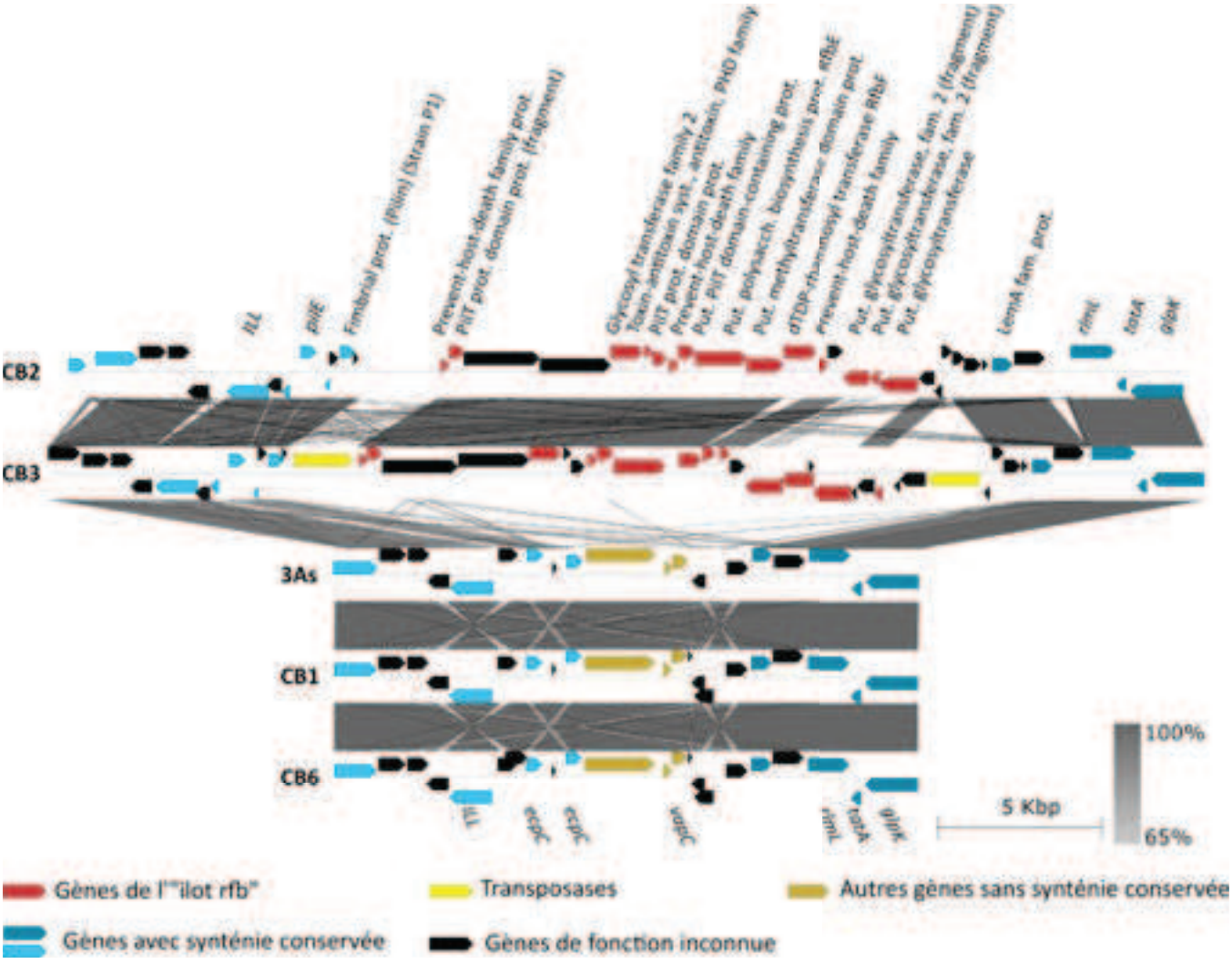


Figure 35 Synténie au niveau de l'îlot "rfb". L'échelle grise indique le % d'identité nucléotidique (BlastN).

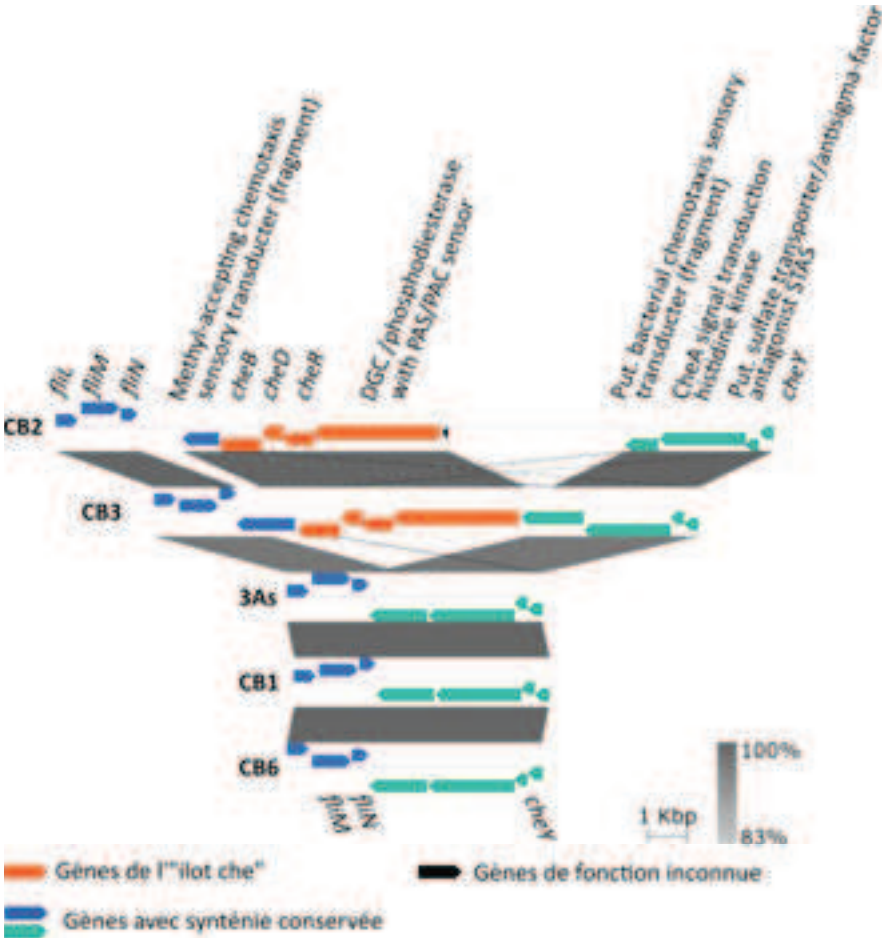


Figure 36 Synténie au niveau de l'îlot "che". L'échelle grise indique le % d'identité nucléotidique (BlastN).

L'absence de l'îlot "eps" chez *Tm.* spp. CB2 et CB3 et l'absence des îlots "rfb" et "che" de 3As, CB1 et CB6 pourrait en partie expliquer les différences observées entre ces deux groupes de souches. La coloration du biofilm au crystal violet et les tests de mobilité ont en effet permis de discerner ces deux groupes : *Tm.* spp. 3As, CB1 et CB6 étant mobiles en présence d'As(III) et produisant relativement peu de biofilm à l'inverse de *Tm.* spp. CB2 et CB3 non mobiles en gélose molle et produisant plus de biofilm. Il a cependant été remarqué en microscopie confocale que la souche *Tm.* sp. CB1 produisait une quantité de biofilm comparable à *Tm.* spp. CB2 et CB3 à 24h. Une des hypothèses est que *Tm.* sp. CB1 produirait un biofilm qualitativement différent des deux autres souches, plus sensible aux opérations de rinçage (crystal violet) et de renouvellement du milieu. Les éléments de la matrice produits par les gènes *eps* et par les gènes *rfb* pourraient en effet avoir des propriétés différentes. Les gènes *che* et celui codant la diguanylate cyclase sont des gènes impliqués dans la régulation de la mobilité et de la formation de biofilm, et seraient donc potentiellement impliqués dans les différences de mobilité observées sur gélose molle ou de synthèse de biofilms étudiées en microscopie confocale.

4 - Transcriptomique (RNAseq)

Dans le but de mettre en évidence des gènes potentiellement impliqués dans le développement en biofilm chez *Tm.* sp. CB2, notamment la réponse à l'As(III), une expérience de transcriptomique par RNAseq a été réalisée. Différentes conditions ont été comparées : 24 h, 48 h et 72 h en présence ou absence de 5,33 mM d'As(III). Les données obtenues étant en cours d'analyse, seules celles concernant les gènes potentiellement impliqués dans la formation de biofilms sont brièvement présentées ici. De manière intéressante, le profil d'expression au cours du temps des gènes des deux îlots portant les gènes *rfb* et *che* semble impacté par la présence d'As(III).

Le gène *rfbE* dont l'expression diminue au cours du temps en absence d'As(III) a au contraire son expression qui augmente entre 24 h et 48 h en présence d'As(III). De plus son expression est toujours supérieure en présence qu'en absence d'As(III). L'expression des glycosyl transferases est aussi augmentée en présence d'As(III). Ceci pourrait en partie expliquer pourquoi la quantité de matrice est plus importante en présence de cet élément chez *Tm.* sp. CB2.

Au niveau de l'îlot "che", les gènes sont d'une manière générale surexprimés en présence d'As(III) à l'exception de *cheD* et *cheR* sous exprimés à 48 h en présence de cet élément. L'expression de *cheB* et *cheR* diminue au cours du temps mais augmente entre 48 h et 72 h en présence d'As(III). Les gènes de ces deux îlots pourraient donc être en partie responsables des variations observées en présence d'As(III) chez *Tm.* sp. CB2. Afin de savoir si ces gènes sont également impliqués dans la formation de biofilm chez *Tm.* sp. CB3 et si les gènes de l'îlot "eps" possèdent un profil d'expression intéressant chez les souches *Tm. arsenitoxydans* 3As et *Tm.* spp. CB1 et CB6, il serait intéressant d'étendre l'expérience de RNAseq à ces souches.

IV - Conclusion

Les biofilms permettent de coloniser des niches écologiques via leurs propriétés adhésives, et de limiter les contraintes environnementales en protégeant les cellules. Cependant, la biosynthèse de biofilm est un processus complexe et connu pour être souvent variable d'une espèce à l'autre, ou même d'une souche à l'autre au sein d'une même espèce (Hall-Stoodley et al., 2004; O'Toole et al., 2000). Ce caractère variable reflète l'importance des biofilms dans l'adaptation à une niche particulière.

Les cinq souches de *Thiomonas* étudiées ici présentent des profils différents en termes de production de biofilms et de mobilité. Deux profils principaux peuvent être distingués. Un premier groupe comprend les souches *Tm.* spp. 3As, CB1 et CB6, mobiles sur gélose molle en présence d'As(III) et ne semblant pas être très influencées par l'As(III) pour la production de biofilm. Cependant, il n'est pas exclu que ces souches possèdent un biofilm "fragile", qui serait éliminé lors des expérimentations. Le deuxième groupe comprend les souches *Tm.* spp. CB2 et CB3, non-mobiles sur gélose molle, dont la production de biofilms est influencée par l'As(III). Le type de biofilms semble différent de celui du premier groupe dans la mesure où son aspect est différent en microscopie confocale. Leur régulation semble aussi différente, les volumes en réponse à l'As(III) suivant des tendances inverses. Les souches de ces deux groupes ont en commun des GEIs probablement impliqués dans la biosynthèse de biofilms, la mobilité et leur régulation. Ces GEIs diffèrent entre les deux groupes et ces différences sont cohérentes avec l'hypothèse selon laquelle les biofilms seraient de nature différente, dans le cas où leur synthèse impliquerait les gènes de l'îlot "eps" chez *Tm.* spp. 3As, CB1 et CB6, et des 'îlots "rfb" et "che" chez *Tm.* spp. CB2 et CB3.

Les différentes capacités de ces *Thiomonas* spp. à former des biofilms sont donc probablement en partie liées aux trois îlots identifiés. Les premières données du RNAseq montrent que chez CB2, l'As(III) semblent influencer l'expression des gènes des GEIs "che" et rfb". La confirmation de l'implication de ces derniers dans les différences observées expérimentalement nécessiterait de réaliser une expérience de RNAseq chez les autres souches et de développer des outils génétiques. Ceux-ci permettront par exemple de réaliser des expériences de complémentation fonctionnelle entre les différentes souches, ou d'effectuer de la mutagenèse ciblée. L'exploitation des données de transcriptomique devrait aussi permettre de mettre en évidence d'autres gènes impliqués dans le développement du biofilm, y compris des gènes dont la fonction n'est pas encore connue.

Conclusion - Chapitre 1

Au sein du DMA de Carnoulès coexistent des bactéries du genre *Thiomonas* spp. du groupe I et appartenant à la même espèce. Le séquençage de plusieurs souches isolées de ce site a permis de définir plusieurs gènes conservés et de construire ainsi un arbre phylogénétique avec une résolution importante. Les souches CB2 et CB3 divergent des souches *Tm. arsenitoxydans* 3As, et *Tm.* sp. CB1 et CB6. La souche CB2, dont le génome a été assemblé, a subi plusieurs remaniements génomiques de grande envergure comme l'a montré la comparaison de sa synténie avec les souches 3As et *Tm. intermedia* K12. La structure du génome des *Thiomonas* spp. peut donc évoluer indépendamment de la divergence phylogénétique par des remaniements génomiques.

La comparaison de ces génomes a confirmé la présence de nombreux îlots génomiques. Les cas décrits dans les trois parties précédentes montrent que les différences observées entre les souches concernant les GEIs sont corrélées avec des différences fonctionnelles. L'oxydation de l'As(III) et la dégradation de l'urée sont en relation avec la présence de ces deux composés sur le site. La présence de plusieurs copies de l'opéron *aioBA* est corrélée avec une meilleure résistance à l'As(III). L'avantage conféré par la dégradation de l'urée concerne à la fois la résistance à l'acidité et la précipitation des métaux du milieu dans les conditions de laboratoire. Les souches diffèrent enfin de par leur capacité à former des biofilms, ce qui pourrait également leur conférer des avantages dans certaines conditions, et correspondre à des adaptations à des niches particulières.

Les *Thiomonas* spp. ont donc évolué différemment via la perte et l'acquisition de GEIs mais également via d'importants remaniements de la structure de leur génome. La diversité fonctionnelle qui semble en résulter peut être interprétée comme des adaptations à des conditions particulières sur le site, ou comme un réservoir de diversité permettant de faire face à des changements des conditions du milieu. Dans les deux cas, il semble exister un lien entre les GEIs et le potentiel évolutif des *Thiomonas* spp. La suite des travaux présentée dans le chapitre 2 concerne des facteurs influençant ce potentiel adaptatif chez la souche *Tm.* sp. CB2.

Chapitre II - Influence de l'As(III) et des biofilms sur la variabilité génomique

I - Introduction

Les biofilms confèrent une protection face à divers stress et à la variabilité des conditions environnementales. En plus de former des biofilms, les bactéries sont également capables d'adapter leur physiologie via des processus de réponses au stress (Marles-Wright and Lewis, 2007). La réponse au stress en général et la formation de biofilms ont en commun le fait d'avoir un impact sur la variabilité génomique des bactéries. En réponse au stress, les bactéries expriment par exemple l'ADN polymérase IV dont le taux d'erreurs est plus élevé que l'ADN polymérase III (Foster, 2005; Galhardo et al., 2007) Dans le cas de la réponse SOS induite lors de cassures de l'ADN, la réparation peut faire intervenir des recombinaisons homologues susceptibles de causer des réarrangements génomiques (Baharoglu and Mazel, 2014). Il a été démontré qu'au sein des biofilms peut exister un stress oxydant intrinsèque entraînant des réarrangements (Boles and Singh, 2008). Les biofilms sont également susceptibles de favoriser des transferts génétiques horizontaux à cause de l'importante densité cellulaire favorisant les échanges par conjugaison (Madsen et al., 2012), et de la présence d'ADN extracellulaire dans la matrice qui peut être le substrat de processus de transformation (Hannan et al., 2010).

Au sein du DMA de Carnoulès, la présence d'importante concentration d'As(III) peut potentiellement causer de la variabilité génomique. L'As(III) est décrit comme étant à l'origine de stress oxydant, notamment à cause de son affinité pour les groupements dithiols présents sur les thioredoxines, glutathions et glutaredoxines. Dans une étude précédente, il avait été démontré qu'au sein des biofilms de *Tm.* sp. CB2 cultivée en présence d'As(III), pouvait émerger une sous-population de variants ayant acquis une résistance supérieure à la population de départ (Marchal et al., 2011). Les travaux présentés dans ce chapitre font directement suite à cette observation. Dans un premier temps, j'ai optimisé un protocole mis en place par Kelle Freel afin de déterminer l'influence des biofilms et de l'As(III) sur la fréquence d'apparition des variants. Ensuite, certains d'entre eux ont été isolés sur des concentrations croissantes d'As(III) à partir de biofilms cultivés en présence de différentes concentrations d'As(III). L'ensemble des isolements a été réalisé par K. Freel. Parmi les variants isolés, certains ont été sélectionnés sur la base de leur résistance à l'As(III) ou de leur capacité à former des biofilms. Ce travail de sélection ainsi que la détermination de la CMI des variants sélectionnés ont été effectués par Martina Péres au cours de son stage de master que j'ai supervisé. J'ai ensuite étudié les biofilms de trois variants en microscopie confocale afin de déterminer l'évolution de leur développement au cours du temps et l'effet de l'As(III) sur ce développement. Cette expérience au cours de laquelle j'ai effectué la préparation des cultures et l'analyse des résultats ont été réalisés en

partenariat avec l'UMR1319 (Micalis). Le génome de deux variants a ensuite été envoyé à séquencer puis q été intégré à la plateforme MaGe (Génoscope). Ceci a permis la comparaison de leur génome avec la souche originelle *Tm.* sp. CB2 afin d'identifier les mutations et réarrangements survenus et susceptibles d'être impliqués dans les nouvelles caractéristiques des variants. Enfin, dans le but d'identifier des gènes potentiellement impliqués dans l'apparition de ces variants, une étude de transcriptomique par RNAseq différentiel de *Tm.* sp. CB2 cultivée en biofilm en absence ou en présence de 5,33 mM d'As(III) à différents temps a été réalisée. Cette étude, a été réalisée avec la plateforme "Transcriptome et EpiGenome" de l'institut Pasteur. J'ai participé à cette expérience en préparant l'ensemble des cultures.

II - Matériels et méthodes

1 - Mesure du pourcentage taux de variants "super-résistants" à l'As(III)

Afin de déterminer l'influence de la croissance en biofilm et de la présence d'As(III) sur la variabilité génomique, le taux d'apparition de variants résistants à une concentration en As(III) supérieure ou égale à la CMI de départ a été mesuré. L'expérience a consisté à cultiver la souche *Thiomonas* sp. CB2 en absence ou en présence de différentes concentrations d'As(III), en condition planctonique ou en biofilm. A l'issue de ces cultures, les cellules ont été étalées sur milieu solide en absence d'As(III) pour connaître le nombre total de cellules vivantes, et en présence de 10,6 mM d'As(III) (CMI de *Tm.* sp. CB2) pour connaître le nombre de variants. Le rapport du nombre de variants par le nombre de cellules totales a ensuite permis de déterminer le pourcentage de variants dans les cultures.

a. Cultures

Une culture planctonique de la souche *Tm.* sp. CB2 a été réalisée en inoculant 20 mL de milieu m126 liquide avec une pré-culture de manière à obtenir une $D.O_{600nm}$ de 0,002. Cette culture a ensuite été incubée en condition agitée à 30 °C pendant 40 h, temps au bout duquel la culture est en phase exponentielle de croissance avec une $D.O_{600nm}$ d'environ 0,2 (Figure 37).

A partir de cette culture, des dilutions à $D.O_{600nm}$ = 0,002 ont été réalisées dans du m126 supplémenté ou non avec 2,67 mM ou 5,33 mM d'As(III). Ces dilutions ont servi à inoculer trois plaques 12 puits en polystyrène (Nunc) à raison de 1 mL/puits et de façon à obtenir 3 puits de chaque concentration en As(III) pour chaque plaque. Ces dernières ont ensuite été incubées à 30 °C en condition statique durant un temps différent : 24 h, 48 h ou 72 h à 30 °C. Dans les cas d'incubation de 48 h ou 72 h, le milieu a été renouvelé au bout de 24 h. En parallèle, la même culture planctonique de départ a servi à inoculer 9 cultures planctoniques supplémentées avec 0 mM, 2,67 mM ou 5,33 mM d'As(III), à raison de 3 cultures par concentration, toujours en diluants à $D.O_{600nm}$ = 0,002. Ces cultures ont également été incubées à 30 °C durant 24 h, 48 h ou 72 h.

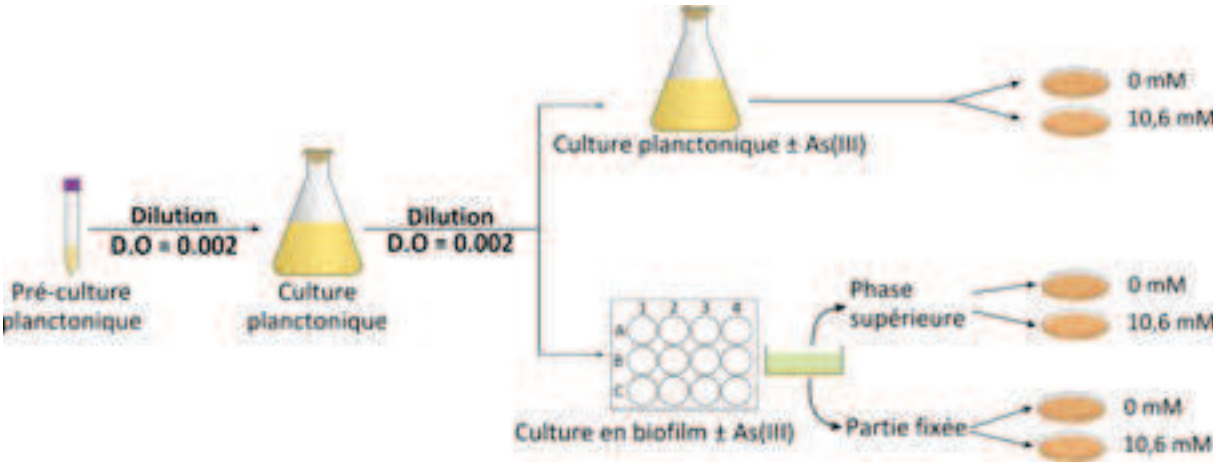


Figure 37 Schéma du protocole utilisé pour la mesure du taux d'apparition de variants "super-résistants" à l'As(III). Une culture planctonique inoculée à partir d'une pré-culture a été incubée à 30 °C pendant 40 h puis a elle-même servie à inoculer des cultures planctoniques et des cultures en biofilm après dilution à D.O = 0,002. Ces cultures ont été réalisées en absence ou en présence de 2,67 ou 5,33 mM d'As(III). Pour les biofilms, il y avait trois puits par concentration, et pour les cultures planctoniques, 3 cultures par concentration. Comme représenté par le schéma du puits en coupe à droite de la plaque 12 puits, le pourcentage de variants a été mesuré à la fois dans la phase supérieure et la partie fixée du biofilm. Des dilutions ont été réalisées avant l'ensemencement des milieux solides.

b. Mesures du pourcentage d'apparition de variants

Après incubation (24 h, 48 h ou 72 h), les phases liquides supérieures (phases supérieures) des cultures en biofilm ont été délicatement retirées des puits et transférées dans des tubes eppendorf de 5 mL. Dans chaque tube de 5 mL étaient mélangées les phases supérieures des trois puits d'une même concentration en As(III). Ensuite, 1 mL d'eau physiologique a été ajouté dans chaque puits et les biofilms ont été mis en suspension en les décrochant à l'aide d'un cône de 1 mL. Les suspensions obtenues pour chaque concentration ont également été transférées dans des tubes eppendorf de 5 mL. Enfin, 1 mL de chaque culture planctonique d'une concentration en As(III) donnée a également été transféré dans un tube eppendorf de 5 mL.

A partir des suspensions cellulaires ainsi obtenues, des dilutions ont été réalisées puis 40 µL de ces dilutions ont été étalées sur milieu m126 solide dans de petites boîtes de Pétri (diamètre : 60 mm) sans As(III) ou supplémenté avec 10,6 mM d'As(III). Cette concentration correspondant à la CMI en As(III) pour *Tm.* sp. CB2. Dans le cas des cultures de 24 h, et de la phase supérieure des biofilms cultivés avec As(III), les dilutions étalées étaient de 1/10 000. Dans les autres cas, les dilutions étaient de 1/100 000. Les boîtes ont ensuite été incubées à 30

°C pendant 10 jours.

Le pourcentage de variants résistants à la CMI d'As(III) de la population de départ a été calculé en faisant le rapport du nombre de colonies comptées sur les boîtes contenant 10,6 mM d'As(III) par le nombre de colonies comptées sur les boîtes sans As(III) puis en multipliant le tout par 100 :

(Nb. de colonies sur 10,6 mM d'As(III)) / (Nb. de colonies sans As(III)) x 100

L'expérience décrite ici a été réalisée en triplicats, c'est-à-dire avec trois cultures indépendantes au départ.

2 - Isolements de variants "super-résistants" à l'As(III)

Des variants ayant acquis une résistance à l'As(III) supérieure à celle de la population de départ après culture en biofilm ont été isolés afin de les caractériser. Dans le but d'obtenir des variants résistant à des concentrations élevées, plusieurs cycles de culture en biofilm ont été réalisés, chaque nouveau cycle étant démarré avec un variant ayant acquis une résistance plus élevée que celle des variants du cycle précédent. Comme dans le cas de la mesure du taux d'apparition de variants "super-résistants" à l'As(III), une culture planctonique inoculée avec une pré-culture a servi à inoculer une plaque 12 puits (Figure 38). La culture planctonique a été diluée à D.O = 0,02 et 1 mL/puits a été ajouté dans la plaque, supplémenté ou non avec 1,33 mM, 2,67 mM, 4 mM ou 5,33 mM d'As(III). Deux puits de chaque concentration ont été inoculés : A1 et A2 = 0 mM ; B1 et B2 = 1,33 mM ; C1 et C2 = 2,67 mM ; A3 et A4 = 4 mM ; B3 et B4 = 5,33 mM. Le milieu a été renouvelé au bout de 24 h. Comme précédemment, la partie supérieure des biofilms et celle fixée au fond des puits ont été séparées. Des dilutions ont ensuite été réalisées puis 40 µL ont été étalés sur m126 solide supplémenté avec 10,6 mM d'As(III) lors du premier cycle, puis avec des concentrations croissantes (allant jusqu'à 26,6 mM) lors de cycles suivants. Les variants isolés ont été remis en culture, puis utilisés pour démarrer un nouveau cycle et/ou stockés dans des cryotubes à - 80°C (après ajout de 7 % de DMSO).

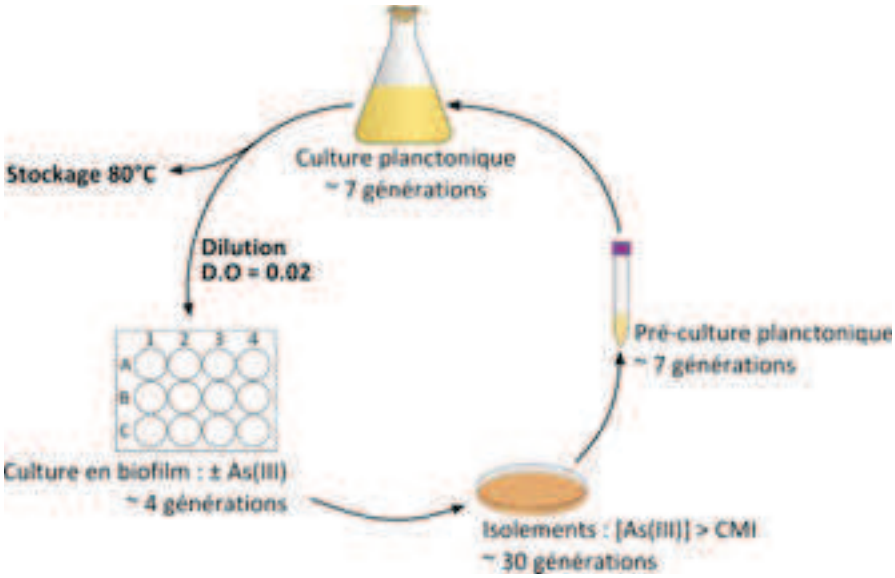


Figure 38 Schéma des cycles de biofilm réalisés pour l'isolement des variants "super-résistants" à l'As(III). Comme pour la mesure du taux d'apparition de variants, la culture en biofilm a été inoculée à partir d'une culture planctonique de 40 h, elle-même inoculée à partir d'une pré-culture. Les biofilms ont été cultivés pendant 72 h avec renouvellement du milieu à 24 h et des concentrations en As(III) de 0 mM, 1,33 mM, 2,67 mM, 4 mM, ou 5, 33 mM, à raison de 2 puits par concentration. La phase supérieure et la partie fixée des biofilms ont été séparées avant les isolements (non indiqué sur le schéma). Les variants isolés ont été remis en culture pour être ensuite stockés à - 80 °C et/ou démarrer un nouveau cycle. Le nombre de générations a été déduit des mesures des D.O pour les cultures planctoniques et en biofilms. Concernant les cultures sur milieu solide, l'estimation a été faite en considérant qu'une colonie contenait 10^9 cellules.

3 - Sélection et caractérisation de variants

Un nombre important de variants a été isolé sur différentes concentrations d'As(III) rendant impossible la caractérisation de tous. Une première étape a donc consisté à sélectionner ceux présentant un intérêt du point de vue de leur résistance accrue à l'As(III) ou de leur production de biofilms. Ce travail a été réalisé par Martina Péres lors de son stage de master 2 que j'ai encadré.

a. Sélection sur la base de la résistance à l'As(III)

Dans un premier temps, 18 variants conservés à - 80 °C ont été choisis sur la base de la haute concentration en As(III) sur laquelle ils avaient été isolés, suggérant une résistance accrue à l'As(III), et 17 autres de par l'aspect de leurs colonies, laissant présager d'éventuelles différences dans leur capacité à former des biofilms. Afin d'effectuer un premier tri concernant la résistance à l'As(III), les 35 variants ont tout d'abord été directement cultivés sur milieu m126

solide sans As(III) ou supplémenté avec une gamme de concentration allant de 10,6 mM à 26,6 mM avec un intervalle de 2,67 mM. Les cultures ont été incubées à 30°C pendant 12 jours puis la concentration maximale sur laquelle une croissance était visible a été notée.

b. Sélection sur base de la capacité à former des biofilms et de la mobilité

Certains variants ont été choisis pour une quantification de leur formation de biofilms par coloration au crystal violet. Parmi eux, certains ont été sélectionnés pour leur résistance à l'As(III) lors de l'étape précédente, et d'autres sur la base de la morphologie de leurs colonies. Une culture en milieu m126 liquide a tout d'abord été réalisée et incubée 40 h à 30°C. Des dilutions des cultures à $D.O_{600nm}$ = 0,002 ont été effectuées et supplémentées avec 0 mM, 2,67 mM et 5,33 mM d'As(III), puis 1 mL de ces dilutions a été transféré dans les puits d'une plaque 12 puits de manière à obtenir trois puits de chaque concentration. Les plaques ont été incubées en condition statique pendant 72 h avec renouvellement du milieu au bout de 24 h. Au bout de 72 h d'incubation, le milieu a été retiré des puits et les biofilms ont été délicatement rincés avec 1 mL d'eau physiologique (NaCl = 9 $g.L^{-1}$). Les plaques ont ensuite été séchées par incubation à 30°C pendant 20 min environ, puis 1 mL de crystal violet à 0,1 % a été ajouté dans chaque puits. Les plaques ont alors été à nouveau incubées à 30°C pendant 20 min puis le crystal violet a été retiré et l'excès enlevé par un rinçage avec 1 mL d'eau physiologique. 1 mL/puits d'éthanol à 95 % a été ajouté pour resuspendre le crystal violet fixé dans les biofilms puis des dilutions ¼ ont été réalisées dans des plaques 96 puits (transfert de 50 µL des puits de la plaque 12 puits dans 150 µL d'éthanol 95 % dans la plaque 96 puits). La $D.O_{595nm}$ de la plaque 96 puits a ensuite été mesurée avec un lecteur de plaque.

Les dilutions à $D.O_{600nm}$ = 0,002 sans As(III) utilisées pour l'inoculation des plaques 12 puits ont également été utilisées pour comparer la mobilité sur gélose à 0,3 % d'agar des variants entre eux et avec les souches *Tm.* spp. CB1 et CB2. Pour cela, trois dépôts de 1 µl de dilution ont été effectués sur des milieux m126 à 0,3% d'agar et supplémentés de 0 mM, 2,67 mM ou 5,33 mM d'As(III). Les milieux ont ensuite été mis à incuber à 30°C pendant 7 jours. En cas de mobilité, un halo était visible autour des dépôts à l'issue de l'incubation.

c. Détermination de la concentration minimale inhibitrice en As(III)

La CMI d'As(III) des variants choisis après les étapes de sélection précédentes a été mesurée plus précisément. Pour cela, des cultures planctoniques en phase exponentielle (incubées à 30 °C pendant 40 h) ont d'abord été diluées 10 fois dans la première colonne d'une plaque 96 puits, et une dilution en série de dix en dix a été réalisée dans les puits adjacents

jusqu'à 10^{-5}. Ces dilutions ont ensuite été répliquées sur m126 solide supplémenté ou non avec 13,3 mM, 16 mM, 18,6 mM, 20 mM, 21,3 mM, 22,6 mM ou 23,9 mM. Les boîtes ont ensuite été incubées à 30 °C pendant 10 jours.

d. Etude des biofilms des variants en microscopie confocale

Trois variants ont été choisis sur la base de leur profil de formation de biofilm déterminé par coloration au crystal violet. Une étude des biofilms de ces variants a été réalisée en microscopie confocale de la même manière que décrit dans la troisième partie du chapitre I.

Les biofilms ont été cultivés dans des plaques 12 puits à fond en verre de la même manière que décrit plus haut pour les colorations en crystal violet. Deux triplicats biologiques (trois cultures indépendantes) ont été réalisés : un pour les observations à 24 h, l'autre pour les observations à 72 h (avec renouvellement du milieu à 24 h). Les biofilms ont été marqués avec trois colorants différents : 2,5 µM de SYTO9 green (485/498) pour révéler les cellules viables ; 2,5 µM de iodure de propidium (305-538/617) pour révéler les cellules mortes ; 100 $\mu g.mL^{-1}$ de tetramethylrhodamine conjuguée à de la concanavaline A (tetramethylrhodamine-ConA) (555/580) pour révéler la matrice. Après ajout des colorants, les plaques ont été incubées à l'obscurité pendant 30 min.

L'observation et l'acquisition des images ont ensuite été réalisées à l'aide d'un microscope confocal (Leica TCS SP8, objectif 63 X à immersion). Deux z-stacks (série d'images sur un plan vertical) avec un intervalle de 1 µm ont été acquis pour chaque puits. Les longueurs d'ondes d'excitation étaient de 488 nm (SYTO9 green et iodure de propidium) et 561 nm (tetramethylrhodamine-ConA). Les plages de longueurs d'ondes émises étaient de 495 nm à 547 nm pour le SYTO9 green, de 657 nm à 757 nm pour l'iodure de propidium et de 567 nm à 598 nm pour la tetramethylrhodamine-ConA.

Les images ont ensuite été traitées pour obtenir une représentation en trois dimensions avec le logiciel IMARIS (Bitplane, Zürich, Suisse). Les biovolumes de cellules (SYTO9 green) et de matrice (tetramethylrhodamine-ConA) ont été quantifiés à partir des séries d'images avec le logiciel ICY (http://icy.bioimageanalysis.org).

4 - Génomique comparée des variants

Deux variants ont été jugés intéressants à la fois sur le critère de leur résistance accrue à l'As(III) et, pour l'un des deux, des différences observées concernant son habilité à former des biofilms. J'ai préparé l'ADN de ces variants et les ai envoyés au Génoscope. Les séquences de deux variants ont ensuite été intégrées à la plateforme MaGe et ont pu être comparées entre elles ainsi qu'avec celle de la souche originelle *Tm.* sp. CB2 grâce à l'outil Paloma. Cet outil permet de lister les différentes mutations accumulées par chaque variant (SNPs, insertions et délétions) et de savoir si elles sont communes ou spécifiques aux variants. La profondeur de lecture du séquençage et l'emplacement des mutations peuvent également être visualisés.

5 - Transcriptomique (RNAseq) chez *Tm.* sp. CB2

Afin de rechercher des processus pouvant être impliqués dans l'apparition des variants au sein des biofilms de *Thiomonas* sp. CB2, une expérience de RNAseq a été réalisée chez cette souche.

a. Cultures

Les biofilms ont été cultivés dans des plaques 6 puits sans As(III) ou avec 5,33 mM d'As(III) pendant 24 h, 48 h et 72 h, avec renouvellement du milieu après 24 h. Au total, 8 plaques ont été utilisées pour chaque condition et l'expérience a été réalisée en triplicat (3 cultures indépendantes). Après l'incubation à 30°C, le milieu a été délicatement retiré des puits et remplacé par 1,5 mL d'eau physiologique à 4°C. Les biofilms ont alors été décollés du fond des puits à l'aide de grattoirs à cellules (Biologix) afin de mettre en suspension les cellules dans l'eau physiologique. Cette étape a été réalisée sur glace pour éviter la dégradation des ARN. Les suspensions ont ensuite été centrifugées à 8000 *x g* pendant 15 min à 4°C et après élimination des surnageants, les culots cellulaires ont été stockés à -80°C.

b. Extractions d'ARN et séquençage

Les extractions d'ARN ont été réalisées par Sandrine Koechler. Après décongélation à température ambiante, 60 µL d'EDTA (0,5 M ; pH 7,6) à 4°C ont été ajoutés dans chaque tube. Les cellules ont ensuite été reprises dans 400 µL de solution de resuspension (20 % glucose ; 25 mM Tris ; 10 mM (v/v) EDTA) afin de lyser les cellules par choc osmotique. Les suspensions ont ensuite été transférées dans des tubes contenant 2 mL de phénol, 0,5g de microbilles et 500 µL de phénol acide. Les tubes ont été agités à l'aide d'un broyeur (FastPrep) durant deux cycles

de 30 secondes, puis centrifugés à 16 000 *x g* pendant 5 minutes à 4°C. Les phases aqueuses ont alors été transférées dans de nouveaux tubes puis 1 mL de trizol a été ajouté. La solution obtenue a été homogénéisée à l'aide d'une pipette jusqu'à que l'aspect devienne translucide, et 200 µL de chloroforme / alcool isoamylique ont ensuite été ajoutés. Le mélange obtenu a de nouveau été homogénéisé à la pipette, puis après avoir reposé 3 min à température ambiante, les tubes ont été centrifugés à 16 000 *x g* pendant 5 min à 4°C. La phase aqueuse a alors été transférée dans de nouveaux tubes et 200 µL de chloroforme / alcool isoamylique ont à nouveau été ajoutés. Le contenu des tubes a été homogénéisé à l'aide d'une pipette puis les tubes ont été centrifugés à 16 000 *x g* pendant 5 min à 4 °C. Les surnageants ont été transférés dans de nouveaux tubes et les ARN contenus dans cette phase ont été purifiés sur colonne (Kit Qiagen RNeasy Plus Mini Kit) puis envoyés à séquencer au Génoscope.

III - Résultats et discussion

1 - Mesure du pourcentage de variants "super-résistant" à l'As(III)

En absence d'As(III) à 24 h, le pourcentage de variant est < 2 % quelle que soit la condition de culture (Figure 39). Au bout de 48 h, cette proportion passe à 25 % dans le cas des cultures planctoniques et de la partie fixée du biofilm, et à 50 % dans la phase supérieure mais avec d'importantes variations entre les réplicats. A 72 h, la proportion de variants en absence d'As(III) augmente à 30 % pour la culture planctonique mais diminue à 15 % pour les deux phases de la culture en biofilm. De manière intéressante, au bout de 72 h, la proportion de variants en absence d'As(III) est plus importante dans les cultures planctoniques qu'en biofilm. Il y a donc une émergence naturelle de variants dans la population de *Thiomonas* sp. CB2 indépendamment de la présence d'As(III) et du type de culture.

En présence d'As(III), le pourcentage de variants à 24 h des cultures planctoniques est nul ou quasi nul. Dans la partie fixée des biofilms, il augmente à 50 % en présence d'au moins 5,33 mM, avec cependant une variabilité importante. Dans la phase supérieure du biofilm, le pourcentage de variants est supérieur avec 2,67 mM d'As(III) (~20 %) et atteint plus de 75 % avec 5,33 mM. Au bout de 48 h, la culture planctonique contient une proportion de variants aux alentours de 20 % avec une concentration en As(III) de 2,67 mM et 30 % avec 5,33 mM. Dans la partie fixée du biofilm, le taux est d'environ 50 % quelle que soit la concentration, mais avec une variabilité importante. Dans la phase supérieure, les proportions de variants sont respectivement d'environ 30 % et 40 % pour 2,67 mM et 5,33 mM. A 72 h, le taux de variants augmente légèrement dans la phase planctonique, avec environ 30 % et 45 % pour 2,67 mM et 5,33 mM d'As(III). Il est d'environ 45- 50 % en présence de 2,67 mM et 5,33 mM d'As(III) dans la partie fixée du biofilm. Enfin, dans la phase supérieure, le taux de variants en présence de 2,67 mM est similaire à celui mesuré à 48 h (environ 30 %), mais est supérieur à 75 % en présence de 5,33 mM d'As(III).

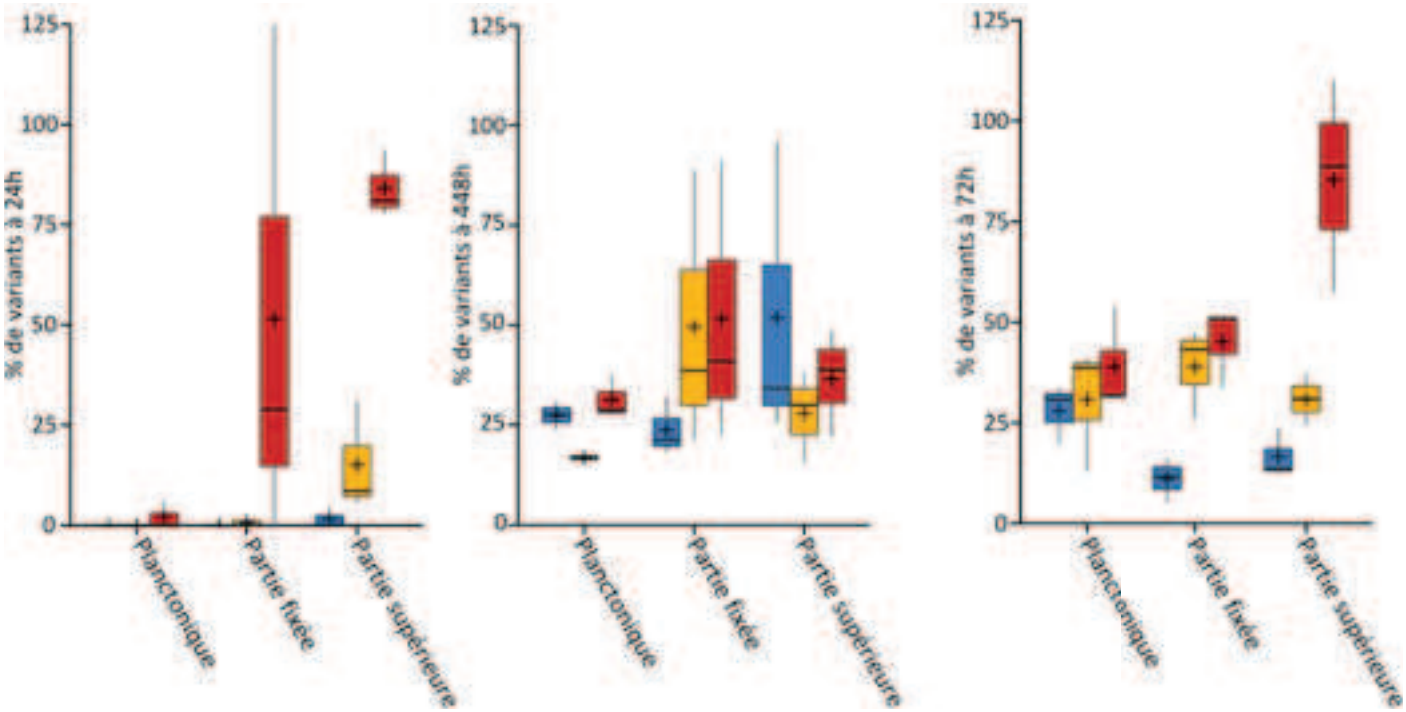


Figure 39 Proportion (%) d'apparition de variants "super-résistants" à l'As(III) chez Tm.CB2 à 24 h (gauche), 48 h (milieu) et 72 h (droite) en condition planctonique, et biofilm (partie fixée et phase mobile du biofilm). Bleu = 0 mM d'As(III) ; Jaune = 2,67 mM d'As(III) ; Rouge = 5,33 mM d'As(III). La croix "+" indique la moyenne.

D'une manière générale la présence d'As(III) est positivement corrélée avec le pourcentage de variants dans les deux types de culture, avec cependant des profils différents. Dans la culture planctonique, les variants apparaissent à 48 h puis leur pourcentage stagne et l'As(III) ne semble pas avoir beaucoup d'effet. Dans la partie fixée des biofilms, les pourcentages mesurés à 2,67 mM et 5,33 mM d'As(III) sont supérieurs à ceux mesurés sans As(III). A 72 h, une diminution de la proportion de variants en absence de cet élément dans la partie fixée et au niveau de la phase supérieure du biofilm pourrait soit être due à un taux d'apparition de ces variants plus faible en biofilm sans As(III), soit à plus de difficultés pour les variants à s'imposer dans la population. De la même manière, les pourcentages plus importants en présence d'As(III) pourraient soit résulter d'une fréquence d'apparition plus importante liée par exemple par l'induction de variabilité génomique, ou bien résulter d'une sélection de ces variants. Dans la partie supérieure du biofilm, une diminution est observée à 48 h à 5,33 mM d'As(III), puis une nouvelle augmentation se produit à 72 h. Celle-ci peut facilement s'expliquer par l'élimination des cellules lors du renouvellement du milieu à 24 h.

Une partie des cellules de cette phase supérieure pourrait avoir pour origine la dissémination de cellules de la partie fixée. Dans ce cas, les cellules disséminées depuis la partie

fixée pourraient être une sous-population redevenue planctonique et seraient préférentiellement des cellules variantes. Dans une étude précédente des biofilms de *Tm.* CB2, il avait était observé qu'en présence d'As(III) des cavités se formaient au centre des biofilms par des évènements de mort cellulaire et que ces cavités étaient ensuite colonisées par des cellules mobiles (Marchal et al., 2011). L'hypothèse avait été émise que ces cellules mobiles pourraient correspondre aux variants identifiés. Les résultats obtenus ici semblent aller dans le sens de cette hypothèse. En conclusion, même si l'apparition des variants n'est pas spécifique des biofilms, ces résultats montrent que leur pourcentage augmente en présence d'As(III) dans les biofilms.

2 - Isolement de variants "super-résistants" à l'As(III)

Afin d'étudier les processus impliqués dans l'apparition de variants ayant acquis une résistance accrue à l'As(III) en biofilm, et ainsi de mieux comprendre le potentiel adaptatif de *Tm.* sp. CB2 face à l'As(III), quelques variants ont été isolés et caractérisés. Pour cela, plusieurs cycles de biofilms ont été effectués et au terme de chaque cycle, des isolements ont été réalisés sur des concentrations croissantes en As(III). Un total de quatre cycles de biofilms a été réalisé, au terme desquels 107 variants ont été isolés sur des concentrations allant de 10,6 mM (CMI de *Tm.* sp. CB2) à 26,6 mM (Figure 40 et Tableau 2). Les variants isolés ont été nommés selon la partie du biofilm de laquelle ils sont issus, du numéro du biofilm et du puits duquel ils sont issus. Ainsi, les préfixes "Sup" et "Bio" indiquent qu'ils sont issus de la partie supérieure ou fixée du biofilm, le chiffre qui suit indique le numéro du biofilm et enfin le chiffre et la lettre indique le puits. Par exemple, Bio17 B3 signifie "issu de la partie fixée du puit B3 du biofilm N°17". Un chiffre précédé d'un "#" est parfois ajouté dans le cas où plusieurs variants ont été isolés du même puits.

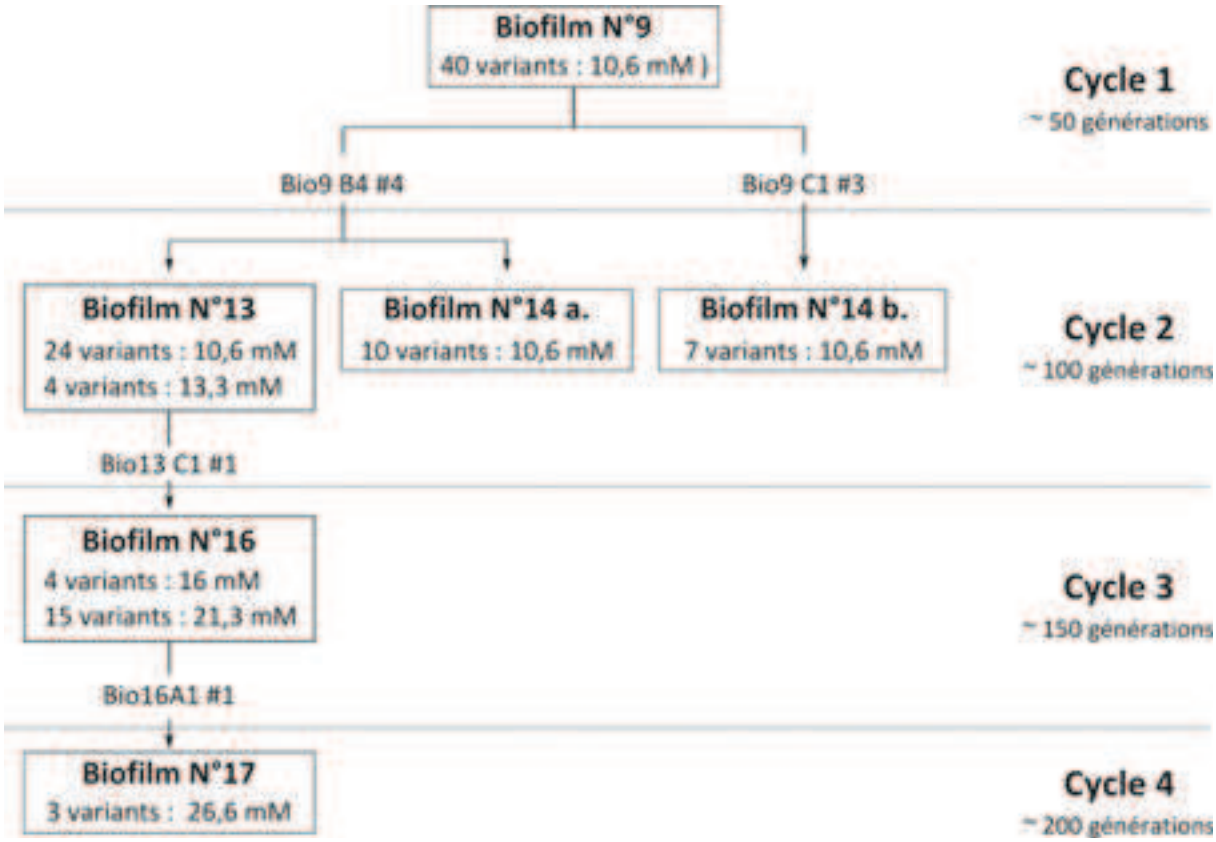


Figure 40 Schéma des variants isolés. Quatre cycles de biofilms ont été réalisés. Les premiers isolements ont été réalisés sur 10,6 mM d'As(III) (Biofilms N°9 et N°14). A partir du biofilm N°13, les isolements ont été réalisés sur des concentrations croissantes en As(III). Un variant isolé sur la concentration la plus élevée a alors été utilisé pour démarrer un nouveau cycle (Bio13 C1 #1 isolé sur 13,3 mM et Bio16A1 #1 isolé sur 21,3mM). Les trois derniers variants ont été isolés sur 26,6 mM lors du quatrième cycle.

Lors du 1er cycle (Biofilm N°9), 40 variants ont été isolés de la partie fixée des biofilms sur 10,6 mM d'As(III), et deux ont été utilisés pour démarrer un deuxième cycle de biofilm. "Bio9 B4 #4" a servi à inoculer les biofilms N° 13 et N°14a, alors que "Bio9 C1 #3", a servi à inoculer le biofilm N°14b.

Lors du 2ème cycle, 10 variants ont été isolés sur des boîtes contenant 10,6 mM d'As(III) à partir du biofilm N°14a., dont 7 isolés à partir de la phase supérieure du biofilm. Dans le biofilm N°14b, 7 variants ont été isolés de la partie fixée du biofilm sur des boîtes contenant 10,6 mM d'As(III), tous ces variants ont été isolés à partir de la partie fixée du biofilm. Aucun des variants des biofilms N°14 n'a été utilisé pour démarrer d'autres cycles. Cependant, l'aspect des colonies de ces biofilms paraissait différent des autres, leurs tailles étant légèrement plus grandes et certaines étaient irrégulières. Ces caractéristiques pouvant suggérer des différences dans leur capacité à former des biofilms, ces variants ont été conservés. Dans le biofilm N°13, 24 ont été isolés sur 10,6 mM d'As(III) et 4 sur 13,3 mM d'As(III), tous isolés à partir de la

partie fixée du biofilm. Le variant" Bio13 C1 #1" isolé sur 13,3 mM a été utilisé pour démarrer un troisième cycle de biofilm, le biofilm N°16.

Lors du 3ème cycle, 7 variants ont été isolés de la partie fixée du biofilm dont 4 sur 16 mM et 3 sur 21,3 mM d'As(III), et 12 variants ont été isolés de la partie supérieure des biofilms sur 21,3 mM d'As(III). Le variant "Bio16 A1 #1" isolé sur 16 mM a été utilisé pour un 4ème et dernier cycle de biofilms, le biofilm N°17. Dans ce dernier biofilm, trois variants ont été isolés sur 26,6 mM : "Sup17 B1", "Bio17 A3" et "Bio17 B3".

A chaque cycle de biofilms, plusieurs variants ont pu être isolés sur des concentrations croissantes en As(III). Au bout du troisième cycle, soit environ 150 générations, 15 variants ont pu être isolés sur une concentration égale au double de la CMI de la population de départ. La souche *Tm.* sp. CB2 cultivée en biofilm a donc un important potentiel adaptatif face à l'As(III). Les étapes suivantes ont consisté à sélectionner parmi ces variants ceux présentant le plus d'intérêt pour ensuite les caractériser afin de déterminer leur diversité et l'origine de leur résistance accrue à l'As(III).

Biofilm	Concentration en As(III) dans les puits				
	0 mM	1,33 mM	2,67mM	4 mM	5,33 mM
N°9	Bio9 A1 #1	Bio9 B1 #1	Bio9 C1 #1	Bio9 A3 #1	Bio9 B3 #1
	Bio9 A1 #2	Bio9 B1 #2	Bio9 C1 #2	Bio9 A3 #2	Bio9 B3 #2
	Bio9 A1 #3	Bio9 B1 #3	Bio9 C1 #3	Bio9 A3 #3	Bio9 B3 #3
	Bio9 A1 #4	Bio9 B1 #4	Bio9 C1 #4	Bio9 A3 #4	Bio9 B3 #4
	Bio9 A2 #1	Bio9 B2 #1	Bio9 C2 #1	Bio9 A4 #1	Bio9 B4 #1
	Bio9 A2 #2	Bio9 B2 #2	Bio9 C2 #2	Bio9 A4 #2	Bio9 B4 #2
	Bio9 A2 #3	Bio9 B2 #3	Bio9 C2 #3	Bio9 A4 #3	Bio9 B4 #3
	Bio9 A2 #4	Bio9 B2 #4	Bio9 C2 #4	Bio9 A4 #4	Bio9 B4 #4
N°13	Bio13 A1 #2	Bio13 B1 #1	***Bio13 C1 #1**	***Bio13 A3 #1**	***Bio13 B3 #1**
	Bio13 A1 #3	Bio13 B1 #2	Bio13 C2 #4	Bio13 A4 #1	***Bio13 B3 #2**
	Bio13 A1 #5	Bio13 B1 #3	Bio13 C2 #5	Bio13 A4 #2	Bio13 B3 #6
	Bio13 A1 #6	Bio13 B1 #4	Bio13 C2 #7	Bio13 A4 #4	
	Bio13 A1 #7	Bio13 B1 #5	Bio13 C2 #8	Bio13 A4 #5	
		Bio13 B1 #6		Bio13 A4 #6	
		Bio13 B1 #7		Bio13 A4 #8	
		Bio13 B1 #8			
N°14a.		Bio14 B1 #1	Bio14 C2 #1	Sup14 A3 #1	Sup14 B4 #1
			Sup14 C2 #1	Sup14 A3 #2	Sup14 B4 #2
				Bio14 A3 #1	
				Sup14 A4 #1	
				Sup14 A4 #2	
N°14b.	Bio14 A1 #1		Bio14 C2 #1	Bio14 A3 #1	
	Bio14 A1 #2			Bio14 A3 #2	
	Bio14 A1 #3			Bio14 A3 #4	
N°16	****Bio16 A1 #1**	****Bio16 B1 #1b**	****Bio16 C1 #1**		****Bio16 B3 #1**
	Sup16 A1 #2	Bio16B1 #1	Sup16 C1 #1		Bio16 B3 #1
	Sup16 A1 #3		Sup16 C1 #2		Bio16 B3 #2
			Sup16 C1 #3		Sup16 B3 #1
			Sup16 C1 #5		Sup16 B3 #2
			Sup16 C1 #6		Sup16 B3 #3
					Sup16 B3 #5
					Sup16 B3 #6
N°17		Sup17 B1		Bio17 A3	Bio17 B3

Concentration d'As(III) utilisée pour les isolements :

		Exceptions :
Biofilm N°9 :	10,6 mM	
Biofilm N°13 :	13,3 mM	* 13,3 mM
Biofilms N°14 :	10,6 mM	
Biofilm N°16 :	21,3 mM	** 16 mM
Biofilm N°17 :	26,6 mM	

Tableau 2 Liste des variants isolés dans les biofilms lors des quatre cycles. Chaque colonne correspond à la concentration d'As(III) dans les puits lors de la culture des biofilms. Sous le tableau sont indiquées les concentrations auxquelles ont été isolés les variants.

3 - Sélection et caractérisation de variants

a. Sélection de variants sur la base de leur résistance à l'As(III)

Parmi les 107 variants isolés, certains ont été choisis pour être étudiés plus en détail. Un premier tri a été réalisé avec comme critère la concentration en As(III) sur laquelle les variants ont été isolés : les 15 variants isolés du biofilm N°16 (3ème cycle) sur des boîtes contenant 21,3 mM ainsi que les trois du biofilm N°17 (4ème cycle) isolés sur 26,6 mM ont été choisis. L'ensemble des variants des deux biofilms N°14 (2ème cycle) ont également été choisis à cause de l'aspect différent de leurs colonies. Au total, ce sont donc 35 variants qui ont été sélectionnés.

Parmi les 35 variants choisis, un deuxième crible grossier a consisté à les cultiver directement du stock -80 °C sur milieu gélosé supplémenté de différentes concentrations d'As(III). (Tableau 3) Au bout de 12 jours, tous les variants des biofilms N°14 à l'exception de deux se sont développés jusqu'à 13,3 mM. Trois variants ont été choisis du biofilm N°14a : "Bio14 A3#1", "Sup14a A3 #1" et "Sup14a A4 #1", isolés sur des boîtes contenant 10,6 mM d'As(III). Parmi les variants du biofilm N°16, un s'est développé jusqu'à 26,6 mM, deux jusqu'à 21,3 mM et les autres jusqu'à 13,3 mM. Trois variants ont été choisis : "Sup16 B3 #1", "Sup16 A1 #3" et "Bio16 B1 #1" qui se sont respectivement développés jusqu'à 26,6 mM, 21,3 mM et 13,3 mM d'As(III). Enfin, parmi les trois variants du biofilm N°17, "Bio17 A3" et "Bio17 B3" ont été choisis car ils se sont développés jusqu'à 21,3 mM alors que "Sup17 B1" ne s'est développé que jusqu'à 13,3 mM. Au total, 8 variants ont donc été sélectionnés.

			Temps de culture		
Variants	**Biofilm**	**Cycle**	**J+3**	**J+5 / J+7**	**J+12**
Sup14 a.1			0 mM	0 mM	13,3 mM
Sup14 a.2			0 mM	0 mM	13,3 mM
Sup14 a.3			0 mM	0 mM	13,3 mM
Sup14 a.4			0 mM	0 mM	13,3 mM
Sup14 a.5			n.d	13,3 mM	13,3 mM
Sup14 a.7			n.d	13,3 mM	13,3 mM
Sup14 a.8			n.d	13,3 mM	13,3 mM
Sup14 a.9			n.d	13,3 mM	13,3 mM
Bio14 a.10	**N°14**	**2**	n.d	13,3 mM	13,3 mM
Bio14 b.1			n.d	0 mM	13,3 mM
Bio14 b.2			n.d	0 mM	0 mM
Bio14 b.3			n.d	13,3 mM	13,3 mM
Bio14 b.4			n.d	13,3 mM	13,3 mM
Bio14 b.5			n.d	0 mM	13,3 mM
Bio14 b.6			n.d	13,3 mM	13,3 mM
Bio14 A3 #1			n.d	0 mM	13,3 mM
Bio14 C2 #1			n.d	n.d	n.d
Bio16 B1			13,3 mM	13,3 mM	13,3 mM
Bio16 B3 #1			13,3 mM	13,3 mM	13,3 mM
Bio16 B3 #2 800			13,3 mM	13,3 mM	13,3 mM
Sup16 A1 #2			n.d	13,3 mM	21,3 mM
Sup16 A1 #3			n.d	21,3 mM	21,3 mM
Sup16 B3 #1			n.d	13,3 mM	26,6 mM
Sup16 B3 #2 m126			n.d	13,3 mM	13,3 mM
Sup16 B3 #3	**N°16**	**3**	n.d	13,3 mM	13,3 mM
Sup16 B3 #5			n.d	13,3 mM	13,3 mM
Sup16 B3 #6			n.d	13,3 mM	13,3 mM
Sup16 C1 #1			n.d	13,3 mM	13,3 mM
Sup16 C1 #2			n.d	13,3 mM	13,3 mM
Sup16 C1 #3			n.d	13,3 mM	13,3 mM
Sup16 C1 #5			n.d	13,3 mM	13,3 mM
Sup16 C1 #6			n.d	13,3 mM	13,3 mM
Sup17 B1			13,3 mM	13,3 mM	13,3 mM
Bio17 B3	**N°17**	**4**	13,3 mM	13,3 mM	21,3 mM
Bio17 A3			13,3 mM	13,3 mM	21,3 mM

Tableau 3 Premier crible grossier après culture des variants directement du stock -80 °C sur des boîtes contenant différentes concentrations d'As(III). La concentration maximale sur laquelle les variants se sont développés après 3, 7 et 12 jours de culture a été notée. n.d = non déterminée.

La CMI en As(III) des huit variants sélectionnés a été déterminée plus précisément par un test "en goutte" consistant à repiquer une série de dilution des variants sur des milieux solides, supplémentés avec des concentrations croissantes en As(III). Comme il pouvait être attendu, les deux variants du biofilm N°17 isolés sur 26,6 mM possèdent la CMI la plus élevée, qui est de 23,9 mM. Cependant, le variant "Sup16 B3#1" issu de biofilm N°16 et isolé sur 21,3 mM a également une CMI de 23,9 mM. De manière plus surprenante, la CMI du variant "Bio16 B1 #1" n'est que de 16 mM. Les variants du biofilm N°14a ont une CMI de 21,3 mM et celui du N°14b de 18,6 mM, c'est-à-dire inférieure aux autres variants mis à part "Bio16 B1 #1". Pour tester la stabilité de leur résistance à l'As(III), les huit variants ont été repiqués sans As(III) puis leur CMI à nouveau mesurée. Après repiquage en milieu liquide sur une trentaine de générations, les CMI étaient identiques et semblent donc stables. La stabilité de cette résistance suggère une origine génétique de celle-ci et non une physiologie particulière des variants.

Variants	Biofilm N°	Cycle N°	Isolé sur (As(III))	CMI As(III) en mM
Sup14a A3 #1	14a	2	10 ,6 mM	21,3
Sup14a A4 #1	14a	2	10, 6 mM	21,3
Bio14b A3 #1	14b	2	10,6 mM	18,6
Sup16 A1 #3	16	3	21,3 mM	23,9
Bio16 B1 #1	16	3	21,3 mM	16
Sup16 B3 #1	16	3	21,3 mM	23,9
Bio17 A3	17	4	26,6 mM	23,9
Bio17 B3	17	4	26,6 mM	23,9

Tableau 4 CMI en As(III) des huit variants sélectionnés pour leur résistance accrue à l'As(III) (Biofilms N°16 et N°17) ou de la morphologie de leurs colonies (biofilms N°14)

b. Quantification de la formation de biofilms et détermination des capacités de nage des variants

Une manière pour les variants de mieux résister à l'As(III) serait de synthétiser plus de biofilms. La capacité à former des biofilms des huit variants sélectionnés a donc été déterminée à différentes concentrations en As(III) par coloration au crystal violet et comparée à la souche originelle *Tm.* sp. CB2, ainsi qu'à *Tm*. sp. CB1 qui possède des capacités différentes.

Trois profils généraux peuvent être distingués parmi ces huit variants (Figure 41). Un premier profil est celui des variants du biofilm N°14, qui produisent plus de biofilms que les autres, notamment en présence de 2,67 mM d'As(III). Ces variants avaient été en partie sélectionnés à cause de l'aspect de leurs colonies qui suggérait justement des capacités de formation de biofilms différentes. Un deuxième profil est celui du variant "Sup16 B3 #1" qui produit très peu de biofilms, de façon comparable à *Tm.* sp. CB1. Un troisième profil est celui des variants restants, similaire à celui de *Tm.* sp. CB2.

Comme la formation de biofilms et la mobilité sont deux processus liés, la mobilité de ces variants a ensuite été testée sur gélose molle (0,3 % d'agar) supplémentée ou non avec 2,67 ou 5,33 mM d'As(III). Aucun des variants ne s'est avéré mobile, indiquant que d'après ce test, les capacités de mobilité de ces variants n'ont pas été affectées en comparaison à *Tm.* CB2.

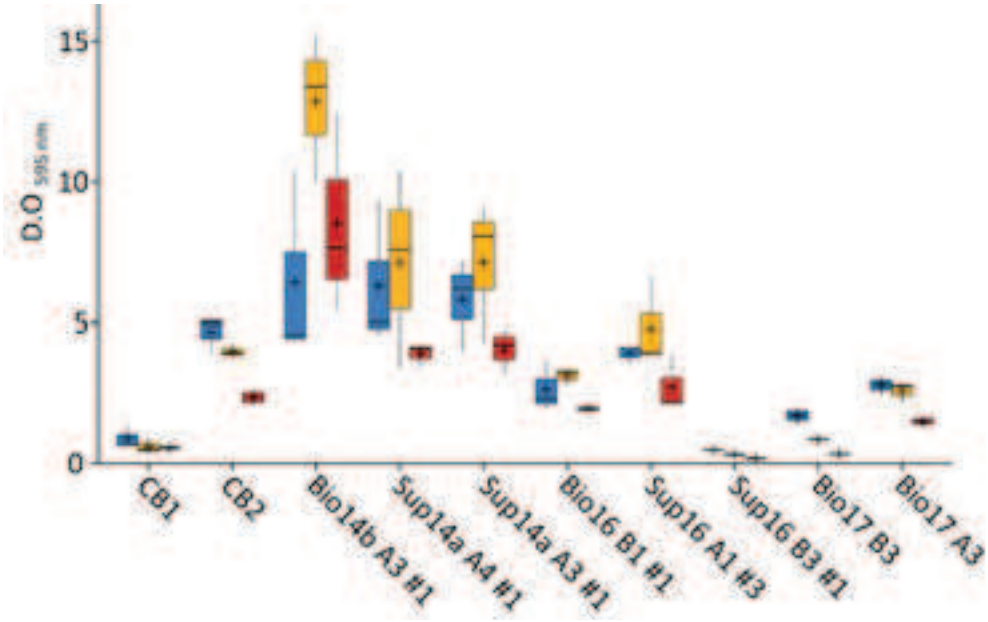


Figure 41 Quantification des biofilms des huit variants sélectionnés et des souches Tm. spp. CB1 et CB2 par coloration au crystal violet. Bleu = 0 mM d'As(III), Jaune = 2,67 mM d'As(III), Rouge = 5,33 mM d'As(II). La croix "+" indique la moyenne.

c. Etude des biofilms en microscopie confocale

Dans le but d'étudier plus en détail les différentes capacités des variants mis en évidence précédemment concernant leur formation de biofilms, une expérience en microscopie confocale permettant de distinguer les cellules et la matrice a été réalisée. Pour cela, un variant de chaque profil déterminé par coloration au crystal violet a été choisi : "Bio14b A3 #1", "Sup16 B3 #1" et "Bio17 B3". Pour plus de clarté, ces trois variants sont respectivement nommés "Bio14", "Sup16" et "Bio17" dans la suite du manuscrit. Les observations ont été réalisées en absence ou en présence de 5,33 mM d'As(III), à 24 h (Figure 42) et 72 h (Figure 43).

Comparé à la souche *Tm.* sp. CB2, le variant Bio14 synthétise plus de matrice à 72 h, notamment en présence d'As(III) (Figure 43). Le variant Sup16 qui semblait ne pas produire beaucoup de biofilm d'après la coloration au crystal violet, forme de nombreux amas de matrice à 72 h, moins volumineux que ceux de Bio14 mais relativement nombreux (Figure 43). Le variant Bio17 produit plus de matrice que les autres, notamment à 24 h (Figure 42). A 72 h, les structures formées par la matrice sont plus volumineuses en présence d'As(III), comme pour Bio14 (Figure 43). Les quantités de biofilms observées pour les variants Sup16 et Bio17 sont donc plus importantes que celles attendues d'après la coloration en crystal violet.

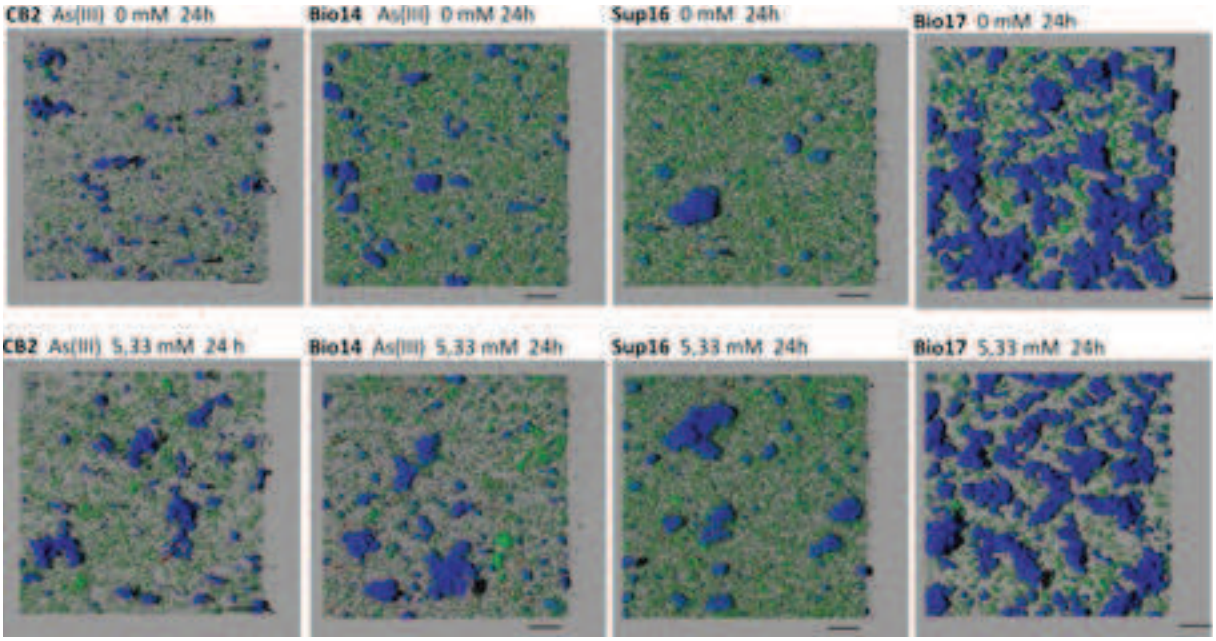


Figure 42 Reconstitution 3D d'observations réalisées au bout de 24 h des biofilms de Tm. sp. CB2 et des variants Bio14, Sup16 et Bio17 cultivées en présence de 2,67 mM d'As(III). Bleu = matrice polysaccharidique ; Vert = Cellules vivantes ; Rouge = Cellules mortes. Echelle = 100 µm.

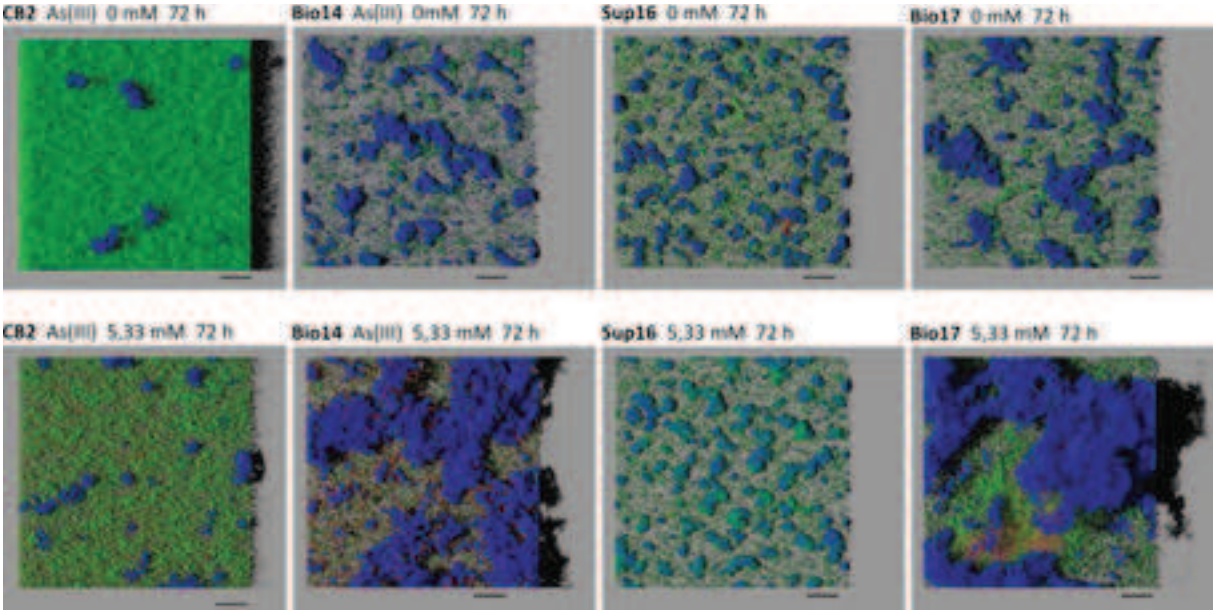


Figure 43 Reconstitution 3D d'observations réalisées au bout de 72 h des biofilms de Tm. sp. CB2 et des variants Bio14, Sup16 et Bio17 cultivées en présence de 2,67 mM d'As(III). Bleu = matrice polysaccharidique ; Vert = Cellules vivantes ; Rouge = Cellules mortes. Echelle = 100 µm.

Les capacités à former des biofilms sont différentes que ce soit entre les variants ou avec *Tm.* sp. CB2. Une caractéristique générale de ces variants est une production de matrice supérieure à la souche originelle. Cette surproduction de matrice est bien visible à 72 h (Figure 45) mais aussi dès 24 h pour Bio17 (Figure 44), et pourrait être en partie impliquée dans la résistance à l'As(III). Chez les variants Bio14 et Bio17, les volumes de matrice et de cellules sont plus importants en présence d'As(III) à 72 h. Ces deux variants ont donc développé une réponse à l'As(III) plus marquée, ce qui va également dans le sens d'une importance des biofilms face à cet élément. Cependant, le cas de Sup16 est intéressant car sa production de matrice est moindre et aucune réponse marquée à l'As(III) n'est visible. Pourtant, sa CMI 23,9 mM, c'est à dire égale à celle de Bio17, et supérieure à celle de Bio14. La production de biofilms n'est donc peut-être pas le facteur essentiel de la résistance à l'As(III) chez les variants, ou bien Sup16 a développé une stratégie de résistance différente des deux autres variants.

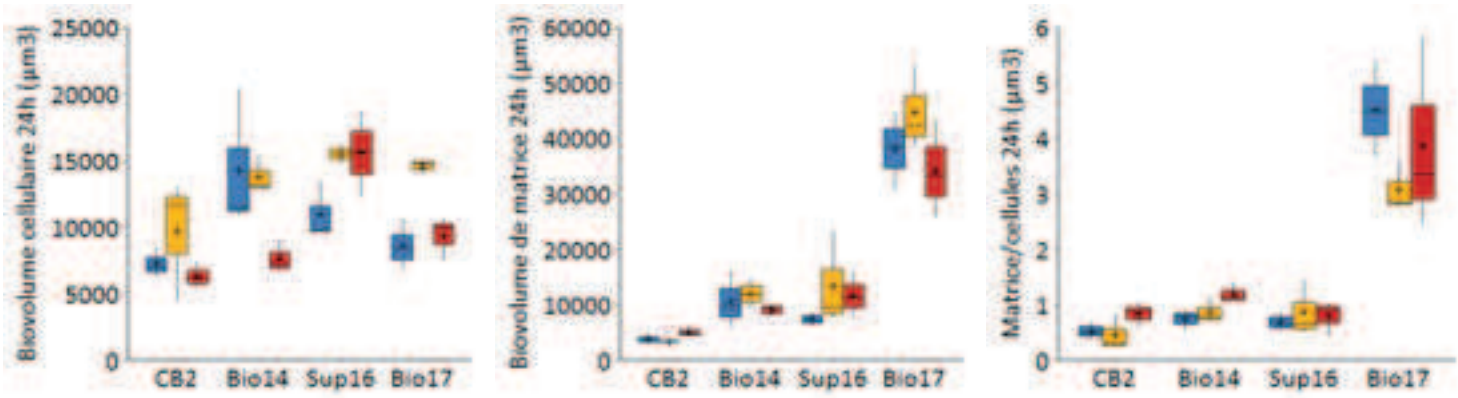


Figure 44 Volume de cellules (à gauche), de matrice (au milieu) et rapport "matrice/cellules" (à droite) de Thiomonas spp. CB2 et des variants Bio14, Sup16 et Bio17 à 24 h suivant différentes concentration d'As(III). Bleu = 0 mM ; Jaune = 2,67 mM ; Rouge = 5,33 mM. La croix "+" indique la moyenne.

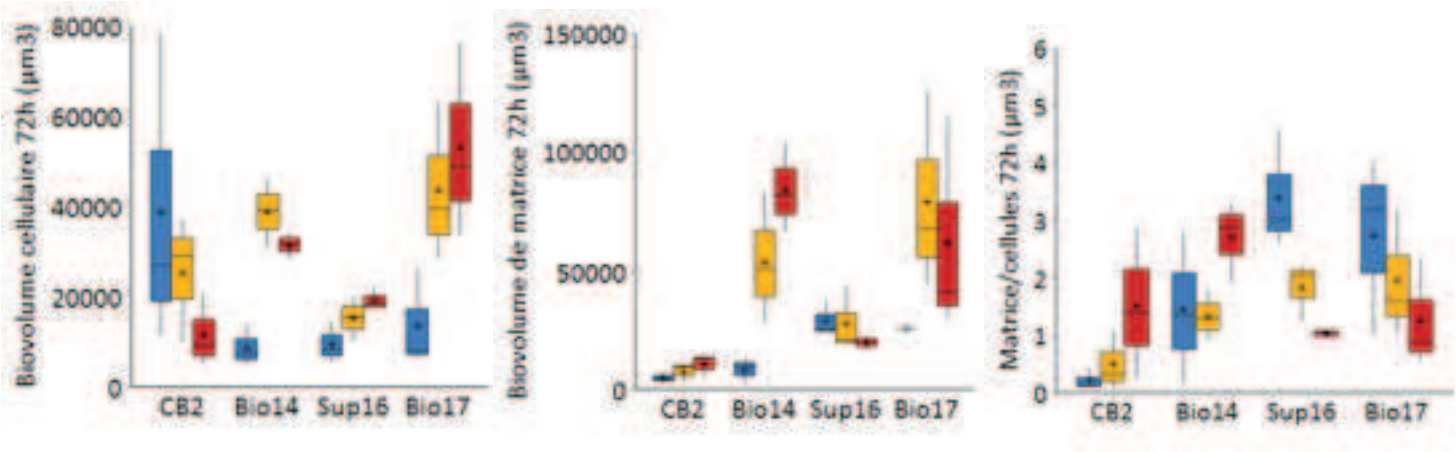


Figure 45 Volume de cellules (à gauche), de matrice (au milieu) et rapport "matrice/cellules" (à droite) de Thiomonas spp. CB2 et des variants Bio14, Sup16 et Bio17 à 72 h suivant différentes concentration d'As(III). Bleu = 0 mM ; Jaune = 2,67 mM ; Rouge = 5,33 mM. La croix "+" indique la moyenne.

4 - Génomique comparée des variants Sup16 et Bio17

Les variants Sup16 et Bio17 sont respectivement issus des biofilms N°16 et N°17 et ont donc comme ancêtre commun le variant "Bio13 C1 #1" (Figure 40 page 180). Ils sont tous deux très résistants à l'As(III) avec une CMI de 23,9 mM, mais ont des capacités de production de biofilm différentes. Ainsi, il a été jugé intéressant de séquencer ces deux variants, mais l'exploration de leur génome n'est qu'à un stade préliminaire. Leur intégration sur la plateforme d'annotation MaGe (Génoscope) a permis de rechercher les mutations et remaniements génomiques.

Plusieurs mutations ponctuelles ont été identifiées dans chaque variant. Une partie de ces mutations est commune aux deux alors que d'autres leurs sont spécifiques. Un total de 99 mutations communes à ces deux variants a été détecté, dont 82 substitutions (SNPs, *Single Nucleotide Polymorphism*s) et 17 insertions ou délétions (indels). Ces mutations en commun ont très probablement été acquises par l'ancêtre commun de ces deux variants "Bio13 C1 #1" qui a servi à inoculer le biofilm N°16 (Figure 40 page 180). Parmi ces mutations figure une substitution dans le gène *aioA* codant la grande sous-unité de l'arsenite oxydase, qui pourrait donc potentiellement expliquer une partie de la résistance de ces variants à l'As(III). Plusieurs mutations sur un gène codant une protéine MCP (Methyl-accepting Chemotaxis Protein) ont également été identifiées, et pourraient être impliquées dans les différences de formation de biofilms. 59 mutations sont spécifiques à Sup16 (47 SNPs et 15 indels) et 62 sont spécifiques à Bio17 (47 SNPs et 15 indels). Chez Sup16, des mutations spécifiques concernant une

diguanylate cyclase, une MCP et une glycosyl transferase ont également été identifiées. Ces gènes étant impliqués dans la synthèse de biofilms et la mobilité, ces mutations pourraient donc être responsables des différences de production de biofilms de ce variant comparé à *Tm.* sp. CB2 et Bio17. Une étude plus approfondie de ces mutations associée au développement d'outils génétiques chez *Tm.* sp. CB2 devrait permettre de vérifier cette hypothèse.

Afin d'identifier d'éventuels réarrangements génomiques chez les variants, leur synténie a été comparée à *Tm.* sp. CB2. Plusieurs réarrangements de grande taille ont ainsi été mis en évidence (Figure 46). Chez Sup16 par exemple, un bloc situé aux alentours de la position 200 000 pb chez *Tm.* sp. CB2 (en rouge sur la figure ci-dessous) a été transposé au niveau de la position 600 000 pb. Chez Bio17, un bloc présent au niveau de la position 600 000 pb de *Tm.* sp. CB2 est retrouvé au niveau de la position 2 000 000 et inversé. De manière intéressante, un bloc plus petit en fin du génome de CB2 (en rose sur la figure ci-dessous) correspond à l'ICE19 identifié dans la Partie 1 du chapitre I. Cet îlot porte notamment une copie de l'opéron *aioBA* codant l'arsénite oxydase. Chez Bio17, cet ICE s'est transposé entre les positions 2 500 000 et 3 000 000. Il semble donc pouvoir s'intégrer à différents sites dans le génome.

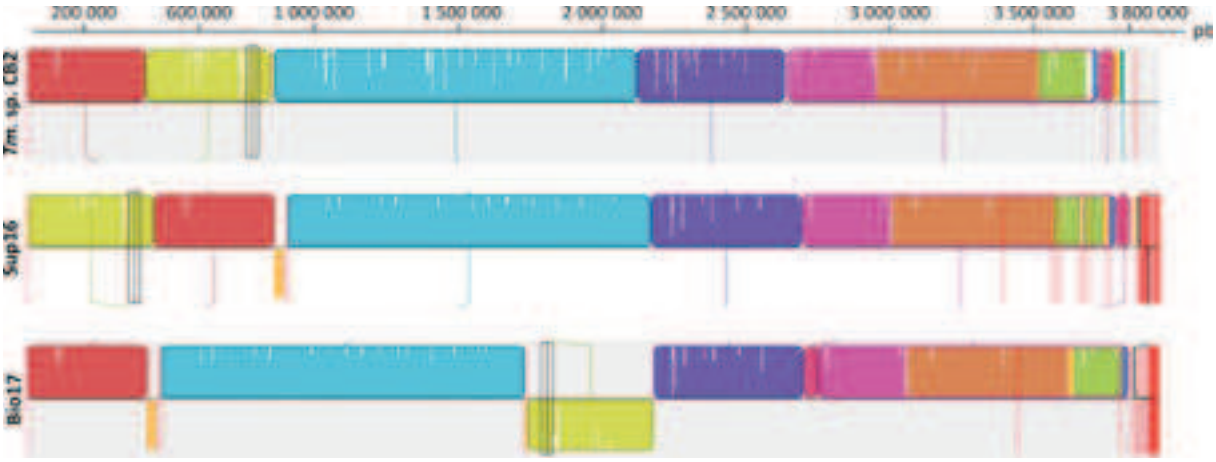


Figure 46 Blocs de synténie déduits (en couleurs) de la comparaison des génomes de Tm. sp. CB2 (haut), Sup16 (milieu) et Bio17 (bas). Sur cette figure sont visibles différents remaniements génomiques ayant eu lieu chez les variants.

En regardant plus en détail la profondeur de lecture de séquençage du génome des variants mappés sur le génome de *Tm.* sp. CB2, il est apparu que la région correspondant à l'ICE19 possède une profondeur de lecture trois fois plus importante que le reste du génome (Figure 47). Les variants possèdent donc visiblement trois copies de l'ICE19. Comme il l'a été remarqué chez Bio17, cet ICE peut s'insérer à au moins deux endroits du génome. Il peut donc

être envisagé que les copies soient insérées à ces différents endroits. Une autre hypothèse est une accrétion en tandem de cet ICE, c'est-à-dire l'insertion des ICEs supplémentaires au niveau des sites *att* de ceux déjà intégrés dans le génome. Il est cependant étonnant que la présence de plusieurs copies dans le génome n'ait pas été révélée lors de l'assemblage du génome. Une dernière hypothèse est donc la présence d'une ou plusieurs copies circulaires, auquel cas cet ICE serait particulier car ces éléments ne sont généralement pas capables de se maintenir dans les cellules sous cette forme. Dans tous les cas, la présence de plusieurs copies de cet élément a probablement un lien avec la résistance accrue à l'As(III), auquel cas il représenterait un moyen de moduler le niveau de cette résistance et donc un potentiel d'adaptation à l'As(III).

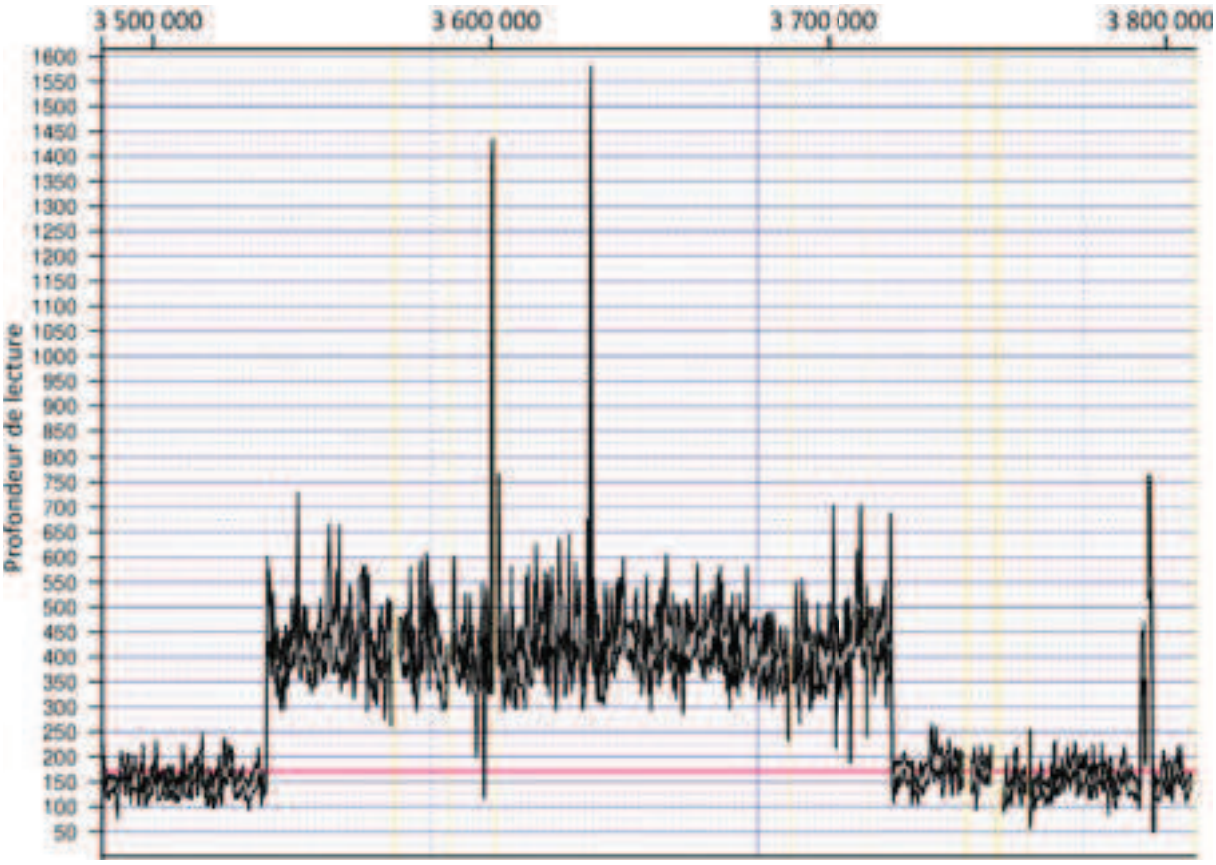


Figure 47 Profondeur de lecture du variant Bio17 dans la région correspondant à l'ICE19. L'ICE19 a une couverture trois fois plus élevée que le reste du génome, indiquant la présence de plusieurs copies de celui-ci chez les variants.

5 - Transcriptomique (RNAseq) chez *Tm.* sp. CB2

Dans le but de rechercher quels processus pourraient être à l'origine de l'apparition des variants dans les biofilms de *Tm.* sp. CB2, une expérience de transcriptomique par RNAseq a été réalisée sur la partie fixée des biofilms de cette souche, à plusieurs temps de culture et avec 0 mM ou 5,33 mM d'As(III). Ces deux concentrations d'As(III) ont été choisies car c'est pour celles-ci que les pourcentages de variants étaient respectivement les moins et les plus élevés.

Ces données récentes n'ont été que partiellement analysées. Globalement, un retard de l'expression d'un certain nombre de gènes a été observé en présence d'As(III) par rapport à son absence. L'As(III) est à l'origine d'un stress oxydant pouvant causer des dégâts au niveau de l'ADN. Les processus de réparation de ces dégâts peuvent être à l'origine de remaniements génomiques. L'expression de gènes impliqués dans la réparation a donc été analysée. Les gènes sont surexprimés en présence d'As(III) et leur expression augmente au cours du temps, à l'image de *recA* et *recX* (Figure 48). Ceci suppose donc l'induction de dommages à l'ADN par l'As(III), ce qui pourrait entrainer des mutations ou remaniements et donc l'apparition de variants.

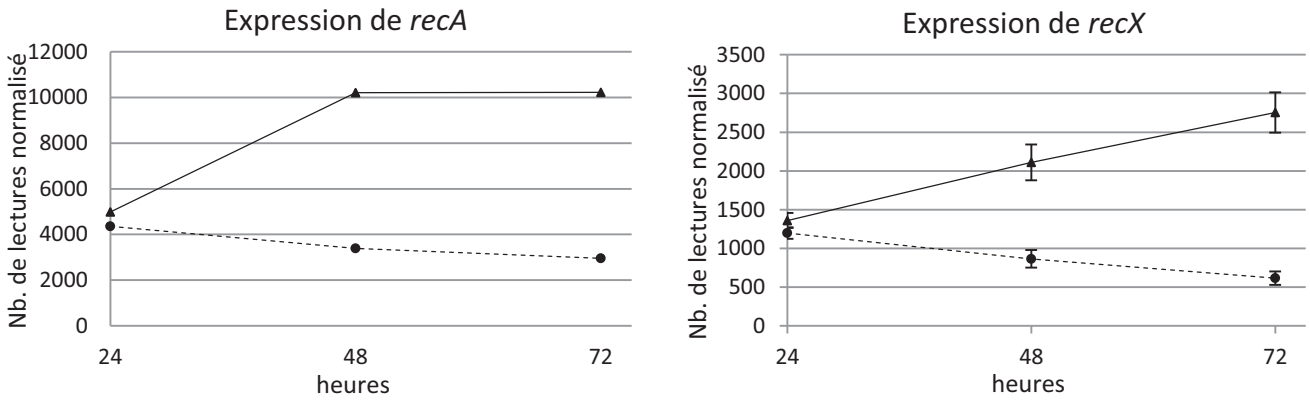


Figure 48 Expression au cours du temps des gènes codant les protéines RecA (à gauche) et RecX (à droite) impliquées dans la recombinaison homologue. L'expression a été déterminée en absence d'As(III) (---●---) et en présence de 5,33 mM d'As(III (—▲—).

Le stress oxydant engendré par l'As(III) peut également occasionner la dénaturation des protéines. L'expression des gènes codant des protéines chaperons impliquées dans le repliement des protéines endommagées, mais aussi celui de gènes impliqués dans la réponse au stress oxydant a donc également été étudiée. De manière surprenante, ces gènes sont sous-exprimés en présence d'As(III), comme c'est par exemple le cas de *groEL* et de *dnaK* (Figure 49). Ceci suggère que les cellules subissent moins ce type de stress en présence d'As(III) qu'en son

absence au temps 72 h dans le biofilm. Ces résultats pourraient être interprétés comme une adaptation du protéome de *Tm.*sp. CB2 au stress causé par la présence d'As(III). Il n'est cependant pas exclus que cette souche exprime une réponse plus spécifique au stress causé par l'As(III) au niveau des protéines.

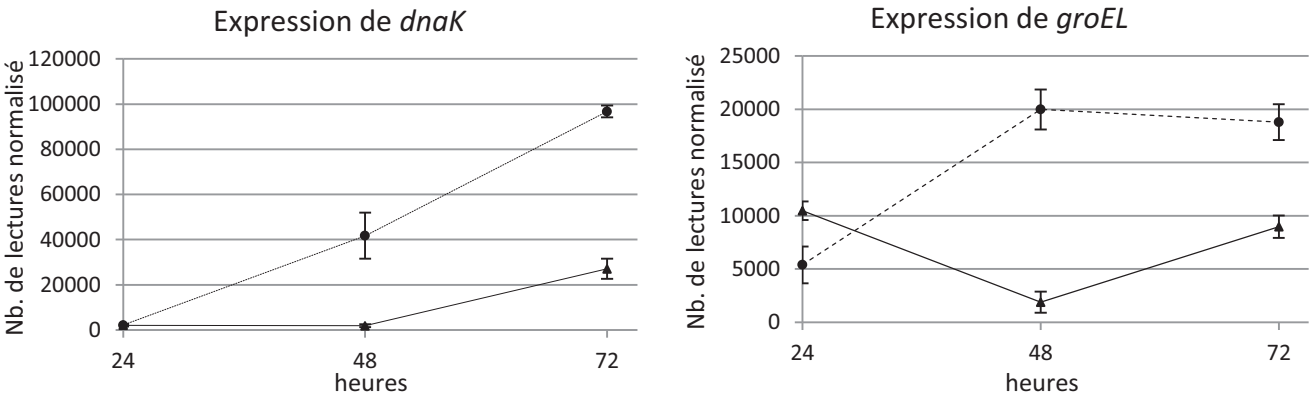


Figure 49 Expression des gènes codant les protéines chaperons DnaK et GroEL. L'expression a été déterminée en absence d'As(III) (---●---) et en présence de 5,33 mM d'As(III (—▲—).

L'émergence des variants en présence d'As(III) dans les biofilms pourrait donc s'expliquer en partie par les dommages causés à l'ADN et les systèmes de réparation mis en place en réponse à ces dégâts qui sont surexprimés en présence d'As(III). Dans ce cas, l'apparition de variants pourrait constituer une réponse indirecte à de fortes concentrations en As(III) permettant l'adaptation de la population à cet élément.

IV - Conclusion

Les résultats présentés dans ce chapitre montrent que des variants "super-résistants" à l'As(III) apparaissent chez la souche *Tm.* sp. CB2 quelles que soient les conditions de cultures testées. L'émergence de ces variants peut donc être considérée comme une caractéristique de cette souche. Ce phénomène est cependant influencé par la présence d'As(III) et la croissance en biofilm, la proportion de variants étant maximale dans la partie supérieure des biofilms en présence de 5,33 mM d'As(III). L'isolement en 200 générations, de variants résistants à une concentration en As(III) deux fois plus élevée que la CMI de départ indique un important potentiel évolutif de *Tm.* sp. CB2 face à cet élément.

Les variants isolés n'ont pas tous les mêmes phénotypes (diffèrents de par leur résistance à l'As(III) et leur capacité à synthétiser des biofilms). Aucune corrélation n'a été établie entre la quantité de biofilm formée et la résistance, mais ces résultats mettent en évidence qu'il existe une diversité chez ces variants et que leur apparition semblerait donc bien refléter une diversification génétique de la population. Le séquençage du génome de deux d'entre eux a montré que ces variants ont subi des remaniements de grande taille et accumulé des mutations dont une partie est spécifique à chacun d'eux. De manière intéressante, plusieurs copies de l'ICE19 contenant l'opéron *aioBA* codant l'arsénite oxydase ont été détectées dans le génome de ces deux variants. Dans la première partie de cette thèse, j'ai montré que la présence de plusieurs copies de cet opéron est corrélée avec une résistance plus importante chez les souches *Tm.* spp. CB2 et CB3 par rapport à *Tm. arsenitoxydans* 3As et *Tm.* spp. CB1 et CB6. La présence de plusieurs copies de cet ICE est donc probablement en partie responsable de la résistance accrue des variants à l'As(III). Cet élément génétique mobile est potentiellement transférable par conjugaison, et pourrait donc être à la fois impliqué dans la dissémination de la résistance dans la population, et dans la modulation du niveau de cette résistance.

Les données de RNAseq montrent qu'en présence d'As(III), les gènes codant des protéines impliquées dans la réparation de l'ADN sont surexprimés. Ceci confirme que l'As(III) cause très probablement des dommages au niveau du génome. L'apparition des variants, ainsi que les mutations et remaniements observés chez ceux-ci, pourraient donc être liés aux systèmes de réparation de l'ADN comme il l'a par exemple été démontré chez *Pseudomonas aeruginosa* (Boles and Singh, 2008). Ces mécanismes, bien qu'impliqués dans le maintien de l'intégrité du génome, produisent des erreurs lors des réparations. Les effets indésirables de l'As(III) sur le génome de *Tm.* sp. CB2 serait donc également en partie à l'origine de l'adaptation de cette souche à cet élément. L'expérience de transcriptomique a également mis en évidence que, de

manière surprenante, les protéines chaperons ne sont pas surexprimées en présence d'As(III). L'intégrité du protéome de *Tm.* sp. CB2 ne semble donc pas être affecté par le stress causé par la présence d'As(III). Ceci pourrait être dû à une composition particulière des protéines de *Tm.* sp. CB2 ou à des protéines spécifiques (ex: Hsc ou Hsp70) et constituerait donc une adaptation de cette souche à l'As(III) ou au stress oxydant causé par ce dernier.

Conclusion & discussion générales

Cette thèse consacrée à l'étude de différentes souches de *Thiomonas* spp. appartenant à la même espèce et isolées du DMA de Carnoulès (Arsene-Ploetze et al., 2010), avait deux objectifs principaux. Le premier objectif était d'inventorier la diversité de ces souches, de caractériser les îlots génomiques et de mieux comprendre leur rôle dans les capacités physiologiques. Un deuxième objectif était de déterminer l'influence de l'arsénite et du développement en biofilm sur la variabilité génomique de ces *Thiomonas* spp.

L'ensemble de ces travaux a tout d'abord mis en évidence une divergence entre ces souches à la fois du point de vue de leur contenu génétique, de leurs capacités fonctionnelles et de la structure de leur génome. Leur phylogénie a été affinée, et il apparait que *Tm. arsenitoxydans* 3As, *Tm.* sp. CB1 et *Tm.* sp. CB6 sont plus proches entre elles qu'elles ne le sont de *Tm.* sp. CB2, *Tm.* sp. CB3 et *Tm. intermedia* K12 (non-issue du DMA de Carnoulès). *Tm.* spp. CB2 et CB3 sont plus proches du groupe formé pas les trois premières souches que de *Tm. intermedia* K12, mais divergent cependant entre elles. Dans une étude précédente (Slyemi et al., 2011), il avait été montré que *Tm. arsenitoxydans* 3As et *Tm. intermedia*^{T} sont deux expèces différentes. Il est à noter ici que les souches de *Tm. intermedia* utilisées ne sont pas les mêmes dans les deux études. Slyemi et al. ont utilisé la souche type isolée par London (London, 1963a) alors que la souche utilisée pendant cette thèse, et dont le génome était disponible (NC_014153.1 ; NC_014154.1 , NC_014155.1), était celle isolée par Milde (Milde et al., 1983). Le regroupement des souches *Tm. arsenitoxydans* 3As, *Tm.* sp. CB1 et *Tm.* sp. CB6, appelé par la suite "groupe 3As", et la divergence des souches *Tm.* sp. CB2 et *Tm.* sp. CB3, est principalement du à la présence d'îlots génomiques qui peuvent biaiser les analyses taxonomiques classiques de type hybridation ADN-ADN (Arsene-Ploetze et al., 2010; Slyemi et al., 2011).

De nombreux îlots génomiques codent des fonctions non-essentielles au fonctionnement des cellules, mais pouvent conférer un avantage dans les conditions du DMA. Deux îlots contenant les gènes impliqués dans le métabolisme de l'arsenic ont été étudiés au cours de cette thèse. Les souches du groupe 3As possèdent une ou deux copie de l'opéron *aioBA* codant l'arsénite oxydase dont une seule est fonctionnelle, alors que les souches *Tm.* spp. CB2 et CB3 en possèdent respectivement deux et trois copies. Le nombre de copies de cet opéron est positivement corrélé avec la CMI d'As(III). D'autres îlots sont à l'origine de la divergence entre les souches. *Tm.* sp. CB2 possède les gènes *ure* codant une activité uréase qui lui permet

d'augmenter le pH du milieu et d'accélérer ainsi la précipitation du fer, de l'aluminium et de l'arsenic dans l'eau du DMA. Le gène *ureC* est absent de *Tm.* sp. CB3 qui ne possède en conséquence aucune activité uréase. Les souches du groupe 3As possèdent des remaniements au niveau de l'îlot génomique portant les gènes *ure*, avec certains gènes vraisemblablement perdus, et ont une activité uréase beaucoup plus faible. Les souches du groupe 3As sont également différentes de *Tm.* spp. CB2 et CB3 concernant la formation de biofilms et la mobilité. Elles ont la particularité d'être mobiles en présence d'As(III), de posséder un îlot portant des gènes *eps* potentiellement impliqués dans la synthèse de biofilms, et forment un biofilms différent de celui des deux autres souches. Les souches *Tm.* sp. CB2 et *Tm.* sp. CB3 possèdent quant à elles deux îlots génomiques contenant des gènes *rfb* potentiellement impliqués dans la synthèse de biofilms et des gènes *che* potentiellement impliqués dans la régulation de la mobilité et la formation de biofilms, tous deux absents du groupe 3As. Ces deux groupes sont clairement différents concernant leur capacité de biosynthèse de biofilms, notamment en réponse à l'As(III). D'autres îlots pourraient être étudiés. Un îlot génomique portant les gènes codant la synthèse de carboxysome est par exemple présent chez le groupe 3As et absent de *Tm.* spp. CB2 et CB3, alors qu'un autre portant des gènes impliqués dans le transport du glycérol est au contraire absent du groupe 3As. La caractérisation de ces îlots serait nécessaire pour mieux comprendre l'écologie de ces différentes souches. De plus, le développement d'outils génétiques chez ces souches permettrait, par des expériences de complémentation fonctionnelle ou de mutagenèse ciblée, de confirmer leur rôle dans les capacités physiologiques de ces souches.

Les résultats obtenus au cours de cette thèse soulèvent des questions plus générales concernant la diversité génomique de souches proches au sein des populations naturelles. Il est notamment difficile de savoir si cette diversité est due à la présence de multiples écotypes, c'est-à-dire à des groupes génotypiques adaptés à des niches écologiques particulières (Cohan, 2002; Cohan and Koeppel, 2008), ou bien à des individus possédant un génome "mosaïque" où la sélection ne fixe que certains loci avantageux (Figure 50). Dans les populations naturelles, l'existence d'écotypes a déjà été rapportée. Un des cas les mieux décrits est celui de la cyanobactérie *Prochlorococcus*, où une expérience de séquençage *single-cell* de la population a permis de mettre en évidence l'existence de nombreuses sous-populations qui résulteraient de pressions de sélection à différentes échelles environnementales (Kashtan et al., 2014). La formation des écotypes, bien que liée à l'occupation de différentes microniches écologiques, impliquerait également un flux génétique restreint avec d'autres groupes similaires. La recombinaison homologue entre souches proches maintient en effet des échanges génétiques au

sein de la population qui empêchent la formation de clusters génotypiques (Fraser et al., 2007). Dans une étude concernant une population de *Leptospirillum* du DMA de Richmond (Californie), Simmons et al. ont montré que la diversité de la population résultait de recombinaisons entre deux populations ancestrales ayant divergé suite à leur isolement géographique (Simmons et al., 2008). Ce type de structure génétique a également été observé dans des populations de *Vibrio* (Shapiro et al., 2012), d'archées thermophiles (Cadillo-Quiroz et al., 2012) et plus récemment au sein de biofilm photosynthétique de *Synechococcus* sp. (Rosen et al., 2015).

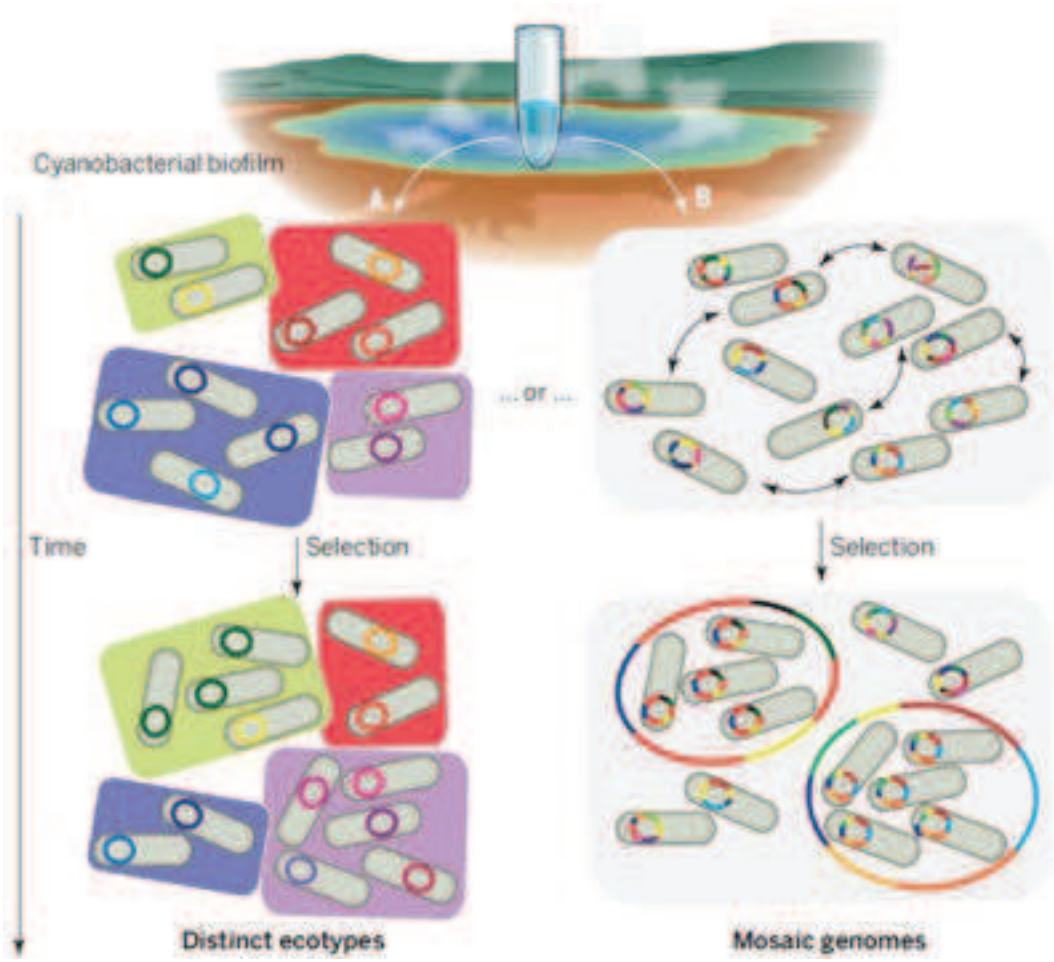


Figure 50 Illustration des deux types de structure génétique d'une population bactérienne. Dans la partie gauche (A), différents écotypes adaptés à des niches écologiques spécifiques forment des clusters génotypiques. L'abondance de chacun varie dans le temps sous l'effet des pressions de sélection. Dans la partie droite (B), les échanges génétiques empêchent la formation d'écotypes et entrainent la formation de génomes mosaïques, même si les pressions de sélection ne fixent que certains loci dans des sous-populations D'après Desai and Walczak, 2015.

Les souches *Tm.* sp. CB1, *Tm.* sp. CB6, mais aussi les nouveaux isolats *Tm.* spp. ACO3 et ACO7, sont très proches de *Tm. arsenitoxydans* 3As à la fois de par leur phylogénie et leur contenu génétique. Le groupe 3As semblerait donc former un cluster génotypique et donc un écotype. La divergence des souches *Tm.* spp. CB2 et CB3 vis à vis du groupe 3As serait plutôt due à des différences de structure de leur génome. En effet et de façon intéressante, alors que la synténie du génome est conservée entre les souches *Tm. arsenitoxydans* 3As et *Tm. intermedia* K12, elle ne l'est pas entre *Tm.* sp. CB2 et *Tm. arsenitoxydans* 3As. Pourtant, *Tm.* sp. CB2 est plus proche de *Tm. arsenitoxydans* 3As que ne l'est *Tm. intermedia* K12, ce qui indique que *Tm.* sp. CB2 a probablement subi plusieurs remaniements de son génome. Les génomes du groupe 3As seraient donc plus stables, confirmant qu'ils ne se recombinent probablement que très peu avec ceux d'autres groupes. En revanche, le génome de la souche *Tm.* sp. CB3 est à la fois différent du groupe 3As et de *Tm.* sp. CB2. Il semble donc que la synténie de son génome ne soit conservée ni avec *Tm. arsenitoxydans* 3As, ni même avec *Tm.* sp. CB2. L'assemblage des génomes de *Tm.* spp. CB1, CB3 et CB6 va être vérifié pour tester cette hypothèse et vérifier que leur synténie est conservée ou non avec *Tm. arsenitoxydans* 3As, et si leur génome parait donc plus stable que celui de *Tm.* sp. CB2.

La population de *Thiomonas* spp. du DMA de Carnoulès semble être un modèle intéressant du point de vue de sa structure génétique puisqu'elle semblerait à la fois composée d'écotypes (groupe 3As) et de sous-populations dont le génome serait beaucoup plus flexibles (*Tm.* spp. CB2 et CB3). Ainsi, il peut être imaginé que l'écotype formé par le groupe 3As serait spécialisé dans une micro-niche et aurait en même temps perdu de la flexibilité génomique et donc du potentiel évolutif. Au contraire, il se pourrait que les souches *Tm.* spp. CB2 et CB3 occupent une niche écologique relativement large, où la sélection serait moins stringente, et que leurs génomes se recombinent fréquemment avec d'autres souches. Dans le cas du génome de *Tm.* sp. CB3, il pourrait être envisagé que sa structure soit variable à cause de l'expansion des séquences IS (Figure 19 page 44), dont la prolifération est liée au relâchement de la pression sélective négative. En effet, son génome est plus grand (4,4 Mb) que les autres souches (~3,7 Mb) et cette souche possèdent environ 200 séquences IS contre une centaine chez le groupe 3As et une cinquantaine chez *Tm.* sp. CB2 (annotation automatique des séquences IS avec ISsaga (www.issaga.biotoul.fr)). De plus, ce génome contient beaucoup de répétitions qui confèrent potentiellement de la variabilité. Cette flexibilité du génome confèrerait un potentiel évolutif plus important à cette souche, ce qui assurerait sa persistance sur le site même face à d'importants changements des conditions environnementales. En cas d'élimination d'écotypes,

une partie de la sous-population "flexible" pourrait coloniser les niches écologiques libérées et se spécialiser. Concernant la souche *Tm.* sp. CB2, les expériences menées sur l'apparition de variants ayant acquis une résistance accrue à l'As(III) chez cette souche ont permis de confirmer le caractère flexible de son génome. En effet, afin de savoir quels sont les facteurs influençant cette flexibilité, le pourcentage de variants a été mesuré chez *Tm.* sp. CB2 cultivé en biofilm ou en culture planctonique en présence ou absence d'As(III). Le pourcentage de variants est plus important dans les biofilms en présence d'As(III), ces deux facteurs semblent donc favoriser le potentiel adaptatif de *Tm.* sp. CB2. D'après les données de RNAseq, l'apparition de remaniements génomiques au sein du biofilm en présence d'As(III) pourrait être en partie due à des mécanismes de réparation de l'ADN à la suite de dommages probablement dus au stress oxydant causé par l'As(III). Chez deux variants dont les génomes ont été séquencés, plusieurs remaniements génomiques ont été identifiés et trois copies de l'ICE19 portant l'opéron *aioBA* ont été détectées. Les copies surnuméraires de cet ICE sont probablement en partie responsables de la résistance de ces variants à l'As(III), et permettrait donc de moduler cette résistance. La flexibilité génomique de *Tm.* sp. CB2 lié aux îlots génomiques lui confèrerait donc un potentiel adaptatif.

De manière générale, les biofilms qui constituent un milieu hétérogène mais sans barrière aux flux de gènes, pourraient favoriser le maintien d'une population très diversifiée et prévenir la formation d'écotype comme suggéré par Rosen et al. (Rosen et al., 2015). Ceci pourrait être mis en parallèle avec les capacités des *Thiomonas* étudiées, *Tm.* spp. CB2 et CB3 semblant former plus de biofilms que les souches du groupe 3As. Pour tester cette hypothèse, le pourçentage de variants sera mesuré chez *Tm. arsenitoxydans* 3As dont le génome parait plus stable, et chez *Tm.* sp. CB3 où il semble au contraire similaire à *Tm.* sp. CB2 en conditions de biofilms. En parallèle, une deuxième expérience de RNAseq sera réalisée chez ces souches dans les mêmes conditions que l'expérience réalisée pendant cette thèse. Ces données transcriptomiques, au-delà de permettre de comparer quels sont les gènes impliqués dans l'apparition de variants et la synthèse de biofilms chez les différentes souches, pourrait permettre de mettre au jour de nouveaux mécanismes ou stratégies d'adaptation à l'As(III).

En conclusion, cette thèse a mis en évidence une diversité génomique et fonctionnelle chez plusieurs souches de *Thiomonas* spp. isolées de Carnoulès due à la présence d'îlots génomiques. Cette diversité est en lien avec une importante flexibilité génomique conférant un potentiel adaptatif à ces souches face à l'As(III) et probablement modulé par le mode de vie en biofilm et les conditions environnementales. Il semblerait également que les populations naturelles de l'AMD de Carnoulès soient à la fois constituées d'écotypes et d'individus au

génome "mosaïque" et plus flexible. Ces derniers pourraient être essentiels à la résilience de la population sur le site. Il serait intéressant d'élargir ce type d'étude à d'autres espèces et d'autres éléments que l'arsenic, que ce soit dans les DMA ou dans d'autres écosystèmes pollués par des métaux. Ceci permettrait de savoir si les résultats sont spécifiques des *Thiomonas* spp. et de l'As(III) ou si ces mécanismes sont plus généraux.

Bibliographie

Ajees, A.A., Yang, J., and Rosen, B.P. (2011). The ArsD As(III) metallochaperone. Biometals Int. J. Role Met. Ions Biol. Biochem. Med. *24*, 391–399.

Akcil, A., and Koldas, S. (2006). Acid Mine Drainage (AMD): causes, treatment and case studies. J. Clean. Prod. *14*, 1139–1145.

Allegrucci, M., and Sauer, K. (2008). Formation of Streptococcus pneumoniae non-phase-variable colony variants is due to increased mutation frequency present under biofilm growth conditions. J. Bacteriol. *190*, 6330–6339.

Arsene-Ploetze, F., Koechler, S., Marchal, M., Coppee, J.-Y., Chandler, M., Bonnefoy, V., Brochier-Armanet, C., Barakat, M., Barbe, V., Battaglia-Brunet, F., et al. (2010). Structure, Function, and Evolution of the Thiomonas spp. Genome. Plos Genet. *6*, e1000859.

Baharoglu, Z., and Mazel, D. (2014). SOS, the formidable strategy of bacteria against aggressions. FEMS Microbiol. Rev. *38*, 1126–1145.

Barreto, M., Jedlicki, E., and Holmes, D.S. (2005). Identification of a gene cluster for the formation of extracellular polysaccharide precursors in the chemolithoautotroph Acidithiobacillus ferrooxidans. Appl. Environ. Microbiol. *71*, 2902–2909.

Battaglia-Brunet, F., Dictor, M.C., Garrido, F., Crouzet, C., Morin, D., Dekeyser, K., Clarens, M., and Baranger, P. (2002). An arsenic(III)-oxidizing bacterial population: selection, characterization, and performance in reactors. J. Appl. Microbiol. *93*, 656–667.

Battaglia-Brunet, F., Joulian, C., Garrido, F., Dictor, M.C., Morin, D., Coupland, K., Johnson, D.B., Hallberg, K.B., and Baranger, P. (2006). Oxidation of arsenite by Thiomonas strains and characterization of Thiomonas arsenivorans sp nov. Antonie Van Leeuwenhoek Int. J. Gen. Mol. Microbiol. *89*, 99–108.

Battesti, A., Majdalani, N., and Gottesman, S. (2011). The RpoS-Mediated General Stress Response in Escherichia coli. Annu. Rev. Microbiol. Vol 65 *65*, 189–213.

Beaber, J.W., Hochhut, B., and Waldor, M.K. (2004). SOS response promotes horizontal dissemination of antibiotic resistance genes. Nature *427*, 72–74.

Bellanger, X., Payot, S., Leblond-Bourget, N., and Guedon, G. (2014). Conjugative and mobilizable genomic islands in bacteria: evolution and diversity. Fems Microbiol. Rev. *38*, 720–760.

Bergthorsson, U., and Ochman, H. (1998). Distribution of chromosome length variation in natural isolates of Escherichia coli. Mol. Biol. Evol. *15*, 6–16.

Bertin, P.N., Heinrich-Salmeron, A., Pelletier, E., Goulhen-Chollet, F., Arsene-Ploetze, F., Gallien, S., Lauga, B., Casiot, C., Calteau, A., Vallenet, D., et al. (2011). Metabolic diversity among main microorganisms inside an arsenic-rich ecosystem revealed by meta- and proteo-genomics. Isme J. *5*, 1735–1747.

Bishop, D.G., and Work, E. (1965). An extracellular glycolipid produced by Escherichia coli grown under lysine-limiting conditions. Biochem. J. *96*, 567–576.

Bobay, L.-M., Rocha, E.P.C., and Touchon, M. (2013). The Adaptation of Temperate Bacteriophages to Their Host Genomes. Mol. Biol. Evol. *30*, 737–751.

Boles, B.R., and Singh, P.K. (2008). Endogenous oxidative stress produces diversity and adaptability in biofilm communities. Proc. Natl. Acad. Sci. U. S. A. *105*, 12503–12508.

Boles, B.R., Thoendel, M., and Singh, P.K. (2004). Self-generated diversity produces "insurance effects" in biofilm communities. Proc. Natl. Acad. Sci. U. S. A. *101*, 16630–16635.

Brochet, M., Da Cunha, V., Couvé, E., Rusniok, C., Trieu-Cuot, P., and Glaser, P. (2009). Atypical association of DDE transposition with conjugation specifies a new family of mobile elements. Mol. Microbiol. *71*, 948–959.

Brown, L., Kessler, A., Cabezas-Sanchez, P., Luque-Garcia, J.L., and Casadevall, A. (2014). Extracellular vesicles produced by the Gram-positive bacterium Bacillus subtilis are disrupted by the lipopeptide surfactin. Mol. Microbiol. *93*, 183–198.

Bruneel, O., Personne, J.C., Casiot, C., Leblanc, M., Elbaz-Poulichet, F., Mahler, B.J., Le Fleche, A., and Grimont, P. a. D. (2003). Mediation of arsenic oxidation by Thiomonas sp in acid-mine drainage (Carnoules, France). J. Appl. Microbiol. *95*, 492–499.

Bruneel, O., Pascault, N., Egal, M., Bancon-Montigny, C., Goni-Urriza, M.S., Elbaz-Poulichet, F., Personne, J.-C., and 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, G.E., Personne, C.J., et al. (2011). Characterization of the Active Bacterial Community Involved in Natural Attenuation Processes in Arsenic-Rich Creek Sediments. Microb. Ecol. *61*, 793–810.

Bryan, C.G., Marchal, M., Battaglia-Brunet, F., Kugler, V., Lemaitre-Guillier, C., Lievremont, D., Bertin, P.N., and Arsene-Ploetze, F. (2009). Carbon and arsenic metabolism in Thiomonas strains: differences revealed diverse adaptation processes. Bmc Microbiol. *9*, 127.

Burrus, V., Pavlovic, G., Decaris, B., and Guedon, G. (2002). Conjugative transposons: the tip of the iceberg. Mol. Microbiol. *46*, 601–610.

Cabezon, E., Ripoll-Rozada, J., Pena, A., de la Cruz, F., and Arechaga, I. (2015). Towards an integrated model of bacterial conjugation. Fems Microbiol. Rev. *39*, 81–95.

Cadillo-Quiroz, H., Didelot, X., Held, N.L., Herrera, A., Darling, A., Reno, M.L., Krause, D.J., and Whitaker, R.J. (2012). Patterns of Gene Flow Define Species of Thermophilic Archaea. PLoS Biol *10*, e1001265.

Cardenas, P.P., Carrasco, B., Soufo, C.D., Cesar, C.E., Herr, K., Kaufenstein, M., Graumann, P.L., and Alonso, J.C. (2012). RecX Facilitates Homologous Recombination by Modulating RecA Activities. Plos Genet. *8*, e1003126.

Casiot, C., Bruneel, O., Personne, J.C., Leblanc, M., and 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.

Casjens, S. (1998). The diverse and dynamic structure of bacterial genomes. Annu. Rev. Genet. *32*, 339.

Casjens, S., and Huang, W.M. (1993). Linear chromosomal physical and genetic map of Borrelia burgdorferi, the Lyme disease agent. Mol. Microbiol. *8*, 967–980.

Cohan, F.M. (2002). What are bacterial species? Annu. Rev. Microbiol. *56*, 457–487.

Cohan, F.M., and Koeppel, A.F. (2008). The Origins of Ecological Diversity in Prokaryotes. Curr. Biol. *18*, R1024–R1034.

Conley, E., Saunders, V., Jackson, V., and Saunders, J. (1986). Mechanism of Intramolecular Recyclization and Deletion Formation Following Transformation of Escherichia-Coli with Linearized Plasmid Dna. Nucleic Acids Res. *14*, 8919–8932.

Costerton, J.W., Lewandowski, Z., Caldwell, D.E., Korber, D.R., and Lappin-Scott, H.M. (1995). Microbial biofilms. Annu. Rev. Microbiol. *49*, 711–745.

De la Cruz, F., Frost, L.S., Meyer, R.J., and Zechner, E.L. (2010). Conjugative DNA metabolism in Gram-negative bacteria. Fems Microbiol. Rev. *34*, 18–40.

Curcio, M.J., and Derbyshire, K.M. (2003). The outs and ins of transposition: from Mu to Kangaroo. Nat. Rev. Mol. Cell Biol. *4*, 865–877.

Delavat, F., Lett, M.-C., and Lievremont, D. (2012). Novel and unexpected bacterial diversity in an arsenic-rich ecosystem revealed by culture-dependent approaches. Biol. Direct *7*, 28.

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

Desai, M.M., and Walczak, A.M. (2015). Flexible gene pools. Science *348*, 977–978.

Dobrindt, U., Hochhut, B., Hentschel, U., and Hacker, J. (2004). Genomic islands in pathogenic and environmental microorganisms. Nat. Rev. Microbiol. *2*, 414–424.

Dorward, D.W., Garon, C.F., and Judd, R.C. (1989). Export and intercellular transfer of DNA via membrane blebs of Neisseria gonorrhoeae. J. Bacteriol. *171*, 2499–2505.

Drees, J.C., Lusetti, S.L., Chitteni-Pattu, S., Inman, R.B., and Cox, M.M. (2004). A RecA filament capping mechanism for RecX protein. Mol. Cell *15*, 789–798.

Driffield, K., Miller, K., Bostock, J.M., O'Neill, A.J., and Chopra, I. (2008). Increased mutability of Pseudomonas aeruginosa in biofilms. J. Antimicrob. Chemother. *61*, 1053–1056.

Dubey, G.P., and Ben-Yehuda, S. (2011). Intercellular nanotubes mediate bacterial communication. Cell *144*, 590–600.

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

Ehrlich, S., Bierne, H., Dalencon, E., Vilette, D., Petranovic, M., Noirot, P., and Michel, B. (1993). Mechanisms of Illegitimate Recombination. Gene *135*, 161–166.

Eisen, J.A., Heidelberg, J.F., White, O., and Salzberg, S.L. (2000). Evidence for symmetric chromosomal inversions around the replication origin in bacteria. Genome Biol. *1*.

Erill, I., Campoy, S., and Barbe, J. (2007). Aeons of distress: an evolutionary perspective on the bacterial SOS response. Fems Microbiol. Rev. *31*, 637–656.

Fahy, A., Giloteaux, L., Bertin, P., Le Paslier, D., Medigue, C., Weissenbach, J., Duran, R., and Lauga, B. (2015). 16S rRNA and As-Related Functional Diversity: Contrasting Fingerprints in Arsenic-Rich Sediments from an Acid Mine Drainage. Microb. Ecol. *70*, 154–167.

Fleischmann, R., Adams, M., White, O., Clayton, R., Kirkness, E., Kerlavage, A., Bult, C., Tomb, J., Dougherty, B., Merrick, J., et al. (1995). Whole-Genome Random Sequencing and Assembly of Haemophilus-Influenzae Rd. Science *269*, 496–512.

Foster, P.L. (2005). Stress responses and genetic variation in bacteria. Mutat. Res.-Fundam. Mol. Mech. Mutagen. *569*, 3–11.

Foster, P.L. (2007). Stress-induced mutagenesis in bacteria. Crit. Rev. Biochem. Mol. Biol. *42*, 373–397.

Foster, P.L., Hanson, A.J., Lee, H., Popodi, E.M., and Tang, H. (2013). On the Mutational Topology of the Bacterial Genome. G3 GenesGenomesGenetics *3*, 399–407.

Fraser, C., Hanage, W.P., and Spratt, B.G. (2007). Recombination and the nature of bacterial speciation. Science *315*, 476–480.

Fraser, C.M., Gocayne, J.D., White, O., Adams, M.D., Clayton, R.A., Fleischmann, R.D., Bult, C.J., Kerlavage, A.R., Sutton, G., Kelley, J.M., et al. (1995). The minimal gene complement of Mycoplasma genitalium. Science *270*, 397–403.

Frost, L.S., Leplae, R., Summers, A.O., and Toussaint, A. (2005). Mobile genetic elements: the agents of open source evolution. Nat. Rev. Microbiol. *3*, 722–732.

Furuta, Y., Abe, K., and Kobayashi, I. (2010). Genome comparison and context analysis reveals putative mobile forms of restriction–modification systems and related rearrangements. Nucleic Acids Res. *38*, 2428–2443.

Galhardo, R.S., Hastings, P.J., and Rosenberg, S.M. (2007). Mutation as a stress response and the regulation of evolvability. Crit. Rev. Biochem. Mol. Biol. *42*, 399–435.

Garcillan-Barcia, M.P., Francia, M.V., and de la Cruz, F. (2009). The diversity of conjugative relaxases and its application in plasmid classification. Fems Microbiol. Rev. *33*, 657–687.

Ghigo, J.M. (2001). Natural conjugative plasmids induce bacterial biofilm development. Nature *412*, 442–445.

Gillings, M.R. (2014). Integrons: Past, Present, and Future. Microbiol. Mol. Biol. Rev. *78*, 257–277.

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

Griffith, F. (1928). The Significance of Pneumococcal Types. J. Hyg. (Lond.) *27*, 113–159.

Grindley, N.D.F., Whiteson, K.L., and Rice, P.A. (2006). Mechanisms of site-specific recombination. In Annual Review of Biochemistry, (Palo Alto: Annual Reviews), pp. 567–605.

Guérillot, R., Da Cunha, V., Sauvage, E., Bouchier, C., and Glaser, P. (2013). Modular evolution of TnGBSs, a new family of integrative and conjugative elements associating

insertion sequence transposition, plasmid replication, and conjugation for their spreading. J. Bacteriol. *195*, 1979–1990.

Guerin, E., Cambray, G., Sanchez-Alberola, N., Campoy, S., Erill, I., Da Re, S., Gonzalez-Zorn, B., Barbé, J., Ploy, M.-C., and Mazel, D. (2009). The SOS response controls integron recombination. Science *324*, 1034.

Guglielmini, J., de la Cruz, F., and Rocha, E.P.C. (2013). Evolution of Conjugation and Type IV Secretion Systems. Mol. Biol. Evol. *30*, 315–331.

Hall, R.M. (2012). Integrons and gene cassettes: hotspots of diversity in bacterial genomes. Ann. N. Y. Acad. Sci. *1267*, 71–78.

Hallberg, K.B., and Johnson, D.B. (2003). Novel acidophiles isolated from moderately acidic mine drainage waters. Hydrometallurgy *71*, 139–148.

Hall-Stoodley, L., Costerton, J.W., and Stoodley, P. (2004). Bacterial biofilms: From the natural environment to infectious diseases. Nat. Rev. Microbiol. *2*, 95–108.

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

Han, K., Li, Z., Peng, R., Zhu, L., Zhou, T., Wang, L., Li, S., Zhang, X., Hu, W., Wu, Z., et al. (2013). Extraordinary expansion of a Sorangium cellulosum genome from an alkaline milieu. Sci. Rep. *3*, 2101.

Hannan, S., Ready, D., Jasni, A.S., Rogers, M., Pratten, J., and Roberts, A.P. (2010). Transfer of antibiotic resistance by transformation with eDNA within oral biofilms. Fems Immunol. Med. Microbiol. *59*, 345–349.

Harrison, J.J., Ceri, H., and Turner, R.J. (2007). Multimetal resistance and tolerance in microbial biofilms. Nat. Rev. Microbiol. *5*, 928–938.

Hayes, F., and Van Melderen, L. (2011). Toxins-antitoxins: diversity, evolution and function. Crit. Rev. Biochem. Mol. Biol. *46*, 386–408.

Heinrich-Salmeron, A., Cordi, A., Brochier-Armanet, C., Halter, D., Pagnout, C., Abbaszadeh-fard, E., Montaut, D., Seby, F., Bertin, P.N., Bauda, P., et al. (2011). Unsuspected Diversity of Arsenite-Oxidizing Bacteria as Revealed by Widespread Distribution of the aoxB Gene in Prokaryotes ▿. Appl. Environ. Microbiol. *77*, 4685–4692.

Hiller, N.L., Janto, B., Hogg, J.S., Boissy, R., Yu, S., Powell, E., Keefe, R., Ehrlich, N.E., Shen, K., Hayes, J., et al. (2007). Comparative genomic analyses of seventeen Streptococcus pneumoniae strains: insights into the pneumococcal supragenome. J. Bacteriol. *189*, 8186–8195.

Holmes, A.J., Holley, M.P., Mahon, A., Nield, B., Gillings, M., and Stokes, H.W. (2003). Recombination activity of a distinctive integron-gene cassette system associated with Pseudomonas stutzeri populations in soil. J. Bacteriol. *185*, 918–928.

Houry, A., Gohar, M., Deschamps, J., Tischenko, E., Aymerich, S., Gruss, A., and Briandet, R. (2012). Bacterial swimmers that infiltrate and take over the biofilm matrix. Proc. Natl. Acad. Sci. U. S. A. *109*, 13088–13093.

Huber, H., and Stetter, K. (1990). Thiobacillus-Cuprinus Sp-Nov, a Novel Facultatively Organotrophic Metal-Mobilizing Bacterium. Appl. Environ. Microbiol. *56*, 315–322.

Jacob, F., and Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. J. Mol. Biol. *3*, 318–356.

Johnson, D.B., and Hallberg, K.B. (2005). Biogeochemistry of the compost bioreactor components of a composite acid mine drainage passive remediation system. Sci. Total Environ. *338*, 81–93.

Johnston, C., Martin, B., Fichant, G., Polard, P., and Claverys, J.-P. (2014). Bacterial transformation: distribution, shared mechanisms and divergent control. Nat. Rev. Microbiol. *12*, 181–196.

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

Kaneko, T., Nakamura, Y., Sato, S., Minamisawa, K., Uchiumi, T., Sasamoto, S., Watanabe, A., Idesawa, K., Iriguchi, M., Kawashima, K., et al. (2002). Complete genomic sequence of nitrogen-fixing symbiotic bacterium Bradyrhizobium japonicum USDA110. DNA Res. Int. J. Rapid Publ. Rep. Genes Genomes *9*, 189–197.

Kashtan, N., Roggensack, S.E., Rodrigue, S., Thompson, J.W., Biller, S.J., Coe, A., Ding, H., Marttinen, P., Malmstrom, R.R., Stocker, R., et al. (2014). Single-Cell Genomics Reveals Hundreds of Coexisting Subpopulations in Wild Prochlorococcus. Science *344*, 416–420.

Katayama, Y., Uchino, Y., Wood, A.P., and Kelly, D.P. (2006). Confirmation of Thiomonas delicata (formerly Thiobacillus delicatus) as a distinct species of the genus Thiomonas Moreira and Amils 1997 with comments on some species currently assigned to the genus. Int. J. Syst. Evol. Microbiol. *56*, 2553–2557.

Katayama-Fujimura, Y., Enokizono, Y., Kaneko, T., and Kuraishi, H. (1983). Deoxyribonucleic-Acid Homologies Among Species of the Genus Thiobacillus. J. Gen. Appl. Microbiol. *29*, 287–295.

Katayama-Fujimura, Y., Kawashima, I., Tsuzaki, N., and Kuraishi, H. (1984). Physiological-Characteristics of the Facultatively Chemolithotrophic Thiobacillus Species Thiobacillus-Delicatus Nom Rev, Emend, Thiobacillus-Perometabolis, and Thiobacillus-Intermedius. Int. J. Syst. Bacteriol. *34*, 139–144.

Kim, J.H., Lee, J., Park, J., and Gho, Y.S. (2015). Gram-negative and Gram-positive bacterial extracellular vesicles. Semin. Cell Dev. Biol. *40*, 97–104.

Kitts, P.A., and Nash, H.A. (1987). Homology-dependent interactions in phage λ site-specific recombination. Nature *329*, 346–348.

Kolling, G.L., and Matthews, K.R. (1999). Export of Virulence Genes and Shiga Toxin by Membrane Vesicles of Escherichia coli O157:H7. Appl. Environ. Microbiol. *65*, 1843–1848.

Koonin, E.V. (2009). Evolution of genome architecture. Int. J. Biochem. Cell Biol. *41*, 298–306.

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

Land, M., Hauser, L., Jun, S.-R., Nookaew, I., Leuze, M.R., Ahn, T.-H., Karpinets, T., Lund, O., Kora, G., Wassenaar, T., et al. (2015). Insights from 20 years of bacterial genome sequencing. Funct. Integr. Genomics *15*, 141–161.

Lang, A.S., Zhaxybayeva, O., and Beatty, J.T. (2012). Gene transfer agents: phage-like elements of genetic exchange. Nat. Rev. Microbiol. *10*, 472–482.

Lapierre, P., and Gogarten, J.P. (2009). Estimating the size of the bacterial pan-genome. Trends Genet. *25*, 107–110.

Lawrence, J.G., and Roth, J.R. (1996). Selfish operons: Horizontal transfer may drive the evolution of gene clusters. Genetics *143*, 1843–1860.

Le, T.B.K., and Laub, M.T. (2014). New approaches to understanding the spatial organization of bacterial genomes. Curr. Opin. Microbiol. *22*, 15–21.

Leblanc, M., Achard, B., Othman, D.B., Luck, J.M., BertrandSarfati, J., and Personne, J.C. (1996). Accumulation of arsenic from acidic mine waters by ferruginous bacterial accretions (stromatolites). Appl. Geochem. *11*, 541 – &.

Lederberg, J. (1952). Cell Genetics and Hereditary Symbiosis. Physiol. Rev. *32*, 403–430.

Lee, E.-Y., Choi, D.-Y., Kim, D.-K., Kim, J.-W., Park, J.O., Kim, S., Kim, S.-H., Desiderio, D.M., Kim, Y.-K., Kim, K.-P., et al. (2009). Gram-positive bacteria produce membrane vesicles: proteomics-based characterization of Staphylococcus aureus-derived membrane vesicles. Proteomics *9*, 5425–5436.

Lin, Y.S., Kieser, H.M., Hopwood, D.A., and Chen, C.W. (1993). The chromosomal DNA of Streptomyces lividans 66 is linear. Mol. Microbiol. *10*, 923–933.

London, J. (1963a). Thiobacillus intermedius Nov.sp. - a Novel Type of Facultative Autotroph. Arch. Mikrobiol. *46*, 329 – &.

London, J. (1963b). Thiobacillus intermedius nov.sp. Arch. Für Mikrobiol. *46*, 329–337.

London, J., and Rittenberg, S.C. (1967). Thiobacillus perometabolis Nov Sp a Non-Autotrophic Thiobacillus. Arch. Mikrobiol. *59*, 218 – &.

Lossouarn, J., Nesbø, C.L., Mercier, C., Zhaxybayeva, O., Johnson, M.S., Charchuck, R., Farasin, J., Bienvenu, N., Baudoux, A.-C., Michoud, G., et al. (2015). “Ménage à trois”: a selfish genetic element uses a virus to propagate within Thermotogales. Environ. Microbiol. *17*, 3278–3288.

Madsen, J.S., Burmolle, M., Hansen, L.H., and Sorensen, S.J. (2012). The interconnection between biofilm formation and horizontal gene transfer. Fems Immunol. Med. Microbiol. *65*, 183–195.

Maiques, E., Ubeda, C., Tormo, M.A., Ferrer, M.D., Lasa, I., Novick, R.P., and Penades, J.R. (2007). Role of staphylococcal phage and SaPI integrase in intra- and interspecies SaPI transfer. J. Bacteriol. *189*, 5608–5616.

Marchal, M., Briandet, R., Halter, D., Koechler, S., DuBow, M.S., Lett, M.-C., and Bertin, P.N. (2011). Subinhibitory Arsenite Concentrations Lead to Population Dispersal in Thiomonas sp. Plos One *6*, e23181.

Marin, I., Amils, R., and Abad, J.P. (1997). Genomic organization of the metal-mobilizing bacterium Thiobacillus cuprinus. Gene *187*, 99–105.

Marles-Wright, J., and Lewis, R.J. (2007). Stress responses of bacteria. Curr. Opin. Struct. Biol. *17*, 755–760.

Masuda, H., Tan, Q., Awano, N., Wu, K.-P., and Inouye, M. (2012). YeeU enhances the bundling of cytoskeletal polymers of MreB and FtsZ, antagonizing the CbtA (YeeV) toxicity in Escherichia coli. Mol. Microbiol. *84*, 979–989.

De Mello Varani, A., Souza, R.C., Nakaya, H.I., de Lima, W.C., Paula de Almeida, L.G., Kitajima, E.W., Chen, J., Civerolo, E., Vasconcelos, A.T.R., and Van Sluys, M.-A. (2008). Origins of the Xylella fastidiosa prophage-like regions and their impact in genome differentiation. PloS One *3*, e4059.

Menouni, R., Hutinet, G., Petit, M.-A., and Ansaldi, M. (2015). Bacterial genome remodeling through bacteriophage recombination. FEMS Microbiol. Lett. *362*, 1–10.

Michel, B., and Ehrlich, S. (1986). Illegitimate Recombination Occurs Between the Replication Origin of the Plasmid Pc194 and a Progressing Replication Fork. Embo J. *5*, 3691–3696.

Milde, K., Sand, W., Wolff, W., and Bock, E. (1983). Thiobacilli of the Corroded Concrete Walls of the Hamburg Sewer System. J. Gen. Microbiol. *129*, 1327–1333.

Mira, A., Ochman, H., and Moran, N.A. (2001). Deletional bias and the evolution of bacterial genomes. Trends Genet. TIG *17*, 589–596.

Mizoguchi, T., Sato, T., and Okabe, T. (1976). New Sulfur-Oxidizing Bacteria Capable of Growing Heterotrophically, Thiobacillus-Rubellus Nov-Sp and Thiobacillus-Delicatus Nov-Sp. J. Ferment. Technol. *54*, 181–191.

Moinier, D., Slyemi, D., Byrne, D., Lignon, S., Lebrun, R., Talla, E., and Bonnefoy, V. (2014). An ArsR/SmtB Family Member Is Involved in the Regulation by Arsenic of the Arsenite Oxidase Operon in Thiomonas arsenitoxydans. Appl. Environ. Microbiol. *80*, 6413–6426.

Molin, S., and Tolker-Nielsen, T. (2003). Gene transfer occurs with enhanced efficiency in biofilms and induces enhanced stabilisation of the biofilm structure. Curr. Opin. Biotechnol. *14*, 255–261.

Moran, N.A., and Plague, G.R. (2004). Genomic changes following host restriction in bacteria. Curr. Opin. Genet. Dev. *14*, 627–633.

Moreira, D., and Amils, R. (1997). Phylogeny of Thiobacillus cuprinus and other mixotrophic thiobacilli: Proposal for Thiomonas gen nov. Int. J. Syst. Bacteriol. *47*, 522–528.

Murray, N.E. (2002). 2001 Fred Griffith review lecture. Immigration control of DNA in bacteria: self versus non-self. Microbiol. Read. Engl. *148*, 3–20.

Norman, A., Hansen, L.H., and Sorensen, S.J. (2009). Conjugative plasmids: vessels of the communal gene pool. Philos. Trans. R. Soc. B-Biol. Sci. *364*, 2275–2289.

Oliveira, P.H., Touchon, M., and Rocha, E.P.C. (2014). The interplay of restriction-modification systems with mobile genetic elements and their prokaryotic hosts. Nucleic Acids Res. *42*, 10618–U803.

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

O'Toole, G., Kaplan, H.B., and Kolter, R. (2000). Biofilm formation as microbial development. Annu. Rev. Microbiol. *54*, 49–79.

Panda, S.K., Jyoti, V., Bhadra, B., Nayak, K.C., Shivaji, S., Rainey, F.A., and Das, S.K. (2009). Thiomonas bhubaneswarensis sp nov., an obligately mixotrophic, moderately thermophilic, thiosulfate-oxidizing bacterium. Int. J. Syst. Evol. Microbiol. *59*, 2171–2175.

Pena, A., Teeling, H., Huerta-Cepas, J., Santos, F., Yarza, P., Brito-Echeverria, J., Lucio, M., Schmitt-Kopplin, P., Meseguer, I., Schenowitz, C., et al. (2010). Fine-scale evolution: genomic, phenotypic and ecological differentiation in two coexisting Salinibacter ruber strains. Isme J. *4*, 882–895.

Penadés, J.R., Chen, J., Quiles-Puchalt, N., Carpena, N., and Novick, R.P. (2015). Bacteriophage-mediated spread of bacterial virulence genes. Curr. Opin. Microbiol. *23*, 171–178.

Petersen, F.C., Tao, L., and Scheie, A.A. (2005). DNA binding-uptake system: a link between cell-to-cell communication and biofilm formation. J. Bacteriol. *187*, 4392–4400.

Pinto, U.M., Pappas, K.M., and Winans, S.C. (2012). The ABCs of plasmid replication and segregation. Nat. Rev. Microbiol. *10*, 755–765.

Price, C., and Bickle, T.A. (1986). A possible role for DNA restriction in bacterial evolution. Microbiol. Sci. *3*, 296–299.

Quiles-Puchalt, N., Carpena, N., Alonso, J.C., Novick, R.P., Marina, A., and Penades, J.R. (2014). Staphylococcal pathogenicity island DNA packaging system involving cos-site packaging and phage-encoded HNH endonucleases. Proc. Natl. Acad. Sci. U. S. A. *111*, 6016–6021.

Rao, V.B., and Feiss, M. (2008). The Bacteriophage DNA Packaging Motor. In Annual Review of Genetics, (Palo Alto: Annual Reviews), pp. 647–681.

Rasmussen, B. (2000). Filamentous microfossils in a 3,235-million-year-old volcanogenic massive sulphide deposit. Nature *405*, 676–679.

Reed, R., and Grindley, N. (1981). Transposon-Mediated Site-Specific Recombination Invitro - Dna Cleavage and Protein-Dna Linkage at the Recombination Site. Cell *25*, 721–728.

Riley, M., and Anilionis, A. (1978). Evolution of Bacterial Genome. Annu. Rev. Microbiol. *32*, 519–560.

Rocha, E.P.C. (2004). The replication-related organization of bacterial genomes. Microbiol.-Sgm *150*, 1609–1627.

Rocha, E.P.C. (2008). The Organization of the Bacterial Genome. In Annual Review of Genetics, (Palo Alto: Annual Reviews), pp. 211–233.

Rocha, E.P.C., Cornet, E., and Michel, B. (2005). Comparative and Evolutionary Analysis of the Bacterial Homologous Recombination Systems. PLoS Genet. *1*.

Rodriguez-Valera, F., and Ussery, D.W. (2012). Is the pan-genome also a pan-selectome? F1000Research *1*, 16.

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

Rosen, M.J., Davison, M., Bhaya, D., and Fisher, D.S. (2015). Fine-scale diversity and extensive recombination in a quasisexual bacterial population occupying a broad niche. Science *348*, 1019–1023.

Rothman, J.L. (1965). Transduction studies on the relation between prophage and host chromosome. J. Mol. Biol. *12*, 892–912.

Schuster, C.F., and Bertram, R. (2013). Toxin-antitoxin systems are ubiquitous and versatile modulators of prokaryotic cell fate. Fems Microbiol. Lett. *340*, 73–85.

Schwechheimer, C., and Kuehn, M.J. (2015). Outer-membrane vesicles from Gram-negative bacteria: biogenesis and functions. Nat. Rev. Microbiol. *13*, 605–619.

Shapiro, B.J., Friedman, J., Cordero, O.X., Preheim, S.P., Timberlake, S.C., Szabó, G., Polz, M.F., and Alm, E.J. (2012). Population Genomics of Early Events in the Ecological Differentiation of Bacteria. Science *336*, 48–51.

Shintani, M., Sanchez, Z.K., and Kimbara, K. (2015). Genomics of microbial plasmids: classification and identification based on replication and transfer systems and host taxonomy. Evol. Genomic Microbiol. *6*, 242.

Shooner, F., Bousquet, J., and Tyagi, R.D. (1996). Isolation, phenotypic characterization, and phylogenetic position of a novel, facultatively autotrophic, moderately thermophilic bacterium, Thiobacillus thermosulfatus sp nov. Int. J. Syst. Bacteriol. *46*, 409–415.

Shuman, S., and Glickman, M.S. (2007). Bacterial DNA repair by non-homologous end joining. Nat. Rev. Microbiol. *5*, 852–861.

Siguier, P., Filée, J., and Chandler, M. (2006). Insertion sequences in prokaryotic genomes. Curr. Opin. Microbiol. *9*, 526–531.

Siguier, P., Gagnevin, L., and Chandler, M. (2009). The new IS1595 family, its relation to IS1 and the frontier between insertion sequences and transposons. Res. Microbiol. *160*, 232–241.

Siguier, P., Gourbeyre, E., and Chandler, M. (2014). Bacterial insertion sequences: their genomic impact and diversity. FEMS Microbiol. Rev. *38*, 865–891.

Simmons, S.L., DiBartolo, G., Denef, V.J., Goltsman, D.S.A., Thelen, M.P., and Banfield, J.F. (2008). Population genomic analysis of strain variation in Leptospirillum group. Plos Biol. *6*, 1427–1442.

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

Slyemi, D., Moinier, D., Brochier-Armanet, C., Bonnefoy, V., and Johnson, D.B. (2011). Characteristics of a phylogenetically ambiguous, arsenic-oxidizing Thiomonas sp., Thiomonas arsenitoxydans strain 3As(T) sp nov. Arch. Microbiol. *193*, 439–449.

Smillie, C., Garcillán-Barcia, M.P., Francia, M.V., Rocha, E.P.C., and Cruz, F. de la (2010). Mobility of Plasmids. Microbiol. Mol. Biol. Rev. *74*, 434–452.

Smyth, D.S., and Robinson, D.A. (2009). Integrative and Sequence Characteristics of a Novel Genetic Element, ICE6013, in Staphylococcus aureus. J. Bacteriol. *191*, 5964–5975.

Stein, D.C., Gunn, J.S., Radlinska, M., and Piekarowicz, A. (1995). Restriction and modification systems of neisseria gonorrhoeae. Gene *157*, 19–22.

Stewart, P.S., and Franklin, M.J. (2008). Physiological heterogeneity in biofilms. Nat. Rev. Microbiol. *6*, 199–210.

Stingl, K., Altendorf, K., and Bakker, E.P. (2002). Acid survival of Helicobacter pylori: how does urease activity trigger cytoplasmic pH homeostasis? Trends Microbiol. *10*, 70–74.

Suttle, C.A. (2007). Marine viruses — major players in the global ecosystem. Nat. Rev. Microbiol. *5*, 801–812.

Szekeres, S., Dauti, M., Wilde, C., Mazel, D., and Rowe-Magnus, D.A. (2007). Chromosomal toxin-antitoxin loci can diminish large-scale genome reductions in the absence of selection. Mol. Microbiol. *63*, 1588–1605.

Tamaki, S., and Frankenberger, W. (1992). Environmental Biochemistry of Arsenic. Rev. Environ. Contam. Toxicol. *124*, 79–110.

Tenaillon, O., Denamur, E., and Matic, I. (2004). Evolutionary significance of stress-induced mutagenesis in bacteria. Trends Microbiol. *12*, 264–270.

Tettelin, H., Masignani, V., Cieslewicz, M.J., Donati, C., Medini, D., Ward, N.L., Angiuoli, S.V., Crabtree, J., Jones, A.L., Durkin, A.S., et al. (2005). Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: Implications for the microbial "pan-genome." Proc. Natl. Acad. Sci. U. S. A. *102*, 13950–13955.

Toussaint, A., and Merlin, C. (2002). Mobile elements as a combination of functional modules. Plasmid *47*, 26–35.

Touzain, F., Petit, M.-A., Schbath, S., and Karoui, M.E. (2011). DNA motifs that sculpt the bacterial chromosome. Nat. Rev. Microbiol. *9*, 15–26.

Treangen, T.J., and Rocha, E.P.C. (2011). Horizontal Transfer, Not Duplication, Drives the Expansion of Protein Families in Prokaryotes. PLoS Genet *7*, e1001284.

Vasu, K., and Nagaraja, V. (2013). Diverse Functions of Restriction-Modification Systems in Addition to Cellular Defense. Microbiol. Mol. Biol. Rev. *77*, 53–72.

Van der Oost, J., Westra, E.R., Jackson, R.N., and Wiedenheft, B. (2014). Unravelling the structural and mechanistic basis of CRISPR-Cas systems. Nat. Rev. Microbiol. *12*, 479–492.
Van der Veen, S., and Abee, T. (2011). Generation of Variants in Listeria monocytogenes Continuous-Flow Biofilms Is Dependent on Radical-Induced DNA Damage and RecA-Mediated Repair. Plos One *6*, e28590.

Vernikos, G.S., and Parkhill, J. (2006). Interpolated variable order motifs for identification of horizontally acquired DNA: revisiting the Salmonella pathogenicity islands. Bioinforma. Oxf. Engl. *22*, 2196–2203.

Vernikos, G., Medini, D., Riley, D.R., and Tettelin, H. (2015). Ten years of pan-genome analyses. Curr. Opin. Microbiol. *23*, 148–154.

Vesteinsdottir, H., Reynisdottir, D.B., and Orlygsson, J. (2011). Thiomonas islandica sp. nov., a moderately thermophilic, hydrogen- and sulfur-oxidizing betaproteobacterium isolated from a hot spring. Int. J. Syst. Evol. Microbiol. *61*, 132–137.

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

Volant, A., Bruneel, O., Desoeuvre, A., Hery, M., Casiot, C., Bru, N., Delpoux, S., Fahy, A., Javerliat, F., Bouchez, O., et al. (2014). Diversity and spatiotemporal dynamics of bacterial communities: physicochemical and other drivers along an acid mine drainage. Fems Microbiol. Ecol. *90*, 247–263.

Waack, S., Keller, O., Asper, R., Brodag, T., Damm, C., Fricke, W.F., Surovcik, K., Meinicke, P., and Merkl, R. (2006). Score-based prediction of genomic islands in prokaryotic genomes using hidden Markov models. BMC Bioinformatics *7*, 142.

Wang, X., Llopis, P.M., and Rudner, D.Z. (2013). Organization and segregation of bacterial chromosomes. Nat. Rev. Genet. *14*, 191–203.

Wen, Y., Behiels, E., and Devreese, B. (2014). Toxin- Antitoxin systems: their role in persistence, biofilm formation, and pathogenicity. Pathog. Dis. *70*, 240–249.

Wiedenbeck, J., and Cohan, F.M. (2011). Origins of bacterial diversity through horizontal genetic transfer and adaptation to new ecological niches. Fems Microbiol. Rev. *35*, 957–976.

Wozniak, R.A.F., and Waldor, M.K. (2010). Integrative and conjugative elements: mosaic mobile genetic elements enabling dynamic lateral gene flow. Nat. Rev. Microbiol. *8*, 552–563.

Zargar, K., Hoeft, S., Oremland, R., and Saltikov, C.W. (2010). Identification of a Novel Arsenite Oxidase Gene, arxA, in the Haloalkaliphilic, Arsenite-Oxidizing Bacterium Alkalilimnicola ehrlichii Strain MLHE-1. J. Bacteriol. *192*, 3755–3762.

Annexes

Annexe 1

Table S3. Details of the RGP found in CB2 genome

RGP Number	Found in…	ID	Gene	Product
RGP1	K12 and 3As in part	THICB2_v3_10001		Single-stranded DNA-binding protein
		THICB2_v3_10002		conserved protein of unknown function
		THICB2_v3_10003		protein of unknown function
		THICB2_v3_10004		exported protein of unknown function
		THICB2_v3_10005		conserved protein of unknown function
		THICB2_v3_10006		conserved protein of unknown function
		THICB2_v3_10007		conserved protein of unknown function
		THICB2_v3_10008		putative nucleoside 2-deoxyribosyltransferase
		THICB2_v3_10009		Integrase family protein
		THICB2_v3_10010		protein of unknown function
		THICB2_v3_10011		Phage integrase family protein
		THICB2_v3_10012		protein of unknown function
		THICB2_v3_10013		protein of unknown function
		THICB2_v3_10014		protein of unknown function
		THICB2_v3_10015		putative Prophage CP4-57 regulatory
		THICB2_v3_10016		protein of unknown function
		THICB2_v3_10017		Primase C 2 (PriCT-2) family
		THICB2_v3_10018		protein of unknown function
		THICB2_v3_10019		protein of unknown function
		THICB2_v3_10020		protein of unknown function
		THICB2_v3_tRNA45		Ser tRNA
RGP2	–	THICB2_v3_100008		protein of unknown function
		THICB2_v3_110001		Prevent-host-death family protein
		THICB2_v3_110002		PilT protein domain protein (fragment)
		THICB2_v3_110003		conserved membrane protein of unknown function
		THICB2_v3_110004		conserved membrane protein of unknown function
		THICB2_v3_110005		Glycosyl transferase family 2
		THICB2_v3_110006		Toxin-antitoxin system, antitoxin component, PHD family
		THICB2_v3_110007		PilT protein domain protein
		THICB2_v3_110008		Prevent-host-death family protein
		THICB2_v3_110009		conserved protein of unknown function
		THICB2_v3_110010		putative Polysaccharide biosynthesis protein
		THICB2_v3_110011		conserved protein of unknown function
		THICB2_v3_110012		dTDP-rhamnosyl transferase RfbF
		THICB2_v3_110013		Prevent-host-death family protein
		THICB2_v3_110014		conserved protein of unknown function
		THICB2_v3_110015		putative Glycosyl transferase family 2 (fragment)
		THICB2_v3_110016		putative Glycosyl transferase, family 2 (fragment)
		THICB2_v3_110017		putative glycosyltransferase

		THICB2_v3_110018		conserved protein of unknown function
		THICB2_v3_110019		conserved protein of unknown function
		THICB2_v3_110020		protein of unknown function
		THICB2_v3_110021		conserved protein of unknown function
		THICB2_v3_110022		conserved protein of unknown function
		THICB2_v3_110023		conserved protein of unknown function
RGP3	_	THICB2_v3_180001	istA	transposase
		THICB2_v3_180002		conserved protein of unknown function
		THICB2_v3_190001		transposase
		THICB2_v3_200001		protein of unknown function
		THICB2_v3_200002		protein of unknown function
		THICB2_v3_200003		protein of unknown function
		THICB2_v3_200004		Phage integrase family protein
		THICB2_v3_200005		conserved protein of unknown function
		THICB2_v3_200006		protein of unknown function
		THICB2_v3_200007		conserved protein of unknown function
		THICB2_v3_200008		conserved protein of unknown function
		THICB2_v3_200009		conserved protein of unknown function
		THICB2_v3_200010		Type III restriction protein res subunit (fragment)
		THICB2_v3_210001		Putative type III restriction enzyme, res subunit (fragment)
		THICB2_v3_210002		DNA methylase N-4/N-6 domain protein
		THICB2_v3_210003		putative phage-associated protein
		THICB2_v3_210004		conserved protein of unknown function
		THICB2_v3_210005		exported protein of unknown function
		THICB2_v3_210006		conserved protein of unknown function
		THICB2_v3_210007		protein of unknown function
		THICB2_v3_210008		protein of unknown function
		THICB2_v3_210009		protein of unknown function
		THICB2_v3_210010		protein of unknown function
		THICB2_v3_210011		protein of unknown function
		THICB2_v3_210012		protein of unknown function
		THICB2_v3_210013		exported protein of unknown function
		THICB2_v3_210014		putative Conjugal transfer protein TraN
		THICB2_v3_210015		protein of unknown function
		THICB2_v3_210016		putative Pilus assembly protein TraH
		THICB2_v3_210017		protein of unknown function
		THICB2_v3_220001		conserved protein of unknown function
		THICB2_v3_220002		protein of unknown function
		THICB2_v3_220003		protein of unknown function
		THICB2_v3_220004		conserved protein of unknown function
		THICB2_v3_220005		conserved protein of unknown function
		THICB2_v3_220006		conserved protein of unknown function
		THICB2_v3_220007		putative Transcriptional regulator, XRE family
		THICB2_v3_220008		putative Transcription repressor protein relE
		THICB2_v3_220009		Major facilitator superfamily MFS_1
		THICB2_v3_220010		conserved protein of unknown function
		THICB2_v3_220011		Transcriptional regulator, AraC family

		THICB2_v3_220012		3-isopropylmalate dehydratase
		THICB2_v3_220013		putative Anti-ECFsigma factor, ChrR (modular protein)
		THICB2_v3_220014		Isochorismatase hydrolase
		THICB2_v3_220015		conserved protein of unknown function
		THICB2_v3_220016	aspC	Aspartate aminotransferase
		THICB2_v3_220017		protein of unknown function
		THICB2_v3_220018		conserved protein of unknown function
		THICB2_v3_220019		Major facilitator superfamily MFS_1
		THICB2_v3_220020		conserved protein of unknown function
		THICB2_v3_220021		putative Taurine catabolism dioxygenase TauD, TfdA family
		THICB2_v3_220022		conserved membrane protein of unknown function
		THICB2_v3_220023		Predicted thioesterase
		THICB2_v3_220024		putative Permease of the drug/metabolite transporter (DMT) superfamily
		THICB2_v3_220025		putative Transcriptional regulator, AraC family
		THICB2_v3_220026		Conjugation TrbI family protein
		THICB2_v3_220027		Conjugal transfer protein TrbG/VirB9/CagX
		THICB2_v3_220028		conserved protein of unknown function
		THICB2_v3_220029		putative conjugal transfer protein
		THICB2_v3_220030		Transcriptional regulator, TetR family
		THICB2_v3_220031		Conserved protein of unknown function
		THICB2_v3_220032		Transcriptional regulator, AraC family
		THICB2_v3_220033		protein of unknown function
		THICB2_v3_220034		Major facilitator superfamily MFS_1
		THICB2_v3_220035		conserved protein of unknown function
		THICB2_v3_220036		protein of unknown function
		THICB2_v3_220037		putative MoaC domain protein (fragment)
		THICB2_v3_220038		Putative O-methyltransferase
		THICB2_v3_220039		protein of unknown function
		THICB2_v3_220040		conserved protein of unknown function
		THICB2_v3_220041		conserved protein of unknown function
		THICB2_v3_220042		P-type conjugative transfer protein TrbL
		THICB2_v3_220043		P-type conjugative transfer protein TrbJ
		THICB2_v3_220044		putative Cell wall surface anchor family protein (modular protein)
RGP4	few genes in K12	THICB2_v3_220056		conserved protein of unknown function
		THICB2_v3_220057		Putative addiction module antidote protein
		THICB2_v3_220058	trbB	putative conjugal transfer protein TrbB
		THICB2_v3_220059		conserved protein of unknown function
		THICB2_v3_220060		protein of unknown function
		THICB2_v3_220061		conserved exported protein of unknown function
		THICB2_v3_220062	cusA	copper/silver efflux system, membrane component
		THICB2_v3_220063		putative Heavy metal RND efflux membrane fusion protein, CzcB family (CusB or silB)
		THICB2_v3_220064		putative Integral outer membrane protein TolC,efflux pump component
		THICB2_v3_220065		protein of unknown function
		THICB2_v3_220066		conserved exported protein of unknown function
		THICB2_v3_220067		Transposase of ISThsp1, IS1182 family (fragment)

		THICB2_v3_220068		Methyltransferase type 11
		THICB2_v3_220069		conserved protein of unknown function
		THICB2_v3_220070	merA	Mercuric reductase
		THICB2_v3_220071	merP	Mercuric transport protein periplasmic component
		THICB2_v3_220072	merT	Mercuric transport protein
		THICB2_v3_220073	merR	Mercuric resistance operon regulatory protein
		THICB2_v3_220074		Transcriptional regulator, LysR family
		THICB2_v3_220075		Major facilitator superfamily MFS_1
		THICB2_v3_220076		Drug resistance transporter, EmrB/QacA subfamily
		THICB2_v3_220077		Transcriptional regulator, LysR family
		THICB2_v3_220078		protein of unknown function
		THICB2_v3_220079		protein of unknown function
		THICB2_v3_220080	insK	IS150 conserved protein InsB
		THICB2_v3_220081		transposase
		THICB2_v3_220082		OsmC family protein (modular protein)
		THICB2_v3_220083		protein of unknown function
		THICB2_v3_230001		GCN5-related N-acetyltransferase (fragment)
		THICB2_v3_230002		conserved protein of unknown function
		THICB2_v3_230003		Relaxase/mobilization nuclease family protein
		THICB2_v3_230004		conserved protein of unknown function
		THICB2_v3_230005		protein of unknown function
		THICB2_v3_230006		protein of unknown function
		THICB2_v3_230007		protein of unknown function
		THICB2_v3_230008		conserved protein of unknown function
		THICB2_v3_230009		TraB family protein
		THICB2_v3_230010		putative Permeases of the drug/metabolite transporter (DMT) superfamily
		THICB2_v3_230011		conserved protein of unknown function
		THICB2_v3_230012		Argininosuccinate synthase
		THICB2_v3_230013		putative HTH-type transcriptional regulator ltrA
		THICB2_v3_230014		protein of unknown function
		THICB2_v3_230015		putative DNA topoisomerase III (TopB) (fragment)
		THICB2_v3_230016		conserved protein of unknown function
		THICB2_v3_230017		conserved protein of unknown function
		THICB2_v3_230018		protein of unknown function
		THICB2_v3_230019		Bkd operon transcriptional regulator
		THICB2_v3_230020		Extracellular ligand-binding receptor
		THICB2_v3_230021		putative SAM-dependent methyltransferases
		THICB2_v3_230022	mmr	Methylenomycin A resistance protein
		THICB2_v3_230023	Ldhd	D-lactate dehydrogenase
		THICB2_v3_230024		Filamentation induced by cAMP protein Fic
		THICB2_v3_230025		DNA topoisomerase III
		THICB2_v3_230026		Lytic transglycosylase catalytic
		THICB2_v3_230027		Chorismate mutase, type II
		THICB2_v3_230028		conserved protein of unknown function
		THICB2_v3_230029		conserved protein of unknown function
		THICB2_v3_230030		protein of unknown function

		THICB2_v3_230031		conserved protein of unknown function
		THICB2_v3_230032		ParB-like partition protein
		THICB2_v3_230033		conserved protein of unknown function
		THICB2_v3_230034		Phage transcriptional regulator, AlpA
		THICB2_v3_230035		putative uncharacterized protein conserved in bacteria, prophage-related
		THICB2_v3_230036		putative phage repressor
		THICB2_v3_230037	ssb	Single-stranded DNA-binding protein
		THICB2_v3_230038		conserved protein of unknown function
		THICB2_v3_230039		conserved protein of unknown function
		THICB2_v3_230040		conserved protein of unknown function
		THICB2_v3_230041		Aldo/keto reductase
		THICB2_v3_230042		protein of unknown function
		THICB2_v3_230043	ycaN	Uncharacterized HTH-type transcriptional regulator LysR family
		THICB2_v3_230044		Putative nucleoside 2-deoxyribosyltransferase (fragment)
		THICB2_v3_230045		Transcriptional regulator, LysR family
		THICB2_v3_230046		Carboxymuconolactone decarboxylase
		THICB2_v3_230047		Alcohol dehydrogenase zinc-binding domain protein
		THICB2_v3_230048		putative permease of the major facilitator superfamily (MFS)
		THICB2_v3_230049		putative Zinc-binding alcohol dehydrogenase family protein (fragment)
		THICB2_v3_230050		Major facilitator superfamily MFS_1, EmrB/QacA subfamily
		THICB2_v3_230051		Transcriptional regulator, LysR family
		THICB2_v3_230052		Zinc-binding alcohol dehydrogenase family protein
		THICB2_v3_230053	ydeP	HTH-type transcriptional regulator HxlR family
		THICB2_v3_230054		conserved protein of unknown function, Darcynin homolog
		THICB2_v3_230055		Transcriptional regulator, TetR family
		THICB2_v3_230056		Major facilitator superfamily MFS_1
		THICB2_v3_230057		Transcriptional regulator, ArsR family
		THICB2_v3_230058		Major facilitator superfamily MFS_1
		THICB2_v3_230059		conserved protein of unknown function
		THICB2_v3_230060		Uncharacterized HTH-type transcriptional regulator lysrR family
		THICB2_v3_230061		protein of unknown function
		THICB2_v3_230062		putative chr0mate reductase
		THICB2_v3_230063		conserved protein of unknown function
		THICB2_v3_230064		Transcriptional regulator, MarR family
		THICB2_v3_230065		protein of unknown function
		THICB2_v3_230066		DNA damage-inducible protein DinB family protein
		THICB2_v3_230067		conserved protein of unknown function
		THICB2_v3_230068		protein of unknown function
		THICB2_v3_230069		Lysine exporter protein (LYSE/YGGA)
		THICB2_v3_230070		Integrase family protein
		THICB2_v3_tRNA41		Ser tRNA
RGP5	_	THICB2_v3_240009		protein of unknown function
		THICB2_v3_240010		putative LmbE family protein
		THICB2_v3_240011		putative GDP-mannose-dependent alpha-(1-6)-phosphatidylinositol dimannoside mannosyltransferase
		THICB2_v3_240012		conserved protein of unknown function
		THICB2_v3_240013		Efflux ABC transporter, ATP-binding protein

		THICB2_v3_240014		ABC-type polysaccharide/polyol phosphate export system, permease component
		THICB2_v3_240015		exported protein of unknown function
		THICB2_v3_240016		exported protein of unknown function
		THICB2_v3_240017		General secretion pathway protein D
		THICB2_v3_240018		protein of unknown function
		THICB2_v3_240019		putative Tetratricopeptide TPR_2 repeat protein
		THICB2_v3_240020		putative SAM-dependent methyltransferases
		THICB2_v3_240021		conserved protein of unknown function
		THICB2_v3_240022		putative General secretion pathway protein M
		THICB2_v3_240023		putative Fimbrial assembly protein
		THICB2_v3_240024		putative General secretion pathway protein K
		THICB2_v3_240025		putative General secretion pathway protein J
		THICB2_v3_240026	xpsI	General secretion pathway protein I
		THICB2_v3_240027		General secretion pathway protein H
		THICB2_v3_240028	gspG	pseudopilin, cryptic, general secretion pathway
		THICB2_v3_240029		Type II secretion system protein
		THICB2_v3_240030	xpsE	Type II secretion system protein E (GspE)
		THICB2_v3_240031		protein of unknown function
		THICB2_v3_240032		protein of unknown function
		THICB2_v3_240033		protein of unknown function
		THICB2_v3_250001		Putative UDP-N-ACETYLGLUCOSAMINE--PEPTIDE N-ACETYLGLUCOSAMINYLTRANSFERASE SUBUNIT(Fragment)
		THICB2_v3_250002		protein of unknown function
RGP6	_	THICB2_v3_280037		3',5'-cyclic-nucleotide phosphodiesterase; CpdA
		THICB2_v3_280038	potA	Polyamine-transporting ATPase
		THICB2_v3_280039		putative ABC-type transport system, permease component
		THICB2_v3_280040		putative Binding-protein-dependent transport systems inner membrane component
		THICB2_v3_280041		Putative ABC-type Fe3+ transport system periplasmic component
		THICB2_v3_280042	potA	Spermidine/putrescine import ATP-binding protein PotA
		THICB2_v3_280043		Putative periplasmic iron-binding protein
		THICB2_v3_280044		putative Permease protein, ABC-type iron transporter
		THICB2_v3_280045	ugpC2	glycerol-3-phosphate transporter subunit ; ATP-binding component of ABC superfamily
		THICB2_v3_280046	ugpB2	sn-glycerol-3-phosphate-binding periplasmic protein UgpB
		THICB2_v3_280047	ugpA2	sn-glycerol-3-phosphate transport system permease protein UgpA
		THICB2_v3_280048	ugpE2	sn-glycerol-3-phosphate transport system permease protein UgpE
		THICB2_v3_280049		Porin Gram-negative type
		THICB2_v3_280050	ugpB	glycerol-3-phosphate transporter subunit ; periplasmic-binding component of ABC superfamily
		THICB2_v3_280051	ugpA	glycerol-3-phosphate transporter subunit ; membrane component of ABC superfamily
		THICB2_v3_280052	ugpE	glycerol-3-phosphate transporter subunit ; membrane component of ABC superfamily
		THICB2_v3_280053	ugpC	glycerol-3-phosphate transporter subunit ; ATP-binding component of ABC superfamily

RGP7	3AS	THICB2_v3_280115		conserved protein of unknown function
		THICB2_v3_280116		putative Bacterial regulatory protein, Crp
		THICB2_v3_280117		putative CsbD protein, yjbJ
		THICB2_v3_280118		conserved protein of unknown function
		THICB2_v3_280119		putative Transport-associated protein
		THICB2_v3_280120	osmB	Osmotically-inducible lipoprotein B
		THICB2_v3_280122		putative Transglutaminase-like protein
		THICB2_v3_280123		Putative universal stress protein family UspA
		THICB2_v3_280124		conserved hypothetical protein
		THICB2_v3_290001		conserved protein of unknown function
		THICB2_v3_290002		putative NdvB protein
		THICB2_v3_290003		conserved protein of unknown function
		THICB2_v3_290004		putative Stomatin protein
		THICB2_v3_290005		conserved protein of unknown function
RGP8	_	THICB2_v3_320001		putative 5-methylcytosine restriction system component-like protein
		THICB2_v3_320002		protein of unknown function
		THICB2_v3_320003		protein of unknown function
		THICB2_v3_320004		conserved protein of unknown function
		THICB2_v3_330001		protein of unknown function
		THICB2_v3_330002		putative Metallophosphoesterase
		THICB2_v3_330003		conserved protein of unknown function
		THICB2_v3_330004		conserved protein of unknown function
		THICB2_v3_330005		conserved protein of unknown function
		THICB2_v3_330006		conserved protein of unknown function
		THICB2_v3_330007		conserved protein of unknown function
		THICB2_v3_330008		Putative DNA helicase
		THICB2_v3_330009		conserved protein of unknown function
		THICB2_v3_330010		protein of unknown function
		THICB2_v3_330011		conserved protein of unknown function
		THICB2_v3_330012		conserved protein of unknown function
		THICB2_v3_330013		protein of unknown function
		THICB2_v3_340001		conserved protein of unknown function
		THICB2_v3_340002		putative transposition helper protein
		THICB2_v3_340003		putative integrase, catalytic region
		THICB2_v3_340004		conserved protein of unknown function
		THICB2_v3_340005		protein of unknown function
		THICB2_v3_340006		conserved protein of unknown function
		THICB2_v3_340007		protein of unknown function
		THICB2_v3_340008		protein of unknown function
		THICB2_v3_340009		conserved protein of unknown function
		THICB2_v3_340010		conserved protein of unknown function
		THICB2_v3_340011		protein of unknown function
		THICB2_v3_340012		protein of unknown function
		THICB2_v3_340013		protein of unknown function
RGP9	3As	THICB2_v3_misc_RNA_3		ykkC-yxkD
		THICB2_v3_360017		putative ABC-type nitrate transport system, periplasmic component
		THICB2_v3_360018		putative Nitrate transport permease protein nrtB

		THICB2_v3_360019		conserved protein of unknown function
		THICB2_v3_360020		conserved protein of unknown function
		THICB2_v3_360021		conserved protein of unknown function
		THICB2_v3_360022		Urea carboxylase
		THICB2_v3_360023		putative Allophanate hydrolase
		THICB2_v3_360024	cynS	Cyanate hydratase
		THICB2_v3_360025		conserved protein of unknown function
		THICB2_v3_360026		putative AHL-lactonase
		THICB2_v3_360027	crnA	Creatinine amidohydrolase
		THICB2_v3_360028		putative Permease of the drug/metabolite transporter (DMT) superfamily
		THICB2_v3_360029	fabG	3-oxoacyl-[acyl-carrier-protein] reductase FabG
		THICB2_v3_360030		putative 3-oxoacyl-[acyl-carrier-protein] reductase
		THICB2_v3_360031		putative fatty acid desaturase
		THICB2_v3_360032		putative ABC-type nitrate/sulfonate/bicarbonate transport system, periplasmic component
		THICB2_v3_360033		conserved protein of unknown function
		THICB2_v3_360034		putative alpha/beta-Hydrolases
		THICB2_v3_360035		putative ABC-type transport system, ATPase component
		THICB2_v3_360036		putative ABC-type transport system, permease component
		THICB2_v3_360037		putative ABC-type transport system, periplasmic component
		THICB2_v3_360038		putative transcriptional regulator, GntR family
		THICB2_v3_360039		conserved protein of unknown function
		THICB2_v3_360040	atzE	Biuret hydrolase
		THICB2_v3_360041		conserved protein of unknown function
		THICB2_v3_360042		putative 5-formyltetrahydrofolate cyclo-ligase
		THICB2_v3_360043		putative Monosaccharide-transporting ATPase (ABC-type transport system)
		THICB2_v3_360044		putative ABC-type transport system, permease component
		THICB2_v3_360045		putative transport protein (ABC superfamily, membrane) (fragment)
		THICB2_v3_370001		Putative ABC-type transport system, permease component (fragment)
		THICB2_v3_370002		putative ABC-type transport system, periplasmic component
		THICB2_v3_370003	amiF	Formamidase
		THICB2_v3_370004		Putative Isochorismatase hydrolase
		THICB2_v3_370005		Putative Isochorismatase hydrolase
		THICB2_v3_370006		putative S-adenosylhomocysteine deaminase
		THICB2_v3_370007		putative NAD-dependent dihydropyrimidine dehydrogenase sunbunit PreT
		THICB2_v3_370008	yeiA	putative oxidoreductase subunit
		THICB2_v3_370009	dht	D-hydantoinase/dihydropyrimidinase
		THICB2_v3_370010	atcC	Amidase, hydantoinase/carbamoylase family
		THICB2_v3_370011		conserved protein of unknown function
		THICB2_v3_370012		putative ABC-type transport system, permease component
		THICB2_v3_370013		putative ABC-type transport system, permease component
		THICB2_v3_370014		putative ABC-type transport system, periplasmic component
		THICB2_v3_370015	rutR	HTH-type transcriptional regulator RutR
		THICB2_v3_370016		protein of unknown function

		THICB2_v3_370017	codA	fragment of cytosine deaminase (part 1)
		THICB2_v3_370018	codA	fragment of cytosine deaminase (part 2)
		THICB2_v3_370019		protein of unknown function
		THICB2_v3_370020		conserved protein of unknown function
		THICB2_v3_370021		putative hydrolase protein
		THICB2_v3_370022		putative transport protein (ABC superfamily, atp_bind)
		THICB2_v3_370023		putative ring-hydroxylating dioxygenase with Rieske 2Fe-2S domain
		THICB2_v3_370024		protein of unknown function
		THICB2_v3_370025		putative binding protein (ABC superfamily, peri_bind)
		THICB2_v3_370026		putative binding protein (ABC superfamily, peri_bind)
		THICB2_v3_370027		putative permease protein (ABC superfamily, membrane)
		THICB2_v3_370028		putative Clavaminate synthase-like
		THICB2_v3_370029		putative Nucleoside-binding outer membrane protein
		THICB2_v3_370030		putative ribonuclease VapC
		THICB2_v3_370031		transposase
		THICB2_v3_370033		putative Sirohydrochlorin cobaltochelatase CbiX
		THICB2_v3_370034		putative Response regulator antiterminator nasT
		THICB2_v3_370035	nasF	Nitrate transporter component
		THICB2_v3_370036	nrtB	Nitrate transporter component
		THICB2_v3_370037	nasD	Nitrate transport protein NasD
		THICB2_v3_370038	nasD	Nitrite reductase [NAD(P)H]
		THICB2_v3_370039	nasE	Assimilatory nitrite reductase [NAD(P)H] small subunit
		THICB2_v3_370040		putative Nitrate reductase
		THICB2_v3_370041		putative anthranilate phosphoribosyltransferase
		THICB2_v3_370042	cobA	Uroporphyrinogen-III C-methyltransferase
		THICB2_v3_370043		putative transposase (fragment)
		THICB2_v3_370044		putative transposase
		THICB2_v3_370045		conserved protein of unknown function
		THICB2_v3_370046		putative DnaJ-class molecular chaperone
		THICB2_v3_370047		putative plasmid stabilization system protein
		THICB2_v3_370048		conserved protein of unknown function
		THICB2_v3_370049	ureD	Urease accessory protein UreD
		THICB2_v3_370050		putative branched-chain amino acid transport protein (ABC superfamily, ATP_bind)
		THICB2_v3_370051		putative branched-chain amino acid transport protein (ABC superfamily, atp_bind)
		THICB2_v3_370052		putative branched-chain amino acid permease protein (ABC superfamily, membrane)
		THICB2_v3_370053		putative branched-chain amino acid permease protein (ABC superfamily, membrane)
		THICB2_v3_370054		putative branched-chain amino acid transport protein (ABC superfamily, peri_bind)
		THICB2_v3_370055		conserved protein of unknown function
		THICB2_v3_370056	ureA	urease gamma subunit
		THICB2_v3_370057	ureB	urease beta subunit
		THICB2_v3_370058	ureC	urease alpha subunit
RGP10	K12 in part	THICB2_v3_380001		putative two-component sensor histidine kinase
		THICB2_v3_380002	ompR	Transcriptional regulatory protein OmpR
		THICB2_v3_390001		protein of unknown function

		THICB2_v3_390002		protein of unknown function
		THICB2_v3_390003	zraR	Transcriptional regulatory protein aioR
		THICB2_v3_390004		Integral membrane sensor signal transduction histidine kinase
		THICB2_v3_390005		Phosphonate-binding periplasmic protein
		THICB2_v3_390006	aioB	Arsenite oxidase subunit AioB
		THICB2_v3_390007	aioA	Arsenite oxidase subunit AioA
		THICB2_v3_390008	cyt	Cytochrome c-552
		THICB2_v3_390009		Nitroreductase-like
		THICB2_v3_390010		Transcriptional regulator, ArsR family
		THICB2_v3_390011		conserved protein of unknown function
		THICB2_v3_390012		Putative tetrathionate reductase, subunit A
		THICB2_v3_390013		putative Tetrathionate reductase subunit C
		THICB2_v3_390014		4Fe-4S ferredoxin iron-sulfur binding domain protein, putative Tetrathionate reductase subunit B
		THICB2_v3_390015		protein of unknown function
		THICB2_v3_390016		conserved membrane protein of unknown function
		THICB2_v3_390017		conserved membrane protein of unknown function
		THICB2_v3_390018		conserved protein of unknown function
		THICB2_v3_390019		Lipid A 3-O-deacylase-related protein
		THICB2_v3_390020		FAD-dependent pyridine nucleotide-disulfide oxidoreductase
		THICB2_v3_390021		Acriflavin resistance protein
		THICB2_v3_390022		Efflux transporter, RND family, MFP subunit
		THICB2_v3_390023		Outer membrane efflux protein
		THICB2_v3_390024		OsmC family protein
		THICB2_v3_390025		ATPase, P-type (Transporting), HAD superfamily, subfamily IC
		THICB2_v3_390026		conserved protein of unknown function
		THICB2_v3_390027		Cobalt-zinc-cadmium resistance protein CzcD
		THICB2_v3_390028		conserved protein of unknown function
		THICB2_v3_390029		Glyoxalase/bleomycin resistance protein/dioxygenase
		THICB2_v3_390030		Major intrinsic protein
		THICB2_v3_390031		conserved protein of unknown function
		THICB2_v3_390032		putative Carboxymuconolactone decarboxylase
		THICB2_v3_390033		Arsenical resistance protein ArsH
		THICB2_v3_390034		conserved exported protein of unknown function
		THICB2_v3_390035		Thiol-disulfide isomerase and thioredoxins family protein
		THICB2_v3_390036		conserved exported protein of unknown function
		THICB2_v3_390037		Permease
		THICB2_v3_390038		RNA polymerase, sigma-24 subunit, ECF subfamily
		THICB2_v3_390040		putative mobA
		THICB2_v3_390041		conserved protein of unknown function
		THICB2_v3_390042		conserved protein of unknown function
		THICB2_v3_390043		conserved protein of unknown function
		THICB2_v3_390044		conserved protein of unknown function
		THICB2_v3_390045		putative DNA primase
		THICB2_v3_390046		conserved protein of unknown function
		THICB2_v3_390047		conserved protein of unknown function
		THICB2_v3_390048	insH	IS5 transposase and trans-activator; CP4-44 prophage

		THICB2_v3_390049		Conserved protein of unknown function
		THICB2_v3_390050		conserved protein of unknown function
		THICB2_v3_390051		Phosphoglycerate mutase
		THICB2_v3_390052		Putative Bis(5'-adenosyl)-triphosphatase
		THICB2_v3_390053		putative Major facilitator superfamily, general substrate transporter
		THICB2_v3_390054		IstB domain protein ATP-binding protein
		THICB2_v3_390055		Integrase catalytic region
		THICB2_v3_390056		putative Major facilitator superfamily MFS_1 (fragment)
		THICB2_v3_390057		protein of unknown function
		THICB2_v3_400001		putative Major facilitator superfamily MFS_1 (fragment)
		THICB2_v3_400002		protein of unknown function
		THICB2_v3_400003		Integrase (fragment)
		THICB2_v3_400004		protein of unknown function
		THICB2_v3_400005		protein of unknown function
		THICB2_v3_400006	hypE	carbamoyl phosphate phosphatase, hydrogenase 3 maturation protein
		THICB2_v3_400007	hypD	protein required for maturation of hydrogenases
		THICB2_v3_400008		putative sedoheptulose 7-phosphate isomerase
		THICB2_v3_400009	hypC	Hydrogenase expression/formation protein hypC
		THICB2_v3_400010	hypF	Carbamoyltransferase hypF
		THICB2_v3_400011		Nickel-dependent hydrogenase large subunit hupV
		THICB2_v3_400012		Uptake hydrogenase accessory protein hupU
		THICB2_v3_400013		Signal transduction histidine kinase HoxJ (hydrogenase regulation)
		THICB2_v3_400014	hoxA	Hydrogenase transcriptional regulatory protein hoxA
		THICB2_v3_400015		protein of unknown function
		THICB2_v3_400016	hypB	GTP hydrolase involved in nickel liganding into hydrogenases
		THICB2_v3_400017	hybF	protein involved with the maturation of hydrogenases 1 and 2
		THICB2_v3_400018		Ni,Fe-hydrogenase I large subunit
		THICB2_v3_400019		Hydrogenase expression/formation protein HoxT
		THICB2_v3_400020	hoxR	Rubredoxin
		THICB2_v3_400021	hoxQ	Hydrogenase expression/formation protein hoxQ
		THICB2_v3_400022	hoxO	Hydrogenase expression/formation protein hoxO
		THICB2_v3_400023	hupN	Hydrogenase expression/formation protein hupN
		THICB2_v3_400024	hupM	Hydrogenase expression/formation protein hupM
		THICB2_v3_400025		protein of unknown function
		THICB2_v3_400026	hyaA	hydrogenase 1, small subunit
		THICB2_v3_400027	hyaB	hydrogenase 1, large subunit
		THICB2_v3_400028	hyaC	hydrogenase 1, b-type cytochrome subunit
		THICB2_v3_400029	nixA	High-affinity nickel-transport protein nixA
		THICB2_v3_400030		transposase (fragment)
		THICB2_v3_400031		putative Site-specific recombinase, phage integrase family
		THICB2_v3_400032	pcl	Fe(2+)/Mn(2+) transporter pcl1
		THICB2_v3_400033		conserved protein of unknown function
		THICB2_v3_400034		membrane protein of unknown function
		THICB2_v3_400035		protein of unknown function
		THICB2_v3_400036		conserved protein of unknown function
		THICB2_v3_400037		Asparagine synthase family amidotransferase

		THICB2_v3_400038		protein of unknown function
		THICB2_v3_410001		conserved protein of unknown function
RGP11	K12	THICB2_v3_490001		protein of unknown function
		THICB2_v3_490002		IstB-like ATP binding N-terminal
		THICB2_v3_490003		transposase
		THICB2_v3_500001		Integrase, catalytic region (fragment)
RGP12	3As in part	THICB2_v3_500032		putative phage integrase/recombinase protein
		THICB2_v3_500033		Protein of unknown function
		THICB2_v3_500034		conserved exported protein of unknown function
		THICB2_v3_500035		conserved protein of unknown function
		THICB2_v3_500036		conserved protein of unknown function
		THICB2_v3_510001		conserved protein of unknown function
		THICB2_v3_510002		protein of unknown function
		THICB2_v3_510003		protein of unknown function
		THICB2_v3_510004		putative Mobilisation protein
		THICB2_v3_510005		protein of unknown function
		THICB2_v3_510006		putative Transposase (fragment)
		THICB2_v3_510007		putative 3-oxoacyl-[acyl-carrier-protein] reductase FabG
		THICB2_v3_510008		Acetamidase/Formamidase
		THICB2_v3_510009		Urea ABC transporter, ATP-binding protein UrtE
		THICB2_v3_510010		Urea ABC transporter, ATPase protein UrtD
		THICB2_v3_510011		Urea ABC transporter, permease protein UrtC
		THICB2_v3_510012		Urea ABC transporter, permease protein UrtB
		THICB2_v3_510013		Branched-chain amino acid ABC transporter, periplasmic amino acid-binding protein
		THICB2_v3_510014		ABC branched chain amino acid family transporter, periplasmic ligand binding protein
		THICB2_v3_510015		ANTAR
		THICB2_v3_510016		Carbonic anhydrase
		THICB2_v3_510017		protein of unknown function
		THICB2_v3_510018		protein of unknown function
		THICB2_v3_510021		protein of unknown function
		THICB2_v3_510023		conserved protein of unknown function
		THICB2_v3_510024		protein of unknown function
		THICB2_v3_520001		transposase
		THICB2_v3_520002		protein of unknown function
		THICB2_v3_520003		putative Phage integrase
		THICB2_v3_tRNA34		Met tRNA
RGP13	3As and K12 in part	THICB2_v3_590112		protein of unknown function
		THICB2_v3_590113		protein of unknown function
		THICB2_v3_590114		conserved protein of unknown function
		THICB2_v3_590115		conserved protein of unknown function
		THICB2_v3_590116		protein of unknown function
		THICB2_v3_590117		protein of unknown function
		THICB2_v3_590118		protein of unknown function
		THICB2_v3_590119		putative Conjugal transfer protein traB

		THICB2_v3_590120		conserved exported protein of unknown function
		THICB2_v3_590121		HicA protein
		THICB2_v3_590122		HicB family protein
		THICB2_v3_590123		exported protein of unknown function
		THICB2_v3_590124		protein of unknown function
		THICB2_v3_590125		protein of unknown function
		THICB2_v3_590126		protein of unknown function
		THICB2_v3_590127		conserved protein of unknown function
		THICB2_v3_590128		protein of unknown function
		THICB2_v3_590129		protein of unknown function
		THICB2_v3_590130		conserved protein of unknown function
		THICB2_v3_590131		conserved membrane protein of unknown function
		THICB2_v3_590132		protein of unknown function
		THICB2_v3_590133		conserved protein of unknown function
		THICB2_v3_600001	cheY	chemotaxis regulator transmitting signal to flagellar motor component
		THICB2_v3_600002		putative Chemotaxis protein cheZ
		THICB2_v3_600003		putative Chemotaxis protein histidine kinase CheW
		THICB2_v3_600004		Putative Methyl-accepting chemotaxis protein
		THICB2_v3_600005		putative beta-lactamase-like
		THICB2_v3_600006		Methyl-accepting chemotaxis sensory transducer
		THICB2_v3_600007		Methyl-accepting chemotaxis sensory transducer (fragment)
		THICB2_v3_610001		conserved protein of unknown function
		THICB2_v3_610002		putative Integrase catalytic region (fragment)
		THICB2_v3_610003		protein of unknown function
		THICB2_v3_610004		Conjugal transfer protein trbF
		THICB2_v3_610005		Conjugal transfer protein TrbG/VirB9/CagX
		THICB2_v3_610006		Conjugation TrbI family protein
		THICB2_v3_610007		protein of unknown function
		THICB2_v3_610008		protein of unknown function
		THICB2_v3_610009		conserved protein of unknown function
		THICB2_v3_610010		conserved membrane protein of unknown function
		THICB2_v3_610011		conserved protein of unknown function
		THICB2_v3_610012		conserved protein of unknown function
		THICB2_v3_610013		conserved protein of unknown function
		THICB2_v3_610014		conserved membrane protein of unknown function
		THICB2_v3_610015		putative Beta-lactamase
		THICB2_v3_610016		conserved protein of unknown function
		THICB2_v3_610017		conserved protein of unknown function
		THICB2_v3_610018		putative Pili subunit, PilE
		THICB2_v3_610019		conserved protein of unknown function
		THICB2_v3_610020		conserved protein of unknown function
		THICB2_v3_610021		putative Integration host factor (IHF)
		THICB2_v3_610022		Putative DNA methyltransferase YeeA (fragment)
		THICB2_v3_620001		exported protein of unknown function
		THICB2_v3_620002		lipoprotein pcp or SlyB
		THICB2_v3_620003	trbE	Conjugal transfer protein TrbE
		THICB2_v3_620004		Type IV secretory pathway, TrbD component

		THICB2_v3_620005		putative conjugal transfer protein TrbC
		THICB2_v3_620006		protein of unknown function
		THICB2_v3_620007		protein of unknown function
		THICB2_v3_620008		conserved protein of unknown function
		THICB2_v3_620009		Transcriptional regulator, XRE family
		THICB2_v3_620010	trbB	putative conjugal transfer protein TrbB
		THICB2_v3_620011		conserved protein of unknown function
		THICB2_v3_630001		protein of unknown function
		THICB2_v3_630002		conserved protein of unknown function
		THICB2_v3_630003		putative Lytic transglycosylase
		THICB2_v3_630004		protein of unknown function
		THICB2_v3_630005		DNA topoisomerase
		THICB2_v3_630006		protein of unknown function
		THICB2_v3_630007		protein of unknown function
		THICB2_v3_630008		putative Phage integrase
		THICB2_v3_630009		protein of unknown function
		THICB2_v3_tRNA31		Leu tRNA
		THICB2_v3_tRNA30		Leu tRNA
RGP14	3As in part	THICB2_v3_tRNA15		Pro tRNA
		THICB2_v3_640043		protein of unknown function
		THICB2_v3_640044		putative Relaxase/mobilization nuclease family protein
		THICB2_v3_640045		conserved protein of unknown function
		THICB2_v3_640046		conserved protein of unknown function
		THICB2_v3_640047		protein of unknown function
		THICB2_v3_640048		conserved protein of unknown function
		THICB2_v3_640049		conserved protein of unknown function
		THICB2_v3_640051		conserved protein of unknown function
		THICB2_v3_640052		conserved protein of unknown function
		THICB2_v3_640053		putative Transcriptional regulator, GntR family with aminotransferase domain
		THICB2_v3_640054		protein of unknown function
		THICB2_v3_640055		Branched-chain amino acid transport protein AzlC
		THICB2_v3_640056		Branched-chain amino acid transport protein AzlD
		THICB2_v3_640057		conserved hypothetical protein iopA
		THICB2_v3_640058		conserved hypothetical protein IopB
		THICB2_v3_640059	strB	Scyllo-inosamine-4-phosphate amidinotransferase
		THICB2_v3_640060		putative lysine exporter
		THICB2_v3_640061		putative Carbamoyltransferase of the NodU/CmcH family
		THICB2_v3_640062		protein of unknown function
		THICB2_v3_640063		putative Adenosylmethionine--8-amino-7-oxononanoate transaminase
		THICB2_v3_640064		conserved hypothetical protein
		THICB2_v3_640065		protein of unknown function
		THICB2_v3_640066		conserved hypothetical protein
		THICB2_v3_640067		Lysine exporter protein (LYSE/YGGA)
		THICB2_v3_640068		conserved protein of unknown function
		THICB2_v3_640069		conserved protein of unknown function

RGP15	K12	THICB2_v3_tRNA19		His tRNA
		THICB2_v3_640116		protein of unknown function
		THICB2_v3_640117		protein of unknown function
		THICB2_v3_640118		protein of unknown function
		THICB2_v3_640119		protein of unknown function
		THICB2_v3_640120		conserved protein of unknown function
		THICB2_v3_640121		protein of unknown function
		THICB2_v3_640122		conserved protein of unknown function
		THICB2_v3_640123		Putative Primase 2
		THICB2_v3_640124		conserved protein of unknown function
		THICB2_v3_640125		protein of unknown function
		THICB2_v3_640126		protein of unknown function
		THICB2_v3_640127		Addiction module antidote protein, HigA family
		THICB2_v3_640128		Plasmid maintenance system killer HigB like
		THICB2_v3_640129	intF	prophage CP4-6 integrase
		THICB2_v3_tRNA20		Pseudo tRNA
RGP16	K12	THICB2_v3_660002	cheB	fused chemotaxis regulator ; protein-glutamate methylesterase in two-component regulatory system with CheA
		THICB2_v3_660003	cheD	chemoreceptor glutamine deamidase CheD 3
		THICB2_v3_660004	cheR	chemotaxis regulator, protein-glutamate methyltransferase
		THICB2_v3_660005		Diguanylate cyclase/phosphodiesterase with PAS/PAC sensor(S)
		THICB2_v3_660006		protein of unknown function
RGP16?		THICB2_v3_920001		Methyl-accepting chemotaxis sensory transducer (fragment)
		THICB2_v3_920002		CheW protein
		THICB2_v3_920003		Methyl-accepting chemotaxis sensory transducer with Cache sensor
RGP17	_	THICB2_v3_misc_RNA_4		tmRNA
		THICB2_v3_700065		putative Phage integrase
		THICB2_v3_700066		putative Phage transcriptional regulator, AlpA
		THICB2_v3_710001		conserved hypothetical protein
		THICB2_v3_710002		conserved protein of unknown function
		THICB2_v3_710003		protein of unknown function
		THICB2_v3_710004		putative Crossover junction endodeoxyribonuclease
		THICB2_v3_710005		protein of unknown function
		THICB2_v3_710006		conserved protein of unknown function
		THICB2_v3_710007		conserved protein of unknown function
		THICB2_v3_710008		conserved protein of unknown function
		THICB2_v3_tRNA28		Ser tRNA
RGP18	_	THICB2_v3_710020		protein of unknown function
		THICB2_v3_710021		putative Site-specific recombinase, phage integrase family
		THICB2_v3_710022	topB	DNA topoisomerase 3
		THICB2_v3_710023		Lytic transglycosylase catalytic
		THICB2_v3_710024		protein of unknown function
		THICB2_v3_710025		conserved protein of unknown function
		THICB2_v3_710026	trbB	putative conjugal transfer protein TrbB
		THICB2_v3_710027		conserved protein of unknown function
		THICB2_v3_710028		conserved protein of unknown function
		THICB2_v3_710029		putative relaxase/mobilization protein mobB
		THICB2_v3_710030		Endonuclease relaxase, MobA/VirD2

		THICB2_v3_710031		putative conjugal transfer protein TrbC
		THICB2_v3_710032		putative Type IV secretory pathway, TrbD component
		THICB2_v3_710033	trbE	Conjugal transfer protein TrbE
		THICB2_v3_710034		17 kDa surface antigen
		THICB2_v3_710035		exported protein of unknown function
		THICB2_v3_720001		Cell wall surface anchor family protein (fragment)
		THICB2_v3_720002		P-type conjugative transfer protein TrbJ
		THICB2_v3_720003		exported protein of unknown function
		THICB2_v3_720004		putative Type IV secretory pathway TrbL components-like protein
		THICB2_v3_720005		protein of unknown function
		THICB2_v3_720006		protein of unknown function
		THICB2_v3_720007		protein of unknown function
		THICB2_v3_720008		conserved protein of unknown function
		THICB2_v3_720009		protein of unknown function
		THICB2_v3_720010		protein of unknown function
		THICB2_v3_720011		conserved protein of unknown function
		THICB2_v3_720012		exported protein of unknown function
		THICB2_v3_720013	traG	Conjugal transfer protein TraG
		THICB2_v3_720014		protein of unknown function
		THICB2_v3_720015	pin	site-specific DNA recombinase; e14 prophage
		THICB2_v3_720016		conserved protein of unknown function
		THICB2_v3_720017		conserved membrane protein of unknown function
		THICB2_v3_720018		conserved protein of unknown function
		THICB2_v3_720019		conserved protein of unknown function
		THICB2_v3_720020		conserved protein of unknown function
		THICB2_v3_720021		conserved protein of unknown function
		THICB2_v3_720022		conserved protein of unknown function, putative Nucleotidyltransferase
		THICB2_v3_730001	tnpA	transposase
		THICB2_v3_730002		putative Type IV secretory pathway, protease TraF
		THICB2_v3_730003		DNA methyltransferase MmeI
		THICB2_v3_730004	ihfB	Integration host factor subunit beta
		THICB2_v3_730005		Conjugation TrbI family protein
		THICB2_v3_730006		Conjugal transfer protein TrbG/VirB9/CagX
		THICB2_v3_730007		Conjugal transfer protein trbF
		THICB2_v3_730008		protein of unknown function
		THICB2_v3_730009		exported protein of unknown function
		THICB2_v3_730010		protein of unknown function
		THICB2_v3_730011		protein of unknown function
		THICB2_v3_730012		conserved protein of unknown function
		THICB2_v3_730013		protein of unknown function
		THICB2_v3_730014		exported protein of unknown function
		THICB2_v3_730015		HicB family protein
		THICB2_v3_730016		HicA protein
		THICB2_v3_730017		putative TraB family protein
		THICB2_v3_730018		conserved protein of unknown function

		THICB2_v3_730019		protein of unknown function
		THICB2_v3_730020		protein of unknown function
		THICB2_v3_730021		protein of unknown function
		THICB2_v3_730022		protein of unknown function
		THICB2_v3_730023		protein of unknown function
RGP19	3As in part	THICB2_v3_tRNA23		Thr tRNA
		THICB2_v3_730162		putative domain of the Conjugation TrbI family protein
		THICB2_v3_730163		putative Conjugal transfer protein TrbG/VirB9/CagX
		THICB2_v3_730164		putative Conjugal transfer protein trbF
		THICB2_v3_730165		conserved protein of unknown function
		THICB2_v3_730166		conserved protein of unknown function
		THICB2_v3_730167		putative P-type conjugative transfer protein TrbL
		THICB2_v3_730168		putative P-type conjugative transfer protein TrbJ
		THICB2_v3_730169		Conserved protein of unknown function
		THICB2_v3_730170		Type IV secretory pathway, VirB4 component (CagE, TrbE, VirB)
		THICB2_v3_730171		putative Type IV secretory pathway, TrbD component
		THICB2_v3_730172		conserved protein of unknown function
		THICB2_v3_730173		putative conjugal transfer protein TrbC
		THICB2_v3_730174		conserved protein of unknown function
		THICB2_v3_730175		conserved protein of unknown function
		THICB2_v3_730176		putative Type IV secretory pathway, protease TraF
		THICB2_v3_730177		conserved protein of unknown function
		THICB2_v3_730178		conserved protein of unknown function
		THICB2_v3_730179		putative Transcriptional regulator, XRE family (fragment)
		THICB2_v3_730180		putative EAL domain protein
		THICB2_v3_730181		TRAG family protein
		THICB2_v3_730182		protein of unknown function
		THICB2_v3_730183		protein of unknown function
		THICB2_v3_730184		protein of unknown function
		THICB2_v3_730185		protein of unknown function
		THICB2_v3_730186		conserved protein of unknown function
		THICB2_v3_730187		conserved protein of unknown function
		THICB2_v3_730188		conserved protein of unknown function
		THICB2_v3_730189		protein of unknown function
		THICB2_v3_730190	trbB	putative conjugal transfer protein TrbB
		THICB2_v3_730191		conserved protein of unknown function
		THICB2_v3_730192		protein of unknown function
		THICB2_v3_730193		protein of unknown function
		THICB2_v3_730194		putative Transcriptional regulator, XRE family
		THICB2_v3_730195		protein of unknown function
		THICB2_v3_730196	traI	Relaxase/Mobilisation nuclease traI
		THICB2_v3_730197		putative Mobilization protein MobB
		THICB2_v3_730198		protein of unknown function
		THICB2_v3_730199		protein of unknown function
		THICB2_v3_730200		protein of unknown function
		THICB2_v3_730201		protein of unknown function
		THICB2_v3_730202		membrane protein of unknown function
		THICB2_v3_740001		conserved protein of unknown function

		THICB2_v3_740002		Type IV secretion pathway protein pilN
		THICB2_v3_740003		putative transglycosylase
		THICB2_v3_740004		protein of unknown function
		THICB2_v3_740005		putative Protein virB1 (fragment)
		THICB2_v3_740006		protein of unknown function
		THICB2_v3_740007		putative phage integrase
		THICB2_v3_740008		conserved protein of unknown function
		THICB2_v3_740009		conserved protein of unknown function
		THICB2_v3_740010		conserved membrane protein of unknown function
		THICB2_v3_740011		putative ATPases of the AAA+ class
		THICB2_v3_740012		putative AAA family ATPase
		THICB2_v3_740013		conserved protein of unknown function
		THICB2_v3_740014		conserved exported protein of unknown function
		THICB2_v3_740015		Transcriptional regulator, LysR family (fragment)
		THICB2_v3_740016		protein of unknown function
		THICB2_v3_740017		transposase IS116/IS110(fragment)
		THICB2_v3_740018		protein of unknown function
		THICB2_v3_750001		Outer membrane efflux protein (fragment)
		THICB2_v3_750002		Transcriptional regulator, ArsR family
		THICB2_v3_750003		conserved exported protein of unknown function
		THICB2_v3_750004		conserved protein of unknown function (fragment)
		THICB2_v3_750005	insI	IS30 transposase; CP4-6 prophage
		THICB2_v3_750006	retA	RNA-directed DNA polymerase (Reverse transcriptase)
		THICB2_v3_750007	merA	Mercuric reductase
		THICB2_v3_750008	merC	Mercuric resistance protein MerC
		THICB2_v3_750009	merP	Mercuric transport protein periplasmic component
		THICB2_v3_750010	merT	Mercuric transport protein
		THICB2_v3_750011	merR	Mercuric resistance operon regulatory protein
		THICB2_v3_750012		putative Cytochrome c biogenesis protein
		THICB2_v3_750013		putative Thioredoxin/glutaredoxin
		THICB2_v3_750014	merR	Mercuric resistance operon regulatory protein
		THICB2_v3_750015		Uncharacterized 9.0 kDa protein in mobE 3'region
		THICB2_v3_750016		conserved protein of unknown function (fragment)
		THICB2_v3_750017		transposase
		THICB2_v3_760001		transposase (fragment)
		THICB2_v3_760002		protein of unknown function
		THICB2_v3_760003		conserved exported protein of unknown function
		THICB2_v3_760004		putative thioredoxin
		THICB2_v3_760005		putative Protein-disulfide reductase dsbD
		THICB2_v3_760006		putative Thiol:disulfide interchange protein DsbD
		THICB2_v3_760007		conserved exported protein of unknown function
		THICB2_v3_760008		conserved exported protein of unknown function
		THICB2_v3_760009		putative Thioredoxin family protein (fragment)
		THICB2_v3_760010		putative blue (Type 1) copper domain protein, oxido-reductase protein involved in Cu(II)/Cu(I) resistance (CopI)
		THICB2_v3_760011		putative Copper resistance protein B (CopB)
		THICB2_v3_760012	pcoA	copper resistance protein A precursor

		THICB2_v3_760013	cusR	DNA-binding response regulator in two-component regulatory system with CusS
		THICB2_v3_760014	cusS	sensory histidine kinase in two-component regulatory system with CusR, senses copper ions
		THICB2_v3_760015		protein of unknown function
		THICB2_v3_760016		protein of unknown function
		THICB2_v3_760017		conserved exported protein of unknown function
		THICB2_v3_760018		conserved protein of unknown function
		THICB2_v3_760019		conserved protein of unknown function
		THICB2_v3_760020	actP	Copper-transporting P-type ATPase
		THICB2_v3_760021		putative Neuraminidase (Sialidase)
		THICB2_v3_760022		conserved membrane protein of unknown function
		THICB2_v3_760023		putative thioredoxin
		THICB2_v3_760024		conserved exported protein of unknown function
		THICB2_v3_760025		putative Voltage-gated potassium channel
		THICB2_v3_760026		conserved membrane protein of unknown function
		THICB2_v3_760027		putative alpha/beta-Hydrolase
		THICB2_v3_760028		conserved membrane protein of unknown function
		THICB2_v3_760029		Cation transporting P-type ATPase
		THICB2_v3_760030		protein of unknown function
		THICB2_v3_760031		putative blue (type 1) copper domain protein, putative oxido-reductase protein involved in Cu(II)/Cu(I) resistance (copI) (fragment)
		THICB2_v3_760032		transposase (fragment)
		THICB2_v3_760033		transposase (fragment)
		THICB2_v3_760034		protein of unknown function
		THICB2_v3_760035		conserved protein of unknown function
		THICB2_v3_760036		putative 3-hydroxybutyryl-CoA dehydratase (Crotonase)
		THICB2_v3_760037		putative Protochlorophyllide reductase
		THICB2_v3_760038		putative protease
		THICB2_v3_760039		transposase (fragment)
		THICB2_v3_760040		transposase
		THICB2_v3_760041		putative Bacterial regulatory protein, MerR family
		THICB2_v3_760042	cadA	Cadmium-transporting ATPase
		THICB2_v3_760043		protein of unknown function
		THICB2_v3_760044		conserved protein of unknown function
		THICB2_v3_760045		conserved protein of unknown function
		THICB2_v3_760046		transposase
		THICB2_v3_760047		conserved exported protein of unknown function
		THICB2_v3_760048	ompR	Transcriptional regulatory protein OmpR
		THICB2_v3_760049		putative two-component sensor histidine kinase
		THICB2_v3_760050		protein of unknown function
		THICB2_v3_760051		conserved exported protein of unknown function
		THICB2_v3_760052		conserved membrane protein of unknown function
		THICB2_v3_760053		exported protein of unknown function
		THICB2_v3_760054		putative phosphatase / Phospholipase C
		THICB2_v3_760055		protein of unknown function
		THICB2_v3_760056		conserved protein of unknown function
		THICB2_v3_760057		putative permease

		THICB2_v3_760058	cysI	Sulfite reductase (NADPH) hemoprotein beta-component
		THICB2_v3_760059		conserved protein of unknown function
		THICB2_v3_760060	cysH	APS reductase
		THICB2_v3_760061	cysD	sulfate adenylyltransferase, subunit 2
		THICB2_v3_760062	cysN	sulfate adenylyltransferase, subunit 1
		THICB2_v3_760063		putative Uroporphyrinogen-III C-methyltransferase cobA
		THICB2_v3_760064	fdxA	7-Fe ferredoxin
		THICB2_v3_760065		Ferredoxin--NADP reductase
		THICB2_v3_760066	arsR	Arsenic operon regulator
		THICB2_v3_760067	arsC2a	Arsenate reductase
		THICB2_v3_760068	arsD	Arsenical resistance operon trans-acting repressor ArsD
		THICB2_v3_760069	arsA	Arsenical pump-driving ATPase
		THICB2_v3_760070	arsC2b	Arsenate reductase
		THICB2_v3_760071		putative FAD-dependent pyridine nucleotide-disulphide oxidoreductase
		THICB2_v3_760072		conserved protein of unknown function
		THICB2_v3_760073	arsB	arsenite/antimonite transporter
		THICB2_v3_760074		protein of unknown function
		THICB2_v3_760075		Transposase of ISThsp18, IS1182 family (fragment)
		THICB2_v3_760076		putative transcriptional regulator crp family
		THICB2_v3_760077		putative transcriptional regulator
		THICB2_v3_760078	moeA	Molybdopterin molybdenumtransferase
		THICB2_v3_760079	moaA	Cyclic pyranopterin monophosphate synthase
		THICB2_v3_760080		putative bacterial regulatory protein, ArsR
		THICB2_v3_760081		conserved protein of unknown function
		THICB2_v3_760082		putative nitroreductase
		THICB2_v3_760083		putative Cytochrome c, class ID (fragment)
		THICB2_v3_760084		putative Bacterial regulatory protein, ArsR
		THICB2_v3_760085	nirM	Cytochrome c-551
		THICB2_v3_760086	aioA	Arsenite oxidase subunit AioA
		THICB2_v3_760087	aioB	Arsenite oxidase subunit AioB
		THICB2_v3_760088		putative phosphite transport system-binding protein PtxB
		THICB2_v3_760089		putative Signal transduction protein (AoxS)
		THICB2_v3_760090	aoxR	AoxR regulatory protein
		THICB2_v3_760091		putative Prophage antirepressor
		THICB2_v3_760092		protein of unknown function
		THICB2_v3_760093		conserved protein of unknown function
		THICB2_v3_760094		conserved protein of unknown function
		THICB2_v3_760095		protein of unknown function
		THICB2_v3_760096		exported protein of unknown function
		THICB2_v3_760097		transposase
		THICB2_v3_760098		conserved protein of unknown function
		THICB2_v3_760099		conserved protein of unknown function
		THICB2_v3_760100		conserved protein of unknown function
		THICB2_v3_760101		protein of unknown function
		THICB2_v3_760102		protein of unknown function
		THICB2_v3_760103		conserved protein of unknown function

		THICB2_v3_760104		Phage transcriptional regulator, AlpA
		THICB2_v3_770001		conserved exported protein of unknown function
		THICB2_v3_770002		Single-stranded DNA-binding protein
		THICB2_v3_770003		protein of unknown function
		THICB2_v3_770004		putative Metal dependent phosphohydrolase
		THICB2_v3_770005		exported protein of unknown function
		THICB2_v3_770006		conserved protein of unknown function
		THICB2_v3_770007		conserved protein of unknown function
		THICB2_v3_770008		exported protein of unknown function
		THICB2_v3_770009		protein of unknown function
		THICB2_v3_770010		exported protein of unknown function
		THICB2_v3_770011		putative superfamily S-adenosyl-L-methionine-dependent methyltransferase
		THICB2_v3_770012		protein of unknown function
		THICB2_v3_770013		protein of unknown function
		THICB2_v3_770014		conserved protein of unknown function
		THICB2_v3_770015		Prophage CP4-like integrase
		THICB2_v3_770016		Integrase family protein
		THICB2_v3_770017		conserved exported protein of unknown function
		THICB2_v3_770018		conserved protein of unknown function
		THICB2_v3_770019		protein of unknown function
		THICB2_v3_770020		exported protein of unknown function
		THICB2_v3_770021		exported protein of unknown function
		THICB2_v3_770022		exported protein of unknown function
		THICB2_v3_770023		conserved protein of unknown function
		THICB2_v3_770024		protein of unknown function
		THICB2_v3_770025		Mobilization protein TriK (fragment)
		THICB2_v3_780001		exported protein of unknown function
		THICB2_v3_780002		putative mobC
		THICB2_v3_780003		protein of unknown function
		THICB2_v3_780004		exported protein of unknown function
		THICB2_v3_780005		conserved protein of unknown function
		THICB2_v3_780006		conserved protein of unknown function
		THICB2_v3_780007		exported protein of unknown function
		THICB2_v3_780008		exported protein of unknown function
		THICB2_v3_780009		protein of unknown function
		THICB2_v3_780010		protein of unknown function
		THICB2_v3_780012		conserved protein of unknown function
		THICB2_v3_780013		conserved protein of unknown function
		THICB2_v3_780014		protein of unknown function
		THICB2_v3_780015		conserved protein of unknown function
		THICB2_v3_780016		Phage Gp37Gp68 family protein
		THICB2_v3_780017		Conjugation TrbI family protein
		THICB2_v3_780018		Conjugal transfer protein TrbG/VirB9/CagX
		THICB2_v3_780019		conserved protein of unknown function
		THICB2_v3_780020		putative Conjugal transfer protein TrbF
		THICB2_v3_780021		Addiction module toxin, RelE/StbE family
		THICB2_v3_780022		Transcriptional regulator, XRE family

		THICB2_v3_780023		conserved protein of unknown function
		THICB2_v3_780024		P-type conjugative transfer protein TrbL
		THICB2_v3_780025		P-type conjugative transfer protein TrbJ
		THICB2_v3_780026		exported protein of unknown function
		THICB2_v3_780027	trbE	Conjugal transfer protein TrbE
		THICB2_v3_780028		putative Type IV secretory pathway, TrbD component
		THICB2_v3_780029		conserved protein of unknown function
		THICB2_v3_780030		putative conjugal transfer protein TrbC
		THICB2_v3_780031		putative Type IV secretory pathway, protease TraF
		THICB2_v3_780032	traG	Conjugal transfer protein TraG
		THICB2_v3_780033		conserved protein of unknown function
		THICB2_v3_780034		putative Peptidase S9 prolyl oligopeptidase active site domain protein
		THICB2_v3_780035		Putative DNA primase (fragment)
		THICB2_v3_790001		Putative DNA primase (fragment)
		THICB2_v3_790002		putative Co/Zn/Cd efflux system component
		THICB2_v3_790003		protein of unknown function
		THICB2_v3_790004		conserved protein of unknown function
		THICB2_v3_790005		putative membrane protein
		THICB2_v3_790006		conserved protein of unknown function
		THICB2_v3_800001		conserved protein of unknown function
		THICB2_v3_800002		protein of unknown function
		THICB2_v3_800003		putative Chemotaxis signal transduction protein (fragment)
		THICB2_v3_800004		putative Methyl-accepting chemotaxis protein
		THICB2_v3_800005		conserved membrane protein of unknown function
		THICB2_v3_800006		putative Permease of the major facilitator superfamily
		THICB2_v3_800007		conserved protein of unknown function
		THICB2_v3_800008		conserved protein of unknown function
		THICB2_v3_800009		conserved protein of unknown function
		THICB2_v3_810001		Transposase IS21 family (fragment)
		THICB2_v3_810002		protein of unknown function
		THICB2_v3_810003	bioI	Biotin biosynthesis cytochrome P450-like enzyme
		THICB2_v3_810004	bioB	biotin synthase
		THICB2_v3_810005	bioD	dethiobiotin synthetase
		THICB2_v3_810006	bioF	8-amino-7-oxononanoate synthase
		THICB2_v3_810007	bioA	Adenosylmethionine-8-amino-7-oxononanoate aminotransferase
		THICB2_v3_810008		transposase (fragment)
		THICB2_v3_810009		conserved exported protein of unknown function
		THICB2_v3_810010	panD	Aspartate 1-decarboxylase precursor (Aspartate alpha-decarboxylase) [Contains: Aspartate 1-decarboxylase beta chain; Aspartate 1-decarboxylase alpha chain]
		THICB2_v3_810011	trbB	putative conjugal transfer protein TrbB
		THICB2_v3_810012		conserved protein of unknown function
		THICB2_v3_810013		protein of unknown function
		THICB2_v3_810014		conserved protein of unknown function
		THICB2_v3_810015		conserved protein of unknown function
		THICB2_v3_810016	traI	Relaxase/Mobilisation nuclease

		THICB2_v3_810017		conserved protein of unknown function
		THICB2_v3_810018		protein of unknown function
		THICB2_v3_810019		protein of unknown function
		THICB2_v3_810020		conserved protein of unknown function
		THICB2_v3_810021		conserved protein of unknown function
		THICB2_v3_810022		TraB family protein
		THICB2_v3_810023		protein of unknown function
		THICB2_v3_810024		putative Type II restriction enzyme, methylase subunits
		THICB2_v3_810025		Transcriptional regulator, AbrB family
		THICB2_v3_810026		PilT protein domain protein
		THICB2_v3_810027		putative DNA helicase, restriction/modification system component YeeB
		THICB2_v3_810028		putative YeeC-like protein
		THICB2_v3_810029		protein of unknown function
		THICB2_v3_810030		conserved protein of unknown function
		THICB2_v3_810031	topB	DNA topoisomerase 3
		THICB2_v3_810032		Lytic transglycosylase catalytic
		THICB2_v3_810033		conserved protein of unknown function
		THICB2_v3_810034		conserved protein of unknown function
		THICB2_v3_810035		conserved protein of unknown function
		THICB2_v3_810036		Conserved protein of unknown function
		THICB2_v3_810037		conserved protein of unknown function
		THICB2_v3_810038		transposase
		THICB2_v3_820001		protein of unknown function
		THICB2_v3_820002		putative phage repressor
		THICB2_v3_830001		putative Phage transcriptional regulator, AlpA
		THICB2_v3_830002		conserved protein of unknown function
RGP20 or more?	–	THICB2_v3_840001		protein of unknown function
		THICB2_v3_840002		protein of unknown function
		THICB2_v3_840003		conserved protein of unknown function
		THICB2_v3_840004		protein of unknown function
		THICB2_v3_840005		Lysine exporter protein (LYSE/YGGA)
		THICB2_v3_840006		protein of unknown function
		THICB2_v3_840007		putative permease of the major facilitator superfamily
		THICB2_v3_840008		Metallophosphoesterase
		THICB2_v3_840009		conserved protein of unknown function
		THICB2_v3_840010		putative peptidase family M4
		THICB2_v3_840011		N-6 DNA methylase
		THICB2_v3_850001		protein of unknown function
		THICB2_v3_850002		conserved protein of unknown function
		THICB2_v3_850003		Type I restriction-modification system, restriction subunit R
		THICB2_v3_850004		conserved protein of unknown function
		THICB2_v3_850005		conserved protein of unknown function
		THICB2_v3_850006		conserved protein of unknown function
		THICB2_v3_850007		protein of unknown function

Annexe 2

Spatio-temporal detection of the minor Thiomonas population and its crucial arsenite oxidase involved in arsenite natural attenuation in the Carnoulès Acid Mine Drainage

Agnès Hovasse[1], Odile Bruneel[2], Corinne Casiot[2], Angélique Desoeuvre[2], Julien Farasin[3], Marina Hery[2], Alain Van Dorsselaer[1], Christine Carapito[1], Florence Arsene-Ploetze[3*]

[1]Laboratoire de Spectrométrie de Masse BioOrganique (LSMBO), UMR7178, IPHC, Université de Strasbourg, CNRS,, France, [2]Laboratoire HydroSciences Montpellier, UMR HSM 5569, CNRS, Université Montpellier, IRD, France, [3]Laboratoire Génétique Moléculaire, Génomique et Microbiologie, UMR7156, , Université de Strasbourg, CNRS, France

Submitted to Journal:
Frontiers in Cell and Developmental Biology

Specialty Section:
Cellular Biochemistry

Article type:
Original Research Article

Manuscript ID:
169472

Received on:
18 Sep 2015

Frontiers website link:
www.frontiersin.org

Conflict of interest statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest

Author contribution Statement

O. B., M. H., A. V. D., Ch.C., and. F.A.-P. designed the research project; A. H., O. B., Co. C., A. D., J. F., M. H., Ch.C., and F.A.-P. performed the research; A. H., O. B., Co. C., A. D., M. H., Ch.C., and F.A.-P.. wrote the paper.

Keywords

Community proteomics, targeted proteomics, LC-SRM, acid mine drainage, pyrosequencing, fish

Abstract

Word count: 174

The acid mine drainage (AMD) impacted creek of the Carnoulès mine (south of France) is characterized by acid waters containing high concentrations of heavy metals. The microbial community found in this AMD has been extensively studied using isolation, metagenomic and metaproteomic technologies, revealing that a natural attenuation of arsenic and iron occurred at this site and involved the arsenite oxidase activity of different Thiomonas strains. In this study, a sensitive and quantitative Selected Reaction Monitoring (SRM)-based proteomic approach was developed to detect and quantify the two subunits of the arsenite oxidase and RpoA of two different Thiomonas groups. Using this approach combined with 16S rRNA sequence analysis by pyrosequencing and FISH, we demonstrated for the first time that these Thiomonas strains are ubiquitously present in minor proportions in this AMD and that they express the key enzymes involved in natural remediation, at different locations and time points in this AMD. Besides this demonstration, this study also highlighted the power of targeted proteomics applied at the community level, to detect low abundant proteins in situ.

Spatio-temporal detection of the minor *Thiomonas* population and its crucial arsenite oxidase involved in arsenite natural attenuation in the Carnoulès Acid Mine Drainage

Agnès Hovasse[1], Odile Bruneel[2], Corinne Casiot[2], Angélique Desoeuvre[2], Julien Farasin[3], Marina Hery[2], Alain Van Dorsselaer [1], Christine Carapito[1], Florence Arsène-Ploetze[3*]

[1] Laboratoire de Spectrométrie de Masse BioOrganique (LSMBO), IPHC, Université de Strasbourg, CNRS, UMR7178, Strasbourg, France
[2] Laboratoire HydroSciences Montpellier, UMR HSM 5569, (CNRS, Université Montpellier, IRD, Université Montpellier, CC0057 (MSE), Montpellier, France
[3] Laboratoire Génétique Moléculaire, Génomique et Microbiologie, UMR7156, CNRS-Université de Strasbourg, Département Microorganismes, Génomes, Environnement, Equipe Ecophysiologie Moléculaire des Microorganismes, Strasbourg, France

*correspondence :
Dr. Florence Arsène-Ploetze
Laboratoire Génétique Moléculaire, Génomique et Microbiologie, UMR7156
Université de Strasbourg, Institut de Botanique,
28 rue Goethe, 67083 Strasbourg, France
ploetze@unistra.fr

Running Title : The *Thiomonas* arsenite oxidase is expressed in a sustainable way in the Carnoulès AMD

Abstract

The acid mine drainage (AMD) impacted creek of the Carnoulès mine (south of France) is characterized by acid waters containing high concentrations of heavy metals. The microbial community found in this AMD has been extensively studied using isolation, metagenomic and metaproteomic technologies, revealing that a natural attenuation of arsenic and iron occurred at this site and involved the arsenite oxidase activity of different *Thiomonas* strains. In this study, a sensitive and quantitative Selected Reaction Monitoring (SRM)-based proteomic approach was developed to detect and quantify the two subunits of the arsenite oxidase and RpoA of two different *Thiomonas* groups. Using this approach combined with 16S rRNA sequence analysis by pyrosequencing and FISH, we demonstrated for the first time that these *Thiomonas* strains are ubiquitously present in minor proportions in this AMD and that they express the key enzymes involved in natural remediation, at different locations and time points in this AMD. Besides this demonstration, this study also highlighted the power of targeted proteomics applied at the community level, to detect low abundant proteins *in situ*.

Keywords : Community proteomics, Targeted proteomics, LC-SRM, Acid Mine Drainage, Pyrosequencing, FISH

Introduction

Arsenic contamination has been identified as a major risk to human health in various parts of the world (Mandal and Suzuki, 2002; Sharma and Sohn, 2009) with more than 50 million people exposed to high concentrations worldwide (Kruger et al., 2013; Oremland and Stolz, 2003; Slyemi and Bonnefoy, 2012). Acid mine drainage (AMD) environments are known to be highly toxic to most living organisms, due to the presence of many toxic elements such as arsenic, and because of the acidic pH, usually below 3. The sulfurous wastes of the former mine of Carnoulès, Gard (France), contain As-rich pyrite permanently leached by subsurface waters emerging at the base of a dam, forming the spring of the small Reigous Creek. This creek contains between 50 and 350 mg/L of soluble arsenic, mainly in the form of As(III) (Casiot et al., 2003; Egal et al., 2010; Morin et al., 2003). Arsenic concentration decreases by 95% between the source of the Reigous and its confluence with the downstream river Amous.

Despite these harsh conditions, a microbial community composed of bacteria, archaea and eukaryotes is present in this AMD and has been described for several years (Bruneel et al., 2006, 2008, 2011; Volant et al., 2012, 2014). Microorganisms have been isolated from this site and analyzed in laboratory conditions to determine their metabolic function in toxic ecosystems (Bruneel et al., 2003; Bryan et al., 2009; Casiot et al., 2003, 2004; Duquesne et al., 2003, 2008; Halter et al., 2012; Morin et al., 2003). Moreover, using environmental genomic approaches, seven bacterial genomes were reconstructed (Bertin et al., 2011). These approaches completed by geochemical studies (Casiot et al., 2003; Egal et al., 2010; Morin et al., 2003), showed that the observed natural attenuation process in the Reigous creek was mainly due to microbial metabolism. In particular, it was proposed that *Acidithiobacillus* and *Gallionella* genera are involved in the oxidation of iron into Fe(III), whereas *Thiomonas* strains oxidize arsenite (Battaglia-Brunet et al., 2011b; Bertin et al., 2011; Bruneel et al., 2006; Casiot et al., 2003, 2005; Duquesne et al., 2003; Morin et al., 2003).

Bacterial As(III) oxidation is of particular interest for natural remediation processes since As(III) is more soluble and more toxic than As(V) (Kruger et al., 2013; Oremland and Stolz, 2003). The arsenite oxidase involved in this process (AioA/B) belongs to the dimethyl sulfoxide (DMSO) reductase of the molybdenum family (Oremland and Stolz, 2003; Silver and Phung, 2005). These enzymes are heterodimers comprising a large subunit with a molybdenum center and a [3Fe-4S] cluster associated with a small subunit with a Rieske-type [2Fe-2S] cluster. *aioA* and *aioB* genes have been found in several phylogenetically diverse bacteria distributed among 25 genera, isolated from various arsenic-rich environments (Heinrich-Salmeron et al., 2011; Quéméneur et al., 2008, 2010; Slyemi and Bonnefoy, 2012).

Eight strains, *Thiomonas* sp. X19, CB1, CB2, CB3, CB6, ACO3, ACO7 and *Thiomonas arsenitoxydans* (strain 3As) were isolated from the acidic waters draining tailings of the Carnoulès former mine (Bryan et al., 2009; Delavat et al., 2012; Duquesne et al., 2008; Freel et al., 2015). *Thiomonas arsenivorans* was isolated from another arsenic-rich mine residue at the Cheni former gold mine, Limousin, France (Battaglia-Brunet et al., 2006). The genomes of seven of these *Thiomonas* strains have been recently characterized (Freel et al., 2015) and metagenomic data allowed characterizing the reconstructed genome of an additional bacteria belonging to the *Thiomonas* genus called Carn2 (Bertin et al., 2011). Phylogenetically, this genus is a member of the *Betaproteobacteria* subclass. It was proposed to group *Thiomonas* sp. CB1, CB2, CB3, CB6, ACO3, ACO7 with *Tm. arsenitoxydans*, in the same clade, Group I, while *Thiomonas* sp. X19 and Carn2 belongs to another clade, the Group II, together with *Tm. arsenivorans* isolated from the Cheni former gold mine (Bertin et al., 2011; Bryan et al., 2009; Freel et al., 2015). These *Thiomonas* strains present interesting physiological and phylogenetic differences, for example, adaptation to arsenic differs between *Thiomonas* strains even if some of them were isolated from the same environment (Arsène-

Ploetze et al., 2010; Battaglia-Brunet et al., 2006, 2011a; Bryan et al., 2009; Freel et al., 2015). Interestingly, in several genomes of the *Thiomonas* strains isolated from the Reigous creek, at least two *aio* operons were carried on two genomic islands. Some of these *Thiomonas* were able to express at least two copies under laboratory conditions, and were highly resistant to arsenic (Freel et al., 2015). The expression of these genes is observed only in the presence of arsenic (Duquesne et al., 2008; Freel et al., 2015; Slyemi et al., 2013). Environmental genomic approaches revealed that at least one *Thiomonas* belonging to the Group II (Carn2) expressed the arsenite oxidase in Mai 2007 at one sample location in the Reigous creek (COWG). However, it was not known if these different *Thiomonas* strains were stably present at different locations in the Reigous creek and if they expressed in a sustainable way their arsenite oxidase enzyme. In this study we developed environmental proteomic and genomic approaches to analyze which type of *Thiomonas* strains are present and express the arsenite oxidase, at different locations and time points.

Material and methods

1. Bacterial strains, growing conditions and cell lysis.

Thiomonas sp. CB2, CB1, CB3 and CB6 were isolated from the Reigous creek in the Carnoulès AMD (France) (Arsène-Ploetze et al., 2010; Bryan et al., 2009; Duquesne et al., 2008). The *Thiomonas* strains were routinely grown on gelled or liquid m126 medium (modified 126 medium). Medium m126 contains: (g L^{-1}) yeast extract (YE; 0.5); $Na_2S_2O_3$ (5.0); KH_2PO_4 (1.5); Na_2HPO_4 (4.5); $MgSO_4 \cdot 7H_2O$ (0.1); $(NH_4)Cl$ (0.3), adjusted to pH 5.0 with H_2SO_4 prior to sterilization. Arsenite (As(III)) was added to media to the desired concentration from sterile stocks of 667.4 mM of the metalloid ion in deionized H_2O (ddH_20), from $NaAsO_2$ (Prolabo). Cells (100 mL at OD_{600nm} of 0.2) were harvested by centrifugation (7,000 *g*, 10 min, 4°C). Pellets were washed once with NaCl (9g. L^{-1}) and suspended in 600 µL of ultra-pure water supplemented with 2 µL Benzonase Nuclease (Sigma) and 4 µL of Protease Inhibitor Mix (GE Healthcare). Cell suspensions were disrupted by adding glass beads (1 g) and vortexing 3 times 1 min with 30 s intervals. Cellular debris were removed by two centrifugations, the first at 10,000 g for 5 min and the second at 16,000 g for 90 min. Protein concentrations were measured using the Qubit® Fluorometric Quantitation, according to the manufacturer's instructions (Life technologies).

2. Sampling procedure

Samples were collected in June 2011 and January 2012 along the Reigous creek, at the spring (S1), 30 m downstream from the spring (COWG), and 1500 m downstream from the spring (CONF), which corresponded to a few meters upstream from the confluence between the Reigous creek and the Amous river (Bruneel et al., 2003). Sediments were collected using a sterile spatula in triplicates at the bottom of the creek. The triplicates were pooled, homogenized and kept on ice in the field. Sediments were conserved at 4°C for 7 days or frozen in the presence of one volume of 50% glycerol solution until further analysis.

The main physicochemical parameters of the running water covering the sediments (pH, temperature, and dissolved oxygen concentrations) were measured *in situ* at each sampling point (Table 1). Arsenic speciation, Fe(II) and sulfate analyses were performed as previously described (Bruneel et al., 2008; Egal et al., 2010). Briefly, water samples (500 mL) were immediately filtered through 0.22 µm Millipore membranes fitted on Sartorius polycarbonate filter holders. For total Fe and As determination, the filtered water was acidified to pH 1 with HNO_3 (14.5 M) and stored at 4°C in polyethylene bottles until analysis. A 20 µL aliquot of filtered water sample was added to either a mixture of acetic acid and EDTA for As speciation (Samanta and Clifford, 2005) or to a mixture of 0.5 mL acetate buffer (pH 4.5) and 1 mL of 1,10-phenanthrolinium chloride solution for Fe(II) determination (Rodier et al., 1996). The vials were completed to 10 mL with deionized water. The samples for Fe and As speciation and sulfate determination were stored in the dark and analyzed within 24 h. Chemical analysis were carried out as previously described (Bruneel et al., 2011).

3. DNA extraction and pyrosequencing of bacterial 16S rRNA gene fragments

Triplicate genomic DNA was extracted from sediments using the UltraClean Soil DNA Isolation Kit according to the manufacturer's recommendations (MoBio Laboratories Inc., Carlsbad, CA, USA). These triplicates were pooled before PCR amplification. All extracted genomic DNA samples were stored at -20°C until further analysis. The 16S rRNA genes were amplified by PCR for multiplexed pyrosequencing using barcoded primers. The primers pairs used, targeting the V3 to V5 variable regions of the 16S rRNA gene, were 343F (5'-AxxxTACGGRAGGCAGCAG-3') and 806R (5'-Bxxx GGACTACCAGGGTATCTAAT-3') where A and B represent the two FLX Titanium

adapters (A adapter sequence: 5'- CCATCTCATCCCTGCGTGTCTCCGAC -3'; B adapter sequence: 5'- CCTATCCCCTGTGTGCCTTGGCAGTC -3') and xxx represent the sample specific barcode sequence. Two PCR amplifications with two distinct barcodes were performed for each sample (n1 and n2). The reaction mixture contained 5 to 10 µL of DNA, 5 µL of 10X PCR buffer, 240 µM of dNTP, 1.5 mM of $MgCl_2$, 1 µL of each primers (10 µM) and 1U of taq polymerase. Sterile distilled water was added to reach a final volume of 50 µL. After a step of denaturation at 95°C for 10 min, 30 cycles were done as follow: 95°C for 30 s (denaturation), 50°C for 30 s (annealing) and 72°C for 1 min (extension), followed by 10 min at 72°C. Several PCR were pooled at the purification step given the poor performance of PCRs. Pyrosequencing was performed by a Roche 454 Life Sciences Genome Sequencer FLX Titanium sequencer at Genoscreen (Lille, France).

4. Processing of pyrosequencing data, diversity estimation, and taxonomic classification

The raw data generated by 454 pyrosequencing were processed and analyzed using the 1.32.1 version of mothur software (updated 10/16/13, http://www.mothur.org) (Schloss et al., 2009). These sequences were processed using the commands shhh.flows using the algorithm PyroNoise (Quince et al., 2009). Preprocessing of unaligned sequences included also removing of short sequences (<200 bp), all sequences containing ambiguous characters as well as sequences with more than 8 homopolymers. We further removed sequences that did not align over the same region. Identical sequences were then grouped and representative sequences were aligned against the SILVA bacterial and archaeal reference database using the Needleman-Wunsch algorithm (Needleman and Wunsch, 1970). Chimeric sequences were detected and removed using the implementation of Chimera Uchime (Edgar et al., 2011). A further screening step (pre-cluster) was applied to reduce sequencing noise by clustering reads differing by only 1 base every 100 bases (Huse et al., 2010). The remaining high-quality reads were used to generate a distance matrix and clustering into Operational Taxonomic Units (OTUs) defined at 97% cutoff using the "Average Neighbor" algorithm. The OTUs were then phylogenetically classified to the genus level using the naïve Bayesian classifier (80% confidence threshold) trained on the RDP's taxonomic outline and implemented in Mothur. Nonparametric Chao1 and Shannon alpha diversity estimates as well as the coverage and rarefaction curves were calculated for normalized data for each sample with Mothur software.

5. Fluorescence in situ hybridization (FISH) analysis:

Sediments were fixed in freshly prepared 4% paraformaldehyde for 2 h 30 at 4°C, washed with phosphate buffered saline (PBS), and stored at -20°C in 50% PBS/ethanol solution within 8 h of collection. Fixed samples were sonicated 3 min, incubated 20 min in bovine serum albumin (BSA) (0.4 mg/mL, Amresco, US), spread on 10-well Teflon coated glass slides (Menzel gläser, Germany) and dried 10 min at 26°C. The samples were dehydrated in 50, 80, and 90% ethanol solutions for 3 min each. Bacteria were enumerated with a FISH protocol (Brown et al., 2011; Hugenholtz et al., 2002) that was slightly modified for this study. Sediments were hybridized using the following specific fluorescent probes: EUB338 (Amann et al., 1990) for most Bacteria, NON 338 (Wallner et al., 1993) for negative control, TM1G0138, the specific probe targeting *Thiomonas* group 1 defined by (Hallberg et al., 2006), corresponding to the Group II defined by (Bryan et al., 2009), and TM2G0138 for *Thiomonas* group 2 (Hallberg et al., 2006), corresponding to the Group I defined by (Bryan et al., 2009). Hybridization conditions were previously optimized by testing each probe against a collection of different bacterial species maintained in the laboratory (Hallberg et al., 2006). *In situ* hybridizations were performed at 46°C for 2 h in 0.9 M NaCl, 20 mM Tris-HCl at pH 8, 5 mM EDTA, 0.01% sodium dodecyl sulfate (SDS), 3-5 ng/µL of fluorescently labeled probes,

and 30% formamide. Washing was performed at 48°C for 20 min in 20 mM Tris-HCl at pH 8, 0.01% SDS, 5 mM EDTA and 0.101 M NaCl. Samples were rinsed in distilled water, air dried, counterstained with 4′,6′-diamidino-2-phenylindole (DAPI, 1 µg/mL, 2 min) and mounted with Citifluor AF1 (Biovalley, France). Oligonucleotides labeled with the cyanine dye (Cy5) or Fluorescein isothiocyanate (FITC) at the 5' were purchased from Eurogentec (Belgium). Microscopic counting of at least 1,000 cells per experiments was done using a confocal laser scanning microscope (LSM 510; Zeiss, Germany) and analyzed with Image J program. All experiments were performed in three independent replicates. We compared several methods to separate bacteria from the substratum, improve the hybridization and decrease the background (e.g., vortexing, sonication, manual scratching, using of low melting agarose, lysozyme, BSA). Bacterial separation was optimized by sonication and the background seemed to be lower after BSA incubation. Optimizations were done with EUB338 probe. For the determination of the nonspecific binding of FISH-label probes, the samples were also hybridized with an antisense EUB338, NON338. Very few or no probe-positive cells were detected with the negative control.

6. Protein extraction from the microbial community of Reigous creek sediments.

To recover microbial cells, sediments were subjected to a Nycodenz® gradient. Ten times 15 mL of sediments were washed in 15 mL of solution 1 (in g/L: $Na_2SO_4.10H_2O$, 0.15; $(NH_4)_2SO_4$, 0.45; KCl, 0.05; $MgSO_4.7H_20$, 0.5; KH_2PO_4, 0.05; $Ca(NO_3)_2.4H_2O$, 0.014) and agitated one night at 4°C. After 10 minutes decantation, 15 mL of supernatant were added without mixing on 35 mL of a 65% Nycodenz® solution (Axis-Shield, Oslo Norway), and then centrifuged 1 hour at 10,000 *g*. The cellular fraction (upper fraction) was removed and washed by adding 2 volumes of solution 1 and centrifuged 15 min at 10,000 *g* and 4°C. Pellets were suspended in 600 µL of milliQ water supplemented with 1 µL Benzonase Nuclease (Sigma) and 4 µL of Protease Inhibitor Mix (GE Healthcare). Cell suspensions were disrupted by adding glass beads (1 g) and vortexing 3 times 1 min, with 30 s intervals. Cellular debris were removed by two centrifugations, the first at 10,000 *g* for 5 min and the second at 16,000 *g* for 90 min. Protein concentrations were measured using the Qubit® Fluorometric Quantitation, according to the manufacturer's instructions (Life technologies).

7. Liquid Chromatography (LC) Selected Reaction Monitoring (SRM) assay development

7. 1. Targeted proteins and peptides selection : The targeted assay was developed to target arsenite oxidase (AioA/B) and RpoA serving as an ubiquitously expressed control protein. We compared the AioA/B and RpoA protein sequences derived from the sequenced genomes of 7 selected *Thiomonas* strains and the reconstructed Carn2 genome (Arsene-Ploetze et al., 2012; Bertin et al., 2011; Freel et al., 2015), by aligning these sequences with Clustal W2 (http://www.ebi.ac.uk/Tools/msa/clustalw2/). Using the "Peptidemass" tool from Expasy (http://web.expasy.org/peptide_mass/), peptide sequences resulting from an *in silico* trypsin digestion and with masses comprised between 500 to 1600 Da were selected. Peptides containing cysteine or methionine residues were avoided. To test if the selected peptides are proteotypic of the target protein, the peptide sequences were searched in all genomes available in October 2012 on the MicroScope (Microbial Genome Annotation & Analysis) Platform (https://www.genoscope.cns.fr/agc/microscope/home/index.php) using the pattern match search tool, taking into account the fact that Q, E and K, or N and D or I and L have non distinctive masses. Only peptides that were specific to the proteins from *Thiomonas* strains were manually retained. Fifty peptides were selected and synthetized as crude isotopically-labeled standard peptides (PEPotec Peptides, Thermo Fisher Scientific, Ulm,

Germany) and were spiked into protein extracts obtained from four isolated *Thiomonas* strains grown in the laboratory (*Thiomonas* sp. CB2, CB1, CB3 and CB6) in the absence and in the presence of arsenite. Sixteen peptides gave good LC-MS signals and have been validated for application of the assay on sediment samples (Supplementary Tables 1 and 2).

7. 2. Transitions selection and optimisation: Open-source software Skyline (MacLean et al., 2010) was used for building the spectral library and generating the transitions list. An MS/MS spectral library was acquired on the mixture of the sixteen heavy labelled synthetic peptides injected for nanoLC-MS/MS analysis on an Agilent 1100 series HPLC-Chip/MS system (Agilent Technologies, Palo Alto, USA) coupled to an HCT Plus ion trap (Bruker Daltonics, Bremen, Germany). The column used was a chip C18 SB-ZORBAX, 300 Å (75 µm x 150 mm, 5 µm).The solvent system consisted of 2% acetonitrile, 0.1% formic acid in water (solvent A) and 2% water, 0.1% formic acid in acetonitrile (solvent B). Elution was performed at a flow rate of 300 nL/min with a 8-40% gradient (solvent B) over 30 first minutes followed by a 70% stage (solvent B) over 3 min before the reconditioning of the column at 92% of solvent A. The voltage applied to the capillary cap was optimized to -1,750 V. The system was operated with automatic switching between MS and MS/MS modes. The three most abundant peptides, preferring doubly charged ions, were selected on each MS spectrum for further isolation and fragmentation. The MS/MS scanning was performed in the ultrascan resolution mode at a scan rate of 26.000m/z per second. A total of 6 scans were averaged to obtain an MS/MS spectrum. The complete system was fully controlled by ChemStation (Agilent Technologies) and EsquireControl (Bruker Daltonics) software.

Mass data collected during nano-LC-MS/MS analysis were processed, converted into .mgf files, and interpreted using a local Mascot server (version 2.4.1, Matrix Science, London, U.K.). Searches were performed against a database containing all peptides of interest assembled as a chimeric fake protein. Searches were performed with a tolerance on mass measurements of 0.25 Da in both MS and MS/MS modes. In Skyline, precursor and product ion masses were selected in the spectral library with the following settings : monoisotopic masses, precursor charges +2 and +3, product ion charges +1, +2, from m/z > precursor and up to 6 y- or b- ions, N-terminal fragments to proline were allowed. In order to keep only the most intense transitions and minimize interferences, this large list of transitions was refined using a pool of bacterial strains proteins spiked with the heavy peptides mixture analyzed by microLC-SRM in exactly the same conditions as the samples. These analyzes also allowed to determine peptides retention time. At the end, for each peptide, at least three transitions were monitored in order to identify the peptide and the quantification was done only on non-interfered transitions. The collision energies were also experimentally and individually optimized using the heavy standard peptides mixture injected in microLC-SRM by testing nine values centred on the calculated value suggested by the supplier. The complete list of measured transitions with their retention times and optimized collision energies is provided in Supplementary Table 1.

7. 3. General experimental workflow: After reduction and alkylation, protein samples were digested with trypsin. The tryptic peptides obtained were desalted by Solid Phase Extraction (SPE) before microLC-SRM analysis. Twenty-four isolated bacterial strains samples (eight different samples corresponding to four strains (*Thiomonas* sp. CB2, CB1, CB3 and CB6) cultivated with and without arsenite prepared in biological triplicates) were randomly analyzed in injection triplicates (5 µg). Five randomly chosen bacterial strain samples were pooled, prepared as described below, and injected after each triplicate (10 µg) during the whole course of the experiment as a quality control (QC) sample. Regarding

sediments samples, replicates were performed whenever possible but too little protein material was available for a few samples.

7. 4. Sample preparation and protein digestion: Bacterial strains samples were randomly distributed in 5 batches extemporaneously digested before LC-SRM analysis. 50 µg of each sample were denatured with 100 µL of 8 M urea and 0.1 M NH_4HCO_3 and spiked with a mixture of crude heavy labeled peptides in order to obtain comparable signal intensities between light and heavy transitions (relative volumes of each heavy peptide were experimentally determined in order to constitute a concentration-balanced standard peptides mixture). Reduction was made with 1.75 µL of 700 mM dithiothreitol (DTT) for 30 min at 37°C, followed by alkylation with 6 µL of 700 mM iodoacetamide (IAA) for 1 h at room temperature, in the dark. In order to quench residual IAA, new 1.75 µL of DTT were added. Then samples were diluted to 1 M urea with 750 µL of 0.1 M fresh NH_4HCO_3. Digestion was performed overnight at 37°C with 0.5 µg of trypsin, and stopped with 5 µL HCOOH. All quantities were multiplied by two in order to constitute the QC sample. Sediments samples contained only a maximum of 15 µg of proteins in 300 µL. Their volume was first reduced to 100 µL under speed vacuum. Proteins were then solubilized in 200 µL of 8 M urea and 0.1 M NH_4HCO_3 and spiked with the heavy labeled standard peptides mixture. Reduction was performed with 3.5 µL of 700 mM dithiothreitol (DTT) for 30 min at 37°C, followed by alkylation with 12 µL of 700 mM iodoacetamide (IAA) for 1 h at room temperature, in the dark. In order to quench residual IAA, new 3.5 µL of DTT were added. Then samples were diluted to 1 M urea with 1500 µL of fresh 0.1 M NH_4HCO_3. Digestion was made overnight at 37°C with 0.2 µg of trypsin, and stopped with 10 µL HCOOH. All digested samples were desalted on Sep-Pak C18 50 mg cartridges (Waters, Milford, MA, USA). Cartridges were washed with 2 mL of methanol, 2 mL of acetonitrile (ACN) and equilibrated with 3 mL of 0.1% HCOOH. Samples were loaded, washed with 3 mL of 0.1% HCOOH and eluted with 600 µL of 60% ACN, 0.1% HCOOH (v/v). Bacterial strains samples' volume was reduced to 50 µL with speed vacuum. Sediments samples' volume was reduced to less than 7 µL in order to inject the entire sample in one time.

7. 5. MicroLC-SRM parameters: LC-SRM analyses were performed on a Dionex Ultimate 3000 MicroLC coupled to a TSQ Vantage triple-quadrupole (Thermo Scientific). The system was fully controlled by Xcalibur 2.1 (Thermo Scientific). Peptides were first trapped on a µ-Precolumn (Zorbax C18 stable bond, 5 µm, 1.0 × 17 mm, Agilent Technologies, Palo Alto, CA, USA), and then separated on a Thermo Scientific Acclaim capillary column (PepMap100, C18, 300 µm, 15 cm, 3 µm, 100 Å). The solvent system consisted of 2% acetonitrile and 0.1 % HCOOH in water (v/v/v, solvent A) and 2% water and 0.1% HCOOH in ACN (v/v/v, solvent B). Trapping was performed for 5 min at 30 µL/min with solvent A. Elution was performed at a flow rate of 5 µL/min, using a 1-25 % gradient (solvent B) over 80 min at 35°C. Gradient was followed by a 5 min step of 80% B and returned to initial conditions. The mass spectrometer was equipped with an HESI source operating in positive mode. General instrument operating parameters were used. The isolation width for both Q1 and Q3 was set to 0.7 m/z unit. Scheduled SRM was used for data acquisition; each transition was monitored during a 10 or 15 min time window centered at previously determined peptide retention times for bacterial strains and sediments samples, respectively. Cycle time was set to 2.5 s and minimal dwell times were set to 25 ms and 17 ms, respectively.

7. 6. Data interpretation: Skyline was used to visualize the LC-SRM data, perform peak picking and integration of transition peak areas. Area intensity ratios of the heavy and

the light forms of each peptide were checked manually. Quality control of the experiment was performed by examining the stability of the SRM-MS signal over time on QC samples. The QC sample was analyzed thirteen times over the whole course of the experiment. Peptide relative quantification and testing for differential peptide expression were performed using the R-package MSstats (Chang et al., 2012). The acceptance criteria for statistically different protein abundance changes were set at a p-value lower than 0.05 and a fold change higher than 2.

Results and discussion

1. Spatio-temporal quantification of *Thiomonas* strains using pyrosequencing and Fluorescence in situ hybridization (FISH)

Quantification of *Thiomonas* strains was performed at three different locations along the Reigous creek (Source, COWG and CONF) and at two different time points (summer point in June 2011 and winter point in January 2012). First, the presence and abundance of *Thiomonas* bacteria was evaluated by 16S rRNA sequencing on DNA extracted from sediment samples. A total of 277 871 sequence reads with a median length of around 440 bp were generated in a single run of 454 pyrosequencing from the 12 independent 16S rRNA gene libraries (6 samples in duplicate). After trimming and processing with Mothur, 143 129 reads remained (251 bp average median length) and their clustering led to the identification of 1 497 OTUs (including 522 singletons) defined at 97% of identity (Table 2). Rarefaction curves tend to reach an asymptote for most of the samples (Figure 1), suggesting that the majority of the bacterial phylotypes present in each sample were identified. This was confirmed by the high Good's coverage of all duplicates (ranging from 92 to 99%, Table 2), which provides an estimate of sampling completeness using a probability calculation with randomly selected amplicon sequences. The non-parametric estimator Chao1 varied along the Reigous, the lowest OTUs richness were found in S1 (69 to 178) and the highest were recorded in CONF (439 to 620) (Table 2). The same trend was observed with the Shannon index, which showed the lowest value of diversity in S1 (from 1.06 to 2.94) and the highest values in CONF (4.72 to 5.40) (Table 2).

At a confidence threshold of 80%, 349 and 234 sequences, respectively for the first and second duplicate (n1 and n2), belonged to OTUs which could be assigned to the *Thiomonas* genus, representing 0.42% of the total of sequences for n1 and n2 (data not shown). However, to gain more insight inside this genus, taxonomic classification was performed directly on the sequences (Table 3). By this way, 1008 and 873 sequences (respectively for n1 and n2) could be assigned to the *Thiomonas* genus, representing 0.7 and 0.6% of the total of sequences. These sequences could be assigned to different strains belonging to the phylogenetic Group I (*Thiomonas* Ynys3 and *Thiomonas intermedia*) and Group II (*Thiomonas* PK44, *Thiomonas arsenivorans, Thiomonas cuprina* and *Thiomonas* sp. DM-Cd4) according to (Bryan et al., 2009) or to unclassified *Thiomonas* species (Table 3). Altogether, the high-throughput sequencing used in this study enabled the detection of sequences related to the *Thiomonas* genus and confirmed the persistence of this genus in the Reigous creek since its first isolation in 2001 (Bruneel et al., 2003). Sequences affiliated with the *Thiomonas* genus accounted for a small proportion of the total bacterial community (less than 1% of the total sequences). However, due to technical bias inherent to this technique and the size of the generated fragments, pyrosequencing could only be considered, as a semi quantitative method.

To gain more insight into the actual abundance of *Thiomonas* of each phylogenetic group, a microscopic analysis of samples labeled with FISH probes was performed, using two probes targeting *Thiomonas* Group I or *Thiomonas* Group II (Bryan et al., 2009; Hallberg et al., 2006) (see material and methods). In parallel, the number of total bacteria was also estimated. This approach revealed that 79.6% to 90.1% of DAPI-stained cells were stained with the bacteria-specific probes EUB338, the DAPI-stained cells included some fungal and minor group of bacteria that did not hybridize against EUB338 (Figure 2). *Thiomonas* specific probes indicated that the phylogenetic Groups I and II are present in June 2011 and January 2012 in all the sampled sites along the Reigous creek (S1, COWG and CONF) (Figures 2 and 3). The abundances varied from 2.61% ± 0.43% to 5.55% ± 1.1% for the Group II and 1.19% ± 0.37% to 5.15% ± 0.61% for the Group I and 4.99% to 8.73% for the Group I and II

together (Figure 3). This abundance is slightly higher than the one obtained based on pyrosequencing but FISH is considered to be a more suitable quantitative method than molecular methods based on PCR amplification. Both approaches demonstrated the presence of the *Thiomonas* Group I and Group II in summer 2011 and winter 2012 along the Reigous but as a minority of the total bacterial community (less than 9%). Such results corroborate the observations made in previous studies performed both in Carnoulès water and sediments (Table 4), revealing that *Thiomonas* were present in a sustainable way at this site. Additionnally, our study demonstrates that several phylogenetic groups of *Thiomonas* are present as a minor population and stable since more than a decade, at different sampling points along the Reigous creek. Interestingly, despite their low abundance, this genus was proposed to play a crucial role in the efficient remediation process observed along this creek by favoring the oxidation of As(III) into the less mobile and less toxic form As(V), and its co-precipitation with Fe(III) (Bertin et al., 2011; Casiot et al., 2003; Morin et al., 2003). Therefore, to test if these *Thiomonas* spp. are actually active and express key enzymes, a targeted proteomic approach was developed to detect the presence of *Thiomonas* arsenite oxidases.

2. Spatio-temporal detection of the *Thiomonas* arsenite oxidase using a targeted proteomics approach based on liquid chromatography selected reaction monitoring mass spectrometry (LC-SRM)

The activity of *Thiomonas* spp. in Carnoulès ecosystem was highlighted in two previous global metaproteomic studies (Bertin et al., 2011; Bruneel et al., 2011). In the first study (Bruneel et al., 2011), the activity of *Thiomonas* was suggested at the COWG sampling point, as their GroEL and 50S ribosomal L1 proteins were identified in a 2D-gel based metaproteomic approach (Table 4). In a second SDS-PAGE-based global metaproteomic study (Bertin et al., 2011), 89 proteins among which 3 arsenite oxidases (AioA/B) of the *Thiomonas*-like Carn2 cells were identified at the COWG sampling point revealing for the first time that such bacteria indeed expressed these crucial enzymes for the remediation process *in situ* (Table 4). In this study, to test if the minor population detected at several sampling and different time points always expressed the AioA/B proteins, we developed a targeted method using an LC-SRM approach. This approach offers the best sensitivity and specificity among proteomics approaches and we therefore attempted to identify the *Thiomonas* arsenite oxidase proteins within the complex protein mixture extracted from the Reigous creek sediment's community, without any sample fractionation (Arsene-Ploetze et al., 2012; Arsène-Ploetze et al., 2014).

We selected three proteins of interest to target: AioA and AioB, as they are the two subunits of the arsenite oxydase and RpoA, selected as a control ubiquitously expressed protein. RpoA also presents the interest of including peptides allowing the discrimination between *Thiomonas* Groups I and II. The success of a SRM-based targeted assay resides in the appropriate choice of peptides that will be used as specific tracers for each protein of interest, named proteotypic peptides (Arsene-Ploetze et al., 2012; Arsène-Ploetze et al., 2014; Gallien et al., 2011). Therefore, extensive *in silico* and experimental analyses have been performed to choose best suited proteotypic peptides (unique for the protein and visible in MS) to be able to discriminate with highest sensitivity proteins from *Thiomonas* Group I and Group II (Supplementary Table 2). *In silico* screen for unicity was performed in all genomes available on the MicroScope (Microbial Genome Annotation & Analysis) Platform. Only peptides that were specific to the proteins from *Thiomonas* Groups I and/or II were manually validated. Then, isotopically heavy labeled peptides were synthetized for the selected peptides and used to determine the best transitions, i.e. number and fragment types to be followed for each of the selected peptides, and to optimize the MS instrument parameters. Sixteen peptides

showed good MS-signals and allowed optimization of 136 specific light/heavy transitions (Supplementary Tables 1 and 2). A concentration-balanced mixture of the heavy labeled peptides was made in order to obtain homogeneous signals for all peptides and this mixture was spiked into the samples in order to be able to quantify the endogenous peptides of interest by calculating heavy/light ratios.

First, a test experiment was set up to detect the selected peptides in total cell lysates obtained from four *Thiomonas* strains CB1, CB2, CB3 and CB6 (Group I *Thiomonas*) cultivated in laboratory conditions in the presence or in the absence of arsenite. This experiment has been performed both in biological and injection triplicates. Among the 16 peptides, 8 peptides could be detected and quantified, 6 peptides specific to AioA (peptides 1, 5, 6, 9, 10 and 11), one specific to AioB (peptide 12) and one specific to RpoA (peptide 13) (Table 5, Figure 4 and Supplementary Table 3). The comparison of light/heavy ratios allowed drawing the conclusion that the expression of AioA was higher when CB1, CB2, CB3, and CB6 were cultivated in the presence of As(III), as compared to the absence of As(III) (Figure 4 and Supplementary Table 3). These results were in agreement with previous results showing that expression of this protein is induced in *Thiomonas* cells in the presence of arsenite (Freel et al., 2015). As awaited, peptides 2 and 15 could not be detected in this experiment performed on proteins extracted from Group I *Thiomonas*, as they are specific of Group II *Thiomonas* (Table 5). Nevertheless, these 2 peptides have been identified in previous global metaproteomics studies and were therefore maintained for analysing sediment samples (Bertin et al., 2011) (Table 5).

Then, the optimized LC-SRM assay was tested on a metaproteomic sample extracted in 2007 at the COWG site on which the arsenite oxidases had already been detected (Bertin et al., 2011). This experiment has been performed with duplicate biological replicates and 2 or 3 injection replicates (Figure 5). Surprisingly, peptides 2 and 15, which are specific to the Group II *Thiomonas* were not detected while they were identified in the previous global analysis performed on the same samples (Bertin et al., 2011). This could suggest that these proteins are present at very low levels, in such a way that the peptide amount is just at the limit of detection. In total, eight peptides which were not specific to either Group I or II *Thiomonas* (peptides 1, 5, 6, 9, 10, 11, 12 and 13) could be detected in this metaproteomic test sample (Table 5, Figure 5) confirming that *Thiomonas* were active in these sediments and expressed the arsenite oxidases.

Finally, we have applied the optimized LC-SRM assay to attempt detecting the targeted proteins in sediments extracted at different time and space points in June 2011 and January 2012, the same sampling points at which the FISH and pyrosequencing analyses were performed. Enough protein material could be extracted only from COWG sites in 2011 and 2012. The sediment nature was clearly different in COWG and CONF (data not shown), which could also influence the Nycodenz-based cells extraction. Additionnal samples were included that had been extracted in 2009 at CONF and COWG sites, in which *Thiomonas* cells had been detected and for which physical and chemical characteristics of water samples are available (Delavat, Lett, et Lièvremont 2012; Volant et al. 2014) (Table 4). For all these samples, the protein amounts were not sufficient to perform a robust quantification assay with appropriate numbers of replicates, but could nevertheless be used to detect the proteins of interest.

One peptide specific to RpoA (peptide 15), unique of Group II *Thiomonas* was detected and thus confirmed the stable presence of *Thiomonas* strains from Group II at COWG and at CONF (Table 5 and Figure S1). In contrast, as no peptide specific to Group I *Thiomonas* could be detected, their activity could not be ensured. We were able to detect peptides 6, 10, 11 (AioA), 12 (AioB), all shared among Group I and II, on samples obtained from COWG in 2009, 2011 and 2012 and from CONF sampled in 2009, where it was

previously shown that arsenic concentration in water decreased below 6 mg/L (Volant et al., 2014). These results demonstrate that the *Thiomonas* arsenite oxidase is synthetized at three different times, at the source and downstream of the Reigous creek, even at a point where the As concentration is low (Table 5 and Figure S1). More generally, the physicochemical parameters, in particular the pH, the temperature, the redox potential and the arsenic concentration, and also the diversity and the spatiotemporal distribution of bacterial communities vary along the creek and during the time (Table 1 and 4) (Egal et al., 2010; Volant et al., 2014). Despite such variations, *Thiomonas* strains are ubiquitously present and expressed the arsenite oxidase over a fairly wide range of concentration in the Reigous creek.

As a conclusion, the optimized SRM assay developed in this study and applied without sample fractionation on very complex protein mixtures extracted from a microbial community was sensitive enough to detect AioA/B expressed by minor bacteria representing less than 9% of the community. This detection efficiency suggests that the arsenite oxidase protein was very strongly expressed by these *Thiomonas* and that crucial arsenite oxidation activity involved in the natural attenuation process probably occurred permanently in the Reigous creek, from 2007 to 2012.

Conflict of Interest Statement

The authors declare that no conflicts of interest were involved in this study.

Author and Contributors

O. B., M. H., A. V. D., Ch.C., and. F.A.-P. designed the research project; A. H., O. B., Co. C., A. D., J. F., M. H., Ch.C., and F.A.-P. performed the research; A. H., O. B., Co. C., A. D., M. H., Ch.C., and F.A.-P.. wrote the paper.

Funding / Acknowledgments

This research was supported by grants from the Université de Strasbourg (UdS), the Centre National de la Recherche Scientifique (CNRS) and the “Agence Nationale pour la Recherche (ANR)”. This study was financed by the EC2CO program (INEE, CNRS), the "Observatoire de Recherche Méditerranéen de l’Environnement-Observatoire des Sciences de l’Univers" (OSU OREME), and the THIOFILM (ANR-12-ADAP-0013) projects. Julien Farasin was supported by the ANR THIOFILM project (ANR-12-ADAP-0013). Florence Arsène-Ploetze was supported by the CNRS ("mise en délégation pour activité de recherche au CNRS"). We acknowledge the “Fondation pour la Recherche Médicale (FRM)” for funding the triple quadrupole instrument for targeted proteomics experiments. This work was also supported by the Proteomic French Infrastructure (ProFI; ANR-10-INSB-08-03).

References :

Amann, R. I., Binder, B. J., Olson, R. J., Chisholm, S. W., Devereux, R., and Stahl, D. A. (1990). Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. *Appl. Environ. Microbiol.* 56, 1919–1925.

Arsène-Ploetze, F., Bertin, P. N., and Carapito, C. (2014). Proteomic tools to decipher microbial community structure and functioning. *Environ. Sci. Pollut. Res. Int.* doi:10.1007/s11356-014-3898-0.

Arsene-Ploetze, F., Carapito, C., Plewniak, F., and N., P. (2012). "Proteomics as a Tool for the Characterization of Microbial Isolates and Complex Communities," in *Proteomic Applications in Biology*, ed. J. Heazlewood (InTech). ISBN: 978-953-307-613-3, DOI: 10.5772/29689.

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. doi:10.1371/journal.pgen.1000859.

Battaglia-Brunet, F., Achbouni, H. El, Quemeneur, M., Hallberg, K. B., Kelly, D. P., and Joulian, C. (2011a). Proposal that the arsenite-oxidizing organisms, *Thiomonas cuprina* and "*Thiomonas arsenivorans*" be reclassified as strains of *Thiomonas delicata*. *Int. J. Syst. Evol. Microbiol.* doi:10.1099/ijs.0.023408-0.

Battaglia-Brunet, F., Crouzet, C., Breeze, D., Tris, H., and Morin, D. (2011b). Decreased leachability of arsenic linked to biological oxidation of As(III) in solid wastes from bioleaching liquors. *Hydrometallurgy* 107, 34–39.

Battaglia-Brunet, F., Joulian, C., Garrido, F., Dictor, M.-C., Morin, D., Coupland, K., et al. (2006). Oxidation of arsenite by *Thiomonas* strains and characterization of *Thiomonas arsenivorans* sp. nov. *Antonie Van Leeuwenhoek* 89, 99–108. doi:10.1007/s10482-005-9013-2.

Bertin, P. N., Heinrich-Salmeron, A., Pelletier, E., Goulhen-Chollet, F., Arsène-Ploetze, F., Gallien, S., et al. (2011). Metabolic diversity among main microorganisms inside an arsenic-rich ecosystem revealed by meta- and proteo-genomics. *ISME J.* 5, 1735–47. doi:10.1038/ismej.2011.51.

Brown, J. F., Jones, D. S., Mills, D. B., Macalady, J. L., and Burgos, W. D. (2011). Application of a depositional facies model to an acid mine drainage site. *Appl. Environ. Microbiol.* 77, 545–554. doi:10.1128/AEM.01550-10.

Bruneel, Duran, R., Casiot, C., Elbaz-Poulichet, F., and Personné, J.-C. (2006). Diversity of microorganisms in Fe-As-rich acid mine drainage waters of Carnoulès, France. *Appl. Environ. Microbiol.* 72, 551–556. doi:10.1128/AEM.72.1.551-556.2006.

Bruneel, Pascault, N., Egal, M., Bancon-Montigny, C., Goñi-Urriza, M. S., Elbaz-Poulichet, F., et al. (2008). Archaeal diversity in a Fe-As rich acid mine drainage at Carnoulès (France). *Extrem. Life Extreme Cond.* 12, 563–571. doi:10.1007/s00792-008-0160-z.

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

590 Bruneel, Volant, A., Gallien, S., Chaumande, B., Casiot, C., Carapito, C., et al. (2011). Characterization
591 of the Active Bacterial Community Involved in Natural Attenuation Processes in Arsenic-Rich
592 Creek Sediments. *Microb. Ecol.* 61, 793–810. doi:10.1007/s00248-011-9808-9.

593 Bryan, Marchal, M., Battaglia-Brunet, F., Kugler, V., Lemaitre-Guillier, C., Lièvremont, D., et al. (2009).
594 Carbon and arsenic metabolism in *Thiomonas* strains: differences revealed diverse
595 adaptation processes. *BMC Microbiol.* 9, 127. doi:10.1186/1471-2180-9-127.

596 Casiot, C., Bruneel, O., Personné, J.-C., Leblanc, M., and Elbaz-Poulichet, F. (2004). Arsenic oxidation
597 and bioaccumulation by the acidophilic protozoan, *Euglena mutabilis*, in acid mine drainage
598 (Carnoulès, France). *Sci. Total Environ.* 320, 259–267. doi:10.1016/j.scitotenv.2003.08.004.

599 Casiot, C., Lebrun, S., Morin, G., Bruneel, O., Personné, J. C., and Elbaz-Poulichet, F. (2005). Sorption
600 and redox processes controlling arsenic fate and transport in a stream impacted by acid mine
601 drainage. *Sci. Total Environ.* 347, 122–130. doi:10.1016/j.scitotenv.2004.12.039.

602 Casiot, C., Morin, G., Juillot, F., Bruneel, O., Personné, J. C., Leblanc, M., et al. (2003). Bacterial
603 immobilization and oxidation of arsenic in acid mine drainage (Carnoulès creek, France).
604 *Water Res.* 37, 2929–2936. doi:10.1016/S0043-1354(03)00080-0.

605 Chang, C.-Y., Picotti, P., Hüttenhain, R., Heinzelmann-Schwarz, V., Jovanovic, M., Aebersold, R., et al.
606 (2012). Protein significance analysis in selected reaction monitoring (SRM) measurements.
607 *Mol. Cell. Proteomics MCP* 11, M111.014662. doi:10.1074/mcp.M111.014662.

608 Delavat, F., Lett, M.-C., and Lièvremont, D. (2012). Novel and unexpected bacterial diversity in an
609 arsenic-rich ecosystem revealed by culture-dependent approaches. *Biol. Direct* 7, 28.
610 doi:10.1186/1745-6150-7-28.

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

614 Duquesne, K., Lieutaud, A., Ratouchniak, J., Muller, D., Lett, M.-C., and Bonnefoy, V. (2008). Arsenite
615 oxidation by a chemoautotrophic moderately acidophilic *Thiomonas* sp.: from the strain
616 isolation to the gene study. *Environ. Microbiol.* 10, 228–237. doi:10.1111/j.1462-
617 2920.2007.01447.x.

618 Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., and Knight, R. (2011). UCHIME improves
619 sensitivity and speed of chimera detection. *Bioinformatics* 27, 2194–2200.
620 doi:10.1093/bioinformatics/btr381.

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

624 Freel, K., Krueger, M., Farasin, J., Brochier-Armanet, C., Barbe, V., Andres, J., et al. (2015). Adaptation
625 in toxic environments: arsenic genomic islands in the bacterial genus *Thiomonas*. *PLoS ONE* in
626 press.

627 Gallien, S., Duriez, E., and Domon, B. (2011). Selected reaction monitoring applied to proteomics. *J.
628 Mass Spectrom. JMS* 46, 298–312. doi:10.1002/jms.1895.

629 Hallberg, K. B., Coupland, K., Kimura, S., and Johnson, D. B. (2006). Macroscopic streamer growths in
630 acidic, metal-rich mine waters in north wales consist of novel and remarkably simple
631 bacterial communities. *Appl. Environ. Microbiol.* 72, 2022–2030.
632 doi:10.1128/AEM.72.3.2022-2030.2006.

633 Halter, Casiot, C., Heipieper, H. J., Plewniak, F., Marchal, M., Simon, S., et al. (2012). Surface
634 properties and intracellular speciation revealed an original adaptive mechanism to arsenic in
635 the acid mine drainage bio-indicator *Euglena mutabilis*. *Appl. Microbiol. Biotechnol.* 93,
636 1735–1744. doi:10.1007/s00253-011-3493-y.

637 Heinrich-Salmeron, A., Cordi, A., Brochier-Armanet, C., Halter, D., Pagnout, C., Abbaszadeh-Fard, E.,
638 et al. (2011). Unsuspected Diversity of Arsenite-Oxidizing Bacteria as Revealed by
639 Widespread Distribution of the *aoxB* Gene in Prokaryotes. *Appl. Environ. Microbiol.* 77, 4685–
640 4692. doi:10.1128/AEM.02884-10.

641 Hugenholtz, P., Tyson, G. W., and Blackall, L. L. (2002). Design and evaluation of 16S rRNA-targeted
642 oligonucleotide probes for fluorescence in situ hybridization. *Methods Mol. Biol. Clifton NJ*
643 179, 29–42.

644 Huse, S. M., Welch, D. M., Morrison, H. G., and Sogin, M. L. (2010). Ironing out the wrinkles in the
645 rare biosphere through improved OTU clustering. *Environ. Microbiol.* 12, 1889–1898.
646 doi:10.1111/j.1462-2920.2010.02193.x.

647 Kruger, M. C., Bertin, P. N., Heipieper, H. J., and Arsène-Ploetze, F. (2013). Bacterial metabolism of
648 environmental arsenic—mechanisms and biotechnological applications. *Appl. Microbiol.*
649 *Biotechnol.* 97, 3827–3841. doi:10.1007/s00253-013-4838-5.

650 MacLean, B., Tomazela, D. M., Shulman, N., Chambers, M., Finney, G. L., Frewen, B., et al. (2010).
651 Skyline: an open source document editor for creating and analyzing targeted proteomics
652 experiments. *Bioinforma. Oxf. Engl.* 26, 966–968. doi:10.1093/bioinformatics/btq054.

653 Mandal, B. K., and Suzuki, K. T. (2002). Arsenic round the world: a review. *Talanta* 58, 201–235.

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

657 Needleman, S. B., and Wunsch, C. D. (1970). A general method applicable to the search for
658 similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* 48, 443–453.

659 Oremland, R. S., and Stolz, J. F. (2003). The ecology of arsenic. *Science* 300, 939–944.
660 doi:10.1126/science.1081903.

661 Quéméneur, M., Cébron, A., Billard, P., Battaglia-Brunet, F., Garrido, F., Leyval, C., et al. (2010).
662 Population structure and abundance of arsenite-oxidizing bacteria along an arsenic pollution
663 gradient in waters of the upper isle River Basin, France. *Appl. Environ. Microbiol.* 76, 4566–
664 4570. doi:10.1128/AEM.03104-09.

665 Quéméneur, M., Heinrich-Salmeron, A., Muller, D., Lièvremont, D., Jauzein, M., Bertin, P. N., et al.
666 (2008). Diversity surveys and evolutionary relationships of *aoxB* genes in aerobic arsenite-
667 oxidizing bacteria. *Appl. Environ. Microbiol.* 74, 4567–4573. doi:10.1128/AEM.02851-07.

668 Quince, C., Lanzén, A., Curtis, T. P., Davenport, R. J., Hall, N., Head, I. M., et al. (2009). Accurate
669 determination of microbial diversity from 454 pyrosequencing data. *Nat. Methods* 6, 639–
670 641. doi:10.1038/nmeth.1361.

671 Rodier, J., Bazin, C., Broutin, J. P., Chambon, P., Champsaur, H., and Rodi, L. (1996). *L'analyse de*
672 *l'eau: eaux naturelles, eaux résiduaires, eau de mer*. 8e éd. Paris: Dunod.

673 Samanta, G., and Clifford, D. A. (2005). Preservation of inorganic arsenic species in groundwater.
674 *Environ. Sci. Technol.* 39, 8877–8882.

675 Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., et al. (2009).
676 Introducing mothur: open-source, platform-independent, community-supported software for
677 describing and comparing microbial communities. *Appl. Environ. Microbiol.* 75, 7537–7541.
678 doi:10.1128/AEM.01541-09.

679 Sharma, V. K., and Sohn, M. (2009). Aquatic arsenic: Toxicity, speciation, transformations, and
680 remediation. *Environ. Int.* 35, 743–759. doi:10.1016/j.envint.2009.01.005.

681 Silver, S., and Phung, L. T. (2005). Genes and enzymes involved in bacterial oxidation and reduction of
682 inorganic arsenic. *Appl. Environ. Microbiol.* 71, 599–608. doi:10.1128/AEM.71.2.599-
683 608.2005.

684 Slyemi, D., and Bonnefoy, V. (2012). How prokaryotes deal with arsenic. *Environ. Microbiol. Rep.* 4,
685 571–586. doi:10.1111/j.1758-2229.2011.00300.x.

686 Slyemi, D., Moinier, D., Talla, E., and Bonnefoy, V. (2013). Organization and regulation of the arsenite
687 oxidase operon of the moderately acidophilic and facultative chemoautotrophic *Thiomonas*
688 *arsenitoxydans*. *Extrem. Life Extreme Cond.* 17, 911–920. doi:10.1007/s00792-013-0573-1.

689 Volant, A., Bruneel, O., Desoeuvre, A., Héry, M., Casiot, C., Bru, N., et al. (2014). Diversity and
690 spatiotemporal dynamics of bacterial communities: physicochemical and other drivers along
691 an acid mine drainage. *FEMS Microbiol. Ecol.* 90, 247–263. doi:10.1111/1574-6941.12394.

692 Volant, A., Desoeuvre, A., Casiot, C., Lauga, B., Delpoux, S., Morin, G., et al. (2012). Archaeal diversity:
693 temporal variation in the arsenic-rich creek sediments of Carnoulès Mine, France. *Extrem.*
694 *Life Extreme Cond.* 16, 645–657. doi:10.1007/s00792-012-0466-8.

695 Wallner, G., Amann, R., and Beisker, W. (1993). Optimizing fluorescent in situ hybridization with
696 rRNA-targeted oligonucleotide probes for flow cytometric identification of microorganisms.
697 *Cytometry* 14, 136–143. doi:10.1002/cyto.990140205.

698

699 **Table and Figure legends**
700
701 **Figure 1.** Rarefaction curves of the bacterial 16S rRNA gene sequences obtained by
702 pyrosequencing from Reigous creek sediments based on OTUs calculated at 97% identity for
703 the full dataset. The total number of sequences analyzed is plotted against the number of
704 OTUs observed in the same library.
705
706 **Figure 2.** FISH micrographs of cells labeled with the oligonucleotide probe TM1G0138
707 detecting *Thiomonas* bacteria from the Group II, in COWG sediment (January 2012). Most
708 bacteria are shown with the DAPI staining (blue), and cells targeted with TM1G0138 are
709 labelled with Cy5 (purple). Scale bars = 5 μm
710
711 **Figure 3.** Estimated proportions of *Thiomonas* bacteria hybridized with the oligonucleotide
712 probe TM1G0138 for the Group II according to (Bryan et al., 2009), with the oligonucleotide
713 probe TM2G0138 for the Group I according to (Bryan et al., 2009) and proportions of
714 *Thiomonas* Groups I and II together. Scale bar: SEM.
715
716 **Figure 4.** Quantification of the arsenite oxidase specific peptides (average light/heavy ratios
717 calculated from biological and injection triplicates) in *Thiomonas* CB1, CB2, CB3 and CB6
718 cells cultivated in the absence or in the presence of As(III) in m126 medium.
719
720 **Figure 5.** Detection and quantification (average light/heavy ratios calculated from triplicate
721 injections for two biological replicates) of the *Thiomonas* arsenite oxidase and RpoA specific
722 peptides in sediments of the Reigous creek sampled in 2007 (Bertin et al., 2011). Scale bar:
723 SEM
724

Table 1: Physical and chemical characteristics of water samples.

Sampling date	Station	Physico-chemical parameter			Concentrations of the constituants of interest and redox As species					
		T (°C)	pH	Conductivity (µS/cm)	Oxygen concentration (mg/L)	Sulfate (mg/L)	Fe(II) (mg/L)	Total As (mg/L)	Proportion of As(III) (%)	Proportion of As(V) (%)
30/06/2011	S1	14.8	3.15	4000	0.45	3837	1539	180.7	91.5	8.5
	COWG	15.5	3.23	3590	7.37	2798	1377	160.3	86.1	13.9
	CONF	16.8	6.07	1366	8.56	1081	5	0.02	25.6	74.4
24/01/2012	S1	13.7	3.79	3140	1.07	3479	955	154.9	75.9	24.1
	COWG	11.2	3.68	2640	8.51	2991	691	115.2	89.7	10.3
	CONF	7.6	5.14	971	9.57	806	8	10.3	87.0	13.0

725

Table 2. Estimated OTU richness, diversity indices and estimated sample coverage for each 16S rRNA libraries using pyrosequencing analysis. Results are presented for the normalized data, randomly resampled to have an equal sample size.

Sampling stations	No. of sequences after quality filtering	No. of Normalized sequences	No. of OTUs[a]	Singletons	Coverage[b]	Shannon[c]	Chao1
S1 Jun. 2011 n1	12284	2296	96	26	98	2.94 (2.88 ; 3.00)	178 (131 ; 286)
S1 Jun. 2011 n2	2296	2296	52	11	99	1.94 (1.88 ; 2.00)	88 (65 ; 155)
S1 Jan. 2012 n1	10133	2296	51	9	99	1.11 (1.05 ; 1.18)	97 (67 ; 183)
S1 Jan. 2012 n2	13721	2296	48	9	99	1.06 (0.99 ; 1.13)	69 (55 ; 110)
COWG Jun. 2011 n1	10049	2296	133	27	97	3.61 (3.56 ; 3.66)	223 (178 ; 312)
COWG Jun. 2011 n2	13311	2296	120	32	98	3.51 (3.46 ; 3.56)	273 (189 ; 461)
COWG Jan. 2012 n1	12662	2296	135	33	97	2.39 (2.31 ; 2.47)	264 (204 ; 376)
COWG Jan. 2012 n2	11965	2296	82	16	99	1.65 (1.57 ; 1.72)	117 (97 ; 167)
CONF Jun. 2011 n1	18508	2296	391	71	95	5.40 (5.36 ; 5.44)	479 (446 ; 534)
CONF Jun. 2011 n2	9676	2296	369	64	96	5.35 (5.31 ; 5.39)	439 (411 ; 486)
CONF Jan. 2012 n1	14116	2296	397	108	92	4.87 (4.80 ; 4.93)	620 (547 ; 727)
CONF Jan. 2012 n2	14408	2296	334	69	94	4.72 (4.65 ; 4.78)	448 (407 ; 514)

[a] OTUs were defined at 97% cutoff

[b] Coverage: sum of probabilities of observed classes calculated as (1 - (n/N)), where n is the number of singleton sequences and N is the total number of sequences

[c] Takes into account the number and evenness of species

Values in brackets are 95% confidence intervals

734 **Table 3.** Number of sequences assigned to the genus *Thiomonas* using pyrosequencing analyses.

Taxon	Group	Total n1	S1 Jun. 2011 n1	S1 Jan. 2012 n1	COWG Jun. 2011 n1	COWG Jan. 2012 n1	CONFJ un. 2011 n1	CONF Jan. 2012 n1	Total n2	S1 Jun. 2011 n2	S1 Jan. 2012 n2	COWG Jun. 2011 n2	COWG Jan. 2012 n2	CONF Jun. 2011 n2	CONF Jan. 2012 n2
Total		1008	190	84	105	114	274	241	873	42	115	165	118	98	335
Tm. PK44	II	1	0	0	0	0	0	0	573	40	115	150	118	19	131
Tm. Ynys3	I	4	0	0	0	0	0	1	1	0	0	0	0	0	1
Tm. arsenivorans	II	3	0	0	0	0	0	3	15	0	0	0	0	0	15
Tm. cuprina	II	25	0	0	0	0	0	0	18	0	0	0	0	18	0
Tm.intermedia	I	184	7	1	5	0	11	1	8	1	0	7	0	0	0
Tm. sp. DM Cd4	II	979	0	0	2	1	104	52	99	0	0	4	0	52	43
Unclassified	?	820	183	83	98	113	159	184	159	1	0	4	0	9	145

735

736 **Table 4:** Studies revealing the presence of *Thiomonas* in the Reigous creek.

Date	Sampling sites	Type of *Thiomonas*	Abondance	Approach used	Reference
Every 3–4 weeks from June 2001 to August 2002	Four locations along the first 150 m (water) including S1 et COWG	Group I (B1; B2; B3)	≈ 0,5%	cultivation	(Bruneel et al., 2003).
October 2002 and January 2003	S1, COWA and COWG (water)	None detected	0	terminal-restriction fragment length polymorphism (T-RFLP) and 16S rRNA gene library analyses	(Bruneel et al., 2006)
April 2006	COWG (water and sediments)	Group II	6% sediments, 3% water	16S rRNA gene library analyses; metaproteomic	(Bruneel et al., 2011)
May 2007	COWG (sediments)	Group II	Not quantified	metagenomic, metaproteomic	(Bertin et al., 2011)
November 2009	Sediment at S1, COWG, GAL and CONF	Group II (X19)	Not quantified	cultivation	(Delavat et al., 2012)
From November 2007 to March 2010	Water at S5 (inside the tailing), S1, COWG, GAL and CONF	?	<1%	pyrosequencing	(Volant et al., 2014)
June 2011	Sediments at S1, COWG, CONF	Group I and II	<1%	pyrosequencing	This study
			5-6%	FISH	
January 2012	Sediments at S1, COWG, CONF	Group I and II	<1%	pyrosequencing	
			7-9%	FISH	

737

738 **Table 5:** Peptides allowing to detect AioA/Band RpoA from *Thiomonas* in protein extracts from cultures of isolated strains and from the
739 community of the Reigous creek sediments at different space and time points

Targeted protein	Peptide number	Sequence	Specificity Tm.I / Tm.II[#]	Global proteomics 2007 study (Bertin et al., 2011)	Cell cultures	2007 COWG	2009 COWG	2009 CONF	2011 COWG	2012 COWG
AioA	1	ACVVNQGLSSTR	Tm.I / Tm.II	X	CB1, CB2, CB3, CB6	X	-	-	-	-
	2	IQIFPAK	Tm.II (Carn2_0821 only)*	X	-	-	-	-	-	-
	3	DFIANHTEGFDAAVK	Tm.II (Carn2_1469 only)*	-	-	-	-	-	-	-
	4	DFIANHTEGFEAAVK	Tm.II (Carn2_0821, Carn2_1330)	-	-	-	-	-	-	-
	5	FWINNGR	Tm.I / Tm.II	X	CB1, CB2, CB6	X	-	-	-	-
	6	YPAADFPIPR	Tm.I / Tm.II	X	CB1, CB2, CB3, CB6	X	X	X	X	X
	7	YPASSVPIPR	*T. intermedia* K12 (Tm.I)	-	-	-	-	-	-	-
	8	DFIAQHTEGFEAAVK	Tm.I	-	-	-	-	-	-	-
	9	IQVFPAK	Tm.I / Tm.II	-	CB1, CB2, CB3, CB6	X	-	-	-	-
	10	ITGVPVAQIK	Tm.I / Tm.II	X	CB1, CB2, CB3, CB6	X	X	X	X	X
	11	LIFTGIQTPTVR	Tm.I / Tm.II	X	CB2, CB3, CB6	X	X	X	X	X
AioB	12	AVAVTGLIYGR	Tm.I / Tm.II	X	CB1, CB6	X	X	X	X	X
RpoA	13	SETELLK	Tm.I / Tm.II	-	CB1, CB2, CB3, CB6	X	-	-	-	-
	14	SIGHIVLDASFSPVR	Tm.I	-	-	-	-	-	-	-
	15	SIQVEALGPLR	Tm.II	X	-	-	X	X	X	X
	16	SIQVESLGHNR	Tm.I	-	-	-	-	-	-	-

740 [#] Tm.I and Tm.II: specific to *Thiomonas* Group I and Group II, respectively, according to (Bryan et al., 2009); *This peptide is specific to one of
741 the three AioA proteins found in the *Thiomonas*-like CARN2, Group II (Bertin et al., 2011)

Supplementary Figures

Figure S1. Detection of the AioA/B and RpoA peptides using the SRM assay. Chromatograms were extracted with Skyline (MacLean et al., 2010). Three time points (B, C and D) and two locations (A and B) were analysed. A : CONF 2009; B : COWG 2009; C : COWG 2011; D : COWG 2012. Each transition is represented by one color. At the right side, in the peak area windows, a dotp close to 1 indicates a very good correlation between transitions and relative fragment intensities of the endogenous peptides with the reference library spectra.

Figure 1.TIF

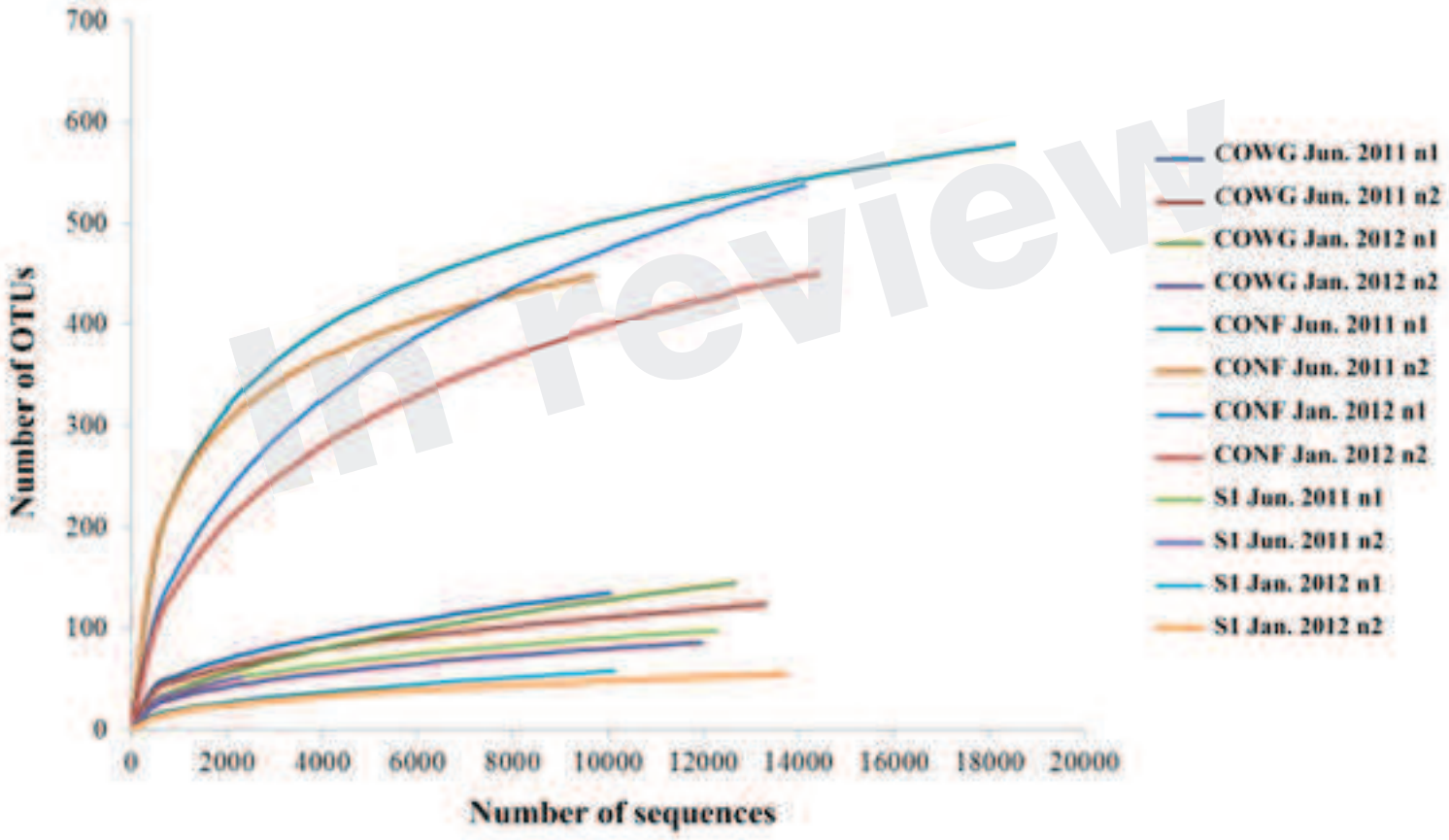


Figure 2.TIF

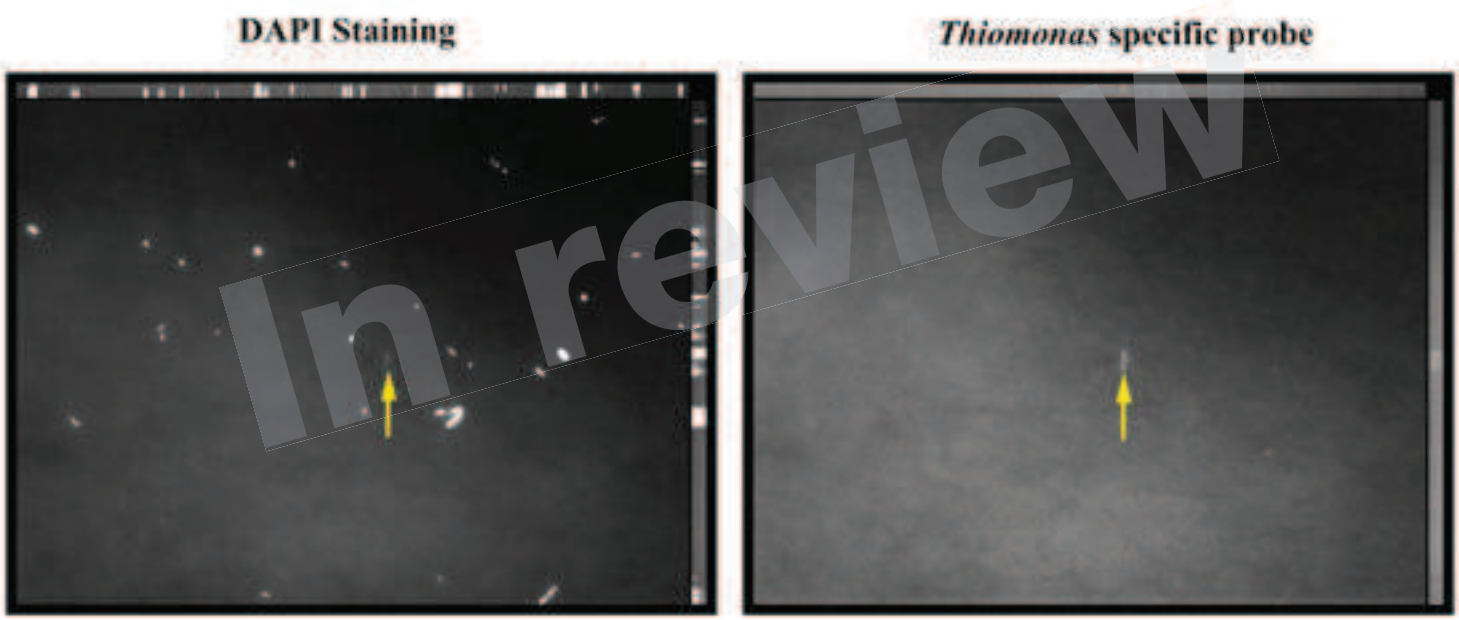


Figure 3.TIF

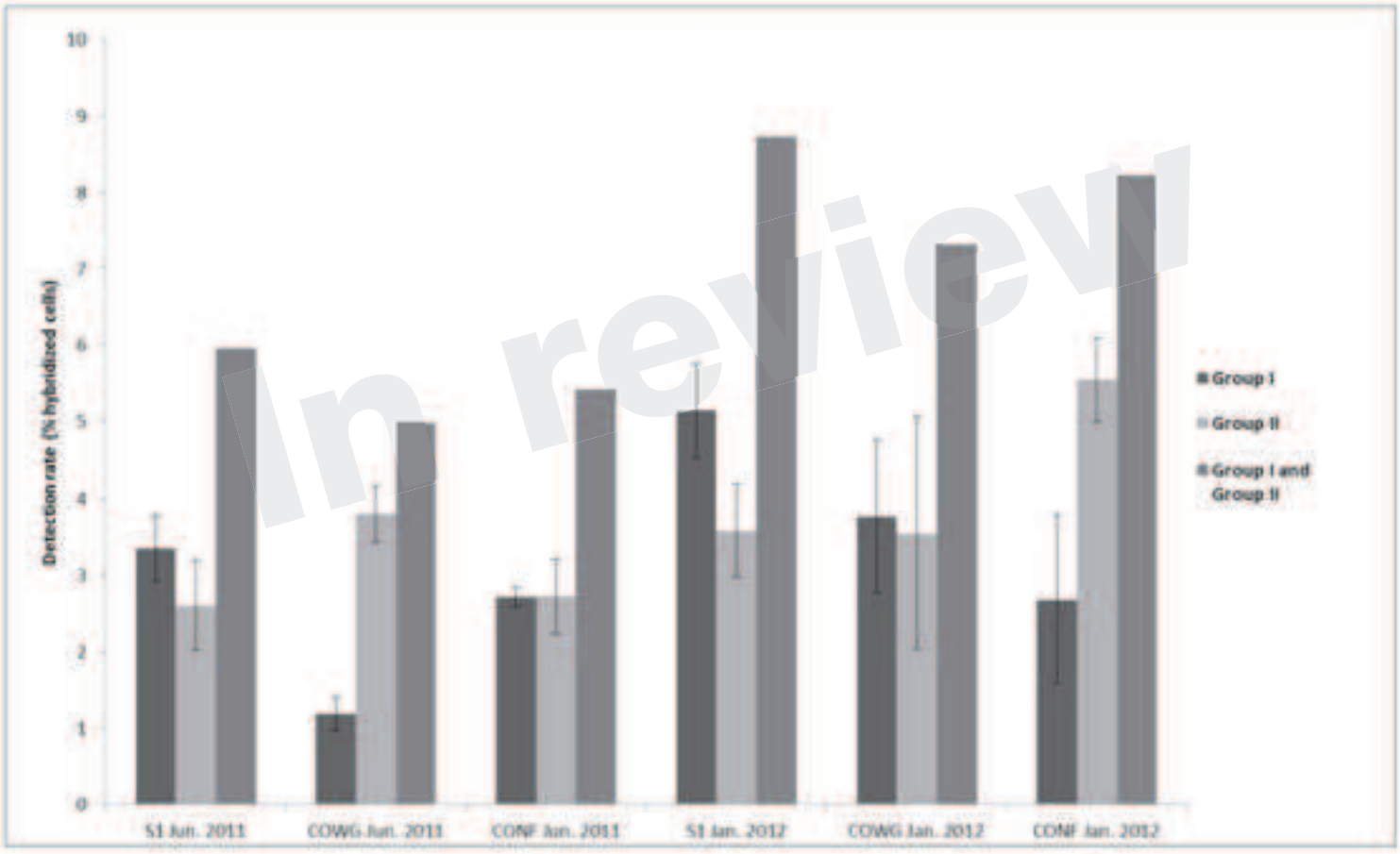


Figure 4.TIF

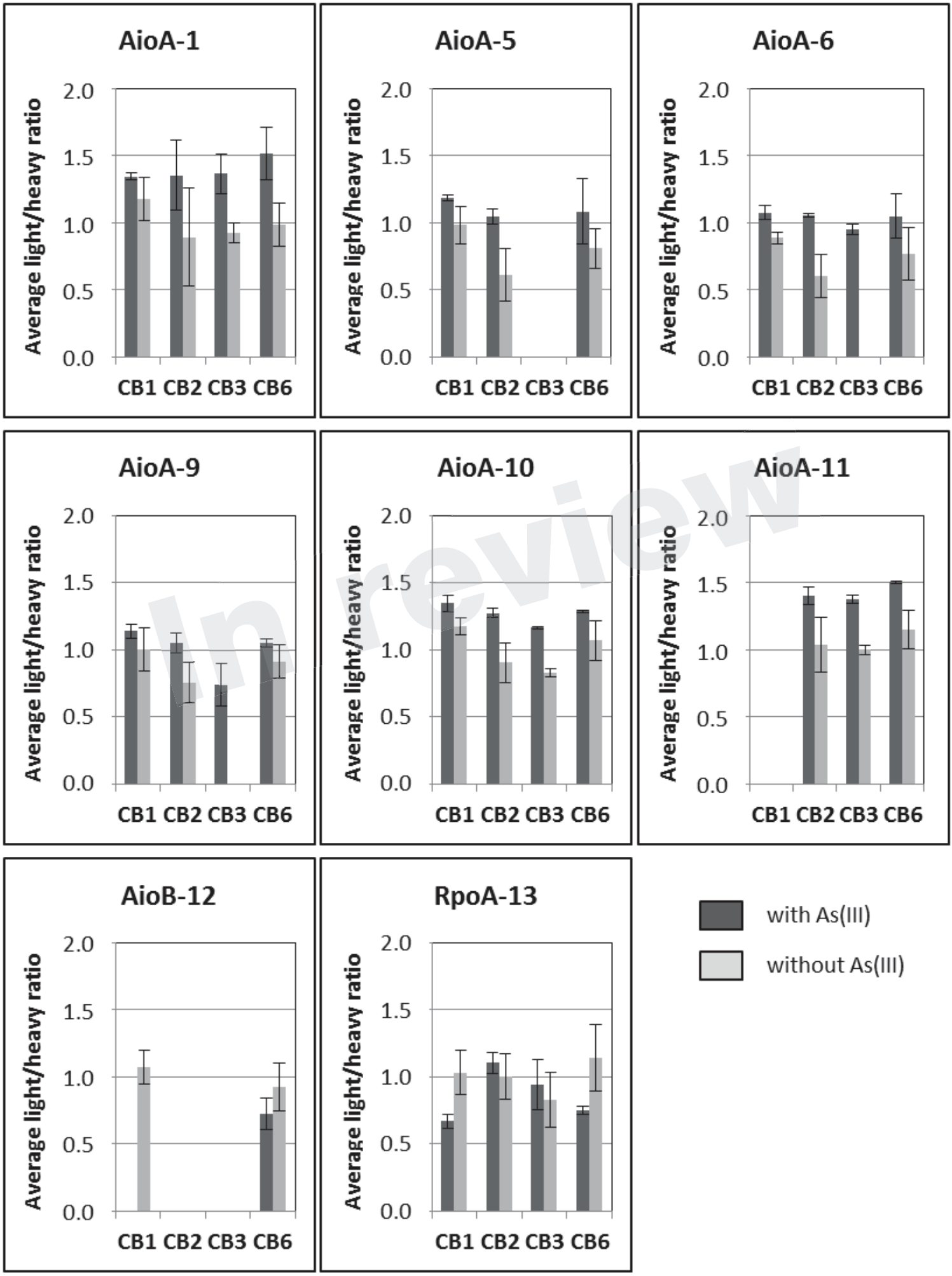
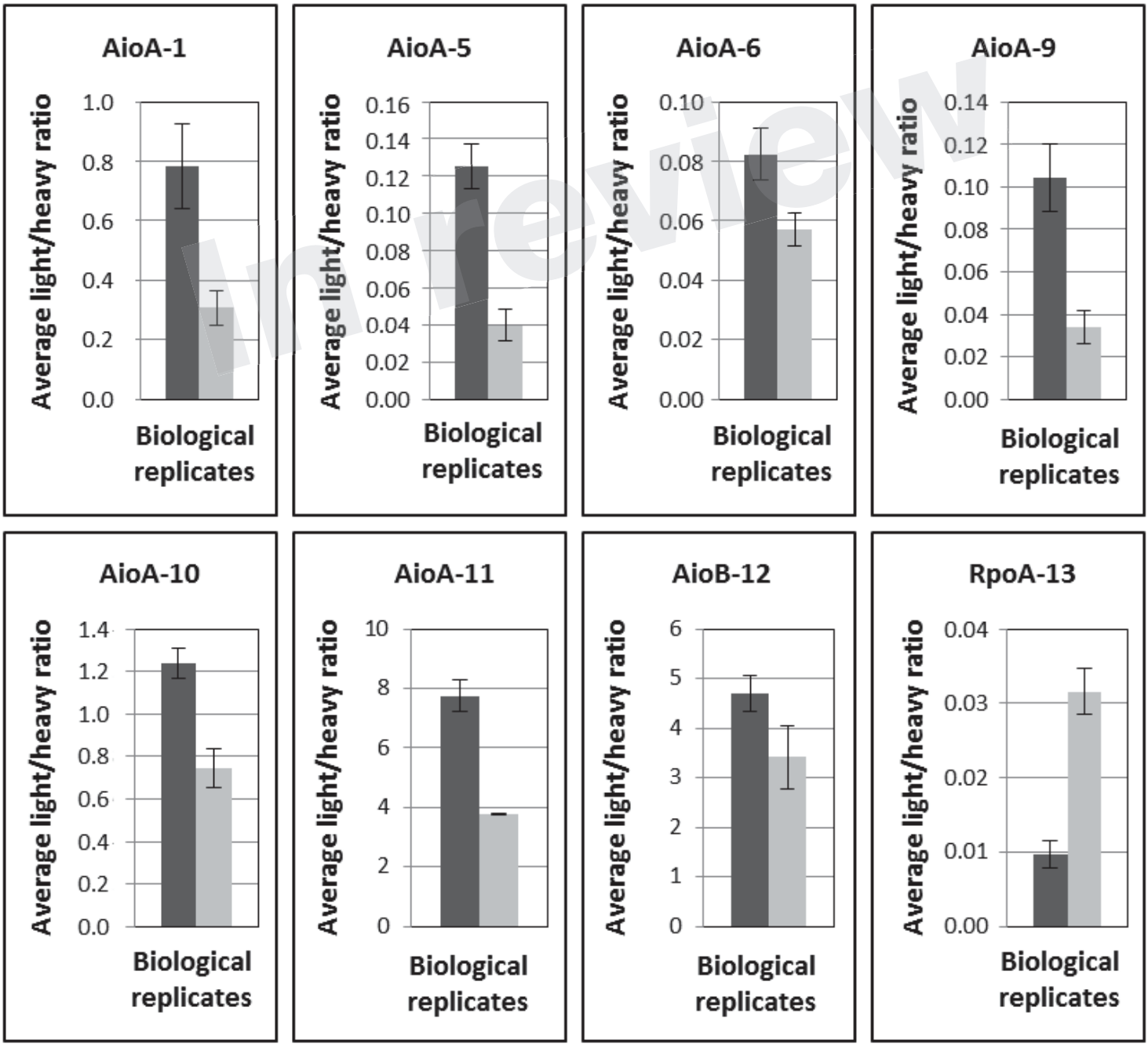


Figure 5.TIF



Annexe 3

Supplementary Table 1. List of the genes present in the urea island of CB2.

						Compositionnal bias				Correspondance in : [5]				
Label	Begin	End	Type	Product	Gene	GC Region[1]	SIGI[2]	IVOM[3]	C.A.I[4]	3As	K12	CB1	CB6	CB3
THICB2_v3_misc_RNA_3	1538398	1538508	misc_RNA	ykkC-yxkD	_	-	-	+	_	-	-	-	-	-
THICB2_v3_360017	1538516	1539547	CDS	putative ABC-type nitrate transport system, periplasmic component	_	+1SD	-	+	0.551665	-	-	-	-	-
THICB2_v3_360018	1539583	1540440	CDS	putative Nitrate transport permease protein nrtB	_	-	-	+	0.634238	-	-	-	-	-
THICB2_v3_360019	1540437	1541297	CDS	conserved protein of unknown function	_	-	-	+	0.597967	+	-	+	+	+
THICB2_v3_360020	1541366	1542103	CDS	conserved protein of unknown function	_	-	-	+	0.568542	-	-	-	-	-
THICB2_v3_360021	1542100	1542834	CDS	conserved protein of unknown function	_	-	-	+	0.581185	-	-	-	-	-
THICB2_v3_360022	1542909	1546538	CDS	Urea carboxylase	_	-	-	-	0.618229	+	+	+	+	+
THICB2_v3_360023	1546539	1548329	CDS	putative Allophanate hydrolase	_	-	-	-	0.629366	+	-	+	+	-
THICB2_v3_360024	1548368	1548811	CDS	Cyanate hydratase	cynS	-	-	-	0.804462	-	-	-	-	-
THICB2_v3_360025	1548821	1549552	CDS	conserved protein of unknown function	_	-	-	-	0.502841	-	-	-	-	-
THICB2_v3_360026	1549549	1550331	CDS	putative AHL-lactonase	_	+1SD	-	-	0.543508	-	-	-	-	-
THICB2_v3_360027	1550328	1551113	CDS	Creatinine amidohydrolase	crnA	-	-	-	0.55664	-	-	-	-	-
THICB2_v3_360028	1551125	1552012	CDS	putative Permease of the drug/metabolite transporter (DMT) superfamily	_	-	-	-	0.451108	+	-	+	+	-
THICB2_v3_360029	1552019	1552771	CDS	3-oxoacyl-[acyl-carrier-protein] reductase FabG	fabG	-	-	+	0.565649	+	+	+	+	+
THICB2_v3_360030	1552783	1553568	CDS	putative 3-oxoacyl-[acyl-carrier-protein] reductase	_	-	-	+	0.579694	+	+	+	+	+
THICB2_v3_360031	1553574	1554629	CDS	putative fatty acid desaturase	_	+1SD	-	+	0.552697	-	-	-	-	-
THICB2_v3_360032	1554661	1555671	CDS	putative ABC-type nitrate/sulfonate/bicarbonate transport system, periplasmic component	_	+1SD	-	+	0.589331	-	-	-	-	-
THICB2_v3_360033	1555821	1556003	CDS	conserved protein of unknown function	_	+1SD	-	+	0.348595	-	-	-	-	-
THICB2_v3_360034	1556000	1556836	CDS	putative alpha/beta-Hydrolases	_	-	-	+	0.509167	+	+	+	+	+
THICB2_v3_360035	1556833	1557681	CDS	putative ABC-type transport system, ATPase component	_	-	-	+	0.53665	+	-	+	+	+
THICB2_v3_360036	1557678	1558424	CDS	putative ABC-type transport system, permease component	_	+1SD	-	+	0.567209	-	-	-	-	-
THICB2_v3_360037	1558436	1559446	CDS	putative ABC-type transport system, periplasmic component	_	+1SD	-	+	0.53396	-	-	-	-	-

THICB2_v3_360038	1559666	1560415	CDS	putative transcriptional regulator, GntR family	_	-	-	+	0.589869	-	-	-	-	-
THICB2_v3_360039	1560412	1560801	CDS	conserved protein of unknown function	_	-	-	-	0.592038	-	-	-	-	-
THICB2_v3_360040	1560827	1562206	CDS	Biuret hydrolase	atzE	-	-	-	0.555097	+	+	+	+	+
THICB2_v3_360041	1562257	1562463	CDS	conserved protein of unknown function	_	-	-	-	0.534453	-	-	-	-	-
THICB2_v3_360042	1562977	1563810	CDS	putative 5-formyltetrahydrofolate cyclo-ligase	_	-	-	-	0.511886	-	-	-	-	-
THICB2_v3_360043	1563964	1565448	CDS	putative Monosaccharide-transporting ATPase (ABC-type transport system)	_	-	-	-	0.611755	+	-	+	+	-
THICB2_v3_360044	1565445	1566491	CDS	putative ABC-type transport system, permease component	_	-	-	-	0.617693	-	-	-	-	-
THICB2_v3_360045	1566488	1567171	CDS	Putative glucose ABC transporter permease protein TsgC13 (fragment)	_	-	-	-	0.535169	+	-	+	+	-
THICB2_v3_370001	1568279	1568473	CDS	Putative ABC-type transport system, permease component (fragment)	_	-	-	-	0.549735	-	-	-	-	-
THICB2_v3_370002	1568470	1569591	CDS	putative ABC-type transport system, periplasmic component	_	-	-	-	0.623661	+	-	+	+	-
THICB2_v3_370003	1569648	1570670	CDS	Formamidase	amiF	-	-	-	0.613995	-	-	-	-	-
THICB2_v3_370004	1570670	1571353	CDS	Putative Isochorismatase hydrolase	_	-	-	-	0.650758	+	-	+	+	-
THICB2_v3_370005	1571350	1572039	CDS	Putative Isochorismatase hydrolase	_	-	-	-	0.645811	+	-	+	+	+
THICB2_v3_370006	1572047	1573579	CDS	putative S-adenosylhomocysteine deaminase	_	-	-	-	0.539489	+	-	+	+	-
THICB2_v3_370007	1573965	1575335	CDS	putative NAD-dependent dihydropyrimidine dehydrogenase sunbunit PreT	_	-1SD	-	-	0.634196	+	-	+	+	-
THICB2_v3_370008	1575328	1576644	CDS	putative oxidoreductase subunit	yeiA	-	-	-	0.72048	-	-	-	-	-
THICB2_v3_370009	1576691	1578136	CDS	D-hydantoinase/dihydropyrimidinase	dht	-	-	-	0.754505	-	-	-	-	-
THICB2_v3_370010	1578165	1579439	CDS	Amidase, hydantoinase/carbamoylase family	atcC	-	-	-	0.736135	+	+	+	+	-
THICB2_v3_370011	1579509	1580327	CDS	conserved protein of unknown function	_	-	-	-	0.757501	+	+	+	+	+
THICB2_v3_370012	1580335	1581270	CDS	putative ABC-type transport system, permease component	_	-1SD	-	-	0.656745	+	-	+	+	-
THICB2_v3_370013	1581267	1582460	CDS	putative ABC-type transport system, permease component	_	-1SD	-	-	0.644853	+	-	+	+	-
THICB2_v3_370014	1582493	1583527	CDS	putative ABC-type transport system, periplasmic component	_	-	-	-	0.724908	-	-	-	-	-
THICB2_v3_370015	1583547	1584242	CDS	HTH-type transcriptional regulator RutR	rutR	-	-	-	0.649552	-	-	-	-	-
THICB2_v3_370016	1584301	1584429	CDS	protein of unknown function	_	-	-	-	0.560174	+	+	+	+	-
THICB2_v3_370017	1584459	1585205	fCDS	fragment of cytosine deaminase (part 1)	codA	-	-	-	0.729664	-	-	-	-	-
THICB2_v3_370018	1585215	1585556	fCDS	fragment of cytosine deaminase (part 2)	codA	-1SD	-	-	0.686135	-	-	-	-	-
THICB2_v3_370019	1585609	1585686	CDS	protein of unknown function	_	-	-	-	0.500334	-	-	-	-	-
THICB2_v3_370020	1585683	1586210	CDS	conserved protein of unknown function	_	-	-	-	0.595289	-	-	-	-	-
THICB2_v3_370021	1586217	1587674	CDS	putative hydrolase protein	_	-1SD	-	-	0.657582	-	-	-	-	-

THICB2_v3_370022	1587675	1588514	CDS	putative transport protein (ABC superfamily, atp_bind)	_	-	-	-	0.645022	+	-	+	+	+
THICB2_v3_370023	1588529	1589590	CDS	putative ring-hydroxylating dioxygenase with Rieske 2Fe-2S domain	_	-	-	-	0.593176	-	-	-	-	-
THICB2_v3_370024	1589529	1589792	CDS	protein of unknown function	_	+1SD	-	-	0.329535	-	-	-	-	-
THICB2_v3_370025	1589946	1590953	CDS	putative ABC-type transport system, periplasmic component	_	-	-	-	0.682419	+	-	+	+	-
THICB2_v3_370026	1590987	1591991	CDS	putative ABC-type transport system, periplasmic component	_	-	-	-	0.705269	+	-	+	+	-
THICB2_v3_370027	1592010	1592840	CDS	putative permease protein (ABC superfamily, membrane)	_	-	-	-	0.632108	+	-	+	+	-
THICB2_v3_370028	1592847	1593752	CDS	putative Clavaminate synthase-like	_	-	-	-	0.627482	-	-	-	-	-
THICB2_v3_370029	1593882	1594748	CDS	putative Nucleoside-binding outer membrane protein	_	-	-	-	0.655602	-	-	-	-	-
THICB2_v3_370030	1595061	1595456	CDS	putative ribonuclease VapC	_	+1SD	-	-	0.466013	+	+	+	+	+
THICB2_v3_370031	1595501	1596487	CDS	transposase	_	-	-	-	0.686224	+	-	+	+	+
THICB2_v3_370032	1596484	1596750	CDS	conserved protein of unknown function	_	+1SD	-	-	0.375754	-	-	-	-	-
THICB2_v3_370033	1596749	1597159	CDS	putative Sirohydrochlorin cobaltochelatase CbiX	_	-	-	-	0.413282	-	-	-	-	-
THICB2_v3_370034	1597236	1597871	CDS	putative Response regulator antiterminator nasT	_	-	-	-	0.597961	-	-	-	-	-
THICB2_v3_370035	1598167	1599447	CDS	Nitrate transporter component	nasF	-	-	-	0.7448	-	-	-	-	-
THICB2_v3_370036	1599479	1600393	CDS	Nitrate transporter component	nrtB	-	-	-	0.694609	-	-	-	-	-
THICB2_v3_370037	1600429	1601229	CDS	Nitrate transport protein NasD	nasD	-	-	-	0.722438	+	-	+	+	+
THICB2_v3_370038	1601262	1603721	CDS	Nitrite reductase [NAD(P)H]	nasD	-	-	-	0.738606	-	-	-	-	-
THICB2_v3_370039	1603718	1604107	CDS	Assimilatory nitrite reductase [NAD(P)H] small subunit	nasE	-1SD	-	-	0.702986	-	-	-	-	-
THICB2_v3_370040	1604107	1606962	CDS	putative Nitrate reductase	_	-1SD	-	-	0.703052	-	-	-	-	-
THICB2_v3_370041	1606996	1608066	CDS	putative anthranilate phosphoribosyltransferase	_	-	-	+	0.554895	+	-	+	+	+
THICB2_v3_370042	1608050	1608874	CDS	Uroporphyrinogen-III C-methyltransferase	cobA	-	-	+	0.44202	+	-	+	+	+
THICB2_v3_370043	1609150	1609497	CDS	putative transposase (fragment)	_	-	-	+	0.52767	+	-	+	+	-
THICB2_v3_370044	1609364	1609867	CDS	putative transposase	_	-	-	+	0.602139	+	-	+	+	-
THICB2_v3_370045	1609912	1610286	CDS	putative Permeases of the major facilitator superfamily	_	-	-	+	0.439381	+	-	+	+	-
THICB2_v3_370046	1610520	1611623	CDS	Conserved protein of unknown function	_	-	-	+	0.601528	+	-	+	+	-
THICB2_v3_370047	1612175	1612516	CDS	putative Toxin-antitoxin system, RelE/ParE toxin family	_	+1SD	-	+	0.511663	+	-	+	+	-
THICB2_v3_370048	1612522	1612791	CDS	putative type II toxin-antitoxin system, antitoxin Phd/YefM	_	+1SD	-	-	0.497572	+	-	+	+	+
THICB2_v3_370049	1612989	1613825	CDS	Urease accessory protein UreD	ureD	-1SD	-	-	0.627108	+	-	+	+	-
THICB2_v3_370050	1614132	1614824	CDS	branched-chain amino acid transport protein (ABC superfamily, ATP_bind)	UrtE	-	-	-	0.718682	+	+	+	+	+

THICB2_v3_370051	1614852	1615700	CDS	branched-chain amino acid transport protein (ABC superfamily, atp_bind)	urtD	-	-	-	0.737107	+	+	+	+	+
THICB2_v3_370052	1615693	1616826	CDS	branched-chain amino acid permease protein (ABC superfamily, membrane)	urtC	-	-	-	0.727396	+	+	+	+	+
THICB2_v3_370053	1616833	1618473	CDS	branched-chain amino acid permease protein (ABC superfamily, membrane)	urtB	-	-	-	0.738714	+	+	+	+	+
THICB2_v3_370054	1618615	1619856	CDS	branched-chain amino acid transport protein (ABC superfamily, peri_bind)	urtA	-	-	-	0.767728	+	-	+	+	+
THICB2_v3_370055	1619939	1620133	CDS	conserved protein of unknown function	_	-	-	-	0.465632	+	-	-	-	-
THICB2_v3_370056	1620201	1620503	CDS	urease gamma subunit	ureA	-	-	-	0.829737	+	-	+	+	-
THICB2_v3_370057	1620527	1620832	CDS	urease beta subunit	ureB	-	-	-	0.635517	+	-	+	+	-
THICB2_v3_370058	1620832	1622571	CDS	urease alpha subunit	ureC	-	-	-	0.780676	+	-	+	+	-
THICB2_v3_370059	1622595	1623227	CDS	Lysine exporter protein (LYSE/YGGA)	lysE	-	-	-	0.529893	+	+	+	+	+
THICB2_v3_370060	1623322	1623705	CDS	death-on-curing protein Doc	doc	-	-	-	0.549416	+	+	+	+	+
THICB2_v3_370061	1623702	1623926	CDS	conserved protein of unknown function	_	-	-	-	0.68384	+	+	+	+	+
THICB2_v3_370062	1624028	1624552	CDS	Urease accessory protein UreE	ureE	-	-	-	0.654117	+	+	+	+	+
THICB2_v3_370063	1624491	1625246	CDS	Urease accessory protein UreF	ureF	-	-	-	0.629131	+	-	+	+	+
THICB2_v3_370064	1625243	1625920	CDS	urease accessory protein	ureG	-	-	-	0.767491	-	-	+	+	-
THICB2_v3_370065	1625924	1626574	CDS	High-affinity nickel transport protein (fragment)	_	-1SD	-	-	0.618839	-	-	+	+	+

Résumé

Les liens entre la diversité et l'adaptation des populations bactériennes à leur environnement constituent une problématique importante en écologie microbienne. L'accès à de nombreux génomes permet aujourd'hui des avancées intéressantes dans ce domaine. Plusieurs souches du genre *Thiomonas* et appartenant à la même espèce, *Tm. arsenitoxydans* 3As et *Tm.* spp. CB1, CB2, CB3 et CB6, ont été isolées d'un drainage minier acide (DMA) à Carnoulès (Gard). La comparaison de leur génome a permis d'affiner leur phylogénie et de mettre au jour des différences de contenu génétique liées à des îlots génomiques. Certaines de ces différences ont été corrélées expérimentalement avec des différences fonctionnelles concernant l'oxydation de l'arsénite (As(III)), la dégradation de l'urée et la biosynthèse de biofilm, et confèrent potentiellement un avantage sur le site (meilleure résistance à l'As(III), précipitation des métaux et augmentation du pH, protection des cellules). La comparaison de la synténie des génomes de *Tm. arsenitoxydans* 3As avec *Tm.* sp CB2 et *Tm. intermedia* K12 (non isolée de ce DMA) a montré que le génome de *Tm.* sp. CB2 a subi plusieurs remaniements importants. Ces types de réarrangements pourraient être en partie à l'origine de l'apparition de variants "super-résistants" à l'As(III) dans la population. En particulier, plusieurs copies d'un élément intégratif et conjugatif portant l'opéron *aioBA* codant l'arsénite oxydase ont été détectées chez deux variants, ce qui pourrait en partie expliquer leur résistance accrue à l'As(III). La proportion de variants est plus importante en présence d'As(III) au sein des biofilms, et des données de transcriptomique ont en effet montré que ces remaniements seraient en partie causés par des systèmes de réparation de l'ADN suite à des dommages liés au stress oxydant induit par l'As(III). Le développement en biofilm et la présence d'As(III) modulerait donc la flexibilité génomique et le potentiel adaptatif de *Tm.* sp. CB2. Ces données suggèrent que cette souche, avec *Tm.* sp. CB3, possèdent un génome visiblement plus flexible que les autres. Les souches *Tm. arsenitoxydans* 3As et *Tm.* spp. CB1 et CB6, contrairement à *Tm.* spp. CB2 et CB3, forment un groupe distinct d'un point de vue phylogénétique et fonctionnel ("groupe 3As"), suggérant qu'elles occupent une niche écologique spécifique et pourraient constituer un écotype. La population de *Thiomonas* de ce DMA serait donc composée d'au moins un écotype stable et de souches au génome plus instable, dont le potentiel adaptatif plus important serait influencé par l'As(III) et le développement en biofilm.

Abstract

Understanding the link between diversity and adaptation in natural bacterial populations represents an important issue in microbial ecology. The amount of whole genome sequencing data currently available has allowed for interesting advances in this field. Several strains from the genus *Thiomonas* belonging to the species, *Tm. arsenitoxydans* (3As) and *Tm.* spp. (CB1, CB2, CB3 and CB6), were isolated from the acid mine drainage (AMD) at Carnoulès (Gard, France). Comparison among genomes allowed for a better definition of their phylogenetic relationships and highlighted differences in genetic content, which is essentially due to the presence of genomic islands. Some of these differences were experimentally correlated with functional traits concerning arsenite oxidation, urea degradation, and biofilm biosynthesis, and are potentially beneficial *in situ* (leading to an enhanced resistance to arsenite (As(III)), metal precipitation, and an increase in pH, ultimately protecting cells). The comparison of genome synteny of *Tm. arsenitoxydans* 3As, *Tm.* sp CB2, and *Tm. intermedia* K12 (not isolated from this AMD) show several important genomic rearrangements exist in *Tm.* sp CB2. This type of rearrangements could be involved in the emergence of arsenite "super-resistant" variants in the *Tm.* sp CB2 population. In particular, several copies of an integrative and conjugative element (ICE) containing the *aioBA* operon coding arsenite oxidase were detected in the genomes of two variants, which could explain their higher levels of resistance to As(III). The percentage of variants in biofilm culture is higher when grown in the presence of As(III), and transcriptomic data suggests that genomic rearrangements probably occurred through DNA repair systems following damage caused by As(III) induced oxidative stress. Therefore, biofilm development and As(III) appear to allow for the adaptation of *Tm.* sp. CB2 genome flexibility and evolutionary potential. These data suggest that CB2 and *Tm.* sp. CB3 have more flexible genomes than the other strains. *Tm. arsenitoxydans* 3As and *Tm.* spp. CB1 and CB6 form both a phylogenetic and functional cluster ("3As group") suggesting that they occupy a specific ecological niche and therefore could represent an ecotype of the genus *Thiomonas*. The *Thiomonas* population from the Carnoulès AMD might therefore consist of at least one stable ecotype as well as other strains with more instable genomes, whose higher adaptive potential could be affected by As(III) and biofilm development.