



HAL
open science

Étude de la structuration des communautés procaryotes des drainages miniers acides par le séquençage haut débit

Fabien Javerliat

► **To cite this version:**

Fabien Javerliat. Étude de la structuration des communautés procaryotes des drainages miniers acides par le séquençage haut débit. Biodiversité et Ecologie. Université de Pau et des Pays de l'Adour, 2014. Français. NNT: . tel-01111065

HAL Id: tel-01111065

<https://theses.hal.science/tel-01111065>

Submitted on 29 Jan 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

N° d'ordre :

UNIVERSITÉ DE PAU ET DES PAYS DE L'ADOUR
École doctorale des sciences exactes et de leurs applications



THÈSE

Étude de la structuration des
communautés procaryotes des
drainages miniers acides par le
séquençage haut débit

Auteur :

Fabien Javerliat

*Une thèse soumise pour l'obtention du
grade de Docteur en Microbiologie*

Soutenue publiquement le 3 Novembre 2014, devant le jury composé de :

Membres du jury :

Pr. Pascale BAUDA (Président, rapporteur)

Laboratoire Interdisciplinaire des Environnements Continentaux –Université de Lorraine– Metz

Dr. Fabienne BATTAGLIA-BRUNET (rapporteur)

Unité Biogéochimie Environnementale –BRGM, Environnement et Procédés– Orléans

Dr. Frantz DEPAULIS (examinateur)

Ecology & Evolution laboratory –Ecole Normale Supérieure– Paris

Directeurs de thèse

Pr. Béatrice LAUGA

Équipe environnement et microbiologie –Université de Pau et des pays de l'Adour– Pau

Pr. Robert DURAN

Équipe environnement et microbiologie –Université de Pau et des pays de l'Adour– Pau

Remerciements

Ces travaux ont été réalisés dans l'Équipe environnement et microbiologie de l'Institut des sciences analytiques et de physico-chimie pour l'environnement et les matériaux. Je remercie la DGA d'avoir financé ma thèse, ainsi que la FRB pour avoir financé le projet MIGRAMD.

Je remercie également les rapporteurs de travail de thèse, Pascale Bauda et Fabienne Battaglia-Brunet, d'avoir acceptées d'évaluer ces travaux et pour leur commentaires constructifs. Je remercie aussi Frantz Depaulis d'avoir accepté de faire parti du jury comme examinateur.

Je remercie mes directeurs de thèse, Béatrice Lauga et Robert Duran, tout d'abord pour m'avoir permis de réaliser mes travaux dans les meilleurs conditions. Je les remercie aussi pour les rencontres qu'ils m'ont permis de faire au travers des collaborations, invités et congrès. Enfin de façon globale pour avoir participé à ma maturation scientifique.

Un grand merci aux personnes avec qui j'ai eu l'occasion de travailler. Solange qui était "l'archiviste" du labo qui a fait beaucoup pour que je puisse travailler le plus rapidement possible. Je remercie aussi Claire pour m'avoir suivie sur les T-RFLP. Un grand merci à Christian Paroissin pour son apport mathématique précieux, sans qui une bonne partie de cette thèse n'existerait pas. Je remercie Sandra Barantal pour m'avoir initiée et éclairée sur une science obscure "les statistiques" et nos discussions d'écologie. Enfin un grand merci à Jean Christophe pour sa disponibilité et son soutien.

Un merci particulier à mes collègues thésards qui sont devenus des amis. Une convivialité, une disponibilité et un soutien sans faille. Il y a les filles Julie, Johanne, Fanny, Arantxa, Justine, Vanessa, Mathilde, Élisabeth, Paola ; et heureusement des garçons Thomas, Théo, Boussad. Il y a des post-doc aussi Hélène et Paul, merci pour ces moments de vie partagés.

Je remercie les amis de toujours, ceux qui m'ont permis de garder les pieds sur Terre. Je remercie aussi les nouveaux amis et notamment ceux du MTP qui ont aussi su être là quand

il fallait.

Je terminerai en remerciant ma famille. Je commencerai par remercier Laure, celle qui partage ma vie et qui a donc partagée tous les bons et mauvais moments de ces 3 dernières années. Elle a été mon pilier et mon point de repère me soutenant sans faille. Je souhaite aussi remercier mes parents pour m'avoir porté jusque là. Enfin je remercierai mes sœurs pour le réconfort qu'elles m'ont apportées.

Table des matières

| | |
|---|-----------|
| Remerciements | v |
| Table des matières | vii |
| Introduction générale | 3 |
| I Exposé bibliographique et procédures expérimentales | 7 |
| 1 Exposé bibliographique | 9 |
| 1.1 Le drainage minier acide | 9 |
| 1.1.1 Contexte géologique des drainages miniers acides | 9 |
| 1.1.2 Formation des drainages miniers acides | 10 |
| 1.1.3 Facteurs abiotiques influençant la formation du DMA | 12 |
| 1.1.4 Facteurs biotiques influençant la formation du DMA | 13 |
| 1.2 Les communautés microbiennes dans les DMA | 15 |
| 1.2.1 Des producteurs primaires et secondaires | 16 |
| 1.2.2 Le soufre et le fer comme source d'énergie | 23 |
| 1.2.3 Transformation de l'arsenic par les procaryotes | 28 |
| 1.2.4 Les <i>Archaea</i> dans les DMA | 31 |
| 1.2.5 Les interactions au sein des communautés microbiennes des DMA . . | 33 |
| 1.3 L'assemblage des communautés : une question majeure de l'écologie | 37 |
| 1.4 Caractérisation des communautés par le séquençage haut débit | 38 |
| 1.4.1 Le pyroséquençage 454 | 39 |
| 1.4.2 Analyse des données issues du séquençage | 40 |
| 2 Procédures expérimentales | 47 |
| 2.1 Échantillonnage | 47 |
| 2.2 Mesures physico-chimiques | 49 |
| 2.3 Extraction ADN | 49 |
| 2.4 Réaction de polymérisation en chaine | 50 |
| 2.5 Pyroséquençage | 53 |
| 2.6 Construction des banques de clones et séquençage Sanger | 53 |
| 2.7 Traitement des données brutes issues du séquençage | 54 |
| 2.7.1 Données issues du pyroséquençage | 54 |
| 2.7.2 Données issues du séquençage Sanger | 55 |

| | | |
|-------|--|----|
| 2.8 | Construction des OTU | 56 |
| 2.8.1 | Construction des OTU des séquences ARNr 16S | 56 |
| 2.8.2 | Construction des OTU des séquences des transporteurs d'arsenic | 57 |
| 2.9 | Analyse des données de séquençage | 57 |
| 2.9.1 | Analyse des données ARNr 16S | 57 |
| 2.9.2 | Analyse des données des transporteurs d'arsenic | 58 |
| 2.10 | Fouille de données | 58 |
| 2.11 | Construction du réseau | 59 |

II Principaux résultats 61

3 In pursuit of drivers of communities assemblage in AMD : a biogeographic approach 63

| | | |
|-----|-------------------------------------|-----|
| 3.1 | Introduction | 63 |
| 3.2 | Contribution scientifique | 64 |
| 3.3 | Conclusion | 107 |

4 Structure des communautés *Archaea* dans les DMA 109

| | | |
|-------|--|-----|
| 4.1 | Introduction | 109 |
| 4.2 | Résultats et discussion | 110 |
| 4.2.1 | Diversité et structure des communautés dans les DMA | 110 |
| 4.2.2 | Taxa <i>Archaea</i> retrouvés dans les DMA | 113 |
| 4.2.3 | Spécificité des <i>Archaea</i> des DMA | 115 |
| 4.2.4 | Communautés <i>Archaea</i> structurées par les paramètres environnementaux 118 | |
| 4.3 | Conclusion | 118 |

5 Diversité des transporteurs d'arsenic dans les DMA 123

| | | |
|-------|---|-----|
| 5.1 | Introduction | 123 |
| 5.2 | Résultats et discussion | 124 |
| 5.2.1 | Phylogénie des séquences ArsB des DMA | 125 |
| 5.2.2 | Phylogénie des séquences Acr3.1 des DMA | 127 |
| 5.2.3 | Phylogénie des séquences Acr3.2 des DMA | 129 |
| 5.3 | Conclusion | 131 |

Conclusion générale 137

Bibliographie 143

Table des figures 153

Liste des tableaux 155

Liste des abréviations 157

Liste des symboles 159

| | | |
|--|--|-----|
| Annexe A | | |
| Production scientifique | | 161 |
| Annexe B | | |
| Script commenté du traitement des séquences d'ARNr 16S | | 179 |
| Annexe C | | |
| Scripts de soumission sur les clusters de calcul | | 181 |
| Annexe D | | |
| Scripts Biopython pour la fouille de données | | 183 |
| Annexe E | | |
| Scripts R utilisés pour l'analyse des données | | 189 |

Introduction générale

L'ÉCOLOGIE est la science des interactions entre les organismes et leur environnement ainsi que des organismes entre eux¹. L'écosystème est l'unité fonctionnelle en écologie. Elle est définie par son biotope qui est son environnement géologique, édaphique, hydrologique, climatique et sa biocénose qui représente l'ensemble des êtres vivants de l'écosystème. Ces deux compartiments sont dynamiques et interagissent. Les êtres vivants s'organisent en communautés qui évoluent dans le temps et l'espace. L'étude des processus structurant les communautés d'organismes est donc une question fondamentale d'écologie. De plus, dans un contexte de changement climatique global et de perte de biodiversité, il est important de comprendre comment les communautés sont affectées et comment des mécanismes de résilience se mettent en place.

Depuis des dizaines d'années, plusieurs auteurs ont travaillé à identifier des processus régissant l'assemblage des communautés. Parmi ces travaux, on peut citer ceux de Diamond (1975) qui mettaient en avant l'importance de la compétition pour les ressources dans la détermination des espèces d'une communauté. Cependant, la méthodologie de Diamond fût controversée par Connor et Simberloff (1979) qui montrèrent que les patrons d'assemblage de communauté de Diamond pouvaient être obtenus par des assemblages aléatoires. Puis, d'autres processus ont été avancés, certains déterministes comme les interactions entre espèces, les contraintes du biotope (différentiation des niches) (Bell 2001) ou bien des processus neutres comme la dispersion des espèces à partir d'un pool régional et des événements de naissance et de mort à l'échelle locale (Hubbell 2001). Bien que ces processus aient été vus parfois comme exclusifs, il a été proposé de les considérer comme complémentaires et de les associer dans un concept de communauté intégrée (Lortie *et al.* 2004).

Cette question a été débattue depuis longtemps en microbiologie. En effet, en 1838 Ehrenberg s'appuya sur des processus stochastiques pour expliquer des résultats expérimentaux, alors que Beijerinck en 1913 plaidait en faveur de processus déterministes (de Wit et Bouvier 2006). Cependant leurs interprétations étaient basées sur des méthodes culturelles de caractérisation des micro-organismes, la grande majorité étant réfractaire à la mise en culture (Amann *et al.* 1995). Cela empêche un inventaire exhaustif des communautés microbiennes biaisant fréquemment les résultats vers l'idée que les espèces microbiennes sont ubiquistes (Green et Bohannan 2006). Actuellement le développement des méthodes non culturelles

1. Ernst Haeckel (1866), *Generelle Morphologie der Organismen*, Berlin, Reimer

basées sur la biologie moléculaire, le séquençage haut-débit d'ADN et la puissance du traitement informatique, repoussent les limites de la microbiologie. La possibilité d'obtenir une description quasiment exhaustive d'une communauté est maintenant offerte. Récemment, des avancées formidables ont été faites sur l'influence des processus déterministes et stochastiques dans l'assemblage des communautés procaryotes (Horner-Devine et Bohannan 2006, Pontarp *et al.* 2012, Lekberg *et al.* 2012). Cependant, les interactions entre les espèces sont très peu prises en compte (Chaffron *et al.* 2010, Barberán *et al.* 2012, Eiler *et al.* 2012). Le projet **MIGRAMD** (**M**icrobial **bio**geo**G**raphy of **A**cid **M**ine **D**rainage) qui encadre les travaux présentés dans ce mémoire, s'inscrit dans cette dynamique en étudiant l'impact des fortes contraintes imposées par les drainages miniers acides (DMA) sur les communautés microbiennes.

Les travaux réalisés au cours de cette thèse sont une étude comparative de douze sites échantillonnés dans des drainages miniers acides de trois régions du monde –le sud-est de la France, le sud-ouest de la péninsule ibérique et la Bolivie– et explorent deux niveaux de biodiversité. Le premier est une étude de la diversité des taxons dans les communautés procaryotes des DMA en utilisant comme marqueur moléculaire le gène de l'ARNr 16S. L'accès à ce gène par amplification PCR et séquençage parallèle haut-débit par la technologie de pyroséquençage 454 des produits d'amplification, permet un inventaire exhaustif des communautés. La prise en compte des conditions physico-chimiques des drainages étudiés et leur localisation géographique nous permet d'explorer les mécanismes qui structurent les communautés dans les DMA. Dans une vision plus globale nous souhaiterions renseigner le concept de Baas Becking sur la dispersion des micro-organismes « Everything is everywhere, but the environment selects » (1934)².

Le deuxième niveau de biodiversité étudié est celui de gènes codant pour des protéines ayant une fonction très liée aux conditions des DMA. La concentration en arsenic, qui est un élément toxique, est une forte pression environnementale sur les communautés microbiennes. Nous avons choisi d'étudier les gènes qui codent pour des transporteurs membranaires d'arsenic : ArsB et Acr3, qui par leur fonction critique pour la survie en présence d'arsenic, sont supposés être portés par tous les micro-organismes des DMA. En étudiant la diversité

2. Baas-Becking, Lourens G.M. (1934), *Geobiologie of inleiding tot de milieukunde*, The Hague, the Netherlands : W.P. Van Stockum & Zoon

de ces gènes, nous souhaitons mettre en évidence des patrons de partage de gènes entre les micro-organismes des communautés. Nous souhaitons aussi observer des indices de partage ou de spécificité des individus des communautés entre les différents sites des DMA. Ces observations peuvent se faire au niveau local ou régional renseignant sur la dispersion et la colonisation des DMA par les procaryotes.

Ce manuscrit est composé de deux parties contenant plusieurs chapitres. La première partie comporte deux chapitres. Le premier est un état des connaissances sur le drainage minier acide, notre environnement d'étude, et les communautés procaryotes qui habitent ce milieu. Nous développerons aussi les théories qui entourent l'assemblage des communautés microbiennes, ainsi que l'utilisation de la technologie de séquençage d'ADN haut-débit 454 pour la caractérisation de ces communautés. Le deuxième chapitre contient les procédures expérimentales mises en place pour réaliser les études développées dans la deuxième partie. La deuxième partie comporte trois chapitres, chacun développant des résultats expérimentaux. Le premier chapitre traite de l'identification des paramètres structurant les communautés procaryotes dans les DMA. Le deuxième chapitre est une étude descriptive des communautés *Archaea* dans les DMA car le rôle des micro-organismes de ce domaine est peu documenté, malgré les indices indiquant leur importance. Le troisième chapitre est une étude de la diversité des transporteurs membranaires d'arsenic ArsB et ACR3, et tente d'observer des patrons de mélange de communautés entre les DMA et des indices d'échange de ce gène. Enfin ce manuscrit se terminera par une conclusion générale sur ces travaux où les principaux résultats seront discutés.

Première partie

Exposé bibliographique et procédures expérimentales

Chapitre 1

Exposé bibliographique

1.1 Le drainage minier acide

Le drainage minier acide (DMA) est le modèle d'étude de ces travaux. Cette section décrit cet environnement, son caractère extrême et l'importance des micro-organismes dans sa création et son maintien. Ces caractéristiques font de lui un bon candidat pour répondre à des questions sur la structuration des communautés microbiennes. En effet, la faible diversité (Baker et Banfield 2003) permet de faire un échantillonnage approfondi de la communauté plus facilement. Une richesse plus faible rend aussi plus simple l'appréhension des interactions au sein de la communauté.

1.1.1 Contexte géologique des drainages miniers acides

Les résidus miniers contiennent des sulfures métalliques (Table 1.1) (Blowes *et al.* 1994). Lorsque ces roches sont exposées à l'oxygène, le soufre s'oxyde et produit de l'acide sulfurique. Des fluides acides et riches en métaux se forment lorsque ces roches sont à l'extérieur et exposées aux intempéries (Baker et Banfield 2003). Ce phénomène a lieu naturellement et l'on parle alors de Drainage Acide de Roche (DAR). Cependant, l'excavation minière amplifie le phénomène qui est alors défini par Drainage Minier Acide (DMA). En effet, ces activités placent brutalement en conditions oxydantes des grandes quantités de minéraux sulfurés. Ceci accélère l'oxydation des minéraux qui se fait par des processus chimiques et biologiques, amplifiant l'acidification des effluents et leur concentration en métaux et

TABLE 1.1 – Principaux sulfures métalliques retrouvés dans les résidus miniers classés par ordre croissant de réactivité (Blowes *et al.* 1994).

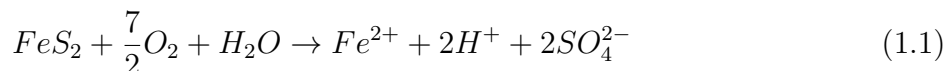
| Minéraux | Formule Chimique |
|--------------|---------------------------|
| Chalcopyrite | CuFeS_2 |
| Arsénopyrite | FeAsS |
| Pyrite | FeS_2 |
| Marcasite | FeS_2 |
| Millerite | NiS |
| Sphalérite | ZnS |
| Chalcocite | CuS_2 |
| Galène | PbS |
| Pyrrhotite | Fe_{1-x}S |

métalloïdes toxiques. Le terme DMA englobe toutes les eaux acides d'origine minière issues de l'oxydation spontanée des minéraux sulfurés : les eaux de galeries, les eaux présentes au fond des mines à ciel ouvert et les eaux de percolation des piles de stockage des matériaux excavés. Les déchets miniers sont un important problème pour l'industrie minière. Pour extraire une ressource minérale, une société minière se doit de déterminer comment traiter les déchets dans le processus d'exploitation. La prise de conscience environnementale s'étant accrue, cette tâche est devenue de plus en plus ardue. Cependant, la prévention de l'impact de ces déchets est à prendre en compte dans l'évaluation initiale des ressources, le développement et la fermeture de la mine. Des actions inappropriées pouvant provoquer des surcoûts encore plus importants, la planification pour le traitement des déchets semble être la meilleure stratégie de gestion (Geology 2003).

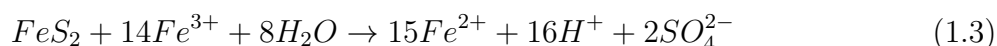
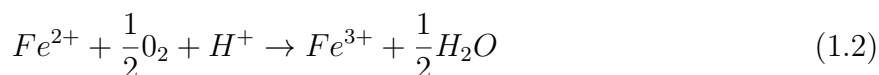
1.1.2 Formation des drainages miniers acides

La pyrite (FeS_2) est un minéral sulfuré très commun dans les drainages miniers. C'est pour cela qu'il est utilisé pour illustrer les réactions chimiques qui mènent à la formation du DMA. Les mécanismes de l'oxydation de la pyrite sont largement connus (Singer et Stumm 1970, Lawson 1982). Une oxydation abiotique a lieu et se produit spontanément, mais aussi une oxydation biotique qui se réalise en présence de microorganismes. L'oxydation de la pyrite peut être réalisée par l'oxygène seulement (oxydation directe), ou par l'oxygène et le fer (oxydation indirecte). Á un pH supérieur à 4,5, la pyrite subit surtout une oxydation

directe et lente telle que la décrit l'équation suivante 1.1.



Du fer ferreux Fe(II) est produit ainsi que des protons, ce qui diminue le pH s'il n'y a pas de minéraux alcalins. Lorsque le pH se trouve entre 4,5 et 3,5 l'oxydation de la pyrite par l'équation 1.1 est accélérée par une baisse de l'énergie d'activation de cette réaction. Puis si le pH du milieu est inférieur à 3,5, le fer ferreux sera oxydé en fer ferrique Fe(III) selon la réaction 1.2. Le fer ferrique va oxyder directement la pyrite (oxydation indirecte) par la réaction 1.3.

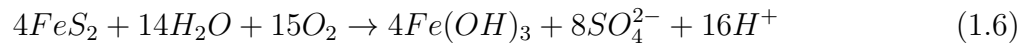


L'oxydation indirecte génère aussi du fer ferreux qui sera à son tour oxydé selon l'équation 1.2. Ainsi les réactions 1.2 et 1.3 constituent un cycle continu d'oxydation du Fe^{2+} en Fe^{3+} , ce qui permet théoriquement une oxydation de la pyrite jusqu'à épuisement des minéraux. À un pH de 3, l'oxydation indirecte de la pyrite par la réaction 1.3 est de 10 à 100 fois plus rapide que l'oxydation directe par l'oxygène (réaction 1.1) . Pour arrêter cette réaction, il faudrait couper l'apport en oxygène ou bien augmenter le pH. En effet, la concentration en fer est directement contrôlée par le pH. La solubilité du Fe^{3+} est très faible à pH neutre et alcalin, car il précipite sous forme d'hydroxyde ($Fe(OH)_3$) et d'oxyhydroxyde ($FeOOH$). Ainsi si le pH est supérieur à 3, ces réactions d'hydrolyse ont lieu :

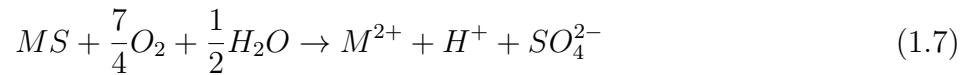


Ces précipités sont de couleur brun-jaune, colorent l'eau et forment ces dépôts caractéristiques des DMA. Cependant ces hydrolyses produisent des protons, rendant la précipitation du fer provisoire, le temps que la solution redevienne suffisamment acide pour poursuivre l'oxydation du fer ferrique en fer ferreux, et l'oxydation indirecte de la pyrite. De façon

globale, on peut résumer la réaction d'oxydation de la pyrite par cette équation :



Nous nous sommes basés sur l'exemple de la pyrite, or d'après le tableau 1.1, d'autres minéraux subissent également ces réactions et libèrent leurs métaux. On peut citer la chalcopryrite ($CuFeS_2$), l'arsénopyrite ($FeAsS$) et la sphalérite (ZnS). Il faut aussi considérer que ces minéraux peuvent contenir d'autres éléments sous forme de traces ou de substitution atomique : Ni, As, Co dans la pyrite; Fe, Mn, Cd, Hg dans la sphalérite. Ces éléments passeront en solution et affecteront la composition du DMA. L'équation 1.1 peut être généralisée en remplaçant le fer par M, représentant un autre métal ou métalloïde lié à un sulfure :



1.1.3 Facteurs abiotiques influençant la formation du DMA

pH

Les équations précédentes montrent que le pH est un paramètre important contrôlant la formation du DMA, car il est critique pour la dissolution du fer. De plus il va conditionner la croissance des micro-organismes, dont l'implication dans le contrôle du DMA sera montrée dans la sous-section 1.1.4.

Oxygène

L'oxydation de la pyrite peut avoir lieu aussi bien à l'air libre que dans l'eau. Cet élément est indispensable pour que la réaction se fasse. Il existe même une forte corrélation entre le taux d'oxydation et la concentration en oxygène du milieu. Enfin, la disponibilité et le transport de l'oxygène sont considérés comme les étapes limitantes de la production du DMA par Ritchie (1994).

Température

L'oxydation des minéraux sulfurés est une réaction exothermique. La chaleur générée par la réaction va se dissiper à l'intérieur des minéraux. Ils sont très peu thermo-conducteurs et leur température va augmenter. Cette chaleur favorise l'oxydation de la pyrite par les micro-organismes, qui pour la plupart sont mésophiles ou thermophiles.

Teneur en eau

Le rôle de l'eau dans la formation du DMA ne fait pas consensus. Elle est nécessaire au développement des micro-organismes, mais elle limite aussi la disponibilité de l'oxygène pour le minéral. Finalement, certains la considèrent comme essentielle dans l'oxydation de la pyrite (Rose et Cravotta 1999), et d'autres comme un milieu réactionnel (Stumm et Morgan 1981).

Taille des particules, porosité

Plus la pyrite exposera de surface à l'oxygène ou aux micro-organismes plus son oxydation sera rapide.

Cristallographie

La présence d'impuretés dans le réseau cristallin de la pyrite peut fragiliser le minerai et le rendre plus vulnérable aux attaques chimiques. Par exemple l'arsenic réduit la résistance à l'oxydation de la pyrite (Plumlee *et al.* 1999).

Effets galvaniques entre sulfures

En plus de la pyrite, d'autres sulfures sont présents dans les drainages. Les contacts directs entre au moins deux sulfures différents peuvent entraîner des déplacements d'électrons de l'un vers l'autre. Le sulfure le plus électronégatif résistera mieux à l'oxydation en rendant l'autre plus sensible.

1.1.4 Facteurs biotiques influençant la formation du DMA

Les micro-organismes présents dans les DMA ont un rôle primordial dans leur formation et leur maintien. En effet, certains obtiennent l'énergie nécessaire à leur développement par des réactions d'oxydo-réduction d'éléments métalliques. Ils agissent comme des catalyseurs des réactions chimiques ayant lieu à la surface des minéraux et en solution. Nous avons vu que la vitesse d'oxydation de la pyrite dépend de la disponibilité en Fe^{3+} . Cependant à pH acide ($pH < 4$), la réaction 1.2 est très lente voire négligeable (Stumm et Morgan 1981). À de tels pH les micro-organismes acidophiles ferro-oxydants peuvent déterminer le taux de dissolution de la pyrite (Baker et Banfield 2003), en augmentant la vitesse d'oxydation du fer jusqu'à 10^5 (Singer et Stumm 1970). L'augmentation de la production de fer, par la dissolution de la pyrite, induite par *Leptospirillum ferriphilum* a été déterminée par Edwards *et al.* (1998). Elle est semblable à celle mesurée chez *Acidithiobacillus ferrooxidans* par Santelli *et al.* (2001).

C'est à partir de ces mesures de dissolution de la pyrite, normalisés par cellule, que Edwards *et al.* (2000) estimèrent l'activité microbienne dans les DMA et qu'elle représentait environ 75 % de la production de cet écosystème.

Cependant les mécanismes par lesquels la pyrite est attaquée par les micro-organismes font débats. Baker et Banfield (2003) expliquèrent les différentes théories à ce sujet : la pyrite peut être attaquée par des mécanismes directs et indirects. Les voies indirectes illustrent l'augmentation de la dissolution de la pyrite induite par l'oxydation du Fe^{2+} soluble. La catalyse "directe" est une attaque enzymatique directe des ions liés à la surface de la pyrite. L'oxydation enzymatique de la pyrite nécessite que l'organisme soit très proche voire attaché à la surface de la pyrite. Plusieurs travaux qui montrent que la fixation des micro-organismes au minéral favorise sa dissolution, alimentent les spéculations en faveur de la voie directe. L'*Archaea Acidianus brierleyi* a besoin d'être très proche de la pyrite pour avoir une croissance et un taux d'oxydation optimum (Larsson *et al.* 1993). De même des *Ferroplasma acidarmanus* retrouvés dans des puits profonds à la surface d'arsenopyrite indiquent une meilleure dissolution à proximité des cellules (Edwards *et al.* 2001). Cependant des expériences de microscopie électronique ont montré que les creux formés en présence de *A. ferrooxidans* et *F. acidarmanus*, à la surface de la pyrite sont aussi observés lors des incubations abiotiques avec du Fe^{3+} seulement (Edwards *et al.* 2001). Ceci indique que la présence de creux à la surface du minéral n'est pas forcément un indice d'oxydation directe. Fowler *et al.* (1999) ont mené des expériences en réacteur à flux continu avec des concentrations constantes en Fe^{2+} et Fe^{3+} , en présence ou non d'*A. ferrooxidans*. Ceci pour déterminer si cet organisme augmente la dissolution de la pyrite autrement qu'en augmentant la concentration en Fe^{3+} du milieu. Il s'est avéré que le taux de dissolution de la pyrite était plus élevé en présence de l'organisme. Ce qui implique la présence d'enzymes de lixiviation ou une accumulation de Fe^{3+} à la surface de la cellule. Enfin dans leurs travaux Edwards *et al.* (1998; 1999; 2000), montrent que la fixation des bactéries n'est pas aléatoire mais spécifique au minéral. Ils suggèrent aussi l'existence d'un chimiotactisme envers un gradient de concentration en Fe^{2+} . Même si la catalyse directe semble être plus efficace que les catalyses indirectes (Zhang et Fang 2005), le contact avec le minéral n'est pas nécessaire et dépend de l'organisme ainsi que du minéral. En plus des mécanismes directs et indirects, l'utilisation de molécules organiques solubles comme navette d'électrons entre la surface de la pyrite et les

enzymes de la cellule est possible. L'accumulation de Fe^{3+} à la surface des cellules peut aussi accroître localement la dissolution de la pyrite sans action enzymatique (Baker et Banfield 2003).

Certains procaryotes des DMA sont capables d'oxyder les composés soufrés réduits (RISC) produits par la dissolution des minéraux soufrés (sous-section 1.2.2). Les produits de cette oxydation sont notamment des ions sulfates et des protons, contribuant au maintien de l'acidité dans les DMA.

1.2 Les communautés microbiennes dans les DMA

De par la nature du milieu dans lequel ils vivent, les micro-organismes des DMA sont acidophiles. On peut distinguer les acidophiles extrêmes ayant une croissance optimale à $\text{pH} < 3$ des acidophiles modérés ayant une croissance optimale à $\text{pH} 3-5$. Les formes de vie multicellulaires complexes sont généralement absentes des milieux extrêmement acides, à l'exception de quelques macrophytes observés dans des lacs acides (Fyson 2000). Des Eucaryotes sont cependant présents dans les DMA : des rotifères, des protozoaires (*Euglena mutabilis*) et des levures (Johnson et Hallberg 2008, Gómez *et al.* 2002). Les principaux êtres vivants peuplant les DMA sont donc des micro-organismes procaryotes : *Bacteria* et *Archaea* (Table 1.2). Deux caractéristiques physiologiques sont vitales pour la survie des micro-organismes à des pH bas. La première est la capacité à résister à l'agression des protons (H_3O^+), même avec des différences de concentration de l'ordre de $10^4 - 10^6$ entre le cytoplasme et le milieu extérieur. Ce gradient de protons est utilisé pour générer de l'ATP (adénosine 5'-triphosphate) grâce à des ATPases membranaires, cependant le flux de protons doit être contrôlé par un apport d'électrons (e^-) (apport pouvant venir de l'oxydation du fer ferreux). La seconde est la différence de potentiel de membrane qu'il existe entre les acidophiles et les neutrophiles. Les premiers ayant un potentiel de membrane positif, dû à un import actif de cations (par exemple K^+), alors que celui de ces derniers est négatif (Albers *et al.* 2001). Ce potentiel de membrane protège des cations métalliques qui sont en concentration importante, mais rend les acidophiles sensibles à des anions comme les nitrates (Norris et Ingledew 1992). On peut caractériser ces acidophiles en fonction de leur capacités physiologiques, comme le type trophique :

- les chimiolithotrophes (chimioolithoautotrophe) : utilisent du carbone et des donneurs d' électrons inorganiques
- les hétérotrophes (chimioorganohétérotrophe) : utilisent du carbone et des donneurs d' électrons organiques
- les mixotrophes (chimioolithohétérotrophe) : utilisent du carbone organique et des donneurs d' électrons inorganiques.

Cependant, c'est une vision simplifiée de la réalité, car les micro-organismes sont capables d'exprimer des métabolismes variés en fonctions des conditions environnementales. C'est ce que nous découvrirons au cours de cette section. On peut aussi les séparer en fonction de leur température optimale de croissance :

- les mésophiles ont une croissance optimale à moins de 40 °C
- les thermophiles modérés ont une croissance optimale entre 40 et 60 °C
- les thermophiles extrêmes ont une croissance optimale à plus de 60 °C.

Les mésophiles sont généralement des bactéries Gram-négatif alors que les thermophiles modérés sont des Gram-positif et les thermophiles extrêmes des archées. Il existe une corrélation positive entre le pH et la température optimale de croissance chez les bactéries Gram-négatif et les archées. Cela veut dire que la résistance à la température diminue la résistance à un pH faible et inversement. A titre d'exemple, le plus thermophile des acidophiles connus peut croître jusqu'à 90 °C (*Acidianus infernus*) alors que le plus thermophile des neutrophiles croît jusqu'à 120 °C (*Pyrolobus spp.*) (Johnson et Hallberg 2008). Dans les DMA nous aurons affaire plutôt à des mésophiles et des thermophiles modérés (Johnson et Hallberg 2003), alors que les thermophiles extrêmes seront plutôt observés dans les eaux acides thermales.

1.2.1 Des producteurs primaires et secondaires

Comme le montre le tableau 1.2, certains procaryotes sont des autotrophes obligatoires. Ces organismes sont les producteurs primaires des DMA, ils vont fixer le carbone inorganique (CO_2) pour le transformer en carbone organique utilisable par les hétérotrophes. En absence de lumière la photosynthèse n'est pas possible, la source d'énergie des autotrophes acidophiles provient donc de l'oxydation (chimio) du fer et du soufre (litho) d'où le terme chimiolithoautotrophe. Cependant la fixation du CO_2 est difficile à pH acide. En effet, il est présent sous forme gazeuse qui est beaucoup moins soluble que la forme bicarbonate (HCO_3^-).

TABLE 1.2 – Espèces observées dans les DMA et leurs principales caractéristiques physiologiques. Adapté de Johnson et Hallberg (2008), Johnson et Hallberg (2003) et Volant (2012). (a) (Johnson *et al.* 2014); (b) (Hallberg 2010); (c) (Alazard *et al.* 2010).¹ Ne dégradent pas le minéral; ² banque de clones. Hétéro = hétérotrophe strict; Auto = autotrophe strict; Auto/Hétéro = autotrophe facultatif; Méso = mésophile; Mod Thermo = thermophile modéré; Ext Thermo = thermophile extrême.

| Phylum/Classe | Genre/Espèce | Réduction | | Oxydation | | Type Trophique | Classification thermique |
|-------------------------|---|------------------|-------------------------------|------------------|---|----------------|--------------------------|
| | | Fe ³⁺ | SO ₄ ²⁻ | Fe ²⁺ | S | | |
| Proteobacteria | | | | | | | |
| <i>α-proteobacteria</i> | <i>Acidiphilium spp.</i> | + | | + | | Hétéro | Méso |
| | <i>Acidiphilium acidophilum.</i> | + | | + | | Auto/Hétéro | Méso |
| | <i>Acidicaldus organivorans</i> | + | | + | | Hétéro | Mod Thermo |
| | <i>Acidocella spp.</i> ¹ | + | | | | Hétéro | Méso |
| | <i>Acidisphaera rubrifaciens</i> ¹ | + | | | | Hétéro | Méso |
| <i>β-proteobacteria</i> | <i>Ferrovum myxofaciens</i> ^(a) | | | + | | Auto | Méso |
| | <i>Gallionella ferruginea</i> | | | + | | Auto/Hétéro | Méso |
| | <i>Thiomonas spp.</i> | | | + | + | Auto/Hétéro | Méso |
| <i>γ-proteobacteria</i> | <i>Acidithiobacillus caldus</i> | | | | + | Auto | Mod Thermo |
| | <i>Acidithiobacillus ferrivorans</i> | | | + | + | Auto | Méso |
| | <i>Acidithiobacillus ferrooxidans</i> | + | | + | + | Auto | Méso |
| | <i>Acidithiobacillus thiooxidans</i> | | | + | + | Auto | Méso |
| | <i>Halothiobacillus spp.</i> ^(b) | | | | + | NA | NA |
| <i>δ-proteobacteria</i> | <i>Desulfosarcina spp.</i> ² | | + | | | NA | NA |
| | <i>Desulmonile tiedjei</i> ² | | + | | | NA | NA |
| | <i>Desulfuromonas svalbardensis</i> ² | + | + | | | NA | NA |
| | <i>Desulfurella spp.</i> ² | | + | | | NA | NA |
| | <i>Syntrophobacter spp.</i> ² | | + | | | NA | NA |
| Acidobacterium | <i>Acidobacterium capsulatum</i> ¹ | + | | | | Hétéro | Méso |
| Actinobacteria | <i>Acidimicrobium ferrooxidans</i> | + | | + | | Auto/Hétéro | Méso |
| | <i>Ferrimicrobium acidiphilum</i> | + | | + | | Hétéro | Méso |
| | <i>Ferrithrix thermotolerans</i> | + | | + | | Hétéro | Mod Thermo |
| Nitrospira | <i>Leptospirillum ferrooxidans</i> | | | + | | Auto | Méso |
| | <i>Leptospirillum ferriphilum</i> | | | + | | Auto | Méso |
| | <i>L. ferrodiazotrophum</i> | | | + | | Auto | Méso |
| Firmicutes | <i>Alicyclobacillus spp.</i> | | | | | Hétéro | Méso |
| | <i>Alb. disulfidooxidans</i> | | | | + | Auto/Hétéro | Méso |
| | <i>Alb. tolerans</i> | + | | + | + | Auto/Hétéro | Méso |
| | <i>Clostridium aciditolerans</i> | | | | | Hétéro | Méso |
| | <i>Sulfobacillus spp.</i> | + | | + | + | Auto/Hétéro | Méso/Mod Thermo |
| Crenarchaeota | <i>Desulfosporosinus acidiphilus</i> ^(c) | | + | | | Hétéro | Méso |
| | <i>Sulfolobus tokodaii</i> | | | + | + | Hétéro | Ext Thermo |
| | <i>Sulfolobus metallicus</i> | | | + | + | Auto | Ext Thermo |
| | <i>Sulfolobus acidocaldarius</i> | | | | | Hétéro | Ext Thermo |
| | <i>Sulfurisphaera ohwakuensis</i> | | + | | | Auto/Hétéro | Ext Thermo |
| | <i>Stygiolobus azoricus</i> | | + | | | Auto | Ext Thermo |
| | <i>Acidianus spp.</i> | | + | | | Auto | Ext Thermo |
| | <i>Metallosphaera spp.</i> | | | | + | Auto/Hétéro | Ext Thermo |
| Euryarchaeota | <i>Thermoplasma spp.</i> | | + | | | Hétéro | Mod Thermo |
| | <i>Picrophilus spp.</i> | | | | | Hétéro | Mod Thermo |
| | <i>Ferroplasma spp.</i> | + | | + | | Auto/Hétéro | Méso |

De plus, la solubilité du CO_2 diminue avec l'augmentation de la température. Ceci est observable en laboratoire car la croissance des autotrophes acidophiles est accélérée lorsqu'ils sont supplémentés en air (qui contient du CO_2) (Johnson et Hallberg 2008). Chez les procaryotes quatre voies de fixation du CO_2 sont reconnues :

- le cycle de Calvin-Benson-Bassham (CBB) (Calvin 1962) (Figure 1.1)
- le cycle réducteur des acides tricarboxyliques (Antranikian *et al.* 1982)
- la voie de l'acétyl-CoA (Pezacka et Wood 1986)
- le cycle du 3-hydroxypropionate (Strauss et Fuchs 1993) (Figure 1.2).

mais seuls les cycles CBB et du 3-hydroxypropionate ont été actuellement observés dans les DMA. La majorité des bactéries acidophiles autotrophes fixe le CO_2 via le cycle CBB (Fi-

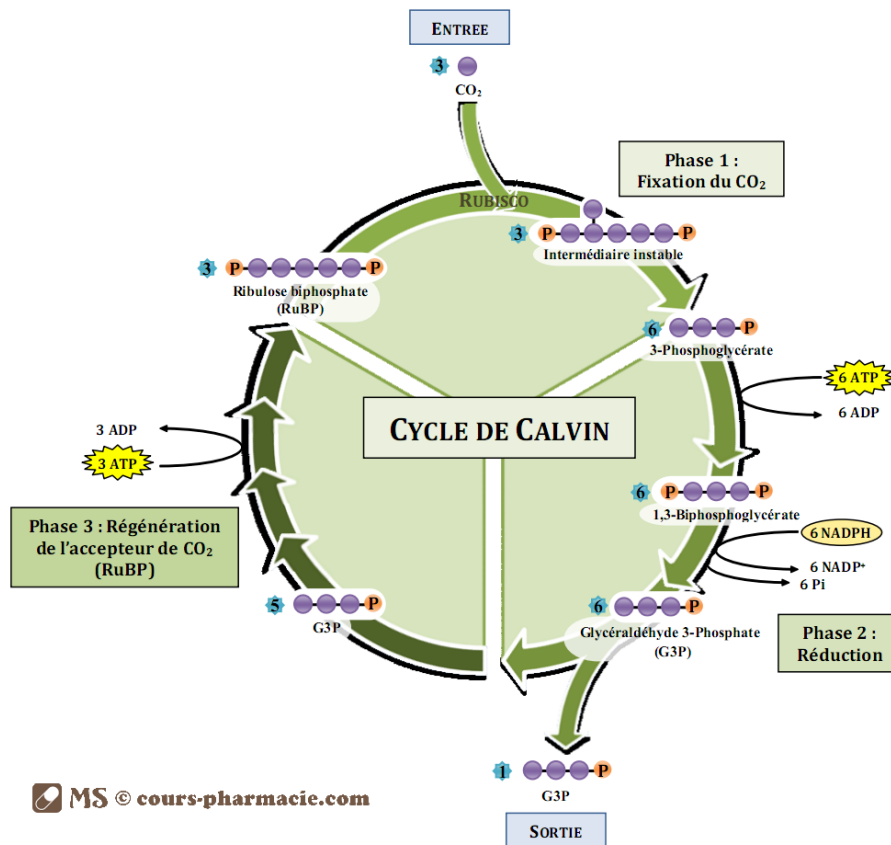


FIGURE 1.1 – Schémas représentant le cycle de Calvin.

gure 1.1) (Johnson et Hallberg 2008). L'enzyme clef de cette voie métabolique est la ribulose biphosphate carboxylase/oxygenase (RUBISCO). Cette enzyme a été identifiée chez *Acidithiobacillus thiooxidans* (Suzuki et Werkman 1958) et *At. ferrooxidans* (Gale et Beck 1967)

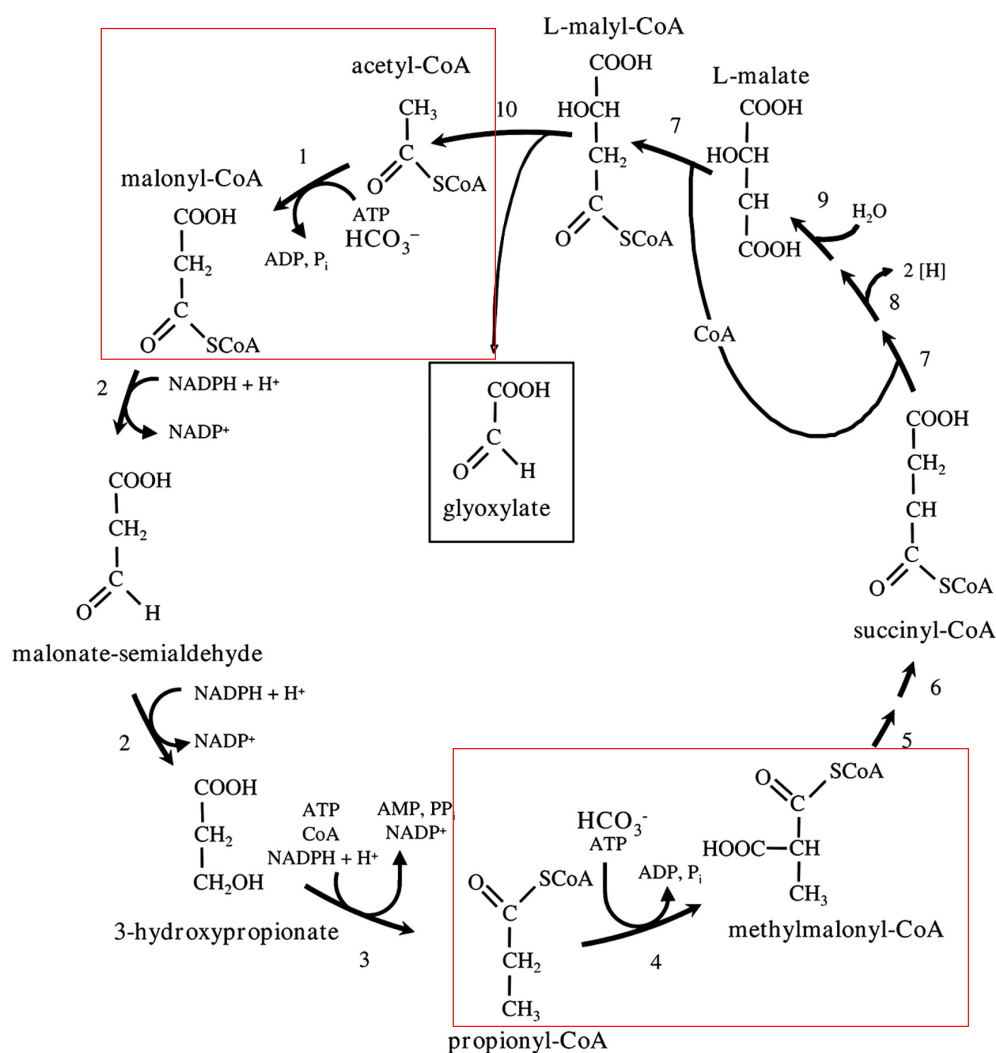


FIGURE 1.2 – Cycle du 3-hydroxypropionate de *Chloroflexus aurantiacus*. Les zones en rouge sont celles où le CO_2 est incorporé. Enzymes : 1, acétyle-CoA carboxylase ; 2, malonyl-CoA réductase (NADPH) ; 3, propionyle-CoA synthase ; 4, propionyle-CoA carboxylase ; 5, méthylmalonyl-CoA épimérase ; 6, méthylmalonyl-CoA mutase ; 7, succinyl-CoA :l-malate-CoA transférase ; 8, succinate déshydrogénase ; 9, fumarate hydratase ; 10, l-malyl-CoA lyase (Alber *et al.* 2006).

regroupés dans des organelles appelés carboxysomes (Shively *et al.* 1998). Ils augmentent l'efficacité de la fixation du CO₂ en augmentant la concentration de HCO₃⁻ dans l'organelle où il est convertie en CO₂ à proximité de la RUBISCO. Cela permet de favoriser la réaction de carboxylation de la RUBISCO sur la réaction d'oxygénation. Des comparaisons de génomes ont montré que les espèces *At. ferrooxidans*, *At. thiooxidans*, *At. caldus* possèdent toutes les enzymes nécessaires pour la fixation du CO₂ par le cycle CBB (Valdès *et al.* 2007). Les bactéries *Leptospirillum ferriphilum*, *L. ferrodiazotrophum* et *Sulfobacillus acidophilus*, *Sb. thermosulfidooxidans* fixent aussi le CO₂ via le cycle CBB (Tyson *et al.* 2004, Caldwell *et al.* 2007), de même que *Ferrovum myxofaciens* isolée récemment (Johnson *et al.* 2014). Contrairement aux bactéries, les archées fixent le CO₂ via le cycle 3-hydroxypropionate (Figure 1.2) (Johnson et Hallberg 2008). Les enzymes clefs de ce cycle : acétyl-CoA carboxylase et propionyle-CoA carboxylase ont été d'abord observés chez *Acidianus bierleyi* (Ishii *et al.* 1997). Puis de telles activités enzymatiques ont été enregistrées chez *A. infernus*, *Metallophaera sedula* et *Sulfolobus metallicus* (Menendez *et al.* 1999). Cependant il a ensuite été montré que *M. sedula* réalise ces deux activités enzymatiques avec une seule enzyme, dans un cycle de fixation du CO₂ modifié. Cette enzyme a aussi été retrouvée chez les espèces *S. solfataricus* et *S. todakii* mais pas chez *Thermoplasma acidophilum* (Hügler *et al.* 2003). Le cas des espèces de *Ferroplasma* est discuté car la souche de référence *Fp. acidiphilum* a été décrite comme autotrophe (Golyshina *et al.* 2000), et des enzymes retrouvés dans le métagénome du DMA de Iron Mountain suggèrent une fixation du CO₂ par la voie de l'acétyl-CoA (Tyson *et al.* 2004). Cependant une autre étude montre que trois souches de *Fp. acidiphilum* ainsi que l'espèce *Fp. acidarmanus* ne fixent pas le ¹⁴CO₂ (Dopson *et al.* 2004). Ces enzymes seraient plutôt impliqués dans la génération d'énergie à partir du formate lors d'une croissance organotrophe (Johnson et Hallberg 2008).

Les producteurs secondaires consomment du carbone organique. Parmi eux on retrouve des hétérotrophes stricts et des mixotrophes. Les premiers procaryotes acidophiles hétérotrophes ont été isolés près de 60 ans après le premier autotrophe *At. thiooxidans*. Les travaux de Arthur Harrison de l'Université du Missouri ont permis de caractériser un grand nombre d'acidophiles hétérotrophes, en créant notamment les milieux de cultures (Harrison 1984). Cependant même si beaucoup de données sont disponibles sur les substrats utilisés par ces organismes, il existe peu d'informations sur les mécanismes biochimiques d'utilisation de ces

substrats (Johnson et Hallberg 2008). Les espèces du genre *Acidiphilium* sont très répandues dans les DMA. Les espèces reconnues sont *Acidiphilium cryptum* qui est l'espèce type, *A. organovorum*, *A. multivorum*, *A. angustum*, *A. rubrum* et *A. acidophilum*. Elles sont hétérotrophes obligatoires à l'exception de la dernière qui est capable d'une croissance autotrophe. *A. organovorum* et *A. multivorum* peuvent croître jusqu'à de fortes densités dans des milieux ayant une concentration en glucose > 10 mM, alors que *Acidiphilium cryptum* est oligotrophe et sa croissance est inhibée à partir de 5,5 mM de glucose. Ces espèces peuvent utiliser une variété de monosaccharides, acides dicarboxyliques (acide malique) au tricarboxylique (acide citrique), mais sont inhibées par les acides monocarboxyliques. De plus *A. multivorum* est capable d'utiliser des alcools (méthanol, éthanol, propanol) (Wakao *et al.* 1994). Les espèces *A. angustum*, *A. rubrum* ont une croissance faible sur des monosaccharides, mais utilisent des acides organiques et du glycérol (Wichlacz *et al.* 1986). Les *Acidiphilium* ne semblent pas utiliser de polymères organiques comme l'amidon, le maltose, le lactose ou la cellobiose (Lobos *et al.* 1986, Wakao *et al.* 1994). Ces espèces sont aérobies, ne fermentent pas et utilisent la réduction dissimilatrice du fer quand l'oxygène est limité. Shuttleworth *et al.* (1985) ont étudié le catabolisme du glucose chez *A. cryptum* et ont conclu que les voies des pentose-phosphates et de Entner-Doudoroff étaient fonctionnelles alors que celle de Emben-Meyerhof-Parnas (voie anaérobie de la glycolyse) était absente. Les *Acidocella* sont d'autres *Alpha-proteobactéria* hétérotrophes. Elles ont une croissance plus rapide que les *Acidiphilium* mais une moindre résistance à l'acidité. Les deux espèces *Ac. facilis* et *Ac. aminolytica* utilisent des monosaccharides et des acides aminés (Johnson et Hallberg 2008). La troisième espèce *Ac. aromatica sp. nov.* utilise le fructose et des composés aromatiques comme les acides benzoïques, le phénol et le naphthalène (Jones *et al.* 2013). On peut aussi citer les genres *Acidisphaera* et *Acidicaldus*. Le premier, dont l'espèce *As. rubrifaciens* se développe bien sur des acides organiques mais de façon sporadique sur les monosaccharides (Hiraishi *et al.* 2000). Le deuxième, dont l'espèce *Acd. organivorans* utilise le phénol aussi bien que le glycérol et d'autres monosaccharides (Johnson *et al.* 2006). Les bactéries acidophiles hétérotrophes Gram positives se retrouvent dans les *phyla Actinobacteria* et *Firmicutes*. Chez les *Firmicutes*, la plupart appartiennent au genre *Alicyclobacillus* qui contient 17 espèces décrites dont 2 sont mixotrophes : *Alb. disulfidooxidans* et *Alb. tolerans*. Ces espèces utilisent une variété de mono, di et polysaccharides (Johnson et Hallberg 2008). Une espèce acidophile

de *Desulfosporosinus* a été récemment décrite sous le nom de, *Desulfosporosinus acidiphilus* sp. nov.. Elle a la particularité d'être anaérobie stricte et utilise des acides organiques et des monosaccharides pour sa croissance (Alazard *et al.* 2010). Chez les *Actinobacteria*, on retrouve les genres *Ferrimicrobium* et *Ferrithrix*, qui utilisent des composés organiques comme source de carbone mais obtiennent des électrons par l'oxydation du fer (Johnson *et al.* 2009). On retrouve trois genres d'*Euryarchaeota* acidophiles : *Thermoplasma*, *Ferroplasma* et *Picrophilus*. Ces organismes ont besoin d'une source de carbone complexe (comme l'extrait de levure) pour soutenir leur croissance. Les *Thermoplasma* sont anaérobies facultatifs alors que les *Picrophilus* sont aérobies strictes. Chez les *Crenarchaeota*, les genres *Sulfolobus* et *Sulfurisphaera* sont hétérotrophes. Les *Sulfolobus* utilisent aussi des sources de carbones complexes mais sont aussi capables de pousser sur des mono et di-saccharides (Johnson et Hallberg 2008). L'espèce *S. solfataricus* est capable d'utiliser le phénol (Izzo *et al.* 2005). L'espèce *Sulfurisphaera ohwakuensis* croît sur extrait de levure et tryptone mais pas sur des mono-saccharides et des acides aminés (Kurosawa *et al.* 1998).

L'utilisation du carbone par les micro-organismes n'est pas un métabolisme de type "noir/blanc". Certains organismes sont mixotrophes vis à vis du carbone : ils peuvent fixer du CO₂ et du carbone organique. L' α -protéobactérie *A. acidophilum* à l'origine s'appelait *Thiobacillus acidophilus* car elle est capable d'une croissance autotrophe et catalyse une oxydation dissimilatrice du soufre. Cependant cette espèce utilise aussi des monosaccharides, des acides aliphatiques et des acides aminés pour sa croissance (Guay et Silver 1975). De plus en culture en milieu discontinu ou continu avec du glucose et du tetrathionate, cette espèce obtient de l'énergie en oxydant ces deux composés et utilise le carbone du glucose et du CO₂ pour sa croissance (Mason et Kelly 1988). Au sein du phylum des *Firmicutes*, les espèces du genre *Sulfobacillus* expriment un métabolisme mixotrophe. *Sb. thermosulfidooxidans* et *Sb. acidophilus* ont besoin d'un donneur d'électron inorganique, et lors d'une croissance en présence de glucose il a été estimé que 20 % du carbone cellulaire provient du CO₂ (Wood et Kelly 1983; 1984). Chez les *Archaea*, les membres des *Sulfolobaceae* sont des autotrophes facultatifs qui utilisent des composés organiques complexes, des mono et di-saccharides pour soutenir une croissance hétérotrophe (Johnson et Hallberg 2008). La capacité de pouvoir utiliser le carbone organique et inorganique est intéressante, mais d'un point de vue énergétique, le carbone organique est plus efficace. Ces organismes ont besoin d'un système de

régulation pour gérer l'assimilation du carbone, cependant ces mécanismes ne sont pas identifiés. On possède en revanche des informations sur la régulation de la RUBISCO chez *At. ferrooxidans*. Elle est dépendante de la quantité de CO₂ disponible, mais aussi du donneur d'électron (fer, sulfure ou acide organique) (Johnson et Hallberg 2008).

1.2.2 Le soufre et le fer comme source d'énergie

Dans les DMA, le fer est souvent présent en forte concentration et il est très soluble au pH typique des drainages. De plus, nous avons vu dans la section 1.1 que le fer est difficilement oxydé par l'oxygène à pH < 4. Les micro-organismes ne sont donc pas en compétition avec l'oxydation abiotique du fer pour leur croissance. La quantité de fer réduit est potentiellement importante mais son apport énergétique est faible. En effet, une seule mole d'électron est libérée par mole de fer oxydé et le potentiel red/ox du couple Fe²⁺/Fe³⁺ est proche de celui O₂/H₂O au pH des drainages (respectivement +770 mV et +820 mV) (Johnson et Hallberg 2008). Le transfert d'e⁻ du fer(II) vers la réduction de l'oxygène et de deux protons pour créer de l'eau est thermodynamiquement favorable; c'est la voie descendante des électrons. La consommation des protons engendrés permet le transfert de protons du périplasma vers le cytoplasme via l'ATPase et de produire de l'ATP. Un modèle moléculaire de cette voie métabolique chez *A. ferrooxidans* ATCC 23270 a été proposé par Quatrini *et al.* (2009) (Figure 1.3) : elle implique un cytochrome c (Cyc2) localisé dans la membrane externe, une protéine "bleu cuivre" périplasmique la rusticyanine (Rus) liée à un cytochrome c₄ (Cyc1) et un cytochrome oxydase terminal de type aa₃ encodé par le cluster de gènes coxBACD. La rusticyanine représente jusqu'à 5 % des protéines de *A. ferrooxidans*. Cependant tous les électrons ne sont pas dirigés vers la réduction de l'oxygène. Les métabolismes cellulaires nécessitent du NAD(P)H, qui est fabriqué par la réduction du NAD(P), dans ce cas à partir d'électrons provenant du fer. Le problème est que le potentiel red/ox du couple NAD(P)/NAD(P)H à pH neutre (celui du cytoplasme) est beaucoup plus bas que celui du couple Fe²⁺/Fe³⁺ (-320 mV contre +770 mV). La réaction de réduction du NAD(P) par Fe²⁺ n'est pas thermodynamiquement favorable, et nécessite une cascade d'enzyme pour se faire. C'est le transport ascendant des électrons vers le NAD(P) : les électrons sont transférés de la rusticyanine à un cytochrome c₄ dihémique (CycA1), puis à un complexe cytochromique bc1 (complexe III, coenzyme Q – cytochrome c réductase), au pool de quinone puis à la NADH

déshydrogénase (complexe I). L'énergie nécessaire à la "montée" des électrons serait fournie par la force protomotrice via l'hydrolyse d'ATP par l'ATP synthase. En ce qui concerne les autres espèces autotrophes, cette chaîne "ascendante" a aussi été observée dans le génome reconstruit de *Leptospirillum* (Tyson *et al.* 2004). En revanche, on ne connaît pas chez les acidophiles hétérotrophes l'utilisation du flux d'électrons provenant de l'oxydation du fer, hormis le fait qu'elle augmente leur taux de croissance (Johnson et Hallberg 2008). Chez les archées cependant, Dopson *et al.* (2005) proposent un modèle de transport d'électrons à partir du fer. Dans ce modèle, les électrons du fer(II) sont transférés à une protéine "bleu cuivre" (sulfocyanine) qui transfère les électrons à une oxydase terminale de type *ccb3*. Cette chaîne de transport plus courte s'explique par l'absence de périplasme chez les archées. Les archées thermophiles (*Sulfolobus*, *Metallosphaera*) utiliseraient d'autres protéines codées par les gènes *fox* (Bathe et Norris 2007).

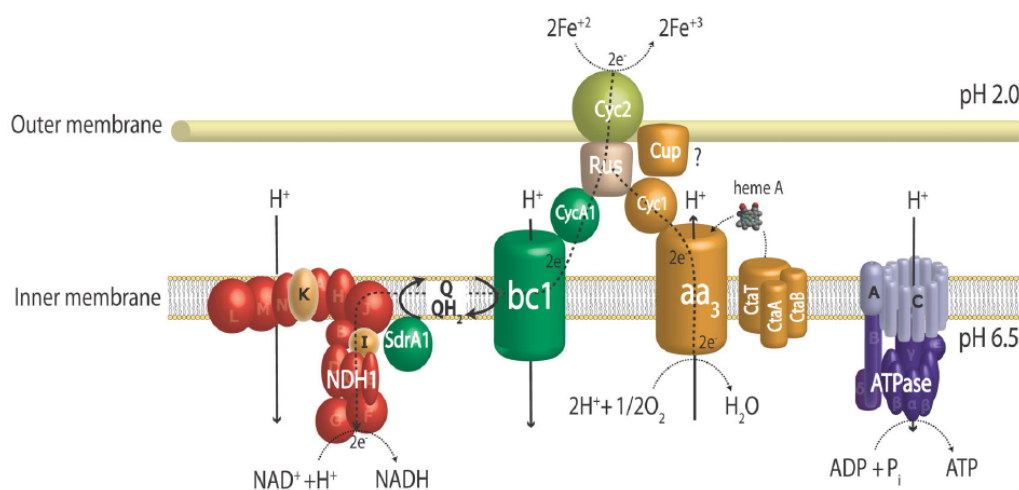


FIGURE 1.3 – Modèle de l'oxydation du Fe(II) chez *At. ferrooxidans* ATCC 23270. Le flux d'électrons est représenté depuis l'oxydation du Fe^{2+} jusqu'à l'oxygène par Cyc2 via le complexe aa_3 (voie descendante) ou bien jusqu'au NAD^+ par le complexe bc1/quinone/NADH (voie ascendante). La voie descendante utilise des protons via le complexe ATPase pour la synthèse d'ATP et la voie ascendante pour la réduction du NAD^+ . La régulation vers la voie ascendante et descendante se fait probablement au niveau de la rusticyanine (Rus) (Quatrini *et al.* 2009).

En plus du fer, les composés réduits soufrés (RISC) sont présents en grande quantité dans les drainages miniers à cause de l'oxydation des minéraux sulfurés. Les RISC sont plus avantageux que le fer d'un point de vue énergétique. D'abord l'oxydation des RISC produit plus d'électrons que l'oxydation du fer (un pour l'oxydation de Fe^{2+} en Fe^{3+} , six pour l'oxy-

dation de S^0 en SO_4^{2-}). Ensuite les électrons provenant de l'oxydation des RISC entrent dans la chaîne de transport à un potentiel red/ox plus élevé. Plus d'ATP est produit par mole de substrat et il n'y a pas besoin de transport ascendant des électrons pour la réduction du NAD. De ce fait, le taux de croissance des cellules est plus important en utilisant les RISC plutôt que le fer. Chez *At. ferrooxidans*, il a été mesuré 0,92 g de matière sèche/mol d' e^- du tétrathionate et 0,23 g/mol d' e^- de fer(II) (Johnson et Hallberg 2008). Cette variété de composés utilisables est intéressante pour les micro-organismes mais renforce la difficulté pour les scientifiques d'identifier les voies métaboliques permettant l'oxydation des RISC. De plus certains RISC peuvent subir des oxydations abiotiques, renforçant encore la difficulté d'identifier les réactions réellement catalysées par les enzymes. Des voies métaboliques d'oxydation du soufre ont été proposées, notamment la voie des protéines Sox très répandue phylogénétiquement et la voie des Sor chez les archées. Cependant, ces voies n'ont pas été détectées chez *At. ferrooxidans*, ainsi que chez les autres acidophiles (Quatrini *et al.* 2009, Johnson et Hallberg 2008). La plupart des connaissances accumulées sur l'oxydation des RISC par *At. ferrooxidans* provient de tests enzymatiques réalisés sur des souches dont la taxonomie n'est pas résolue ou qui n'existent pas en culture pure. Le modèle proposé est donc plutôt une vision du potentiel métabolique du genre *Acidithiobacillus* pour l'oxydation des RISC (Fig 1.4). En fonction de leur niveau d'oxydation, les RISC sont pris en charge par des réductases : la sulfure quinone réductase (SQR), la thiosulfate quinone réductase (TQR) et l'hétérodisulfure réductase (HDR). Les électrons sont transférés via un pool de quinone (QH₂) soit à des oxydases terminales ou au complexe NADH I pour générer du pouvoir réducteur. Des similarités de séquences et de syntenie des gènes de la HDR et de ses protéines accessoires ont été observées chez *Sulfolobus acidocaldarius*, *S. tokodaii*, *S. solfataricus* et *Metallosphaera sedula* (Quatrini *et al.* 2009). Cependant chez *S. metallicus*, un gène codant pour une Soufre Oxygénase Réductase (SOR), absente chez *A. ferrooxidans* est le transcrit majoritaire chez les cellules cultivées sur du soufre, et semble donc impliqué dans l'oxydation du soufre (Quatrini *et al.* 2009). Les voies d'oxydations des RISC chez les acidophiles sont différentes de celles utilisées par les neutrophiles, mais elles sont aussi différentes au sein des acidophiles. De nouvelles études sur d'autres espèces acidophiles sont nécessaires pour mieux les comprendre (Quatrini *et al.* 2009).

Nous venons de voir que le fer et le soufre sont utilisés comme donneur d'électrons. Ils

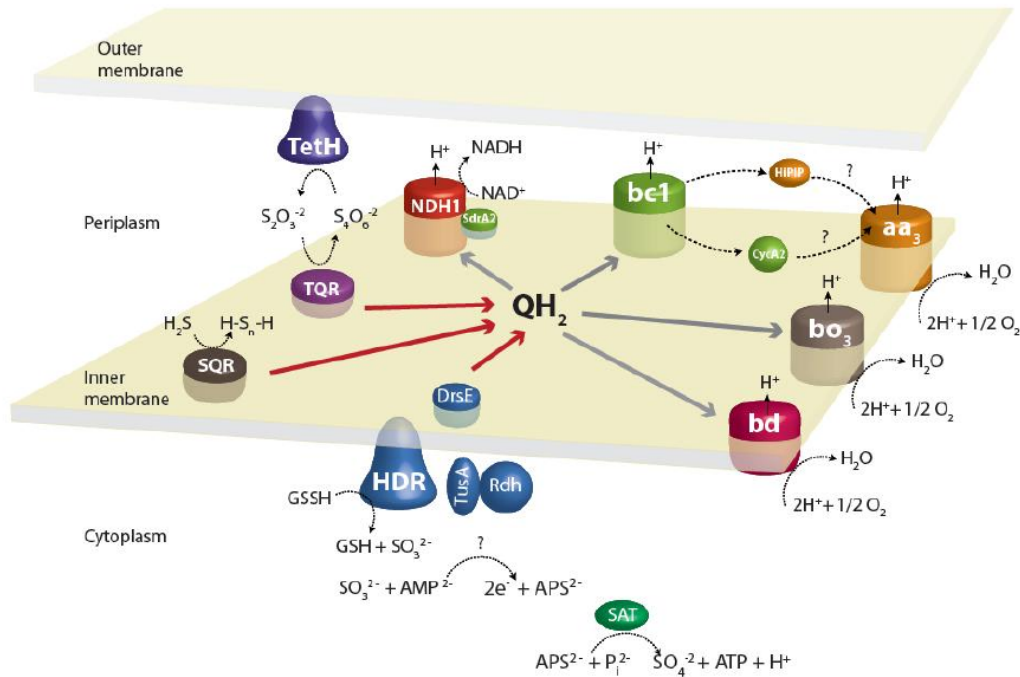


FIGURE 1.4 – Modèle de l'oxydation des sulfures chez *Acidithiobacillus*. L'oxydation des composés inorganiques soufrés réduits (RISC) implique de nombreuses enzymes : dans la membrane externe côté périplasm la tétrathionate hydrolase (TetH), dans le périplasm la protéine à haut potentiel fer-soufre (HiPIP), dans la membrane cytoplasmique côté périplasm le cytochrome c (CycA2), dans la membrane cytoplasmique la sulfure quinone réductase (SQR), la thiosulfate quinone réductase (TQR), le complexe bc1, le complexe NADH I, les oxydases terminales bd et bo3, enfin dans le cytoplasme l'hétérodisulfure réductase (HDR) et l'ATP sulfuryle (SAT). Le soufre insoluble est converti en sulfane (GSSH) qui est ensuite transféré à l'HDR par une cascade de sulfure transférase (DsrE, TusA et Rhd). Les électrons venant des sulfures (H_2S), thiosulfate ($S_2O_3^{2-}$) ou bien sulfane (GSSH) sont transférés par le pool de quinone (QH₂) jusqu'à (1) l'oxydase terminale bd ou BO3 directement, ou indirectement par le complexe bc1 et un cytochrome c (CycA2) ou une protéine fer-soufre à fort potentiel (HiPIP) probablement à l'oxydase aa3 où la réduction de O_2 a lieu (2) complexe NADH I pour générer du pouvoir réducteur (Quatrini *et al.* 2009).

peuvent aussi servir d'accepteur final dans la chaîne respiratoire à la place de l'oxygène pour une croissance anaérobie. *At. ferrooxidans* est capable d'une croissance anaérobie en réduisant le fer et en utilisant les sulfures comme donneurs d'électrons (Johnson et Hallberg 2008). L'hétérotrophe *Acidiphilium cryptum* est capable de co-respirer l'oxygène et le fer, de plus il a été montré qu'il est capable d'accélérer la dissolution de minéraux ferriques (akageneite, goethite, jarosite, natrojarosite et hydroxydes anamorphiques) par des mécanismes indirects (Bridge et Johnson 2000). Un comportement similaire a été observé chez des souches d'*Acidocella* et d'*Acidobacterium capsulatum* envers la schwertmannite (Coulpland et Johnson 2008). *Acidicaldus organivorans* a la particularité de n'utiliser que le fer comme accepteur final d'électrons et d'être strictement anaérobie (Johnson 2006). Beaucoup d'espèces mixotrophes sont capables de réduire le fer pour soutenir leur croissance : les *Sulfo-**bacillus* et *Alicyclobacillus* pour les *Firmicutes*, *Acidimicrobium ferrooxidans*, *Ferrimicrobium acidiphilum* et *Ferrithrix thermotolerans* chez les *Actinobacteria* et *Ferroplasma* chez les archées (Johnson et Hallberg 2008). Même si c'est un phénomène commun, peu de choses sont connues sur les mécanismes de réduction du fer. Il a été suggéré que les espèces pouvant oxyder le fer puissent utiliser les mêmes enzymes en sens inverse (Sugio *et al.* 1987; 1989) mais cela ne peut pas être le cas des organismes hétérotrophes. La réduction du soufre élémentaire est réalisée par quatre genres *Archaea* : *Acidianus*, *Stygiolobus*, *Sulphurisphaera* et *Thermoplasma*. À la différence des trois autres qui sont anaérobies facultatifs, *Stygiolobus* est anaérobie stricte. *Acidianus*, *Stygiolobus* et *Sulphurisphaera* couplent la réduction du soufre avec l'oxydation de l'hydrogène, alors que *Thermoplasma* utilise des donneurs d'électrons organiques (Johnson et Hallberg 2008). Aucune archée acidophile connue ne réduit les sulfates. Bien que des bactéries sulfato-réductrices (BSR) aient été isolées d'environnements acides, anaérobies à basse température, il s'est avéré que les isolats étaient inhibés par des pH < 5. L'hypothèse alors était que des BSR neutrophiles se maintenaient dans des micro-niches ayant un pH plus élevé que le milieu extérieur. Bien que la réduction des sulfates à pH acide soit thermodynamiquement favorable ($E^0 \text{SO}_4^{2-}/\text{HS}^-$ à pH = 3 vaut + 75 mV), il existe deux limitations physiologiques qui peuvent expliquer cette inhibition par le pH. D'abord la plupart des BSR n'oxydent pas complètement leur substrat et accumulent (généralement) de l'acide acétique dans leur milieu. Aux pH optimaux des acidophiles, les acides organiques de faible poids moléculaire sont toxiques car sous une forme non chargée/non dissociée et

lipophiles. Ils traversent donc les membranes biologiques et dans le cytosol (pH 7) se dissocient, conduisant à une acidification du cytoplasme et à la mort cellulaire. La deuxième limitation est la toxicité de l'hydrogène sulfuré produit par la sulfato-réduction. À pH 4 la forme la plus toxique H_2S représente 95 % des sulfures générés. Le fer joue le rôle de puits à sulfure en formant un précipité FeS , mais à pH acide ($< 5,5$) FeS est soluble et ne remplit plus sa fonction (Johnson et Hallberg 2008). Cependant, dans des conditions contrôlées avec un consortium de *Acidocella aromatica* (hétérotrophe) et une nouvelle espèce de *Desulfosporosinus* (isolat M1) Kimura *et al.* (2006) ont maintenu une activité sulfato-réductrice de *Desulfosporosinus* M1 jusqu'à un pH de 3,8.

1.2.3 Transformation de l'arsenic par les procaryotes

L'arsenic est un métalloïde toxique abondant dans la croûte terrestre. Il est principalement retrouvé dans des minéraux sous forme d'anions As_2 ou AsS associé à des métaux comme le fer, le cobalt ou le nickel (par exemple $FeAs_2$, $FeAsS$, $CoAsS$). Il est aussi retrouvé comme impureté dans la pyrite. Il existe aussi des minéraux d'arsenic soufrés simples comme le réalgar (AsS ou As_4S_4), ces minéraux peuvent se dissoudre lors de la formation des DMA, entraînant des concentrations potentiellement très élevées d'arsenic (Morin et Calas 2006). L'arsenic est présent dans différents niveaux d'oxydation : -3, 0, +3 et +5 (Cullen et Reimer 1989). En fonction des conditions physico-chimiques, les composés d'arsenic sont facilement solubilisés dans l'eau, lui conférant une forte bio-disponibilité. La forme sous laquelle se trouve l'arsenic dans les DMA dépend fortement du minerai duquel il est issu (Morin et Calas 2006). La toxicité de l'arsenic est due à ses propriétés chimiques et sa capacité à entrer dans les cellules. L' $As(V)$ est un analogue du phosphate et entre dans les cellules par les transporteurs de phosphate, principalement par le transporteur constitutif Pit (Fig 1.5, point 1) (Rosen et Liu 2009). L' $As(III)$ aux pH $< 9,3$ existe sous la forme non chargée $As(OH)_3$ dont la structure est similaire à celle du glycérol. Sa voie d'entrée préférentielle serait les aquaglycéroporines qui transportent l'eau, le glycérol et des petites molécules non-chargées (Fig 1.5, point 1) (Rosen et Liu 2009). Dans la cellule $As(V)$ se substitue au phosphate et produit des composés instables. Cela interfère avec les processus normaux de phosphorylation et notamment la synthèse de l'ATP, de même que les transports dépendant de l'ATP, la glycolyse, la voie des pentoses phosphates et les voies de transduction du signal (Goyer et

Clarkson 2001). As(III) a une forte affinité pour les groupes sulfhydryles (R-SH), les groupes dithiols comme par exemple les glutathiones, glutarédoxines et thiorédoxines, provoquant des effets cancérogènes (Liu *et al.* 2001). Généralement les procaryotes sont 10 fois plus résistants à l'As(V), qu'à l'As(III) (Tamaki et Frankenberger 1992). Un des mécanismes de résistance à

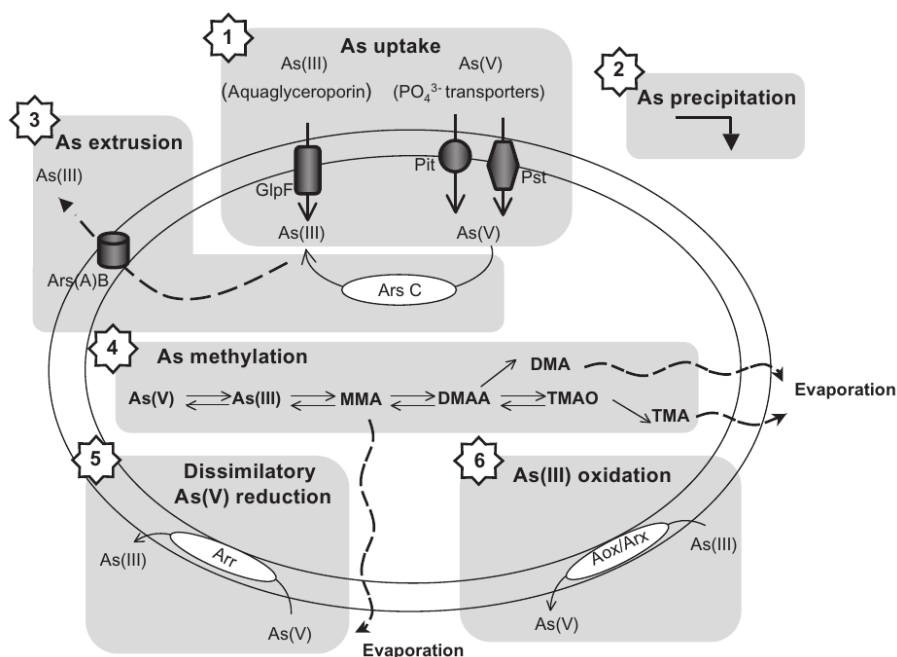


FIGURE 1.5 – Représentation schématique des différentes transformations de l'arsenic par les procaryotes : (1) l'arsenic (As) entre dans les cellules par les transporteurs de phosphate (arsenate, As(V)) ou par les aquaglycéroporines (arsenite, As(III)) ; (2) l'arsenic est immobilisée dans l'environnement extra-cellulaire par précipitation ; (3) dans la cellule, As(V) est réduit par une arsenate réductase, ArsC, en As(III) qui est expulsé de la cellule via des transporteurs Ars(A)B et Acr3 ; (4) l'arsenic inorganique peut être transformé en espèces organiques par une cascade de méthylation ; (5) As(V) est utilisé comme accepteur d'électrons par une arsenate réductase dissimilatrice ArrAB ; (6) As(III) est utilisé comme donneur d'électrons par les oxydases AoxAB (appelée AioAB maintenant (Zargar *et al.* 2012)) ou ArxAB. As, arsenic ; As(III), arsenite ; As(V), arsenate ; Pit, système de transport de phosphate inorganique (PO_4^{3-}) ; Pst, système spécifique de transport de phosphate ; GlpF, GlpF aquaglycéroporine ; ArsC, arsenate réductase détoxifiante ; Ars(A)B, pompe à efflux d'arsenite ; MMA, acide monométhylarsonique ; DMAA, acide diméthylarsinique ; DMA, diméthylarsine ; TMAO, oxyde triméthylarsine ; Aox/Arx, types d'arsenite oxydase ; Arr, arsenate réductase dissimilatrice. Figure modifiée de (Slyemi et Bonnefoy 2012).

l'arsenic est l'extrusion active de la cellule (Fig 1.5, point 3). Seul des transporteurs d'As(III) ont été identifiés jusqu'à présent et appartiennent aux familles ArsB et Acr3. Le système Ars est le mieux caractérisé, il existe sous la forme d'opérons génomique ou plasmidique ArsRBC ou ArsRDABC (Slyemi et Bonnefoy 2012). ArsB est un uniport qui utilise le potentiel de

membrane pour expulser l'arsenite, mais il n'est pas spécifique il extrude aussi les anions d'antimoine III. ArsC est une arsenate réductase cytoplasmique qui réduit As(V) en As(III) (Fig 1.5, point 3). ArsA est une ATPase qui fournit de l'énergie à ArsB pour l'extrusion de As(III), ArsB est fonctionnelle sans ArsA mais son efficacité est moindre (Meng *et al.* 2004). ArsR et ArsD sont des répresseurs qui régulent respectivement les niveaux d'expression basal et maximum de l'opéron (Ajees *et al.* 2011). Une autre pompe existe : Acr3, moins bien caractérisée que ArsB mais plus répandue car présente dans les 3 royaumes du vivant alors que ArsB n'a encore jamais été observée chez les *Eucaryotes* (Mansour *et al.* 2007). On le retrouve aussi sous le nom de Arr3p. Ce transporteur est plus spécifique pour l'As(III) que ArsB et la sortie de As(III) semble associé à la force proto-motrice (Fu *et al.* 2009). La famille des transporteurs Acr3 est divisée en deux groupes : Acr3.1 qui possède un sous-groupe de transporteurs de champignons et Acr3.2 qui ne contient que des procaryotes (Wysocki *et al.* 2003). Une autre stratégie de résistance est la méthylation (Fig 1.5, point (4)) produisant des composés volatiles ou non. Les espèces volatiles dominantes sont des mono-, di- et triméthylarsine (MMA, DMA, TMA), et les non volatiles sont surtout des méthylarsonates et diméthylarsinate. La réaction commence par la réduction de l'As(V) suivi par des cycles de méthylation oxydative et de réduction d'intermédiaires organoarseniés (Dombrowski *et al.* 2007). La réaction de méthylation nécessite des S-adénosylméthionines comme source de groupe méthyle, il est transféré à l'As(III) par une arsenite S-adénosylméthionine méthyltransférase (ArsM) avec la glutathione comme donneur d'électrons. Cette réaction s'exerce sur tous les intermédiaires méthylés jusqu'au TMA (Qin *et al.* 2006, Yuan *et al.* 2008). Le gène ArsM est très répandu, il a été détecté dans plus de 120 procaryotes (dont des *Archaea*) (Stolz *et al.* 2006).

Les réactions d'oxydo-réduction exposées dans les points (5) et (6) de la figure 1.5 ne sont pas des mécanismes de résistance mais des métabolismes d'acquisition d'énergie. La réduction dissimilatrice de As(V) (As(V) est l'accepteur final d'électron) permet la croissance anaérobie de nombreuses espèces procaryotes chez les *Proteobacteria*, *Bacteria* Gram-positifs, *Bacteria* et *Crenarchaeota* thermophiles (Silver et Phung 2005, Stolz *et al.* 2006, Paez-Espino *et al.* 2009, Oremland *et al.* 2009). De même l'arsenite peut être utilisée comme donneur d'électrons. Ce métabolisme a été identifié chez des micro-organismes hétérotrophes, mixotrophes et chimiolithoautotrophe dans des environnements variés (eaux usées, eaux douces et ma-

rines, géothermaux, drainages miniers, dépôts d'arsenic) (Ehrlich 2002, Oremland et Stolz 2003, Stolz *et al.* 2006, Oremland *et al.* 2009).

1.2.4 Les *Archaea* dans les DMA

Les *Archaea* ont d'abord été isolées de milieux extrêmes (pH \approx 0, températures $>$ 100 °C). Dans les DMA elles peuvent constituer la majorité de la population procaryote totale (Edwards 2000b). Cependant beaucoup moins d'informations existent sur la physiologie des archées que celle des bactéries (voir les sous-sections 1.2.1 ; 1.2.2). La plupart des archées de DMA identifiées appartiennent à l'ordre des *Thermoplasmatales* du phylum *Euryarchaeota* où l'on retrouve les genres *Ferroplasma* et *Thermoplasma*. Les archées de cet ordre sont hétérotrophes et seul les organismes du genre *Ferroplasma* sont connus pour réaliser des réactions d'oxydo-réduction du fer (Hallberg 2010). D'autres archées identifiées par des approches moléculaires, appartenant à l'ordre des *Thermoplasmatales*, sont trop éloignées des souches caractérisées pour en déduire des informations sur leur physiologie (Hallberg 2010). Baker et Banfield (2003) ont réalisé un travail de classification de ces *Thermoplasmatales* non cultivé en créant des clades "alphabet-plasma". Des archées méthanogènes, appartenant à l'ordre des *Methanosarcinales* et *Methanobacteriales*, ont aussi été observées dans les sédiments du DMA du Río Tinto (Sanz *et al.* 2011). Les espèces identifiées *Methanosaeta concilii*, *Methanosarcina barkeri* et *Methanobacterium bryantii* se développent dans des micro-niches anaérobies et réductrices des sédiments où l'acidité est moins importante. On retrouve aussi des *Crenarchaeota* dont les espèces sont thermophiles extrêmes (Table 1.2). Le phylum des *Thaumarchaeota* récemment décrit (Brochier-Armanet *et al.* 2008) regroupe des archées mésophiles anciennement affiliées aux *Crenarchaeota*. Récemment Halter *et al.* (2011) ont identifié une espèce de *Thaumarchaeota* : *Candidatus Nitrososphaera gargensis* dans le drainage de S^{te} Marie aux Mines. Bien que celui-ci ne soit pas acide, des séquences d'ARNr 16S proches ont été détectés par séquençage dans le DMA de Carnoulès (Volant *et al.* 2012). Les *Thaumarchaeota* sont impliquées dans une étape clef du cycle de l'azote qui est l'oxydation de l'ammonium (Brochier-Armanet *et al.* 2012), or le cycle de l'azote dans les DMA est quasiment inconnu. Il est fort probable que l'azote dans les DMA soit d'origine atmosphérique et fixé par des bactéries du genre *Acidithiobacillus* et *Leptospirillum*, comme l'indique des données génomiques (Dai *et al.* 2014, Mi *et al.* 2011). Cependant des

Thaumarchaeota observées dans les sources géothermales acides de Yellowstone, ne montrent pas d'activité ammonium oxydase (Beam *et al.* 2014). Leur implication dans l'oxydation de l'ammonium dans les DMA reste donc à confirmer, mais des indices montrent qu'elle est faisable car elle existe dans des sols acides (Gubry-Rangin *et al.* 2011). Enfin il a été récemment observé des archées singulières dans des biofilms de *Richmond Mine* (Baker *et al.* 2006) appelées ARMAN (Archaeal Richmond Mine Nanoorganisms). Elles sont profondément branchées dans le phylum des *Euryarchaeota* à la limite de la division avec les *Crenarchaeota* (Baker *et al.* 2010a). Actuellement 5 types d'ARMAN ont été décrites : ARMAN-1 ; ARMAN-2 (*Candidatus Micrarchaeum*) ; ARMAN-3 ; ARMAN-4 (*Candidatus Parvarchaeum acidophilum*) ; ARMAN-5 (*Candidatus Parvarchaeum acidophilus*) (Baker *et al.* 2010a). Elles sont singulières car leur volume cellulaire (entre 0,009 et 0,062 μm^3) est à la limite du volume théorique pour une vie libre. À titre de comparaison le volume des *Nanoarchaea* varie entre 0,02 et 0,7 μm^3 et sont des parasites obligatoires d'autres archées (Huber *et al.* 2002). Une preuve étayant ce mode de vie pour les ARMAN est l'observation au microscope électronique de la pénétration de la membrane de *Thermoplasmatales* par des ARMAN (Comolli *et al.* 2009). Les ARMAN n'ont pas été isolées et les découvertes sur ces *archaea* sont issues de génomes reconstitués à partir de métagénome et d'enrichissement d'échantillon (fraction inférieure à 500 nm) (Baker *et al.* 2010a). Récemment les avancées des méthodes moléculaires ont posé des questions sur la taxonomie des *Archaea* qui souffre du grand nombre de clades composés de séquences environnementales non caractérisées. Rinke *et al.* (2013) ont réalisé 201 génomes sur une cellule (single-cell genomic) d'*Archaea* et de *Bacteria* des groupes taxonomiques les moins définis. Ils proposent la réunion des phyla d'*Archaea* ayant une petite taille et un petit génome, comme les *Parvarchaeota* et *Micrarchaeota* (ARMAN), en un super-phylum DPANN, car ces caractéristiques définissent un groupe monophylétique. L'étude des *Archaea* est dynamique et progresse constamment. L'avancée des méthodes moléculaires permet de mieux appréhender la taxonomie de ces micro-organismes, ainsi que leur physiologie.

1.2.5 Les interactions au sein des communautés microbiennes des DMA

L'écologie définie par Ernst Haeckel (1869) est la science qui étudie les relations entre les êtres vivants et leur milieu, ainsi que les interrelations des êtres vivants entre eux. Les écosystèmes constituent le siège de ces interactions, où la biocénose définit les organismes et le biotope le milieu physique de l'écosystème. Les relations au sein de la biocénose sont hiérarchisées en réseaux trophiques où la base formée par les producteurs primaires produit la matière organique nécessaire aux consommateurs des rangs supérieurs (Elton 1927). Les relations entre les organismes sont aussi organisées en fonction du gain qu'apporte la relation (Figure 1.6).

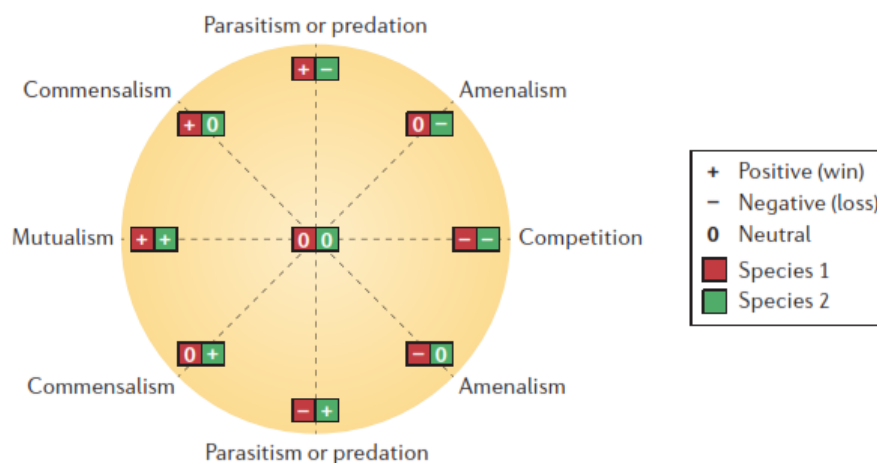


FIGURE 1.6 – Résumé des relations écologiques entre les membres d'espèces différentes. Pour chaque partenaire il y a trois issues possibles : positive (+), négative (-) et neutre (0). Par exemple dans le cas du parasitisme, le parasite tire un bénéfice de la relation (+) au détriment de l'hôte (-); cette relation est donc représenté par la paire de symbole +- (Faust et Raes 2012).

Un exemple de compétition pour le substrat est celui de *At. ferrooxidans* et *L. ferrooxidans*, deux autotrophes ferro-oxydants, pour le Fer(II). Ces organismes peuvent en théorie cohabiter mais des différences physiologiques font que l'un va prendre le pas sur l'autre. *At. ferrooxidans* a une activité oxydative plus importante aux bas potentiels red/ox mais ne peut pas oxyder le fer lorsque le potentiel re/ox est supérieur à +840 mV. *L. ferrooxidans* a une meilleure affinité pour Fe^{2+} que *At. ferrooxidans* ($K_m = 0,25$ mM et 1,34 mM respectivement), et est en plus moins inhibé que *At. ferrooxidans* par le Fe^{3+} . Le résultat est que *L. ferrooxi-*

dans domine dans les systèmes très oxygénés avec une forte concentration en Fe^{3+} , alors que *At. ferrooxidans* domine les systèmes où la forme Fe^{2+} prédomine (Johnson et Hallberg 2008). Des observations similaires ont été faites avec d'autres acidophiles ferro-oxydant, montrant l'adaptation des micro-organismes à différentes niches existant dans les DMA (Norris 1992). Les travaux de Lu *et al.* (2010) montrent aussi l'importance des différentes niches dans les sédiments d'un lac minier acide dans la transformation du fer. En effet dans les sédiments le taux d'oxygène baisse très vite et le pH augmente aussi rapidement avec la profondeur sur les 20 premiers cm des sédiments (Figure 1.7). Ces différentes niches sélectionnent des espèces procaryotes différentes.

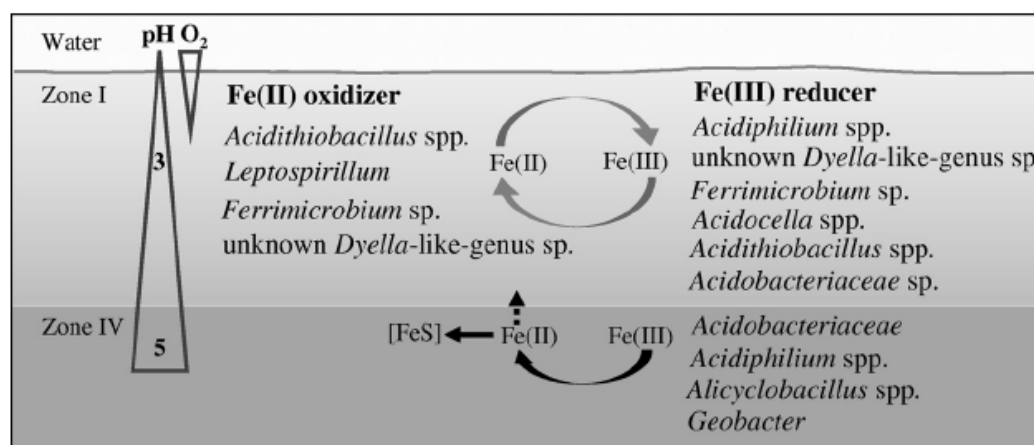


FIGURE 1.7 – Schémas des groupes microbiens impliqués dans le cycle du fer dans le lac minier acide Lake 77. La concentration en oxygène décline rapidement de 0 à 10 mm et le pH augmente de 3 à la surface jusqu'à 5 dans la Zone IV à 20 cm (Lu *et al.* 2010).

La relation syntrophique entre *Desulfosporosinus* M1 et *Acidocella* PFBC est un exemple de synergie qui peut avoir lieu dans les DMA (Figure 1.8). Individuellement *Acidocella* PFBC est décrite comme aérobie obligatoire et *Desulfosporosinus* M1 ne pouvant pas réaliser de sulfato-réduction à pH très acide. Mis en culture ensemble *Acidocella* PFBC est capable d'une croissance anaérobie et *Desulfosporosinus* M1 d'une sulfato-réduction en milieu acide. *Acidocella* PFBC lève l'inhibition due à l'accumulation de l'acétate, mais pour pouvoir le dégrader il a besoin du puits d'hydrogène créé par *Desulfosporosinus* M1 lors de la réduction des sulfates. La réaction acétoclaste qui n'est pas thermodynamiquement favorable, le devient par la consommation rapide de l'hydrogène (Kimura *et al.* 2006). Ceci est un exemple précis, mais les hétérotrophes réalisent ce rôle de façon globale. En effet ils lèvent la toxicité des

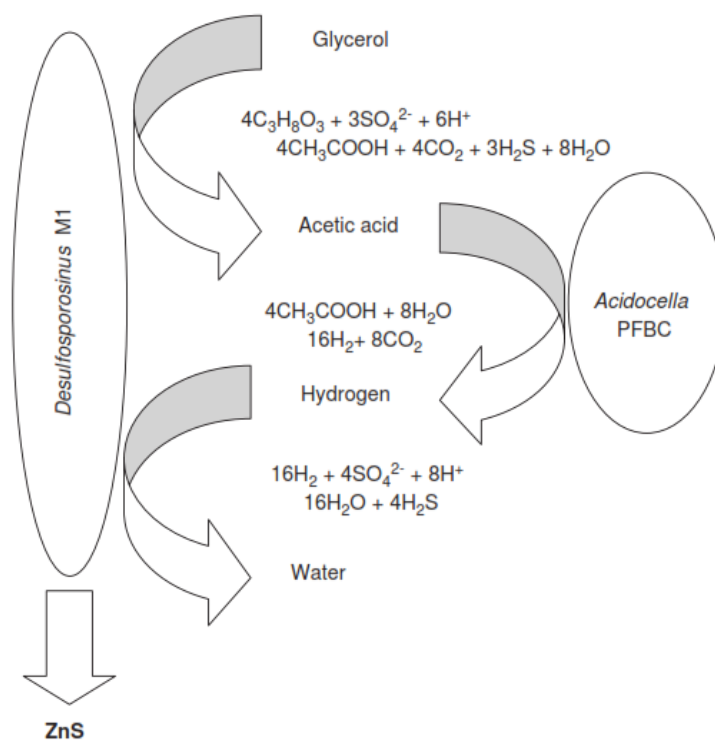


FIGURE 1.8 – Relation syntrophique proposée entre *Desulfosporosinus* M1 et *Acidocella* PFBC en culture anaérobie à un pH < 4 (Johnson et Hallberg 2008).

acides organiques de faible poids moléculaire, et permettent le maintien de l'oligotrophie du milieu qui est nécessaire pour la croissance de beaucoup de procaryotes autotrophes (Fig 1.9) (Johnson et Hallberg 2008).

Au début de cette section, j'évoquais l'existence de protozoaires acidophiles et que des relations de prédation entre les protozoaires et les procaryotes peuvent exister. Cette relation peut être préférentielle : un flagellé proche des *Eutreptia* se nourrit préférentiellement d'*A. ferrooxidans* plutôt que de *L. ferrooxidans* en co-culture (Mc Guinness et Johnson 1992). De plus dans une usine de traitement, la biomasse bactérienne a été drastiquement diminuée par la présence d'un flagellé (Johnson et Rang 1993). Les protozoaires contrôlent la biomasse bactérienne, mais entretiennent aussi des relations symbiotiques avec des procaryotes. En effet lors de l'étude du drainage de la mine Richmond, il a été observé que 4 % des protistes contenait un symbiote proche des *Rickettsiales* (Baker *et al.* 2010b).

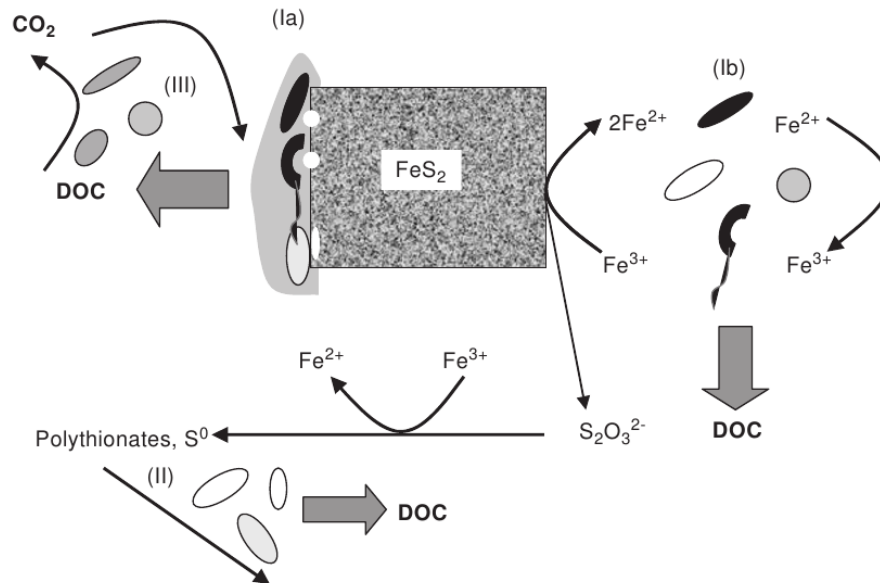


FIGURE 1.9 – Représentation schématique des rôles et des interactions entre les différents groupes physiologiques des procaryotes acidophiles durant la dissolution oxydative de la pyrite (FeS_2). Les procaryotes ferro-oxydants fixés au minéral (groupe Ia) ou non (groupe Ib) ont le premier rôle car ils produisent des ions ferrique (Fe^{3+}) qui sont le principal oxydant chimique de la pyrite. Les procaryotes sulfo-oxydants (groupe II) ont un rôle secondaire dans le sens où ils oxydent les sulfures libérés lors de la dissolution du minéral. Ils produisent de l'acide sulfurique qui maintient l'acidité du milieu. Ces micro-organismes sont principalement autotrophes et libèrent dans le milieu des composés organiques dissous (COD). Il vont être utilisés par le troisième groupe (groupe III) qui sont hétérotrophes et mixotrophes. La minéralisation des COD permet de maintenir un environnement propice aux autotrophes (Johnson et Hallberg 2008).

1.3 L'assemblage des communautés : une question majeure de l'écologie

Il existe un débat animé au sujet de l'importance des interactions écologiques, dans la définition de l'abondance de chaque espèce dans les communautés. Dans son article fondateur sur la distribution des espèces d'oiseaux dans des îles tropicales, Jared Diamond tira des règles d'assemblage de communauté à partir de motifs de présence-absence (Diamond 1975). Brièvement il expliqua que la présence d'une espèce empêcherait une autre espèce d'oiseau adaptée à la même niche de s'installer sur la même île. Sur une matrice de présence-absence si une espèce vaut 1 l'autre aura pour valeur 0 et inversement. Des motifs en damier sur une matrice de présence-absence indiquent une exclusion compétitive. Cette étude amena des discussions sur l'importance de ces règles d'assemblage dans la formation des communautés pour plusieurs décennies. Connor et Simberloff (1979) attaquèrent les théories de Diamond car ils les considéraient comme redondantes et que ses motifs ne résistaient pas aux tests de significativité avec des données simulées. La discussion sur le fait que les motifs de présence-absence ou d'abondance s'expliquent par des phénomènes stochastiques (issus du hasard) ou bien déterministes (processus écologiques) a avancé avec l'utilisation des modèles nuls (Gotelli 2000). Un modèle nul peut être décrit comme un ensemble de règles pour une simulation de données dont le résultat respecterait l'hypothèse nulle. Dans ce contexte il est intéressant de citer la théorie neutre unifiée de la biogéographie et biodiversité ("*The Unified Neutral Theory of Biodiversity and Biogeography*") (Hubbell 2001) que je citerai sous le terme "théorie neutre". Selon cette théorie, la distribution des espèces d'une communauté peut s'expliquer par des processus aléatoires de naissance, de mort et d'immigration à partir d'une "méta-population". Parce qu'elle ne considère pas d'évènements déterministes, la théorie neutre de Hubbell peut servir de modèle nul contre lequel tester l'impact d'interactions écologiques. C'est aussi un outil qui permet de tester le postulat de Baas-Becking "*Everything is everywhere, but the environment selects*" (Baas-Becking 1934). La théorie neutre a été confirmée dans quelques écosystèmes (Woodcock *et al.* 2007) mais aussi infirmée dans d'autres (Wootton 2005). Les communautés seraient donc structurées par des processus neutres et écologiques, l'un et l'autre prenant plus ou moins de poids en fonction des écosystèmes.

Des analyses basées sur la théorie des niches (en utilisant les lois de Diamond) ainsi que

sur la théorie neutre (en utilisant le modèle de Hubell) de motifs d'abondance d'espèces microbiennes ont été réalisées (Horner-Devine *et al.* 2007, Sloan *et al.* 2006). Actuellement, le nombre de "motifs de damier" parfait est significativement plus important dans les communautés microbiennes de sol, sédiments, eau douce et marines, que le nombre de motifs produits par des modèles nuls (Horner-Devine *et al.* 2007). Cela conforterait les lois de Diamond, cependant ces motifs peuvent être créés aussi bien par la compétition que par un filtre de l'habitat. Les travaux Sloan *et al.* (2006) utilisent un modèle de Hubell modifié qui prend en compte différents taux de croissance (ce qui relâche la théorie neutre). Sous cette forme la théorie neutre a été capable de décrire des abondances microbiennes observées dans des lacs, et les voies respiratoires inférieures, mais pas dans les fèces humains. En conclusion les processus écologiques et neutres semblent intervenir dans la structuration des communautés microbiennes et idéalement les deux devraient être pris en compte pour l'étude des assemblages des communautés microbiennes.

1.4 Caractérisation des communautés par le séquençage haut débit

Dans l'environnement les micro-organismes cohabitent en communautés mixtes dont la structuration et la complexité varient dans les différents biotopes. Or la richesse d'une communauté peut varier de 6 taxons dans un biofilm de DMA (Ram *et al.* 2005) à 10^6 par gramme de sol (Wilmes et Bond 2006). Il est pourtant nécessaire pour pouvoir étudier les communautés microbiennes de posséder un inventaire exhaustif des organismes présents. Dans l'accomplissement de cet objectif les méthodes culturales sont vite limitées car la grande majorité des procaryotes sont réfractaires à la mise en culture (Amann *et al.* 1995). Les séquenceurs Sanger à capillaire permettent de produire 96 kb en 3 heures (Morozova et Marra 2008) ce qui peut être suffisant pour la caractérisation de milieux peu riche en micro-organismes. Cependant la préparation des séquences pour le séquençage Sanger nécessite une amplification *in vivo* (clonage) qui est coûteuse en temps et matériel et dont le succès n'est pas garanti. Les technologies de séquençage de seconde génération permettent une plus grande profondeur de séquençage en évitant la contrainte du clonage qui est remplacée par une amplification PCR. Cela permet un échantillonnage beaucoup plus exhaustif

TABLE 1.3 – Pourcentage de la fraction cultivable de la diversité microbienne pour différents écosystèmes. Ce pourcentage représente le nombre d'unités formant des colonies bactériennes par rapport au nombre de bactéries totales dénombrées par comptage direct au microscope (Amann *et al.* 1995).

| Écosystème | Fraction cultivable en % |
|------------------------------|--------------------------|
| Eau de mer | 0,001 – 0,1 |
| Eau douce | 0,25 |
| Lac mésotrophique | 0,1 – 1 |
| Eaux d'estuaires non pollués | 0,1 – 1 |
| Boues activées | 1 – 15 |
| Sédiments | 0,25 |
| Sol | 0,3 |

des communautés même complexe et d'accéder aux procaryotes "rares" (Fuhrman 2009). L'autre avantage de cette technologie est la possibilité de multiplexage, à savoir de pouvoir séquencer plusieurs échantillons en même temps (jusqu'à quelques centaines théoriquement) en intégrant dans les amorces une séquence unique de 6 à 8 bases (code barre) pour chaque échantillon (Meyer *et al.* 2008). Le séquençage en parallèle de plusieurs échantillons réduit encore les coûts de séquençage, mais aussi la profondeur d'échantillonnage.

1.4.1 Le pyroséquençage 454

La première description du pyroséquençage a été faite par Ronaghi *et al.* (1996) . Ce fût la première nouvelle technologie de séquençage depuis la méthode Sanger (Sanger *et al.* 1977). Elle se base sur la synthèse du brin d'ADN plutôt que sur des fragments terminaux. Les désoxyribonucléotides tri-phosphates (dNTP) sont ajoutés séquentiellement par l'ADN polymérase en utilisant le brin à séquencer comme matrice. À chaque base ajoutée le di-phosphate clivé par la polymérase est converti en ATP par une ATP sulfurylase. Les molécules d'ATP générées sont utilisées par une luciférase qui va produire de la lumière. Enfin le signal lumineux est enregistré, et la succession de ces signaux est traduite en séquence par la machine. C'est sur ce procédé que Jonathan Rothberg développa la technologie 454 pour la société 454 Life Sciences Corp. Le premier séquenceur était capable de générer 25 Mb (Mega bases) en 4 heures avec une précision de plus de 99 % (Margulies *et al.* 2005), actuellement le système 454 GS-FLX permet de générer environ 500 Mb par réaction de séquençage avec une taille moyenne de fragment de 450 bases en 12 heures. Je ne développerai

pas le fonctionnement spécifique des séquenceurs 454, il est disponible dans cette publication (Rothberg et Leamon 2008). L'avantage majeur de cette technologie fut la longueur des séquences produites, bien plus importante que les technologies concurrentes comme Illumina. Son inconvénient majeur est sa sensibilité aux erreurs. Le premier type d'erreur est la détection des homopolymères car après une incorporation successive de plus de 8 bases identiques l'intensité du signal lumineux n'est plus linéaire (Margulies *et al.* 2005). La deuxième source d'erreurs est la perte de synchronie, si une base reste dans le puits de séquençage après rinçage elle peut être intégrée lors du cycle suivant (création d'une insertion). Ce phénomène peut aussi arriver si la concentration en base n'est pas suffisante dans le puits, aboutissant à une séquence incomplète. Ces erreurs sont problématiques lors du séquençage de séquences nouvelles et rend cette technologie inadaptée aux études de polymorphisme. Cependant la plate forme et les réactifs ont subi des améliorations continues permettant de diminuer les erreurs tout en allongeant les séquences produites. Pourtant cette technologie a souffert de la concurrence, avec le "pair-end sequencing" la technologie Illumina rattrapa son retard de longueur de séquences et la technologie Ion Torrent (Life Technologies) rendit le séquençage 454 trop cher. En octobre 2013 Roche annonça la fermeture de 454 Life Sciences pour le courant de l'année 2015 et arrêtera le soutien des appareils pour mi-2016.

1.4.2 Analyse des données issues du séquençage

Le but de cette partie n'est pas de faire un inventaire exhaustif des outils disponibles pour l'analyse des données de séquençage haut débit dans le cadre d'une étude de communauté microbienne. Elle est issue de mon expérience acquise au cours de ma thèse et illustre les questions que je me suis posées lors de l'analyse de mes données et les réponses que j'ai obtenues. Le séquençage n'est pas une finalité mais une technologie d'acquisition de données comme peut l'être un chromatographe par exemple. Il est important de connaître ce qui est réellement mesuré par l'appareil pour pouvoir interpréter les données qu'il produit. Le séquenceur 454 mesure un signal lumineux produit par une luciférase lors de l'intégration d'une base par la polymérase. Ce signal est ensuite interprété par la machine et restitué sous la forme d'une séquence (l'appel de bases). L'algorithme *PyroNoise* (Quince *et al.* 2011) permet de réduire les erreurs d'appel de bases et d'identifier les régions de qualité médiocre dans les séquences et les erreurs de lectures (Schloss *et al.* 2011). Lorsque l'on est confiant

de la qualité des séquences obtenues, il n'est pas avéré que les séquences proviennent de l'échantillon étudié et n'ont pas été créées par les réactions PCR antérieures au séquençage. Une source majeure d'erreurs de la PCR est la production de chimères, qui sont des assemblages de fragments de séquences abondantes. *Uchime* (Edgar *et al.* 2011) est un algorithme efficace pour identifier ces chimères qui vont créer une diversité artificielle dans l'échantillon étudié (Schloss *et al.* 2011). Enfin il est aussi nécessaire de se demander lors d'un séquençage parallèle si chaque échantillon sera séquençé de la même façon. Pour s'en assurer, le séquençage d'une communauté témoin (dont la composition est connue) avec ses échantillons est idéal (Pinto et Raskin 2012, Schloss *et al.* 2011), à défaut on peut procéder à un sous échantillonnage de chaque échantillon avec le même nombre de séquence.

Dans cette étude les séquences générées sont celles des gènes de l'ARN ribosomique 16S (ARNr 16S). Ce marqueur moléculaire est utilisé préférentiellement depuis des décennies pour l'identification des procaryotes (Woese et Fox 1977). En effet, il possède les avantages d'une présence universelle, d'une fonction conservée, ainsi que d'être constitué d'une alternance de domaines variables et de domaines conservés permettant d'étudier des divergences anciennes et plus récentes entre les organismes (Stackerbrandt et Rainey 1995). La méthode pour identifier les procaryotes à partir de la séquence du gène de l'ARNr 16S est la comparaison de séquence et le calcul d'un taux d'identité de séquence. Cela nécessite donc la définition de seuils pour lesquels on considère que les deux séquences appartiennent au même groupe phylogénétique (espèce, genre, famille etc...). Stackerbrandt et Goebel (1994) proposèrent que 97 % d'identité de séquence d'ARNr 16S permet de différencier deux espèces procaryotes. Il est admis pour les niveaux taxonomiques supérieurs que les seuils de 95 %, 90 % et 80 % d'identité permettent de différencier respectivement les genres, les familles et les *phyla* procaryotes (Schloss et Handelsman 2004). L'utilisation de ces seuils arbitraires est largement discuté (Ward 1998, Rosselló-Mora 2003), de plus ils sont à adapter en fonction de la région de l'ARNr 16S séquençée (Kim *et al.* 2011). Cependant ils sont très utiles dans un contexte de publication et permettent la comparaison des études (Hughes *et al.* 2001). Une autre caractéristique du gène de l'ARNr 16S est d'être présent en plusieurs copies dans les génomes, avec des cas extrêmes comme *Bacillus subtilis* (10 copies) et *Clostridium paradoxum* (15 copies) (Rainey *et al.* 1996). Pour cette raison on ne peut pas affirmer que deux séquences affiliées à la même espèce appartiennent à deux organismes différents. On

travaille donc avec des Unités Taxonomiques Opérationnelle (OTU), qui sont des groupes de séquences dont les taux d'identités sont supérieurs à un seuil défini (par exemple dans un OTU défini à 97 % de similarité, toutes les séquences qui partagent plus de 97 % d'identité sont considérées comme appartenant à la même espèce). Le calcul de ces OTU est basé sur un alignement des séquences et plusieurs algorithmes existent pour réaliser cela. Cependant, l'ARNr 16S possède une fonction définie dans la cellule qui lui impose une structure spatiale. Certaines parties de la séquence du gène sont donc contraintes par cette structure et cette information doit être prise en compte par l'algorithme pour réaliser un alignement avec la meilleure précision (Schloss 2012). Wang *et al.* (2012) argumentèrent le fait que le gain de précision n'était pas assez avantageux par rapport au coût en ressource informatique demandé. Cependant, lorsque le coût informatique n'est pas limitant, prendre en compte la structure secondaire est une bonne chose.

Les communautés sont donc définies par leur composition en OTU dont on peut calculer des indices pour estimer la richesse spécifique et la diversité. Ils permettent de comparer les communautés en fonction de leur composition en OTU ou bien de leur similarité phylogénétique. Cependant une constante des communautés microbiennes est la présence de quelques OTU majoritaires et d'une multitude d'OTU rares (représentant moins de 1 % de l'effectif total) (Fuhrman 2009). Or les indices de diversité sont plus ou moins sensibles aux effectifs des espèces. L'indice de Shannon (Shannon 1948) favorise les espèces rares alors que Simpson (Simpson 1949) favorisera les espèces dominantes (Shaw *et al.* 2008). De même, les indices d'estimation de la richesse (Chao1 (Chao 1984) et ACE (Chao et Lee 1992)) sous-estiment la richesse réelle lorsque la communauté n'est pas suffisamment échantillonnée (Hughes *et al.* 2001). Ces indices ne sont pas idéaux pour l'étude des communautés microbiennes mais permettent de comparer les études même si l'échantillonnage n'est pas exhaustif (Shaw *et al.* 2008). On peut aussi comparer les différences de structures entre des communautés avec des distances non phylogénétiques (Bray-Curtis (Bray et Curtis 1957)) ou phylogénétiques (UniFrac (Lozupone et Knight 2005)). La distance de Bray-Curtis se base sur les espèces partagées entre les communautés et a le défaut d'être biaisée par les espèces présentant les plus grandes variations d'effectifs entre les communautés. Si quelques espèces concentrent les plus grandes variations d'effectifs alors seul l'effet de ces espèces sur les différentes communautés sera visible (Kindt et Coe 2005). La méthode UniFrac mesure la

distance phylogénétique entre les ensembles de taxons dans un arbre phylogénétique comme la fraction de la longueur de branche de l'arbre qui mène à la descendance de l'un ou l'autre environnement, mais pas les deux. La méthode UniFrac apporte plus d'information que les méthodes utilisant le comptage brut des OTU car elle prend en compte les relations évolutives entre les espèces composant les communautés comparées (Lozupone et Knight 2005). La méthode Bray-Curtis elle, considère que tous les OTU sont également liés les uns aux autres. La distance UniFrac est la mieux adaptée pour séparer des communautés complexes, ayant des échantillonnages différents et/ou n'ayant pas été séquencés sur les mêmes régions de l'ARNr 16S (Lozupone et Knight 2005, Knights *et al.* 2011, Hamady et Knight 2009).

On peut tester par des approches phylogénétiques l'importance des facteurs biotiques et abiotiques sur la structuration des communautés. Les indices NRI (Net Relatedness Index) et NTI (Nearest Taxa Index) mesurent le degré de regroupement des taxons dans un arbre phylogénétique d'un échantillon par rapport au nombre total de taxons (Webb *et al.* 2002). Les valeurs positives indiquent que les membres d'une communauté sont regroupés alors que des valeurs négatives indiquent que les membres d'une communauté sont également distribués à travers les différents taxons. En d'autres termes, une valeur positive illustre une communauté où les membres sont en moyenne plus proches les uns des autres qu'ils ne le sont avec les membres des autres communautés étudiées (Horner-Devine et Bohannan 2006). En considérant que les taxons proches partagent des traits écologiques communs, on peut extraire deux motifs de ces analyses :

- valeur positive et un regroupement phylogénétique qui est le plus souvent attribué à un filtre de l'habitat (biotope) qui structure la communauté ;
- valeur négative et une dispersion phylogénétique qui indique une structuration par des facteurs biotiques (biocénose) comme la prédation, la compétition ou la facilitation (Amaral-Zettler *et al.* 2011).

On peut tester les NRI et NTI des communautés étudiées contre des modèles nuls, des communautés générées aléatoirement à partir des données des communautés observées (Webb *et al.* 2002, Horner-Devine et Bohannan 2006). La génération de milliers de communautés aléatoire permet d'établir une P valeur, permettant d'apprécier si la valeur de NTI ou NRI observée est significativement différente (ou non) de celle des modèles nuls. Si elle ne l'est pas, la communauté serait plutôt structurée par des événements stochastiques (colonisation,

dérive génétique) (Rosindell *et al.* 2011). Une autre méthode, non phylogénétique, pour étudier la structuration des communautés est le calcul du C-score (Stone et Roberts 1990) (voir Chapitre 2, 2.11). Si le C-score calculé sur les données observées est supérieur à celui calculé sur les modèles nuls, alors des espèces co-occurrent moins souvent qu'elles ne le feraient par hasard (il y a une ségrégation des taxons). Au contraire, un C-score inférieur à celui calculé sur les modèles nuls, indique que des espèces co-occurrent plus souvent qu'elles ne le feraient par hasard (il y a une agrégation des taxons). Une ségrégation des taxons peut s'expliquer par la compétition, un filtre de l'habitat alors qu'une agrégation peut être dû à des relations mutualistes ou syntrophiques (Horner-Devine *et al.* 2007).

Après avoir déterminé que les communautés sont structurées par leur environnement, une question intéressante est de savoir quels sont les facteurs structurant ces communautés. Pour cela, il faut réaliser des analyses multivariées. L'analyse canonique des correspondances (ACC) est très utilisée chez les écologues car elle permet une interprétation relativement aisée des assemblages d'espèces. Cela parce que dans la représentation graphique de l'ACC, les espèces sont ordonnées le long des axes canoniques en fonction de leur optima écologique (Borcard *et al.* 2011). Cependant cette analyse donne plus d'importance aux variations des espèces rares qu'à celles des espèces communes. Il est donc indiqué d'utiliser cette analyse lorsque les espèces rares sont bien échantillonnées et considérées comme des potentiels indicateurs de la communauté. Sinon on peut éliminer les espèces rares de l'analyse avant d'effectuer l'ACC (Borcard *et al.* 2011). Une autre méthode est l'arbre de régression multivarié (MRT) (Glenn 2002). Le résultat est un arbre dont les feuilles sont les communautés choisies pour minimiser la somme des carrés du groupe et où chaque partition est définie par une valeur seuil de la variable explicative. La solution retenue est celle qui a le plus grand pouvoir prédictif. Ce qui fait qu'à la différence des méthodes d'ordination qui se focalisent sur le pouvoir explicatif d'une variable, la MRT est indiquée pour des applications de gestion environnementale (Borcard *et al.* 2011). Cette méthode est très versatile et peu être utilisée s'il y a des données manquantes et que les relations entre les variables explicatives et les communautés ne sont pas linéaires.

Enfin si on considère qu'une communauté contient des organismes qui interagissent, il est aussi possible à partir des données de séquençage d'ARNr 16S, d'identifier des micro-organismes présentant des interactions. On représente une communauté sous la forme d'un

réseau de co-occurrence, avec des nœuds représentant les OTU et des arêtes représentant une relation significative entre deux OTU (Faust et Raes 2012). Le calcul des arêtes peut se faire par une matrice de distance ou de similarité (calculé par exemple avec Pearson, Spearman, distribution hypergéométrique et l'indice de Jaccard) pour déterminer des relations deux à deux, ou bien des régressions multiples pour détecter des relations impliquant plus de 2 protagonistes. On réalise ensuite une distribution de valeurs aléatoires en répétant le calcul de la valeur un grand nombre de fois (souvent 1 000) pour obtenir une P valeur permettant de tester la robustesse de la relation. Au regard du grand nombre de valeurs calculées, les P valeurs subissent généralement une correction (Bonferroni ou Benjamini–Hochberg (Benjamini et Hochberg 1995)). Une fois le réseau construit, on peut utiliser des outils mathématiques d'étude des graphes. On peut citer notamment les mesures de connectivité d'un nœud (Node Degree) ou en d'autres termes le nombre de relations qu'il possède. Les réseaux biologiques possèdent beaucoup de nœuds faiblement connectés et quelques-uns fortement connectés appelés *Hub*. La structure de ces réseaux semble robuste à la suppression aléatoire des nœuds mais pas à la suppression des *Hub* (Faust et Raes 2012). Cependant l'interprétation écologique de ces *Hub* est délicate car ils peuvent être perçus comme des espèces clefs. Cependant les effets de la suppression d'un *Hub* peuvent être faibles si celui-ci est l'illustration d'une superposition de niche. Au contraire un nœud faiblement connecté mais étant un membre critique d'une chaîne trophique peut avoir une grande influence sur la structure du réseau (Faust et Raes 2012). Il est possible de détecter dans un réseau des groupes de nœuds ayant plus d'interactions en leur sein qu'avec le reste du réseau (EAGLE, (Shen *et al.* 2009)) et cela même si ces groupes se superposent. Ces groupes peuvent indiquer des organismes partageant des niches communes ou ayant un lien trophique. Le réseau est un outil qui propose beaucoup de possibilités mais son interprétation doit être faite avec énormément de prudence. Tout d'abord, les méthodes d'inférence des réseaux ne sont pas capables de différencier des interactions écologiques (partage de niche ou partage de ressource par exemple). Ensuite, la présence de nombreux zéros dans les matrices d'abondance de communautés microbiennes, notamment dans les taxons rares, pose aussi des problèmes d'interprétation. Le zéro doit-il être interprété comme une absence de ce taxon ou bien une absence de détection. Cela fait que les taxons rares sont souvent écartés des analyses. Enfin le fait que les matrices d'abondance aient peu d'échantillons par rapport au nombre de taxons, affecte le calcul des

corrélations et empêche le calcul classique de corrélation partielle (Faust et Raes 2012) (qui permettrait de détecter des interactions dirigées comme le commensalisme). L'utilisation des réseaux de co-occurrence pour l'étude des communautés microbiennes est très récente et nécessite d'être abordée avec prudence. Il n'est actuellement pas possible de comparer les résultats de réseaux provenant de différentes études car il n'existe pas de méthode standard et l'influence des différents indices sur l'inférence des réseaux est très peu connue (Faust et Raes 2012).

Chapitre 2

Procédures expérimentales

2.1 Échantillonnage

Douze sites ont été considérés (Table 2.1, et Figure 2.1) . Ils étaient situés dans quatre pays de trois régions minières : la ceinture de pyrite ibérique (CPI) en Espagne et au Portugal , les Cévennes en France et la ceinture d'étain en Bolivie.

TABLE 2.1 – Coordonnées GPS et dates des campagnes de prélèvement des 12 sites d'étude.

| Sites | Pays | Coordonnées GPS (Degré décimal) | date de prélèvement |
|---------------|----------|---------------------------------|---------------------|
| BOL_pot | Bolivie | -65,738550° ; -19,626820° | 05/10/2010 |
| BOL_mil | Bolivie | -68,148764° ; -16,319103° | 09/10/2010 |
| BOL_oru0 | Bolivie | -67,133950° ; -17,963917° | 07/10/2010 |
| BOL_oru1 | Bolivie | -67,124056° ; -17,961358° | 06/10/2010 |
| FRA_carn_cowg | France | -4,001670° ; 44,116940° | 06/10/2010 |
| FRA_carn_s1 | France | -4,002220° ; 44,117220° | 06/10/2010 |
| ESP_odiel_st | Espagne | -6,976000° ; 37,798694° | 08/04/2010 |
| ESP_R.T_pt1 | Espagne | -6,551500° ; 37,722111° | 08/04/2010 |
| ESP_R.T_pt2 | Espagne | -6,551056° ; 37,722361° | 08/04/2010 |
| ESP_R.T_ri | Espagne | -6,554333° ; 37,720639° | 08/04/2010 |
| ESP_odiel_th | Espagne | -7,071361° ; 37,599806° | 08/04/2010 |
| PRT_sd | Portugal | -7,504750° ; 37,649361° | 09/04/2010 |

Des sédiments ont été prélevés en trois exemplaires sur chaque site d'étude. Immédiatement après la collecte, les échantillons de sédiment ont été stockés dans de l'azote liquide à -80 °C jusqu'au traitement pour l'extraction d'ADN. Des échantillons d'eau dédiés aux mesures physico-chimiques (500 ml) ont été immédiatement filtrés sur des membranes Millipore de 0,22 μm montées sur des porte-filtres Sartorius en polycarbonate. Pour la mesure du fer

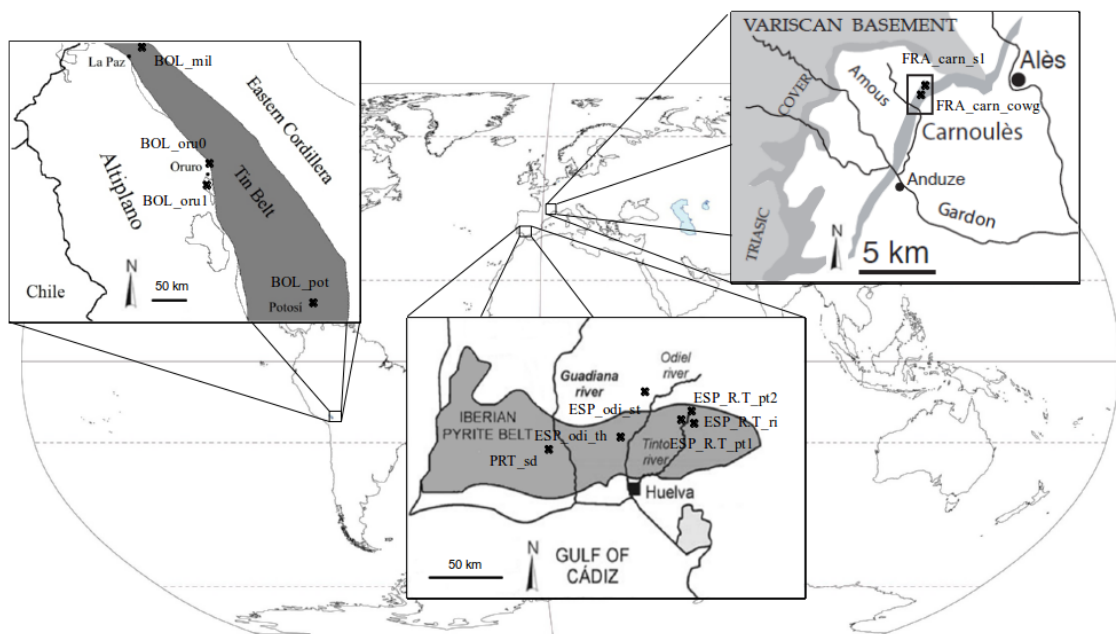


FIGURE 2.1 – Planisphère où sont représentées les localisations des 12 sites d'études.

total, l'eau filtrée a été acidifiée à $\text{pH} = 1$ avec une solution d'acide nitrique (14,5 M) et stocké à 4 °C dans des flacons en polyéthylène jusqu'à l'analyse. La spéciation des espèces d'arsenic a été faite en ajoutant 10 μL d'eau filtrée à 0,5 mL d'une solution d'EDTA à 0,25 M (Bednar *et al.* 2002). Ensuite, les flacons ont été remplis avec 10 mL d'eau déionisée et ont été conservés dans l'obscurité.

2.2 Mesures physico-chimiques

Les paramètres physico-chimiques que sont le pH, la température, la concentration en oxygène dissout, le potentiel red/ox et la conductivité de l'eau libre ont été mesurés comme il a été décrit dans Bruneel *et al.* (2011). La détermination de la concentration en arsenic totale dissoute a été réalisée par ICP-MS en utilisant un Thermo X7 series avec une calibration externe conventionnelle comme cela est décrit dans Bruneel *et al.* (2011). Les analyses des espèces inorganiques d'arsenic (AsIII, AsV) ont été réalisées avec une chromatographie échangeuse d'anions, possédant une colonne Hamilton PRP-X100 de 25 cm x 4,1 mm, avec un système de production de gradient Variant ProStar. Couplée à une spectrométrie d'absorption atomique par génération d'hydrure (VGS 200, FISOONS, France), avec un spectromètre de fluorescence atomique (Excalibur, PSA analytical, GB) (Casiot *et al.* 2005). La limite de détection de l'As III était de 172 ng.L^{-1} et de 458 ng.L^{-1} pour l'As(V). La concentration en fer total a été déterminée par spectrométrie d'absorption atomique de flamme. La concentration en sulfate a été mesurée après précipitation du BaSO_4 avec du BaCl_2 et mesure spectrophotométrique à 650 nm (Rodier *et al.* 1996). Le carbone total a été mesuré par la méthode de la perte au feu. Les échantillons de sédiment ont été séchés à 60 °C durant 48h, puis pyrolysés à 500 °C durant 16h. La différence de poids entre les échantillons secs et pyrolysés donne la quantité de carbone totale du sédiment que nous avons exprimée en g de carbone / g de sédiment sec.

2.3 Extraction ADN

Les sédiments des DMA sont riches en cations métalliques ce qui a pour effet de faire précipiter prématurément l'ADN. Pour augmenter l'efficacité de l'extraction, les sédiments

(des aliquotes d'environ 0,5 g) ont été lavés trois fois avec 1 mL d'une solution d'EDTA à 40 mM préparée dans un tampon Tris-HCl à 100 mM à pH 7, comme cela a été décrit dans Gough et Stahl (2011). Ensuite, le kit DNA Power Soil (Mo Bio Laboratories, Solana Beach, CA, USA) a été utilisé pour extraire l'ADN des sédiments. Les extractions d'ADN ayant servi au séquençage des gènes codants pour l'ARN ribosomique 16S ont été faites en suivant les instructions du kit. Cependant pour les extractions destinées au séquençage des gènes codant pour les transporteurs d'arsenic, une modification au protocole du fournisseur a été apportée. Après le lavage avec la solution d'EDTA, le tampon de lyse du kit a été remplacé par un tampon phosphate $\text{NaH}_2\text{PO}_4/\text{Na}_2\text{HPO}_4$ à 1 M pH 8 avec 15 % (V/V) d'éthanol. Une étape d'incubation à 80 °C durant 40 min a été ajoutée avant la lyse cellulaire. Cela pour augmenter l'efficacité de l'extraction dans des sédiments argileux comme il est décrit dans Direito *et al.* (2012). Les extractions ont été faites sur chaque triplicata d'échantillonnage. Ensuite 10 μL de chaque triplicatas ont été mélangés pour la PCR.

2.4 Réaction de polymérisation en chaine

Les régions V3 à V5 du gène codant pour l'ARNr 16S ont été amplifiées pour les *archaea* et les bactéries. La liste des amorces utilisées est présentée dans le tableau (Table 2.2). En raison du pyroséquençage, des amorces fusions ont été fabriquées. Elle contiennent un adaptateur ainsi qu'un identifiant multiplex de 10 bases (MID) spécifique à chaque échantillon (Table 2.3), en plus de sa composante gène spécifique (Figure 2.2).

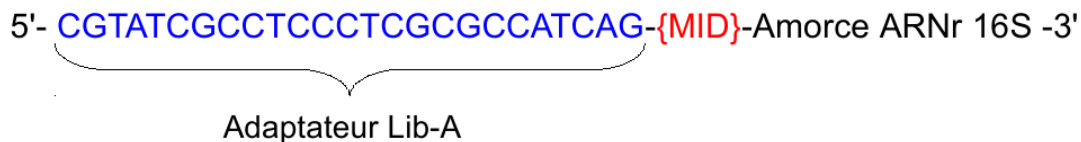


FIGURE 2.2 – Schémas représentant les amorces fusions utilisées pour le séquençage des gènes codant pour l'ARNr 16S.

L'amplification des gènes codant pour l'ARNr 16S a été réalisée à partir de 1 μL pour les bactéries ou de 2 μL pour les *archaea*. La concentration finale des amorces était de 0.4 μM dans une formule commerciale de AmpliTaq Gold®360 (Applied Biosystems) pour un

TABLE 2.2 – Liste des amorces utilisées pour les amplifications PCR.

| Gène cible | Amorces | Référence |
|-----------------------------------|--|---|
| <i>arnr</i> 16S bactérie | 357_F 5'-CCTACGGGAGGCAGCAG-3' 926_R 5'-CCGTCAATTCMTTTRAGT-3' | (Teske <i>et al.</i> 1996) (Lane <i>et al.</i> 1985) |
| <i>arnr</i> 16S <i>archaea</i> | 344_F 5'-ACGGGGYGCAGCAGGCGCGA-3' 915_R 5'-GTGCTCCCCCGCCAATTCCT-3' | (Raskin <i>et al.</i> 1994) (Stahl et Amann 1991) |
| <i>arsB</i> | arsB_F 5'-GAACATCGTCTGGAAYGCNAC-3' arsB_R 5'-GTACACCACCAGRTACATNCC-3' | (Fahy <i>et al. soumis</i>) |
| <i>acr3.1</i> | acr3.1_F 5'-GGCCTGACGTNATGATGTAYCC-3' acr3.1_R 5'-GCGATGGCCAGCTCRAARTTATT-3' | (Fahy <i>et al. soumis</i>) |
| <i>acr3.2</i> | acr3.2_F 5'-TGATCTGGGTCATGATCTTCCCVATGMTGVT-3' acr3.2_R 5'-CGGCCACGGCCAGYTCRAARAARTT-3' | (Achour <i>et al.</i> 2007) |

TABLE 2.3 – Liste des identifiants multiplex utilisés pour les amplifications PCR des gènes ARNr 16S.

| MID standard 454 | Séquences | Échantillon |
|------------------|------------------|---------------|
| MID 14 | 5'-CGAGAGATAC-3' | BOL_pot |
| MID 7 | 5'-CGTGTCTCTA-3' | BOL_mil |
| MID 15 | 5'-ATACGACGTA-3' | BOL_oru0 |
| MID 17 | 5'-CGTCTAGTAC-3' | BOL_oru1 |
| MID 13 | 5'-CATAGTAGTG-3' | FRA_carn_cowg |
| MID 1 | 5'-ACGAGTGCGT-3' | FRA_carn_s1 |
| MID 6 | 5'-ATATCGCGAG-3' | ESP_odiel_st |
| MID 3 | 5'-AGACGCACTC-3' | ESP_R.T_pt1 |
| MID 4 | 5'-AGCACTGTAG-3' | ESP_R.T_pt2 |
| MID 8 | 5'-CTCGCGTGTC-3' | ESP_R.T_ri |
| MID 2 | 5'-ACGCTCGACA-3' | ESP_odiel_th |
| MID 5 | 5'-ATCAGACACG-3' | PRT_sd |

volume final de 50 μL . Pour chaque amplification, nous nous sommes efforcés d'utiliser le minimum de cycles pour éviter les biais de PCR. Les cycles de PCR furent les suivants :

- activation/dénaturation initiale : 95 °C durant 10 min
- 25 cycles (30 pour les archées) :
 - 95 °C durant 45 s
 - 55 °C (61 °C pour les archées) durant 40 s
 - 72 °C durant 1 min
- 72 °C durant 10 min

Les amorces utilisées pour l'amplification des gènes codant pour les transporteurs d'arsenic *arsB* et *acr3.1* proviennent des travaux de Fahy *et al.* (soumis). Les amorces utilisées pour amplifier le gène *acr3.2* sont celles décrites par Achour *et al.* (2007). Ces amplifications ont également été réalisées dans une formule commerciale de AmpliTaq Gold®360 dans un volume finale de 50 μL . Pour ces amplifications, nous avons utilisé 2 μL d'extrait d'ADN génomique et une concentration finale d'amorce de 0,5 μM . Tous les cycles PCR ont été précédés d'une étape d'activation/dénaturation initiale : 94 °C durant 5 min, et ont été finalisés par une étape d'élongation finale : 72 °C durant 5 min. L'amplification des gènes *arsB* et *acr3.2* a été faite dans des conditions classiques :

- 35 cycles :
 - 94 °C durant 45 s
 - 55 °C (*arsB*), 52 °C (*acr3.2*) durant 45 s
 - 72 °C durant 50 s

L'amplification du gène *acr3.1* a nécessité l'utilisation de conditions Touchdown :

- 10 cycles :
 - 94 °C durant 45 s
 - 57 °C durant 45 s (température diminuées de 0,5 °C après chaque cycle)
 - 72 °C durant 50 s
- 30 cycles :
 - 94 °C durant 45 s
 - 52 °C durant 45 s
 - 72 °C durant 50 s

2.5 Pyroséquençage

En vue du séquençage des gènes de l'ARNr 16S, les produits PCR ont été purifiés en utilisant le QIAquick PCR Purification Kit (Qiagen, Hilden, Germany), après excision des amplicons à partir d'un gel d'agarose. La concentration de chaque purification a ensuite été déterminée par une quantification au Picogreen. Enfin, les triplicata de PCR ont été réunis dans des concentrations équimolaires. Le pyroséquençage a été réalisé par la plateforme génomique Genotoul (Toulouse, France). La fusion entre nos amplicons de PCR et les adaptateurs Lib-A 454 FLX a été faite par la plateforme pour le séquençage des ARNr 16S archées, alors que les amorces utilisées pour l'amplification des gènes ARNr 16S bactériens contenaient les adaptateurs. Enfin, le séquençage a été réalisé sur 6 régions (8 au total) d'un séquenceur 454 GS20 FLX (454 Life Sciences). Quatre régions ont été utilisées pour le séquençage des ARNr 16s bactériens et deux pour le séquençage des gènes ARNr 16s archées.

2.6 Construction des banques de clones et séquençage Sanger

Le séquençage des gènes *arsB*, *acr3.1* et *acr3.2* a été réalisé à partir de banques de clones par la méthode Sanger. Les amplicons PCR ont d'abord été purifiés en utilisant le kit Illustra GFXT (GE Healthcare). Ensuite, une banque de 48 clones minimum par échantillon et par gène a été construite. Pour ce faire, des cellules chimio-compétentes *Escherichia coli* JM109 (Genotype : endA1, recA1, gyrA96, thi, hsdR17 (rk⁻, mk⁺), relA1, supE44, Δ (lac-proAB), [F' traD36, proAB, laqIqZ Δ M15]) ont été transformées par des vecteurs recombinés pGEM®-T (Promega) ayant comme insert les amplicons PCR purifiés. La ligation du vecteur dans le plasmide s'est faite en utilisant la thymidine terminale en 3' du vecteur qui s'apparie avec l'adénine terminale générée par la polymérase lors de l'amplification PCR. La réaction de ligation a été faite avec la ligase du phage T4 durant une nuit à une température de 4 °C. Environ 37 ng d'amplicons purifiés ont été utilisés pour les ligations.

Les clones recombinants ont été criblés sur des boîtes de Pétri contenant du milieu LB/agar 15 % (m/v), ampicilline 100 $\mu\text{g.mL}^{-1}$, IPTG 0,5 mM et X-Gal 100 $\mu\text{g.mL}^{-1}$. Les clones recombinants (bleus) ont ensuite été criblés par PCR pour vérifier que le vecteur fai-

sait la taille attendue et repiqués dans des micro-plaques de culture 96 puits contenant du LB/agar 15 % (m/v), ampicilline $100 \mu\text{g.mL}^{-1}$. Le séquençage des inserts a été réalisé par GATC-Biotech (Allemagne), en utilisant l'amorce M13-F (5'-GTTTTCCCAGTCACGAC-3'), qui s'hybride sur près de la région d'insertion du vecteur.

2.7 Traitement des données brutes issues du séquençage

Les données brutes fournies par les séquenceurs ne sont pas directement lisibles. Elles doivent d'abord subir un traitement informatique pour pouvoir être utilisées pour les analyses.

2.7.1 Données issues du pyroséquençage

Les données brutes contenaient 255 153 séquences d'ARNr 16S bactériens et 111 398 séquences archées. Le traitement de ces séquences a été fait en utilisant *Mothur* v.1.28.0 (Schloss *et al.* 2009). Tout d'abord l'implémentation de l'algorithme *PyroNoise* (Quince *et al.* 2011) dans *Mothur* a été utilisé pour réduire le taux d'erreur dû au séquençage en corrigeant le flowgramme original. Ensuite nous avons utilisé les seuils de qualités déterminés par Roche pour le 454 GS20 FLX pour la sélection des flowgrammes, qui doivent contenir entre 360 et 720 flows, comme il est suggéré dans la documentation de *PyroNoise*. Puis nous avons éliminés toutes les séquences présentant une erreur dans le code barre et plus de deux dans les amorces. Nous avons conservé les séquences de bactéries orientées 3'-5', et les séquences archées orientées 5'-3'. Nous avons choisi de ne pas utiliser les séquences sens et le complément inverse des séquences anti-sens car la longueur des séquences fût trop faible et les zones de chevauchement de faible qualité. La priorité a été donnée à la qualité et la longueur des séquences plutôt qu'à la quantité car le milieu étudié est peu riche. Enfin après excision des codes barres et des amorces nous avons sélectionné les séquences ayant une longueur supérieure à 400 bases et ne possédant pas d'homopolymères de plus de 8 bases.

Pour la suite du traitement les séquences ont été alignées. La stratégie d'alignement de *Mothur* consiste à :

1. trouver la séquence la plus proche dans la base de données *SILVA-SEED* (Quast *et al.* 2013) en utilisant une recherche par kmer
2. aligner la séquence candidate et la séquence de référence sans trous avec l'algorithme Needleman-Wunsch (Needleman et Wunsch 1970)
3. enfin réinsérer les trous dans l'alignement avec l'algorithme NAST (DeSantis *et al.* 2006).

Nous avons ensuite pu utiliser l'algorithme *single-linkage pre-clustering* proposé par Huse *et al.* (2010), implémenté dans *Mothur* qui réduit les erreurs liées aux réactions PCR. L'idée de base est que les séquences les plus abondantes sont celles qui ont le plus de chance de produire des erreurs. L'algorithme trie les séquences par abondance et parcourt cette liste en vérifiant s'il existe des séquences plus rares qui possèdent une différence de 4 bases de la séquence d'origine (soit une différence de 1 % dans notre cas). Les séquences répondant à ce critère sont regroupées car considérées comme identiques. Les réactions PCR entraînent la formation de chimères. Pour les détecter nous avons utilisé l'algorithme *Uchime* (Edgar *et al.* 2011). Il utilise les séquences les plus abondantes de chaque échantillon comme référence pour détecter les chimères.

Enfin nous avons attribué une affiliation taxonomique aux séquences retenues en utilisant le classifieur Bayésien de *Mothur* avec la référence taxonomique de *SILVA SEED*. Les affiliations ayant une valeur de bootstrap de plus de 80 % pour 1 000 itérations ont été retenues. Toutes les séquences reconnues comme chloroplastiques ou mitochondriales ont été enlevées.

Notre jeu de données nettoyées contient 83 514 séquences bactériennes et 36 974 séquences archées.

2.7.2 Données issues du séquençage Sanger

Ces données concernent les séquences des gènes codant pour les transporteurs d'arsenic. Elles ont été traitées manuellement en lot, par gène et par échantillon. Tout d'abord nous avons utilisé *VecScreen* proposé par le NCBI, pour identifier les portions de séquences appartenant au vecteur de clonage. Pour visualiser ces séquences nous avons utilisé le logiciel *Jalview* (Waterhouse *et al.* 2009). Puis nous les avons alignées avec l'algorithme *Muscle* (Edgar 2004). Un premier alignement permet de vérifier que les séquences sont dans le même

sens. Ensuite à l'aide de l'alignement et des électrophorégrammes nous avons corrigés les positions ayant des bases indéterminées. Ensuite nous avons utilisé *Uchime* comme précédemment pour rechercher des chimères.

Enfin les séquences nucléotidiques ont été traduites en séquences d'acides aminées avec le logiciel *FrameBot* disponible sur la plateforme *FunGene* (Fish *et al.* 2013). Ce logiciel utilise une base de données de séquences d'acides aminées de la protéine étudiée, pour trouver le cadre de lecture des séquences à traduire. Les séquences traduites des gènes *arsB*, *acr3.1* et *acr3.2* représente le jeu de données qui sera utilisé pour l'étude.

2.8 Construction des OTU

Comme expliqué précédemment les réactions de PCR et le séquençage incorporent des erreurs dans les séquences. De plus un gène peut être présent en plusieurs copies dans un même génome. Pour ces raisons nous travaillons avec des Unités Taxonomiques Opérationnelles (OTU), qui sont des groupes de séquences définis par des seuils de similarité ou de dissimilarité de séquence.

2.8.1 Construction des OTU des séquences ARNr 16S

Le jeu de données nettoyé a été une nouvelle fois aligné avec le logiciel *Infernal* (Cole *et al.* 2009), car il prend en compte la structure secondaire de l'ARNr 16s. Puis nous avons calculé dans *Mothur* une matrice de distance avec un seuil de 0,15. Cette matrice a ensuite servi à construire des OTU avec une distance de 0,05. C'est une distance qui est considérée comme correspondant au niveau taxonomique du genre (Schloss et Handelsman 2004). Les OTU ont été construits pour chaque site échantillonné, et un représentant de chaque OTU par site a été calculé. Puis une taxonomie consensus a été établie pour chaque OTU à partir des affiliations des séquences contenues dans ces OTU. Cependant le classifieur automatique n'a pas pu résoudre la taxonomie de tous les OTU. Pour ces cas particuliers, les séquences contenues dans ces OTU ont été entrés dans le logiciel *ARB* (Ludwig *et al.* 2004) pour tenter de leur donner une classification manuellement.

2.8.2 Construction des OTU des séquences des transporteurs d'arsenic

Les séquences d'acides aminés ont été alignées avec *Muscle*, puis nous avons construit une matrice de similarité de séquence dans *BioEdit* (Hall 1999). Cela car *Mothur* ne lit pas les séquences d'acides aminés. De plus il n'accepte que des matrices de dissimilarité. La matrice *BioEdit* a donc été transformée en matrice de dissimilarité pour pouvoir être lue par *Mothur* et construire les OTU. Il n'existe pas de seuils reconnus pour les séquences d'acides aminés des transporteurs ArsB et Acr3, comme c'est le cas pour le gène codant pour l'ARNr 16S, pour regrouper les séquences appartenant au même groupe taxonomique. Nous avons défini un seuil de 0,20, le choix de ce seuil est développé dans le chapitre 5.

2.9 Analyse des données de séquençage

2.9.1 Analyse des données ARNr 16S

Tout d'abord pour l'étude des structures des communautés microbiennes, le nombre de séquence a été ramené pour chaque site à celui ayant le nombre de séquences le plus faible. Soit 4363 séquences bactériennes et 1204 séquences archées. Pour réaliser ceci nous avons eu recours à un échantillonnage aléatoire. Puis nous avons évalué l'effort d'échantillonnage en calculant l'indice de couverture de *Good* (Good 1953) donnée par $C = 1 - (\frac{n}{N})$, où n est le nombre de séquences présentes seulement une fois dans l'échantillon, et N le nombre total de séquences. D'autres indices de richesse et de diversité des communautés ont été calculés dans *Mothur*, *Chao 1* (Chao 1984) : $\hat{S}_{Chao1} = S_{obs} + F_1^2 / (2F_2)$, *Shannon* (Shannon 1948) : $H' = - \sum_{i=1}^{S_{obs}} p_i * \ln(p_i)$, où S_{obs} est le nombre d'OTU observés, F_1 le nombre d'OTU possédant une seule séquence (singleton) et F_2 le nombre d'OTU possédant 2 séquences et p_i la proportion de séquences dans le $i^{ème}$ OTU, estimé par $\frac{n_i}{N}$ où n_i est le nombre de séquence dans le $i^{ème}$ OTU. Ensuite pour explorer les différences physico-chimiques entre les sites d'études nous avons construit des analyses en composantes principales (ACP) en utilisant la commande *rda()* du paquet *vegan* (Oksanen *et al.* 2012) au sein de l'environnement R. Nous avons aussi testé l'influence de la température, du pH, de la concentration en oxygène dissous, sulfate, fer, zinc, cuivre, aluminium, arsenite et arsenate, sur la structure des communautés microbienne

par une analyse canonique des corrélations (ACC) parcimonieuse avec une procédure pas à pas. Ceci en utilisant le paquet *vegan* et la commande *cca()*. Enfin nous avons calculés les distances phylogénétiques entre les sites d'études, par des distances *Unifrac* pondérées (Lozupone et Knight 2005) à partir d'un arbre construit par une méthode de maximum de vraisemblance dans *ARB* avec les représentants de chaque OTU.

2.9.2 Analyse des données des transporteurs d'arsenic

Nous avons aussi calculé avec *Mothur* l'indice de couverture de *Good*, l'indice de richesse *Chao1* et l'indice de diversité de *Shannon*. Nous avons ensuite fait une analyse phylogénétique des séquences de banques de clones. Un représentant de chaque OTU pour chaque échantillon a été utilisé pour construire les arbres phylogénétiques. Ces arbres ont été construits en utilisant *PhyML* (Guindon et Gascuel 2003), et un modèle *LG* avec une correction γ (4 catégories de site et un paramètre α estimé). L'alignement ayant servi à la construction de l'arbre a été fait avec *Muscle* et nettoyé avec *Gblock* (Talavera et Castresana 2007). Pour avoir des références dans nos arbres phylogénétiques nous avons recherché des homologues des gènes *ArsB* et *Acr3* dans les génomes séquencés disponibles dans NCBI. Le programme *iTOL* (<http://itol.embl.de/>) (Letunic et Bork 2007) nous a permis d'afficher et d'éditer les arbres phylogénétiques.

2.10 Fouille de données

Pour révéler des patrons globaux dans la distribution des micro-organismes présents dans les DMA, nous avons cherché dans les bases de données dans quels autres environnements pouvaient être retrouvés des proches parents des organismes que nous avons identifiés. Pour cela nous avons sélectionné les OTU contenant plus de 5 séquences, puis nous les avons séparés en fonction du nombre de site dans lesquels ils avaient été observés. De là, quatre catégories ont été créées : 1) 1 à 3 sites ; 2) 4 à 6 sites ; 3) 7 à 9 sites et 4) 10 à 12 sites. Les séquences représentant les OTU ont été utilisées comme requête pour une recherche BLAST (Altschul *et al.* 1990) contre la base de données *nt* de NCBI. Nous avons récupéré cinq réponses par requête avec une e-valeur minimale de 10^{-114} . Un script *Python* a été écrit pour trouver le lieu d'échantillonnage de la séquence dans le fichier *Genbank* de

la réponse BLAST. Cependant ce champ de renseignement est à la libre interprétation des contributeurs. Ce qui impose de reprendre ces informations manuellement pour les ranger en catégories normalisées. Nous avons défini 14 termes regroupant la majorité des informations apportés par les fichiers *Genbank* : *AMD* : milieu minier et acide avec une forte concentration d'éléments métalliques ; *acid* : milieu caractérisé par une fortes acidité ; *metals* : milieu contaminé par des éléments métalliques ; *hydrocarbon* : milieu contaminé par des hydrocarbures ; *freshwater* : milieu d'eau douce ; *salty* : milieu avec une forte salinité ; *geothermal* : milieu géothermique ; *animal_host* : associé à des hôtes animaux ; *organic_matter* : milieu avec une forte concentration de matière organique ; *other* : milieu qui n'a pas été clairement défini ou ne correspondant pas aux catégories précédentes ; *not_known* : le milieu n'était pas renseigné.

2.11 Construction du réseau

Nous avons exploré les relations entre les micro-organismes en construisant un réseau de co-occurrence. L'aspect non-aléatoire des corrélations a été testé en calculant le C-score (le score d'échiquier) contre un modèle nul préservant la fréquence des sites (Stone et Roberts 1990). Une unité "d'échiquier" est une matrice 2x2 où les deux OTU n'apparaissent qu'une seule fois mais sur des sites différents. La formule du C-score entre deux OTU est la suivante $C_{ij} = (r_i - S_{ij})(r_j - S_{ij})$, avec S_{ij} le nombre de co-occurrence des deux OTU, r_i le nombre de site où l'OTU1 à 1 et l'OTU2 0, enfin r_j le nombre de site où l'OTU1 a 0 et l'OTU2 1. Pour M OTU le C-score est la moyenne des C-score pour chaque paire d'OTU dans la communauté : $C = \sum_{j=0}^M \sum_{i<j} C_{ij} / P$ avec $P = M(M - 1) / 2$ qui est le nombre de paires d'espèce possibles. Pour l'inférence du réseau nous avons calculé toutes les corrélations des rangs de Spearman entre les OTU contenant plus de cinq séquences (247 OTU) $\rho = 1 - \frac{6 \sum D^2}{n(n^2-1)}$ (n = taille de l'échantillon, $\sum D^2 =$ somme des différences au carré). Cela pour enlever les OTU qui sont peu représentés et réduire la complexité du réseau. Nous avons considéré une corrélation comme robuste si le coefficient de Spearman (ρ) était $> 0,6$ ou $< -0,6$ et statistiquement significatif $p < 0,01$. Un nœud du réseau représente un OTU défini à 95 % de similarité, et les liens représentent une corrélation entre les nœuds. Ces calculs ont été réalisés avec le paquet *vegan* dans l'environnement R. Le réseau a été visualisé avec *Cytoscape* v 3.0 (Shannon *et al.*

2003). L'algorithme de clusterisation de graph *EAGLE* (Shen *et al.* 2009) implémenté dans le greffon *ClusterViz* de *Cytoscape* a été utilisé pour détecter des clusters dans le réseau, avec comme paramètres un "CliqueSize Threshold" de 4 et un "ComplexSize Threshold " de 2.

Deuxième partie

Principaux résultats

Chapitre 3

In pursuit of drivers of communities assemblage in AMD : a biogeographic approach

3.1 Introduction

Les organismes s'organisent en communautés qui sont des assemblages d'individus appartenant à des espèces différentes. La composition des communautés change dans le temps et l'espace, ce qui amène à une question écologique fondamentale qui est la compréhension de l'assemblage des organismes en communautés. C'est un sujet qui est en discussion depuis des décennies. Depuis les travaux de Diamond (1975) qui mettaient en avant la compétition pour les ressources pour déterminer la composition des communautés, puis ceux de Hubbell (2001) qui pointa des mécanismes stochastiques de dispersion, de naissance et de mort des individus; les discussions sont toujours d'actualité. Bien que ces processus aient pu être considérés comme exclusifs, un consensus tend à les combiner dans un "concept de communauté intégré" (Lortie *et al.* 2004). Cette question a été peu documentée jusqu'à récemment et elle a été fréquemment biaisée en faveur de l'ubiquité des micro-organismes à cause des méthodes culturales qui ne sont pas capables de déterminer exhaustivement les communautés microbiennes naturelles (Green et Bohannan 2006). Cependant le développement des méthodes de biologie moléculaire et de génomique a permis l'étude de l'influence des paramètres déterministes et stochastiques dans l'assemblage des communautés microbiennes

(Horner-Devine et Bohannan 2006, Faust et Raes 2012).

Pour évaluer l'impact des paramètres environnementaux et de la dispersion sur la structuration des communautés procaryotes, nous avons choisi les drainages miniers acides comme sites d'étude. En effet ces milieux présentent une importante pression environnementale par une forte acidité et des concentrations élevées en métaux et métalloïdes toxiques (Baker et Banfield 2003). Ces conditions sont rarement retrouvées en dehors des DMA permettant d'évaluer la distribution spatiale entre les DMA de différentes régions.

Nous avons utilisé une procédure d'échantillonnage standardisée de la communauté procaryote de 12 sites de DMA, par un pyroséquençage 454 des régions V3-V5 du gène de l'ARNr 16S. Cela pour obtenir un référencement exhaustif des *Bacteria* et des *Archaea* dans ces 12 sites provenant de trois régions géographiques : sud-est de la France, sud-ouest de la péninsule ibérique et la Bolivie en Amérique du Sud. En comparant les données taxonomiques et environnementales de ces différents sites, nous souhaitons identifier la part des facteurs biotiques et abiotiques dans la structure des communautés procaryotes des DMA. De plus en explorant des bases de données publiques nous voulons mettre en évidence la potentielle spécialisation des procaryotes habitant les DMA.

3.2 Contribution scientifique

Ces travaux sont présentés sous la forme d'un article scientifique qui sera soumis pour publication très prochainement.

In pursuit of drivers of communities assemblage in AMD: a biogeographic approach

Fabien Javerliat¹, Odile Bruneel², Christian Paroissin³, Aurélie Volant², Corinne Casiot², Sandra Laoudi¹, Anne Fahy¹, Solange Karama¹, Aguasanta Miguel Sarmiento⁴, Samanta R. Sánchez A. ⁵, Volga Iñiguez⁵, José Miguel Nieto⁴, Robert Duran¹, and Béatrice Lauga¹

1- Equipe Environnement et Microbiologie, UMR CNRS 5254, IPREM, EEM, Université de Pau et des Pays de l'Adour, IBEAS, BP 1155, 64013 Pau Cedex, France

2- Laboratoire Hydrosociences Montpellier UMR 5569 (CNRS - IRD - Universités Montpellier I et II Université Montpellier II, Bâtiment 40, CCMSE Place Eugène Bataillon 34095 MONTPELLIER cedex 05

3-Laboratoire de Mathématiques et de leurs Applications - PAU UMR CNRS 5142 Bâtiment IPRA - Université de Pau et des Pays de l'Adour Avenue de l'Université, BP 1155, 64013 Pau Cedex, France

4- Departamento de Geología, Universidad de Huelva, Avda. Fuerzas Armadas s/n, 21071 Huelva, Spain

5- Instituto de Biología Molecular y Biotecnología, Facultad de Ciencias Puras, Universidad Mayor de San Andrés, C. 27 Campus Universitario Cota Cota, La Paz, Bolivia

6- INRA Auzéville, Plateforme Génomique Chemin de Borde Rouge 31326 Castanet-Tolosan.

Abstract

This work consists of a comparative analysis of twelve AMD from three regions of the world: France, Iberian pyrite belt and the tin belt of the Bolivian altiplano. They are all characterized by a low pH and high concentrations of several metals and metalloids. We use 454 barcode pyrosequencing of the 16S rRNA genes to describe bacterial and archaeal community of the AMD. Then the use of computer processing, statistical analysis, network approach and data mining we were able to assess microbial dispersion in AMD, highlight environmental parameters shaping AMD communities and reveal specific AMD microorganisms. These data contribute to expand our knowledge on the phylogenetic diversity of prokaryotic communities but more importantly allowed to shed light on the distribution of prokaryotes in AMD-impacted environment at a large scale and give clues to decipher assembly rules in these ecosystems.

Introduction

Organisms on Earth are organized into communities, that are assemblages within a specified place of individuals that belong to different species. The composition of each community changes over space and time. From this intuitive observation of our landscapes and of their biotope was raised one of the most fundamental and oldest quest in ecology: understand how organisms are assembled in a community. This captivating question is the subject matter and the challenge of community ecology. Its resolution is crucial in the context of biodiversity loss to comprehend how environmental changes affect communities and to understand how resilience or restoration take place when communities are sternly affected. Since decades several authors have attempted to identify "assembly rules" to explain species assemblage in communities. One of the pioneer works was presented by Diamond (Chase and Myers 2011; 1975) in a seminal paper that focused on the distribution of the avifauna of New Guinea's islands. Ten years of observations lead this author to conclude about the main role of competition for resources in determining the species composition of communities. However the methodology used by Diamond was controversial and challenged by Connor and Simberloff (1979) who demonstrated that the pattern of assembly of New Guinea's birds could also arise by chance without invoking interspecific competition. It is worthy to note that 40 years later the existence of assembly rules for communities is still an open topics and remains a major focus in ecology (e.g., Bell 2005; Belyea and Lancaster 1999; Chase 2003; HilleRisLambers, et al. 2012; Hubbell 2001; Lortie, et al. 2004; Ricklefs 2004; Weiher, et al. 2011). However during that time the

framework of community ecology has been enlarged and made more complex. Hence in addition to biotic interactions among species (antagonist and synergistic) and to abiotic constraints imposed to organisms by the environment, namely niche-based mechanisms, neutral processes that include dispersal of individuals from a regional pool and births and deaths at a local scale have been proposed in the last decade (Bell 2001; Chust, et al. 2013; Hubbell 2001). Although viewed sometimes as exclusive all these processes have been rapidly proposed as complementary and merged altogether in an *integrated community concept* (Lortie, et al. 2004). It is noteworthy that this underlying question was clearly debated a long time ago by some microbiologists, precursors of microbial ecology. Indeed experimental interpretation relying onto stochastic processes was put forward as soon as 1838 by Ehrenberg whereas Beijerinck in 1913 advocated for deterministic processes against Ehrenberg's standpoint (de Wit and Bouvier 2006 and discussion therein). However due to the difficulties inherent in identifying uncultivated organisms, and thus the difficulties to assess natural microbial community composition, the domain of community ecology has been until recently poorly documented for microorganisms and frequently biased toward the idea that microbial taxa were ubiquitous (Green and Bohannan 2006; Tom 2005). The advances of cultivation independent methods using molecular tools to target phylogenetic information contained in ribosomal RNA (16S or 18S) put off the frontiers of microbial ecology by facilitating the description through metagenetic analysis of the composition and structure of most microbial communities (Pace 1997; Stahl, et al. 1984; Woese 1987, 1994). Nowadays high-throughput nucleic acid sequencing and computer analytical power allow even to go further offering the possibility to get an almost exhaustive description of complex microbial communities although some pitfalls subsist, however. Indeed we have to keep in mind that the description of the real community is still limited by the lack of universal method that would permit to sample any organism from any sample whatever its nature and origin (Delmont, et al. 2012).

Thanks to these analytical developments, exciting avenues of research are now open in microbial ecology. Numerous studies shed light on the composition of microbial communities in miscellaneous habitats and contribute to expand the catalog of microbial diversity (e.g. Benlloch, et al. 1995; Bruneel, et al. 2006; Buckley and Schmidt 2003; Fierer, et al. 2008; Fulthorpe, et al. 2008; Jones, et al. 2009; Lear, et al. 2008; Roh, et al. 2009; Ward, et al. 1990). Furthermore this last decade shows a conspicuous advance in microbial community ecology with the emergence of studies that investigate the influence of deterministic or stochastic factors on microbial

assemblages (e.g. Gobet, et al. ; Horner-Devine and Bohannan 2006; Lekberg, et al. 2011; Newton, et al. 2007; Pommier, et al. 2012b; Pontarp, et al. 2012; Stegen, et al. 2012). However more investigations are still needed to establish the rules of community assembly in the microbial world. In this quest few attentions have been paid to explore among deterministic processes those that involved ecological interactions between members of different species and reports that consider this aspect are still scarce (Barberán, et al. 2012; Chaffron, et al. 2012; Eiler, et al. 2012; Fuhrman 2009; Nemergut, et al. 2013; Steele, et al. 2011; Zhang, et al. 2014a; Zhang, et al. 2014b).

Unravel the pattern of community structure over space among similar habitats provides a valuable approach to get an overall view on the underlying drivers in community assembly for a given type of environment. Acid mine drainages (AMD) provide a good model to evaluate the different facets of community assembly theories. Indeed they are characterized by strong environmental constraints, acidity and high metals concentrations that could limit the range of microorganisms in these habitats. In addition they are fragmented habitats since these conditions are rare outside the mining regions. We thus hypothesized that focusing at the same time on separate sites and regions would allow to highlight the importance of dispersal in assembly processes.

Also in attempt to discern community assembly rules at work for microbial prokaryotes in AMD we prospected and compared both archaeal and bacterial diversity patterns in twelve AMD across different biogeographical regions: Southeast in France, Southwest of the Iberian peninsula and near the South tropic in South America. In addition we explored the global repartition of the AMD prokaryotic inhabitants to disentangle their habitat breadth and highlight their dispersal ability and ubiquity.

Our work that benefits of standardized sampling procedures among the twelve sites, first used data generated with 454 barcode pyrosequencing of the V3-V5 region of the 16S rRNA gene to obtain an exhaustive catalog of the bacterial and archaeal diversities of each AMD. This data set was further used to inspect through database mining the habitat breadth of these microorganisms and underline their potential specialization. We then look for sign of dispersal among sites at any scale and finally prospect, through a comparative approach, into abiotic and biotic factors to identify drivers of community assembly in these ecosystems.

Material and methods

Sample collection

Twelve sampling sites were considered (Figure 1 and Table S1). They were located in three mining regions in four countries: the Iberian Pyrite Belt (IPB) in Spain and Portugal, the Cevennes in France and the tin belt in Bolivia. Sediments were collected in triplicates from two sites of the Carnoulès AMD (France): FRA_carn_cowg and FRA_carn_s1, three sites at the Rio Tinto AMD (Spain): ESP_R.T_pt1/pt2/ri, two sites on the Odiel river (Spain): ESP_odiel_st/th, one site at Saõ Domingos pyrite ore body (Portugal): PRT_sd, and four sites in the Bolivian tin belt: BOL_mil in the Miluni drainage, BOL_oru0 and BOL_oru1 from the surface and the bottom of a mine well still in activity in Oruro and BOL_pot from a drainage in Potosí. Immediately after collection, sediment samples were stored in dry ice and once in the lab at -80°C until DNA extraction. Water samples dedicated to physicochemical analysis (500 ml) were immediately filtered through 0.22 µm Millipore membranes. Some filtered water was immediately acidified to pH 1 with HNO₃ (14.5 M) and stored at 4°C for further total Fe and As determination analysis. For As speciation, an aliquot of 10 µl of filtered water sample was added to 0.5 ml of 5 % (v:v) 0.25 M EDTA solution then vials were completed up to 10 ml with deionized water and stored in the dark and analyzed within 24 h (Bednar, et al. 2002). Determination of As and Fe speciation and sulfate were performed within 24 h.

Physicochemical measurements

The main physicochemical parameters (pH, temperature, dissolved oxygen concentration, the redox potential and conductivity) of the running water at each sampling site were measured *in situ* as described previously (Bruneel, et al. 2011). Total dissolved As was measured by ICP-MS using Thermo X7 series as previously described (Bruneel, et al. 2011). Analyses of inorganic arsenic species (As(III), As(V)) were carried out using anion exchange chromatography (25 cm × 4.1 mm i.d. Hamilton PRP-X100 column with Varian ProStar gradient solvent delivery system) coupled to a Hydride Generation (VGS 200, FISONs, France) with an Atomic Fluorescence Spectrometry detector (Excalibur, PSAAnalytical, GB) (Casiot, et al. 2005). The detection limit obtained was 172 ng.L⁻¹ for As(III) and 458 ng.L⁻¹ for As(V), with a precision better than 5 %. Total dissolved Fe was determined by Flame Atomic Absorption Spectrometry. Sulfate concentration was determined after precipitation of BaSO₄ with BaCl₂ and spectrophotometric measurement at 650 nm (Rodier, et al. 2009).

DNA extraction, amplifications and pyrosequencing

Prior to DNA extraction and in order to improve extraction efficiency by removing divalent cations around 0.5 g aliquots were washed three times with 1 ml of 40 mM sodium EDTA prepared in 100 mM Tris-HCl (pH 7) as described by Gough and Stahl (2011). The DNA Power Soil Kit (Mo Bio Laboratories, Solana Beach, CA, USA) was then used to extract total DNA from the sediments as described by the manufacturer. Extractions were carried out on each triplicate samples. Triplicates were then pooled to amplify by PCR the V3 to V5 region of bacterial and archaeal 16S rRNA genes. Fusion primers containing the 454 FLX Lib-A adaptor were used, and a 10 bases sample-specific multiplex identifier. The 16S bacterial specific primers used were the 357 F (5'-CCTACGGGAGGCAGCAG-3') (Teske, et al. 1996) and the 926 R (5'-CCGTCAATTCMTTTRAGT-3') (Lane 1991). The archaeal 16S specific primers were the 344 F (5'-ACGGGGYGCAGCAGGCGCGA-3') (Raskin, et al. 1994) and the 915 R (5'-GTGCTCCCCCGCCAATTCCT-3') (Stahl and Amann 1991). PCR reactions containing 1 μ L (bacteria) or 2 μ L (archaea) template DNA and 0.4 μ M of primers were performed in triplicate using the AmpliTaq Gold® 360 master mix (Applied Biosystems, according to the protocol of the manufacturer). Cycling conditions were as follow: an initial activation/denaturation step at 95°C for 10 min; followed by 25 cycles (30 for archaea) at 95°C for 45 s, 55°C (61°C for archaea) for 40 s, and 72°C for 1 min; and a final 10 min extension at 72°C. PCR products were then purified using the QIAquick PCR Purification Kit (Qiagen, Hilden, Germany) after excision of the amplicon from an agarose gel. Concentration of DNA of each barcoded purified PCR products was determined using *PicoGreen*® dsDNA quantitation assay (Lifetechnology, Carlsbad, California). PCR triplicates of a same site were pooled into equimolar concentrations and pyrosequenced at the Genotoul genomic platform (Toulouse, France) on six of eight regions (4 for bacteria and 2 for archaeal) of a 454 GS20 FLX (454 Life Sciences).

Sequence processing

Raw data generated from the 454-pyrosequencing contained 255 153 bacterial sequences and 111 398 archaeal sequences. Sequence processing was done using *mothur* v.1.28.0 (Schloss, et al. 2009). We followed the processing pipeline proposed in Schloss et al., 2011. We kept sequences with a minimal length of 400 pb after barcodes and primers were trimmed. Sequences with homopolymers longer than 8 pb were removed. The sequences were aligned using the SILVA reference database implemented in *mothur*, and a single-linkage pre-clustering algorithm (Huse, et al. 2010) was applied. Chimera removal was completed using the *Uchime* algorithm (Edgar, et

al. 2011). The remaining sequences were then classified using the mothur implantation of the Bayesian classifier with the SILVA SEED reference taxonomic outline. Assignments with a bootstrap value over 80% for a 1000 bootstrap iteration were retained. All sequences recognized as chloroplasts or mitochondria were removed. Finally, the cured sequence data set contained 83 514 bacterial and 36 974 archaeal sequences. This data set was used for subsequent analysis.

Data analyses

The cured sequence were aligned with the RDP infernal aligner (Nawrocki, et al. 2009) online to take into account the secondary structure of the ribosomal RNA. Distance matrices were built with a distance cutoff of 0.15 and then used in mothur to create operational taxonomic units (OTUs) with a distance of 0.05. Consensus taxonomy was established for each OTU from the sequences taxonomy obtained with the Bayesian classifier. Classification of OTU with unresolved taxonomy with the Bayesian classifier were finalized in the ARB software (Ludwig, et al. 2004). For the analysis of the community structure the 16S rRNA samples were randomly resampled down to the lowest number of sequences obtained among the 12 sites: 4363 sequences for bacterial sequences and 1204 for archaeal sequences. We evaluated the sampling effort by calculating the Good's coverage (Good 1953) given as $C = 1 - (n/N)$ where n is the number of sequences that occurred only once in the sample. Diversity indices of Chao 1 (Chao 1984) and Shannon (Shannon 1948) as well as rarefaction curves were also calculated with mothur. To explore sample physicochemical differences between samples principal component analysis (PCA) were built using the rda function of the vegan package (Oksanen, et al. 2011) within the R environment. The influence of temperature, pH, dissolved oxygen concentration, carbon, as well as sulfate, iron, zinc, copper, alumina, arsenite and arsenate on community structure was assessed using a parsimonious canonical correspondence analysis (CCA) with a stepwise procedure (Gonzalez, et al. 2008). We compared weighted Unifrac phylogenetic distances (Lozupone and Knight 2005) among communities of each site from a maximum likelihood tree constructed with ARB and containing one representative of each OTU .

Data mining

We prospect in the public databases to reveal global patterns in the distribution of microorganisms from our 12 DMA sites. To do so we selected OTUs which were represented by at least 5 sequences, and we separated them according to the number of sites in which they were observed. Selected OTUs were organized into four categories: OTUs observed in 1 to 3 sites; 4

to 6; 7 to 9 and finally 10 to 12 sites. All the sequences representing these OTUs were then used as query for BLAST search against the "nt" database at NCBI. We retrieved 5 results per query with a minimal e-value of $10e^{-114}$ each. A customized Python script was thereafter used to find the isolation source information from the Genbank files for each BLAST result. Finally a manual data curation allowed to group the different isolation sources into 14 tokens as follow: *AMD*: acid and high metal concentration; *acid*: environment defined as acid; *volcanic*: volcanic environment; *soil*: observed in soils not defined as acidic; *sediment*: observed in sediments; *metals*: environments impacted by metallic elements; *hydrocarbon*: environments impacted by hydrocarbons; *freshwater*: observed in freshwater environments; *salty*: observed in environments of high salinity; *geothermal*: observed in geothermal habitats; *animal_host*: organisms observed in animal host; *organic_matter*: observed in environments with high concentration of organic matters; *other*: sources ambiguously defined or other habitats not already defined as a token; *not_known*: isolation source not defined in the accession.

Network construction

We explored association between microorganisms by building a co-occurrence network. Nonrandom co-occurrence patterns were tested with the checkerboard score (C-score) under a null model preserving site frequencies (Stone and Roberts 1990). For network inference, we calculated all possible Spearman's rank correlations among OTUs with more than five sequences (247 OTUs). This filtering step removed poorly represented OTUs and aimed at reducing the network complexity. We considered a valid co-occurrence event to be a robust correlation if the Spearman's correlation coefficient (r) was both >0.6 or < -0.6 and statistically significant (P-value <0.01). The nodes in the network represented the OTUs at 95% identity, whereas the edges corresponded to a strong and significant correlation between nodes. All these analyses were carried out in the R environment using the *vegan* package. Networks were explored and visualized with Cytoscape V 3.0 (Smoot, et al. 2011). The graph clustering algorithm EAGLE (Shen, et al. 2009) implemented in the Cytoscape plug-in ClusterViz was used to detect clusters in the network, with a clique size threshold of 4 and a complex size threshold of 2.

Results and Discussion

A deep investigation of AMD archaeal and bacterial diversities in three biogeographic regions

The AMD sites were as expected characterized by very acidic pH ranging from 1.02 at ESP_R.T_ri to 3.49 at BOL_mil and by an important contamination in metal among which iron, arsenic, zinc averaging 5.986 ± 7.262 g/L, 47.485 ± 68.569 mg/L and 383 ± 782.765 mg/L respectively as well as sulfate 50.749 ± 72.155 g/L (Table S1). The bacterial and archaeal diversity of the twelve sampling sites were investigated by pyrosequencing. Our strategy allowed to produce 83514 bacterial and 36974 archaeal high quality taxonomically highly informative 16S rRNA sequences (average length of 474 bp for bacteria and 464 bp for archaea). The Good's coverage showed that more than 99 % of the diversity was covered, at an OTU definition of 95 % similarity, in all the sites except the BOL_oru0 site covered at 98,3 %. The sequencing depth obtained with pyrosequencing ensured a confident estimation of the diversity at all sites and moreover allowed to have a good representation of low abundant taxa. As expected the bacterial richness was higher than archaeal richness and highly variable among sites: from 15 to 291 bacterial OTUs for ESP_R.T_pt1 and BOL_oru0 and 6 to 36 archaeal OTUs for BOL_oru1 and FRA_carn_s1 (Tables 1A and 1B). Alpha diversity was low as expected in AMD environments. It was variable among sites and contrasted between the two domains: ESP_R.T_pt1 and ESP_R.T_ri had the lowest bacterial and archaeal diversity respectively, the highest bacterial diversity was observed in the Oruro well mine samples whereas the Saõ Domingos district sample exhibited the highest archaeal diversity. This suggests that representatives of each domain found in these AMD more or less suitable conditions to thrive. It has been suggest that archaea flourished in habitat where chronic energy stress would be the rule (Valentine 2007). Interestingly the extremely acidic site, ESP_R.T_ri, did not harbor particularly high archaeal diversity indicating that the extreme conditions of this site was not the most favorable to the archaeal domain or alternatively that bacteria particularly adapted to these ambient conditions could out-compete with archaea. The majority of the bacteria we observed fall into five phyla: Acidobacteria, Actinobacteria, Deinococcus-Thermus, Firmicutes, Nitrospirae and four classes of the Proteobacteria: Alpha- Beta- Gamma- Deltaproteobacteria (Figure 2a). Similar compositional pattern at this level of resolution was also described by Kuang et al. (2013) in water samples from several AMD across Southeast China. It also reflect previous studies that aimed at describing cultivable prokaryotic AMD diversity (Alazard, et al. 2010; Hallberg 2010;

Johnson and Hallberg 2008; Johnson and Hallberg 2003; Johnson, et al. 2014). It is worthy to note that the distribution pattern of these phyla was different among sites, some were lacking in some samples whereas dominant in other (e.g. Beta- and Gammaproteobacteria, Nitrospirae). The presence of Deinococcus-Thermus representatives, particularly high at ESP_odi_st and PRT_sd is however original in our data set. Focusing on abundant OTUs (representing more than 1 percent of the total community) restricted profoundly the diversity to 18 bacterial and 16 archaeal OTUs that correspond to 95 % and 83 % of the total sequence respectively, 73 % of them are found in more than half of our sites (Table S1). Most of the archaea were represented by two phyla in different proportion according to the site: the Euryarchaeota and the Thaumarchaeota (Figure 2c). The Thaumarchaeota phylum has been proposed as a new archaeal phylum by Brochier-Armanet et al. (2008). Initially encountered in acid soils and acidic hot springs we know today that this phylum is widespread in numerous environments (Erguder, et al. 2009). Its occurrence was also noticed for the first time in an AMD (Carnoulès, France) by Volant et al. (2012). Interestingly Thaumarchaeota as ammonia-oxidizing archaea (AOA) have been demonstrated to be important players in nitrification by controlling the first, but also rate-limiting step, of this process. This property conferring them autotrophy they have been thought to significantly contribute to nitrogen and carbon cycle (Konneke, et al. 2005; Könneke, et al. 2014; Yakimov, et al. 2011). Nevertheless several evidence have also suggested potential metabolic plasticity pointed out that AOA could adopt heterotrophic or mixotrophic lifestyle (Hatzenpichler 2012; Ingalls, et al. 2006; Mußmann, et al. 2011). We showed here that these microorganisms could be represented as much as 30% of the whole archaeal community in three of the Iberian pyrite belt AMD indicating that additional energetic source could exist in these environments along the currently admitting iron and sulfur elements. In addition this suggests that they could play a role in the nitrogen cycle in these sites. This observation would merit further attention in the future to decipher the precise role of these archaea in AMD. As indicated in Tables 1A and 1B samples from the Oruro mine in Bolivia exhibited the largest bacterial diversity without affecting however the global composition at the phylum level compared to the other sites (Figures 2a and 2c). The bacterial diversity at these two sites contrasted greatly with their counterpart and did not reflect the expected diversity in such extreme environments. In order to understand the origin of these patterns we investigated deeply the community composition at the genus level at all the sites. Interestingly our analysis pinpointed the presence of several host-associated taxon such as representatives of *Propionibacterium*, *Acinetobacter*,

Streptococcus genera that accounted to 35 % and 40 % of the whole taxa and were not or scarcely detected in the other AMD (Figure 3). Oruro samples have been collected in a subsurface mine that was still under mining exploitation. Thus these singular patterns emerging there could reflect the superposition at the same place of two (or more) divergent communities, one autochthonous inferred to the typical mine habitat and the other, allochthonous, originating from the intensive anthropization linked to the miner presence and activity. We suggest that this allochthonous community would correspond to transitory assemblage, a thanatocoensis that could be artificially maintained in the milieu by a continuous input from humans and their activity. Although the patterns inferred from metagenomic DNA did not provide enough information about the viability of this allochthonous communities we can suspect that their subsistence in these habitats were limited due to ambient conditions. Thus otherwise this situation reminds that DNA-based approach provide an instantaneous cliché of an environment that should be interpreted with caution.

Niche specialization in AMD

High-throughput sequencing allowed a deep investigation into the composition of microbial communities. However the high-resolution inventories gleaned from unique site are not sufficient to pinpoint the true actors of ecosystem functioning, namely Eltonian specialization, and identify, if any, species intimately linked to the environment investigated due their ecological requirement, namely Grinnellian specialization (Devictor, et al. 2010; Nemergut, et al. 2011; Newton, et al. 2011). Comparative approach at a global scale of species distribution should allow to withdraw independently of spatial and temporal consideration, a robust assessment of ecological specialization in AMD. In addition although AMD are highly specific ecosystems, they are piece of broader landscapes comprised of manifold habitats which harbor different properties from those of mining areas. Also we can suspect that they should not escape to the rule of connectivity with surrounding habitat type. Among important processes that promote connectivity are hydrological cycle and as outlined above animal mediated dispersal. In a first time it is thus necessary to decipher who are the indigenous and specialist organisms in AMD to draw more meaningful conclusion about life-history pattern of AMD communities. In that aim, data mining of public databases was performed to inspect at a larger scale the global distribution pattern of close relatives to the OTUs retrieved in the 12 AMD. The common OTUs (i.e. present in more than 6 sites) represented almost 67 % of the total sequences of our data set. Interestingly we highlighted that more than 86% of their closest relatives were also associated to AMD or

AMD-like environments such as volcanic sites or acid environments (Figure 4). Those that were observed in less than half of the AMD, classified here as uncommon, had also relatives in AMD but to a lesser extent. These microorganisms were most frequently associated to soil or to animal indicating either that they are more generalist species that can live in a broader range of environmental conditions or alternatively that they are vagrant species with no affinity to this environment but were conveyed there, from neighboring or even farther habitats, just by chance via any process of dispersal. Their real activity in the functioning of the AMD needs to be demonstrated. In contrast it is clear that each archaeal and bacterial organism commonly observed in the 12 AMD but also in similar environments found there particular abiotic and biotic conditions to satisfy their ecological requirement (Grinnellian specialization). This implies that these species possessed intrinsic attributes ensuring specific adaptations that allowed them to thrive in these habitats.

Community structure and dispersal in AMD

Dispersal is a fundamental biological process that through the movement of individuals away from their natal origin contribute to shape the composition pattern of communities, their dynamics, their ecological profiles and their evolution (Leibold, et al. 2004; Lowe and McPeck 2014; Slatkin 1987; Wright 1949). To reveal the strength of dispersal processes in community assembly we explored in a first attempt beta diversity through Unifrac distance that take into account phylogenetic diversity among communities (Figure 2b). It is interesting to note that at our scale of investigation (OTUs defined at 95 % of similarity) the relatedness of the 12 communities was not congruent to their proximity, namely the closer sites were not the most similar. Although the most related communities were those from the San-Telmo (ESP_odi_st) and the Saõ Domingos district (PRT_sd), two nearby sites of the Iberic tin belt, the other sample from the Odiel station (ESP_odi_th) showed strong similarity in its phylogenetic composition to the Bolivian sample of Potosí (BOL_pot). In the Rio Tinto basin, the samples issued from the drainages of the tailings (ESP_R.T_pt1/ri) were clustered together, but the sample from the natural source of the Rio Tinto (ESP_R.T_pt2) located a few meters apart clearly showed a distinct community. The community sampled at the bottom of the mine well in Oruro (BOL_oru0) was also distinct from the sample collected at the entrance of the well (BOL_oru1) which clustered with the community of the Miluni district sample (BOL_mil). Nevertheless the Carnoulès communities (FRA_carn_cowg and FRA_carn_s1) that have been sampled in the same continuum, the Reigous creek, were however closely related. To go a step further and

extract more precisely any relationship between geographic distances and community differentiation we investigated distance-decay relationship in our dataset (Figure 5). At our scale of investigation and as suspected through Unifrac distance tree, our data set indicated that the composition of the communities was not dictated by the pool of microorganisms present locally, even at shortest distances. In contrast we highlighted here that communities appeared hermetic to neighboring flux. Several processes can sustain apparent dispersal limitation. One assumption would recognize that dispersal is not a neutral process but instead a property that it is prone to selection and differs among species (Lowe and McPeck 2014; Thomson, et al. 1996). In consequence AMD's microorganisms may possess intrinsic properties that would preclude their dispersal from surrounding locations. Alternatively local conditions at the host site would limit the colonization and settlement of new microorganisms, that is habitat filtering. Finally differentiation among communities could also be explained by priority effect, that is the ability of local residents that arrived first in a community to negatively impact the arrival of new species through mechanism such as niche space or resources preemption or alteration of local conditions (Lockwood, et al. 1997; Mergeay, et al. 2011).

In addition we noticed that pairwise Unifrac distances between sites were quite homogenous (0.412 ± 0.099) indicating that although relatedness was limited, coarse and conserved phylogenetic pattern subsists in AMD. This observation suggests that stochastic processes cannot explain community assembly in AMD and confirmed as proposed above niche specialization.

Community structure and environmental filtering in AMD

In the straight line of Baas Becking paradigm which advocates that environmental factors are the sole determinants of microbial distribution (de Wit and Bouvier 2006; Muller, et al. 2013), evidence of species sorting processes in microbial assembly have been demonstrated in several studies (e.g. Hell, et al. 2013; Logue and Lindstrom 2010; Lozupone and Knight 2007; Pommier, et al. 2012a; Tamames, et al. 2010) including those focusing on AMD (Edwards, et al. 1999; Gonzalez-Toril, et al. 2011; Kuang, et al. 2013; Lear, et al. 2009; Palacios, et al. 2008; Volant, et al. 2014). Taking advantage of our multisite comparison we also explored the relationship between the abundant microorganisms (≥ 0.1 %) (Table 2) and environmental parameters that characterize AMD. To do so we build Canonical Correspondence Analysis (CCA) using eleven physico-chemical parameters (Figure 6). Those parameters were associated to the microbial processes occurring in the AMD. The temperature, the pH, and dissolved oxygen that influence the pyrite dissolution by the microorganisms whereas carbon indicate the resources available in

the environment. Sulfate, iron, zinc, alumina and copper concentrations are linked to the dissolution of pyrite minerals. As well as arsenic which both redox forms were measured. The parsimonious CCA model explains 36.9 % of the variance of the abundant microbial community on the first two axes indicating that others factors not assessed here should contribute to explain community assembly. The CCA revealed that arsenite, carbon, sulfate and pH were key parameters in determining the community structure. The OTUs associated with low pH conditions belong to the *Sulfobacillus* (Otu014), the third group of the genus *Leptospirillum* (Otu015) as well as the first and the second group (Otu002), the *Ferrimicrobium* (Otu019), the *Thermoplasma* (AOtu74) and *Ferroplasma* (AOtu81). Low pH was also closely linked to high sulfate concentration, but the *Ferrimicrobium* and the *Ferroplasma* were more impacted by sulfate than pH. High arsenite and higher pH concentration selects OTUs classified as *Gallionella* (Otu003), *Thiobacillus* (Otu008) closely related to the bin CARN6 reconstructed from metagenomic data of CowG Carnoules site (Bertin, et al. 2011), and two archaeal OTUs, one from the *Thermoplasma* genus and another from the environmental group BSLdp215 of the Thermoplasmatales. Higher carbon content explained community structure of Miluni and Oruro Bolivian AMD and the distribution of *Moraxellaceae* (Otu006), *Propionibacterium* (Otu013) and *Streptococcus* (Otu018) representatives whereas AOtu65 related to the Thaumarchaeota SAGMCG_1 environmental group was determined by lower carbon content. The distribution of another Thaumarchaeota representative (AOtu65) is badly explained with these parameters. Interestingly we showed that arsenite content, an element rarely assessed in AMD studies, along sulfate, carbon and pH are major determinants that contribute to structure the AMD communities. Arsenic can be an important chemical element in AMD (Table S1) where it is concentrated in mineral deposits containing sulfides and potentially liberated by means of the oxidation of sulfide minerals (Morin and Calas 2006; Pérez-Espino, et al. 2009). Its speciation, either as the reduced form, arsenite (As(III)) or oxidized arsenate (As(V)) varied according to local conditions. Although arsenic is generally found as arsenate in AMD, the Carnoulès sites CowG and S1 located near the source of the AMD exhibit exceptionally high concentrations of dissolved arsenite (Casiot, et al. 2005). This element greatly imprinted the community composition at these two sites. Interestingly and as noticed by Kuang et al. (2013) whereas low pH is a feature of AMD, it appeared consistently as a major determinant in AMD community composition (Kuang, et al. 2013; Lear, et al. 2009) even at the species and strain level as demonstrated for laboratory-cultivated acidophilic AMD communities (Belnap, et al. 2011). This

suggests that although all these organisms can thrive in harsh pH conditions they may hold different and specific metabolic functions or regulating systems related to pH that could explain the response range to this factor. Finally our results corroborate data obtained by Kuang et al. (2013) and confirm also the role sulfate in structuring AMD community. Additional factors that were not included in our CCA have also been identified as important determinants of AMD microbial composition. It is the case of temperature, conductivity (Edwards, et al. 1999) and oxygen (Gonzalez-Toril, et al. 2011).

Interspecies interactions in AMD

Microorganisms live in natural habitat almost always in complex communities where each species can be predators, prey as well as commensal, competitors, mutualists, or pathogens (Faust and Raes 2012; Sachs and Hollowell 2012). In an attempt to decipher synecological relationships in AMD we explored a co-occurrence network constructed for our whole dataset. It incorporated information about abundance, distribution and taxonomic affiliation of the OTUs (Figure 7). In addition it do not distinguished space and time and should thus permit to infer biological interactions in AMD ecosystems Observing the network as a whole showed three dense connected areas of low abundant and uncommon OTUs. As indicated after close examination they may more reflect local co-occurring taxa than real interacting taxa as illustrate by the sub-network built with host-associated OTUs of the Bolivian Oruro sites (Figure 8). There was also four sub-networks that were not connected with the main network. Among these one was established with three AMD abundant and common taxa through single link what may indicate frequent association among these organisms without significant exclusion or pairwise association (Figure 8). Interestingly correlation between domains were frequent in the network. It implies *Acidithiobacillus*- or *Ferrimicrobium*-related OTUs and an archaea of the Thermoplasmatales order whereas exclusive association with this latter was suggested with a *Ferrovum*-related OTU that may indicate competition or predation among the taxa (Figure 9). Positive inter-domains co-occurrence was also revealed among *Sulfobacillus*-related OTU and a *Ferroplasma*-related OTU (Figure 9). A striking information from the network emerged with the position of *Acidithiobacillus*. Whereas ubiquitous and highly abundant in the twelve AMD this taxon appeared poorly connected with the other taxa. In addition its distribution pattern overlapped precisely the one of *Leptospirillum* also ubiquitous in the 12

AMD. Our result contrast with the study of Chaffron et al. (2012) that indicated recurring association between the two lineages. This results merits further attention to disentangle this contradictory conclusions. It is interesting to note as well that sister taxa, for instance *Leptospirillum*, may exhibit different pattern of co-occurrence either as negative correlation with *Gallionella* representative or as positive correlation with *Sulfobacillus* representative. This results illustrates that closely related lineages may harbor specific properties that distinguish each other. This observation is particularly true for this genus as previously reported by other authors (Goltsman, et al. 2009; Simmons, et al. 2008) even though it may certainly applied to other lineages. As pointed out by Steele et al. (2011) from their correlation network built from marine microbial communities, it is not yet clear if the co-occurrence observed represent guilds of microorganisms that ensured similar or complementary functions or if they co-occurred because they shared similar environmental conditions. Indeed we have noticed that a highly connected sub-network in this analysis implied relationships between OTUs shared between ESP_odi_st and PRT_sd sites and ESP_R.T_pt2. Overall the network established in this study allowed to raise many hypothesis and putative interactions among AMD microorganisms that would merit further examination. In a first time this can be performed with optimized mathematical and methodological approaches followed in a second time by experimental approaches. As recently out-lined by Berry and Widder (2014) co-occurrence networks have been frequently used in microbial ecology to reveal interactions among community members (e.g. Barberán, et al. 2012; Chaffron, et al. 2012; Eiler, et al. 2012; Steele, et al. 2011) whereas the validation of their ecological interpretation remains scarce. This observation lead these authors to provide a guide of best practices that should help in the construction and interpretation of co-occurrence network from microbial survey datasets and finally draw robust conclusions about biological interactions.

Conclusion

Our comprehensive analysis of AMD microbial communities provide precious insights into the rules that determine community structure in such environments. Although we highlight, along with several authors (Edwards, et al. 1999; Gonzalez-Toril, et al. 2011; Kuang, et al. 2013; Lear, et al. 2009; Palacios, et al. 2008; Volant, et al. 2014) that environmental filters, including arsenite, sulfate, carbon and pH, contribute to shape community in AMD, we additionally

showed that strong niche specialization explained singular assembly in each AMD community even though these environments shared strong physico-chemical features. At our scale of investigation we also demonstrated that dispersal may not be an important determinant of community composition in AMD. Niche specialization may contribute to explain apparent dispersal limitation although other processes such as species intrinsic properties or priority effect cannot be overlooked. Finally we investigated inter-species interactions and provide leads for further investigations.

Interestingly we demonstrate that although acid mine drainages were extensively studied during the last 15 years we showed that the impulse of recent technical development allow to go a step further with these historical models by permitting to explore numerous processes and draw more general conclusion in microbial community ecology.

Acknowledgments

This study was financed by the FRB (Fondation pour la recherche sur la Biodiversité) program blanc AAP-IN-2009-039, the « Observatoire de Recherche Méditerranéen en Environnement » (OSU-OREME). Fabien Javerliat was supported by a grant from the Research Direction Générale de l'Armement (DGA) and Aurélie Volant by a grant from the French Ministry of Education. This work was performed within the framework of the « Groupement de recherche: Métabolisme de l'Arsenic chez les Microorganismes » (GDR2909-CNRS). We gratefully acknowledge the genomic platform GENOTOUL at Toulouse and Olivier Bouchez for generating pyrosequencing data of the MiGrAMD project.

Bibliography:

- Alazard D, Joseph M, Battaglia-Brunet F, Cayol J-L, Ollivier B 2010. *Desulfosporosinus acidiphilus* sp. nov.: a moderately acidophilic sulfate-reducing bacterium isolated from acid mining drainage sediments. *Extremophiles* 14: 305-312. doi: 10.1007/s00792-010-0309-4
- Barberán A, Bates ST, Casamayor EO, Fierer N 2012. Using network analysis to explore co-occurrence patterns in soil microbial communities. *ISME J*: 343-351.
- Bednar AJ, Garbarino JR, Ranville JF, Wildeman TR 2002. Preserving the distribution of inorganic arsenic species in groundwater and acid mine drainage samples. *Environmental Science and Technology* 36: 2213-2218.
- Bell G 2005. The co-distribution of species in relation to the neutral theory of community ecology. *Ecology* 86: 1757-1770.
- Bell G 2001. Neutral macroecology. *Science* 293: 2413-2418.
- Belnap CP, Pan C, Deneff VJ, Samatova NF, Hettich RL, Banfield JF 2011. Quantitative proteomic analyses of the response of acidophilic microbial communities to different pH conditions. *ISME J* 5: 1152-1161.
- Belyea LR, Lancaster J 1999. Assembly rules within a contingent ecology. *Oikos* 86: 402-416.
- Benlloch S, Rodriguez-valera F, Martinez-Murcia AJ 1995. Bacterial diversity in two coastal lagoons deduced from 16S rDNA PCR amplification and partial sequencing. *FEMS Microbiology Ecology* 18: 267-280.
- Berry D, Widder S 2014. Deciphering microbial interactions and detecting keystone species with co-occurrence networks. *Front Microbiol* 5: 219. doi: 10.3389/fmicb.2014.00219
- Bertin PN, Heinrich-Salmeron A, Pelletier E, Goulhen-Chollet F, Arsene-Ploetze F, Gallien S, Lauga B, Casiot C, Calteau A, Vallenet D, Bonnefoy V, Bruneel O, Chane-Woon-Ming B, Cleiss-Arnold J, Duran R, Elbaz-Poulichet F, Fonknechten N, Giloteaux L, Halter D, Koechler S, Marchal M, Mornico D, Schaeffer C, Smith AAT, Van Dorselaer A, Weissenbach J, Medigue C, Le Paslier D 2011. Metabolic diversity among main microorganisms inside an arsenic-rich ecosystem revealed by meta- and proteo-genomics. *ISME Journal* 5: 1735-1747.
- Brochier-Armanet C, Boussau B, Gribaldo S, Forterre P 2008. Mesophilic crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota. *Nat Rev Micro* 6: 245-252.
- Bruneel O, Duran R, Casiot C, Elbaz-Poulichet F, Personne JC 2006. Diversity of microorganisms in Fe-As-rich acid mine drainage waters of Carnoulès, France. *Applied and Environmental Microbiology* 72: 551-556. doi: 10.1128/aem.72.1.551-556.2006

- Bruneel O, Volant A, Gallien S, Chaumande B, Casiot C, Carapito C, Bardil A, Morin G, Brown GE, Personne CJ, Le Paslier D, Schaeffer C, Van Dorsselaer A, Bertin PN, Elbaz-Poulichet F, Arsene-Ploetze F 2011. Characterization of the active bacterial community involved in natural attenuation processes in arsenic-rich creek sediments. *Microbial Ecology* 61: 793-810.
- Buckley DH, Schmidt TM 2003. Diversity and dynamics of microbial communities in soils from agro-ecosystems. *Environmental Microbiology* 5: 441-452.
- Casiot C, Lebrun S, Morin G, Bruneel O, Personné JC, Elbaz-Poulichet F 2005. Sorption and redox processes controlling arsenic fate and transport in a stream impacted by acid mine drainage. *Science of the Total Environment* 347: 122-130.
- Chaffron S, Rehrauer H, Pernthaler J, von Mering C 2012. A global network of coexisting microbes from environmental and whole-genome sequence data. *Genome Research* 20: 947-959. doi: 10.1101/gr.104521.109
- Chao A 1984. Nonparametric estimation of the number of classes in a population. *Scandinavian Journal of Statistics* 11: 265-270.
- Chase JM 2003. Community assembly: When should history matter? *Oecologia* 136: 489-498.
- Chase JM, Myers JA 2011. Disentangling the importance of ecological niches from stochastic processes across scales. *Philosophical Transactions of the Royal Society B: Biological Sciences* 366: 2351-2363. doi: 10.1098/rstb.2011.0063
- Chust G, Irigoien X, Chave J, Harris RP 2013. Latitudinal phytoplankton distribution and the neutral theory of biodiversity. *Global Ecology and Biogeography* 22: 531-543.
- Connor EF, Simberloff D 1979. The Assembly of Species Communities: Chance or Competition? *Ecology* 60: 1132-1140.
- de Wit R, Bouvier T 2006. 'Everything is everywhere, but, the environment selects'; what did Baas Becking and Beijerinck really say? *Environmental Microbiology* 8: 755-758.
- Delmont TO, Simonet P, Vogel TM 2012. Describing microbial communities and performing global comparisons in the 'omic era. *ISME Journal*.
- Devictor V, Clavel J, Julliard R, Lavergne S, Mouillot D, Thuiller W, Venail P, Villeger S, Mouquet N 2010. Defining and measuring ecological specialization. *Journal of Applied Ecology* 47: 15-25. doi: 10.1111/j.1365-2664.2009.01744.x
- Diamond JM. 1975. Assembly of species communities. In: Cody ML, Diamond JM, editors. *Ecology and Evolution of Communities*: Belknap Press of Harvard University Press. p. 342-444.

- Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R 2011. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27: 2194-2200. doi: 10.1093/bioinformatics/btr381
- Edwards KJ, Gihring TM, Banfield JF 1999. Seasonal Variations in Microbial Populations and Environmental Conditions in an Extreme Acid Mine Drainage Environment. *Applied and Environmental Microbiology* 65: 3627-3632.
- Eiler A, Heinrich F, Bertilsson S 2012. Coherent dynamics and association networks among lake bacterioplankton taxa. *ISME J* 6: 330-342.
- Erguder TH, Boon N, Wittebolle L, Marzorati M, Verstraete W 2009. Environmental factors shaping the ecological niches of ammonia-oxidizing archaea. *FEMS Microbiology Reviews* 33: 855-869. doi: 10.1111/j.1574-6976.2009.00179.x
- Faust K, Raes J 2012. Microbial interactions: from networks to models. *Nat Rev Micro* 10: 538-550.
- Fierer N, Hamady M, Lauber CL, Knight R 2008. The influence of sex, handedness, and washing on the diversity of hand surface bacteria. *Proceedings of the National Academy of Sciences* 105: 17994-17999. doi: 10.1073/pnas.0807920105
- Fuhrman JA 2009. Microbial community structure and its functional implications. *Nature* 459: 193-199.
- Fulthorpe RR, Roesch LFW, Riva A, Triplett EW 2008. Distantly sampled soils carry few species in common. *ISME Journal* 2: 901-910.
- Gobet A, Boer SI, Huse SM, van Beusekom JEE, Quince C, Sogin ML, Boetius A, Ramette A 2012. Diversity and dynamics of rare and of resident bacterial populations in coastal sands. *ISME J* 6: 542-553.
- Goltsman DSA, Denev VJ, Singer SW, VerBerkmoes NC, Lefsrud M, Mueller RS, Dick GJ, Sun CL, Wheeler KE, Zemla A, Baker BJ, Hauser L, Land M, Shah MB, Thelen MP, Hettich RL, Banfield JF 2009. Community Genomic and Proteomic Analyses of Chemoautotrophic Iron-Oxidizing "Leptospirillum rubrum" (Group II) and "Leptospirillum ferrodiazotrophum" (Group III) Bacteria in Acid Mine Drainage Biofilms. *Applied and Environmental Microbiology* 75: 4599-4615.
- Gonzalez-Toril E, Aguilera A, Souza-Egipsy V, Lopez Pamo E, Sanchez Espana J, Amils R 2011. Geomicrobiology of La Zarza-Perrunal acid mine effluent (Iberian Pyritic Belt, Spain). *Applied and Environmental Microbiology* 77: 2685-2694.

- Gonzalez I, Dejean S, Martin PGP, Baccini A 2008. CCA: An R package to extend canonical correlation analysis. *Journal of Statistical Software* 23: 1-14.
- Good IJ 1953. The population frequencies of species and the estimation of population parameters. *Biometrika* 40: 237-264. doi: 10.1093/biomet/40.3-4.237
- Gough HL, Stahl DA 2011. Microbial community structures in anoxic freshwater lake sediment along a metal contamination gradient. *ISME J* 5: 543-558.
- Green J, Bohannan BJM 2006. Spatial scaling of microbial biodiversity. *Trends in Ecology & Evolution* 21: 501-507.
- Hallberg KB 2010. New perspectives in acid mine drainage microbiology. *Hydrometallurgy* 104: 448-453.
- Hatzenpichler R 2012. Diversity, Physiology, and Niche Differentiation of Ammonia-Oxidizing Archaea. *Applied and Environmental Microbiology* 78: 7501-7510. doi: 10.1128/aem.01960-12
- Hell K, Edwards A, Zarsky J, Podmirseg SM, Girdwood S, Pachebat JA, Insam H, Sattler B 2013. The dynamic bacterial communities of a melting High Arctic glacier snowpack. *ISME J* 7: 1814-1826. doi: 10.1038/ismej.2013.51
- HilleRisLambers J, Adler PB, Harpole WS, Levine JM, Mayfield MM. 2012. Rethinking community assembly through the lens of coexistence theory. In. *Annual Review of Ecology, Evolution, and Systematics*. p. 227-248.
- Horner-Devine MC, Bohannan BJM 2006. Phylogenetic Clustering and Overdispersion in Bacterial Communities. *Ecology* 87: S100-S108.
- Hubbell SP. 2001. *The Unified Neutral Theory of Biodiversity and Biogeography*. In. Princeton, N.J.: Princeton University Press.
- Huse SM, Welch DM, Morrison HG, Sogin ML 2010. Ironing out the wrinkles in the rare biosphere through improved OTU clustering. *Environmental Microbiology* 12: 1889-1898. doi: EMI2193 [pii]
- 10.1111/j.1462-2920.2010.02193.x [doi]
- Ingalls AE, Shah SR, Hansman RL, Aluwihare LI, Santos GM, Druffel ER, Pearson A 2006. Quantifying archaeal community autotrophy in the mesopelagic ocean using natural radiocarbon. *Proceedings of the National Academy of Sciences* 103: 6442-6447.
- Johnson BD, Hallberg KB. 2008. Carbon, iron and sulfur metabolism in acidophilic microorganisms. In. *Advances in Microbial Physiology*. p. 201-255.

- Johnson DB, Hallberg KB 2003. The microbiology of acidic mine waters. *Research in Microbiology* 154: 466-473.
- Johnson DB, Hallberg KB, Hedrich S 2014. Uncovering a Microbial Enigma: Isolation and Characterization of the Streamer-Generating, Iron-Oxidizing, Acidophilic Bacterium “*Ferrovum myxofaciens*”. *Applied and Environmental Microbiology* 80: 672-680. doi: 10.1128/aem.03230-13
- Jones RT, Robeson MS, Lauber CL, Hamady M, Knight R, Fierer N 2009. A comprehensive survey of soil acidobacterial diversity using pyrosequencing and clone library analyses. *ISME J* 3: 442-453.
- Konneke M, Bernhard AE, de la Torre JR, Walker CB, Waterbury JB, Stahl DA 2005. Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature* 437: 543-546. doi: http://www.nature.com/nature/journal/v437/n7058/supinfo/nature03911_S1.html
- Könneke M, Schubert DM, Brown PC, Hügler M, Standfest S, Schwander T, Schada von Borzyskowski L, Erb TJ, Stahl DA, Berg IA 2014. Ammonia-oxidizing archaea use the most energy-efficient aerobic pathway for CO₂ fixation. *Proceedings of the National Academy of Sciences* 111: 8239-8244. doi: 10.1073/pnas.1402028111
- Kuang J-L, Huang L-N, Chen L-X, Hua Z-S, Li S-J, Hu M, Li J-T, Shu W-S 2013. Contemporary environmental variation determines microbial diversity patterns in acid mine drainage. *ISME J*.
- Lane DJ. 1991. 16S/23S rRNA sequencing. In: Stachenbrady E, Goodfellow M, editors. *Nucleic acid techniques in bacterial systematics*. New York, NY: John Wiley and Sons. p. 115-175.
- Lear G, Anderson MJ, Smith JP, Boxen K, Lewis GD 2008. Spatial and temporal heterogeneity of the bacterial communities in stream epilithic biofilms. *FEMS Microbiology Ecology* 65: 463-473.
- Lear G, Niyogi D, Harding J, Dong Y, Lewis G 2009. Biofilm bacterial community structure in streams affected by acid mine drainage. *Applied and Environmental Microbiology* 75: 3455-3460. doi: 10.1128/aem.00274-09
- Leibold MA, Holyoak M, Mouquet N, Amarasekare P, Chase JM, Hoopes MF, Holt RD, Shurin JB, Law R, Tilman D, Loreau M, Gonzalez A 2004. The metacommunity concept: a framework for multi-scale community ecology. *Ecology Letters* 7: 601-613.

- Lekberg Y, Meadow J, Rohr JR, Redecker D, Zabinski CA 2011. Importance of dispersal and thermal environment for mycorrhizal communities: Lessons from Yellowstone National Park. *Ecology* 92: 1292-1302.
- Lockwood JL, Powell RD, Nott MP, Pimm SL 1997. Assembling ecological communities in time and space. *Oikos* 80: 549-553.
- Loge JB, Lindstrom ES 2010. Species sorting affects bacterioplankton community composition as determined by 16S rDNA and 16S rRNA fingerprints. *ISME Journal* 4: 729-738.
- Lortie CJ, Brooker RW, Choler P, Kikvidze Z, Michalet R, Pugnaire FI, Callaway RM 2004. Rethinking plant community theory. *Oikos* 107: 433-438. doi: 10.1111/j.0030-1299.2004.13250.x
- Lowe WH, McPeck MA 2014. Is dispersal neutral? *Trends in Ecology & Evolution* 29: 444-450. doi: <http://dx.doi.org/10.1016/j.tree.2014.05.009>
- Lozupone C, Knight R 2005. UniFrac: a New Phylogenetic Method for Comparing Microbial Communities. *Applied and Environmental Microbiology* 71: 8228-8235.
- Lozupone CA, Knight R 2007. Global patterns in bacterial diversity. *Proceedings of the National Academy of Sciences* 104: 11436-11440. doi: 10.1073/pnas.0611525104
- Ludwig W, Strunk O, Westram R, Richter L, Meier H, Yadhukumar, Buchner A, Lai T, Steppi S, Jobb G, Förster W, Brettske I, Gerber S, Ginhart AW, Gross O, Grumann S, Hermann S, Jost R, König A, Liss T, Lüßmann R, May M, Nonhoff B, Reichel B, Strehlow R, Stamatakis A, Stuckmann N, Vilbig A, Lenke M, Ludwig T, Bode A, Schleifer KH 2004. ARB: a software environment for sequence data. *Nucleic Acids Research* 32: 1363-1371. doi: 10.1093/nar/gkh293
- Mergeay J, Meester LD, Eggermont H, Verschuren D 2011. Priority effects and species sorting in a long paleoecological record of repeated community assembly through time. *Ecology* 92: 2267-2275.
- Morin G, Calas G 2006. Arsenic in soils, mine tailings, and former industrial sites. *ELEMENTS* 2: 97-101.
- Muller AL, de Rezende JR, Hubert CRJ, Kjeldsen KU, Lagkouvardos I, Berry D, Jorgensen BB, Loy A 2013. Endospores of thermophilic bacteria as tracers of microbial dispersal by ocean currents. *ISME J.* doi: 10.1038/ismej.2013.225
- Mußmann M, Brito I, Pitcher A, Sinninghe Damsté JS, Hatzenpichler R, Richter A, Nielsen JL, Nielsen PH, Müller A, Daims H, Wagner M, Head IM 2011. Thaumarchaeotes abundant in

- refinery nitrifying sludges express amoA but are not obligate autotrophic ammonia oxidizers. *Proceedings of the National Academy of Sciences* 108: 16771-16776. doi: 10.1073/pnas.1106427108
- Nawrocki EP, Kolbe DL, Eddy SR 2009. Infernal 1.0: inference of RNA alignments. *Bioinformatics* 25: 1335-1337. doi: 10.1093/bioinformatics/btp157
- Nemergut DR, Costello EK, Hamady M, Lozupone C, Jiang L, Schmidt SK, Fierer N, Townsend AR, Cleveland CC, Stanish L, Knight R 2011. Global patterns in the biogeography of bacterial taxa. *Environmental Microbiology* 13: 135-144.
- Nemergut DR, Schmidt SK, Fukami T, O'Neill SP, Bilinski TM, Stanish LF, Knelman JE, Darcy JL, Lynch RC, Wickey P, Ferrenberg S 2013. Patterns and Processes of Microbial Community Assembly. *Microbiology and Molecular Biology Reviews* 77: 342-356. doi: 10.1128/mmb.00051-12
- Newton RJ, Jones SE, Eiler A, McMahon KD, Bertilsson S 2011. A guide to the natural history of freshwater lake bacteria. *Microbiology and Molecular Biology Reviews* 75: 14-49.
- Newton RJ, Jones SE, Helmus MR, McMahon KD 2007. Phylogenetic Ecology of the Freshwater Actinobacteria acI Lineage. *Applied and Environmental Microbiology* 73: 7169-7176. doi: 10.1128/aem.00794-07
- Oksanen J, Blanchet FG, Kindt R, Legendre P, O'Hara R, Simpson GL, Solymos P, Stevens MHH, Wagner H 2011. vegan: Community Ecology Package, 2011. R package version: 1.17-10.
- Pace NR 1997. A molecular view of microbial diversity and the biosphere. *Science* 276: 734-740.
- Páez-Espino D, Tamames J, de Lorenzo V, Cánovas D 2009. Microbial responses to environmental arsenic. *Biometals* 22: 117-130.
- Palacios C, Zettler E, Amils R, Amaral-Zettler L 2008. Contrasting Microbial Community Assembly Hypotheses: A Reconciling Tale from the Rio Tinto. *PLoS ONE* 3: e3853.
- Pommier T, Douzery EJ, Mouillot D 2012a. Environment drives high phylogenetic turnover among oceanic bacterial communities. *Biol Lett* 8: 562-566. doi: 10.1098/rsbl.2011.0990
- Pommier T, Douzery EJP, Mouillot D 2012b. Environment drives high phylogenetic turnover among oceanic bacterial communities. *Biology Letters* 8: 562-566. doi: 10.1098/rsbl.2011.0990

- Pontarp M, Canbäck B, Tunlid A, Lundberg P 2012. Phylogenetic analysis suggests that habitat filtering is structuring marine bacterial communities across the globe. *Microbial Ecology* 64: 8-17.
- Raskin L, Stromley JM, Rittmann BE, Stahl DA 1994. Group-specific 16S rRNA hybridization probes to describe natural communities of methanogens. *Applied and Environmental Microbiology* 60: 1232-1240.
- Ricklefs RE 2004. A comprehensive framework for global patterns in biodiversity. *Ecology Letters* 7: 1-15.
- Rodier J, Legube B, Merlet N, Brunet R. 2009 *L'analyse de l'eau, Eaux naturelles, eaux résiduaires, eau de mer*. Paris: Dunod.
- Roh SW, Kim K-H, Nam Y-D, Chang H-W, Park E-J, Bae J-W 2009. Investigation of archaeal and bacterial diversity in fermented seafood using barcoded pyrosequencing. *ISME Journal* 4: 1-16.
- Sachs JL, Hollowell AC 2012. The Origins of Cooperative Bacterial Communities. *mBio* 3. doi: 10.1128/mBio.00099-12
- Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology* 75: 7537-7541. doi: 10.1128/aem.01541-09
- Shannon CE 1948. A mathematical theory of communication. *Bell System Technical Journal*, The 27: 379-423. doi: 10.1002/j.1538-7305.1948.tb01338.x
- Shen H, Cheng X, Cai K, Hu M-B 2009. Detect overlapping and hierarchical community structure in networks. *Physica A: Statistical Mechanics and its Applications* 388: 1706-1712.
- Simmons SL, DiBartolo G, Deneff VJ, Goltsman DSA, Thelen MP, Banfield JF 2008. Population Genomic Analysis of Strain Variation in *Leptospirillum* Group II Bacteria Involved in Acid Mine Drainage Formation. *PLoS Biol* 6: e177.
- Slatkin M 1987. Gene Flow and the Geographic Structure of Natural Populations. *Science* 236: 787-792.
- Smoot ME, Ono K, Ruscheinski J, Wang P-L, Ideker T 2011. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* 27: 431-432. doi: 10.1093/bioinformatics/btq675

- Stahl D, Amann R. 1991. Development and application of nucleic acid probes. In: Stackebrandt E, Goodfellow M, editors. *Nucleic acid techniques in bacterial systematics*. Chichester, England.: John Wiley & Sons. p. 205-248.
- Stahl DA, Lane DJ, Olsen GJ, Pace NR 1984. Analysis of Hydrothermal Vent-Associated Symbionts by Ribosomal RNA Sequences. *Science* 224: 409-411. doi: 10.1126/science.224.4647.409
- Steele JA, Countway PD, Xia L, Vigil PD, Beman JM, Kim DY, Chow C-ET, Sachdeva R, Jones AC, Schwalbach MS, Rose JM, Hewson I, Patel A, Sun F, Caron DA, Fuhrman JA 2011. Marine bacterial, archaeal and protistan association networks reveal ecological linkages. *ISME J* 5: 1414-1425.
- Stegen JC, Lin X, Konopka AE, Fredrickson JK 2012. Stochastic and deterministic assembly processes in subsurface microbial communities. *ISME J* 6: 1653-1664.
- Stone L, Roberts A 1990. The checkerboard score and species distributions. *Oecologia* 85: 74-79. doi: 10.1007/bf00317345
- Tamames J, Abellan JJ, Pignatelli M, Camacho A, Moya A 2010. Environmental distribution of prokaryotic taxa. *BMC Microbiology* 10: 85.
- Teske A, Wawer C, Muyzer G, Ramsing NB 1996. Distribution of sulfate-reducing bacteria in a stratified fjord (Mariager Fjord, Denmark) as evaluated by most-probable-number counts and denaturing gradient gel electrophoresis of PCR-amplified ribosomal DNA fragments. *Applied and Environmental Microbiology* 62: 1405-1415.
- Thomson JD, Weiblen G, Thomson BA, Alfaro S, Legendre P 1996. Untangling Multiple Factors in Spatial Distributions: Lilies, Gophers, and Rocks. *Ecology* 77: 1698-1715. doi: 10.2307/2265776
- Tom F 2005. Cosmopolitan microbes and their 'cryptic' species. *Aquatic Microbial Ecology* 41: 49-54. doi: 10.3354/ame041049
- Valentine DL 2007. Adaptations to energy stress dictate the ecology and evolution of the Archaea. *Nature Reviews Microbiology* 5: 316-323.
- Volant A, Bruneel O, Desoeuvre A, Héry M, Casiot C, Bru N, Delpoux S, Fahy A, Javerliat F, Bouchez O, Duran R, Bertin PN, Elbaz-Poulichet F, Lauga B 2014. Diversity and spatiotemporal dynamics of bacterial communities: physicochemical and others drivers along an acid mine drainage. *FEMS Microbiology Ecology*: n/a-n/a. doi: 10.1111/1574-6941.12394

- Volant A, Desoeuvre A, Casiot C, Lauga B, Delpoux S, Morin G, Personné C, Héry M, Elbaz-Poulichet F, Bertin P, Bruneel O 2012. Archaeal diversity: temporal variation in the arsenic-rich creek sediments of Carnoulès mine, France. *Extremophiles* 16: 645-657.
- Ward DM, Weller R, Bateson MM 1990. 16S rRNA sequences reveal numerous uncultured microorganisms in a natural community. *Nature* 345: 63-65.
- Weiherr E, Freund D, Bunton T, Stefanski A, Lee T, Bentivenga S 2011. Advances, challenges and a developing synthesis of ecological community assembly theory. *Philosophical Transactions of the Royal Society B: Biological Sciences* 366: 2403-2413.
- Woese CR 1987. Bacterial evolution. *Microbiological Reviews* 51: 221-271.
- Woese CR 1994. Microbiology in transition. *Proceedings of the National Academy of Sciences of the United States of America* 91: 1601-1603. doi: 10.1073/pnas.91.5.1601
- Wright S 1949. The genetical structure of populations. *Annals of eugenics* 15: 323-354.
- Yakimov MM, Cono VL, Smedile F, DeLuca TH, Juarez S, Ciordia S, Fernandez M, Albar JP, Ferrer M, Golyshin PN, Giuliano L 2011. Contribution of crenarchaeal autotrophic ammonia oxidizers to the dark primary production in Tyrrhenian deep waters (Central Mediterranean Sea). *ISME J* 5: 945-961. doi: <http://www.nature.com/ismej/journal/v5/n6/supinfo/ismej2010197s1.html>
- Zhang Y, Zhao Z, Dai M, Jiao N, Herndl GJ 2014a. Drivers shaping the diversity and biogeography of total and active bacterial communities in the South China Sea. *Molecular Ecology*: n/a-n/a.
- Zhang Z, Geng J, Tang X, Fan H, Xu J, Wen X, Ma Z, Shi P 2014b. Spatial heterogeneity and co-occurrence patterns of human mucosal-associated intestinal microbiota. *ISME J* 8: 881-893. doi: 10.1038/ismej.2013.185

Captions:

Figure 1: Location of the twelve Acid Mine Drainages sampled in this study. The two french samples from the Carnoulès drainage are FRA_carn_cowg and FRA_carn_s1. In Spain the three samples from the Rio Tinto are ESP_R.T_pt1/pt2/ri. Those from the Odiel River are ESP_odi_st/th. The sample from the Saõ Domingos mine district in Portugal is PRT_sd. In Bolivia the sample from the Miluni drainage is BOL_mil. Those from the active Oruro mine well are BOL_oru0 for the surface and BOL_oru1 for the bottom. Finally the sample from Potosí is BOL_pot.

Figure 2: Relative abundance and comparative structure of bacterial a) and archaeal c) phylum in 12 AMD over three biogeographic regions based on Silva classification. Similarity among the 12 AMD community as assessed by Unifrac distances that take into account phylogenetic distance among sequences b).

Figure 3: Relative abundance of animal-host associated taxa in 12 AMD.

Figure 4: Habitat types of microorganisms detected in 12 AMD inferred from data mining of “nt” database (NCBI). “Common” referred to OTUs observed in more than 6 AMD and “uncommon” referred to OTUs observed in less than 6 sites. The definition of each environments are given in the material and methods section.

Figure 5: Distance decay of AMD communities. Geographic distances were plot against Unifrac similarity distances for each pairwise comparison between any two samples. No significant relationship is highlighted.

Figure 6: Parsimonious Canonical Correspondence Analysis (CCA) showing influence of environmental variables (pH, arsenite, carbon and sulfate) on the abundant OTUs (in red). A

stepwise procedure retained the variables displayed (in blue). Position of sites (centroids) on the triplot are indicated (in black).

Figure 7: Co-occurrence network draw from the whole DMA dataset. Each node represents an OTU defined at 95 % of similarity. The finest taxonomic affiliation of the OTUs is showed. Circles correspond to bacterial taxa and hexagons correspond to archaeal taxa. The size of the geometric figures is proportional to the abundance of the taxon. Figures filled in blue represent the most abundant taxa (see table 2). Common taxon (present in more than 6 sites) are circled by a red line. Presence of an edge between nodes indicate a significant Spearman correlation ($p < 0.01$). Solid lines are positive correlations (> 0.6), i.e. indicative of co-occurring taxa and dot lines negative correlations (< -0.6), i.e. indicative of co-exclusive taxa. The width of the line is proportional to the strength of the Spearman correlation, the widest the line, stronger the correlation is. Black dashed-lined rectangle (A) shows the sub-networks of the Oruro Bolivian sites. Grey dashed-lined rectangle (B) pinpoints four sub-networks that are isolated from the main network.

Figures 8 and 9: Sub-networks extracted from the co-occurrence network of the figure 7. For legend see figure 7.

Table 1: Diversity indices and statistics calculated for the a) bacterial and b) archaeal sequences of the twelve AMD, before and after the sub-sampling procedure done to get an even number of sequence per sample.

Table 2: Taxonomic affiliation, abundance (number of sequences), relative abundance and distribution (+, presence, - absence) of abundant OTUs (≥ 1 % of the total sequences) in the 12 AMD. The OTU are defined at 95 % of sequence identity. "AOtu" are archaeal OTUs and "Otu" are bacterial OTUs. The taxonomy is based on the SILVA SSU database v.111 of July 2012. Nb: number (total archaea: 36974 sequences, total bacteria: 83514) in bold bacteria that were present at all sites.

Table S1: Geographic coordinates in decimal unit and physic-chemicals characteristics of AMD samples. Dissolved O₂, Sulfate and total iron concentrations are in mg.L⁻¹ ; arsenic and other metals concentrations are in µg.L⁻¹ ; red/ox potential in mV ; conductivity in µS and organic carbon in g.g⁻¹ of dry material.

Figure 1

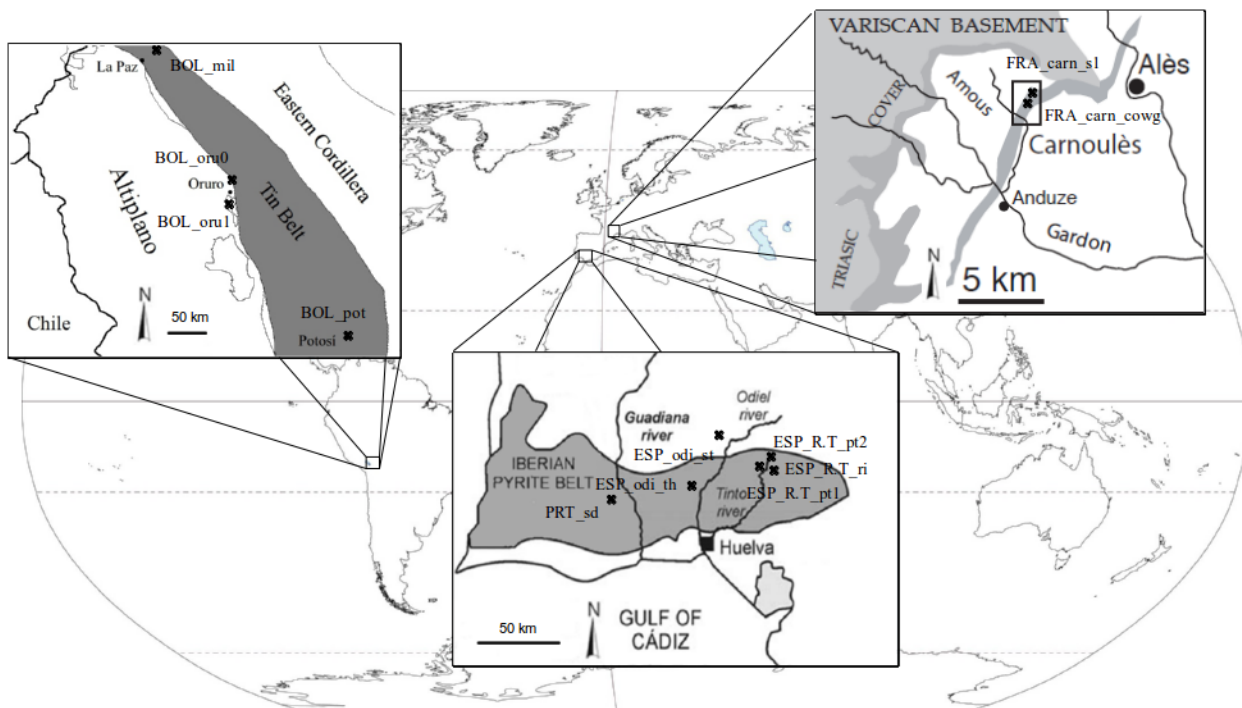


Figure 2

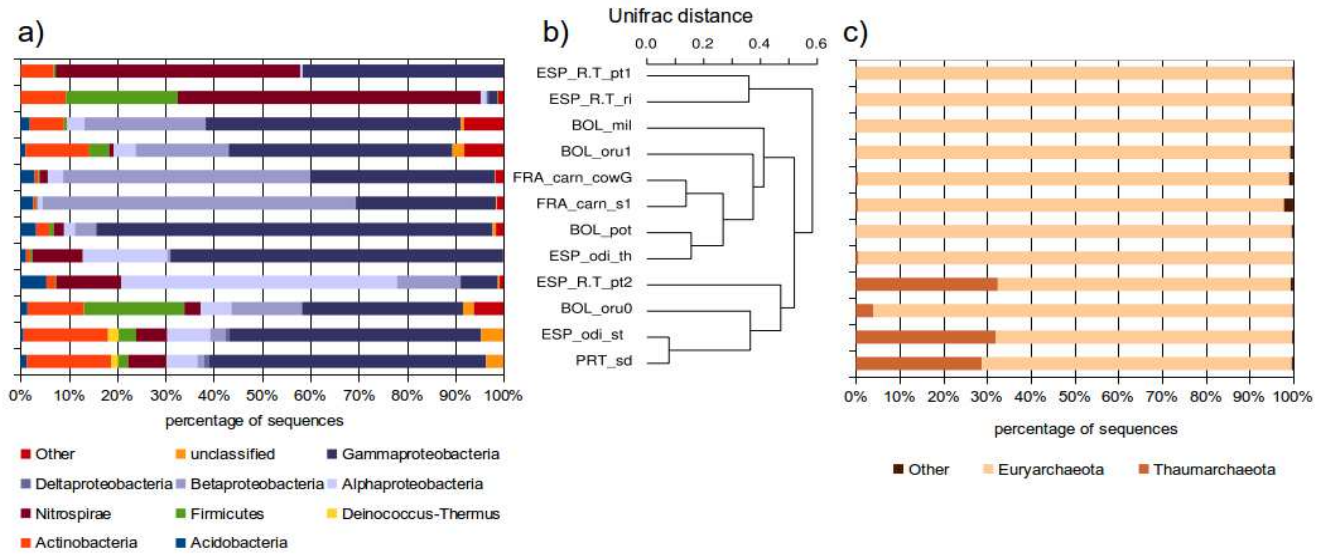


Figure 3

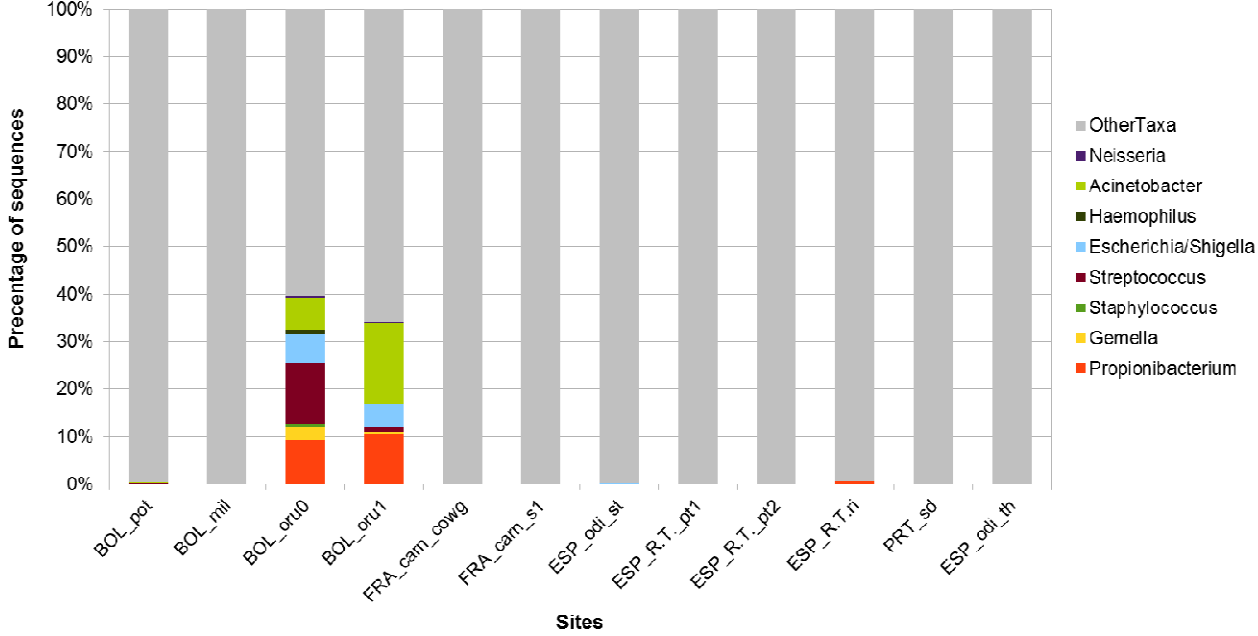


Figure 4

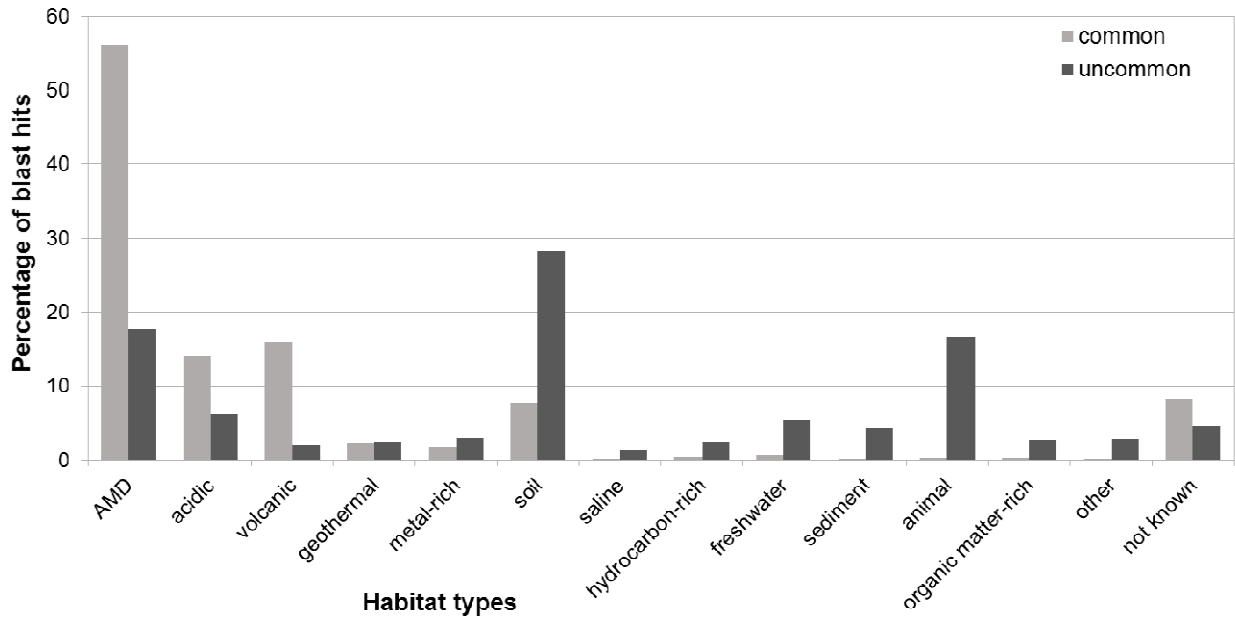


Figure 5

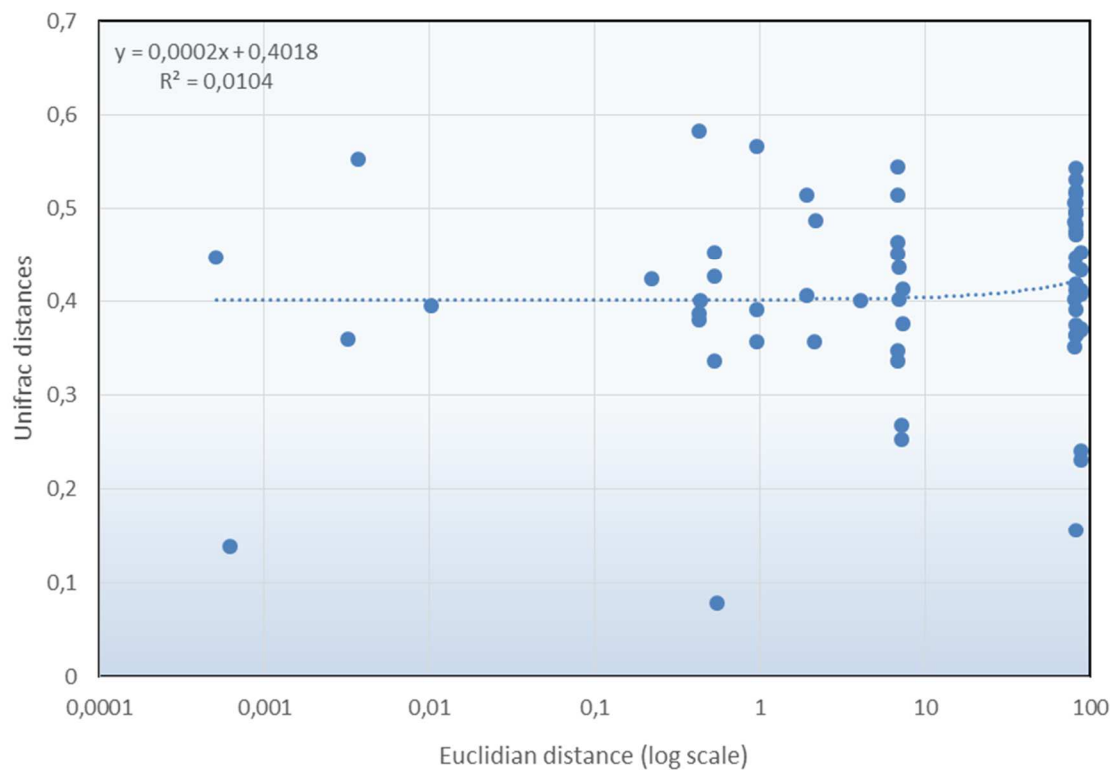


Figure 6

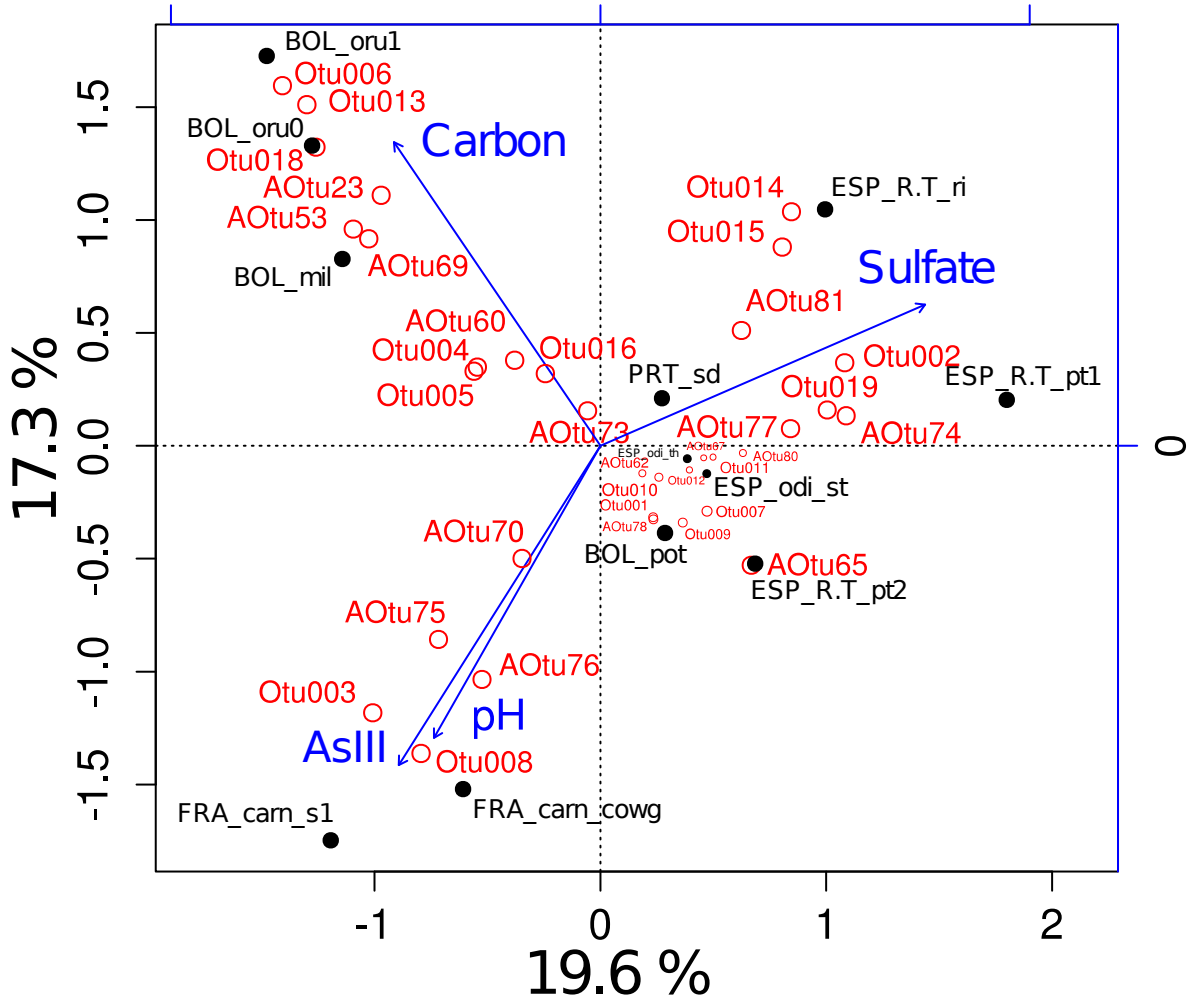


Figure 7

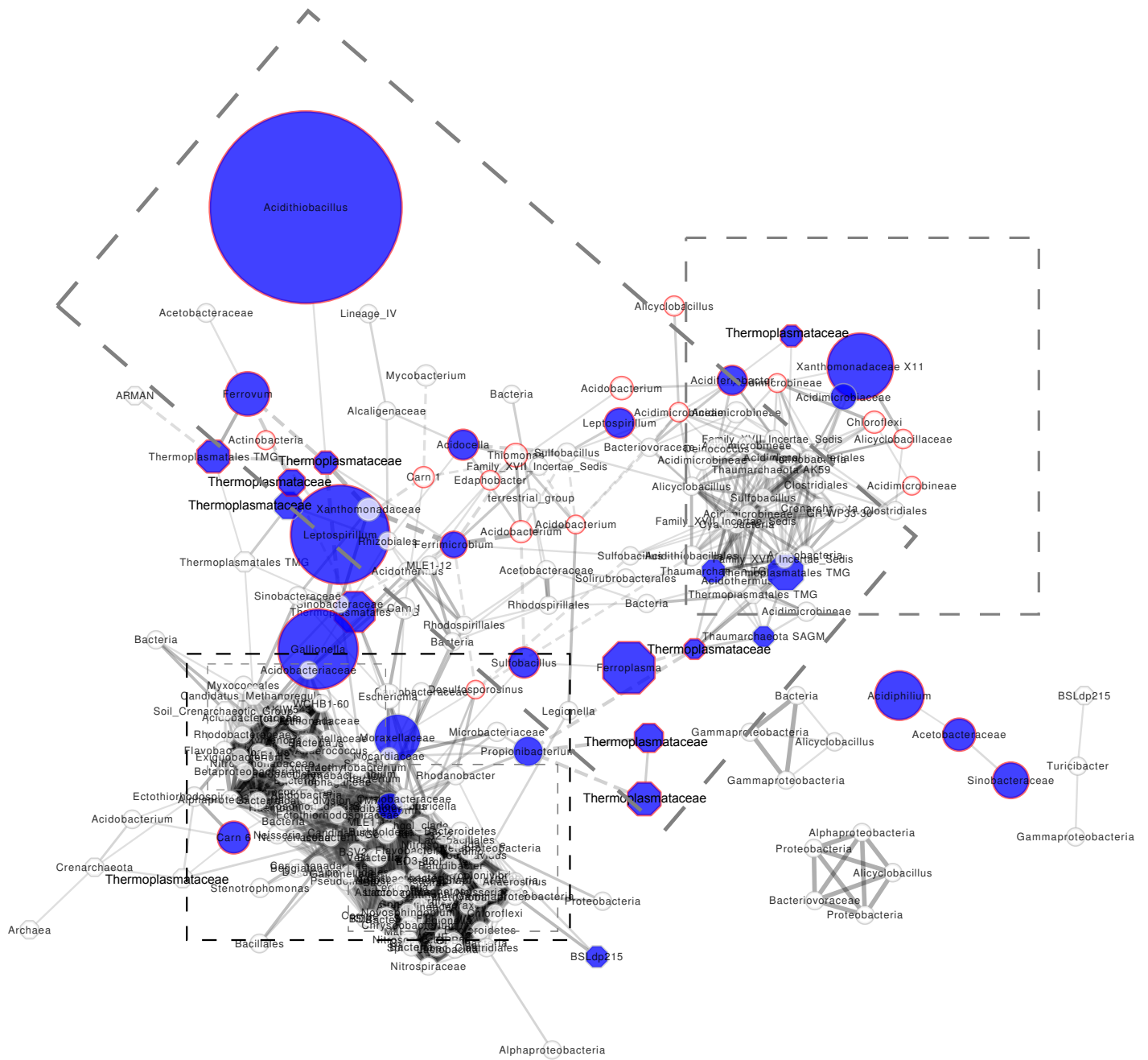


Figure 8

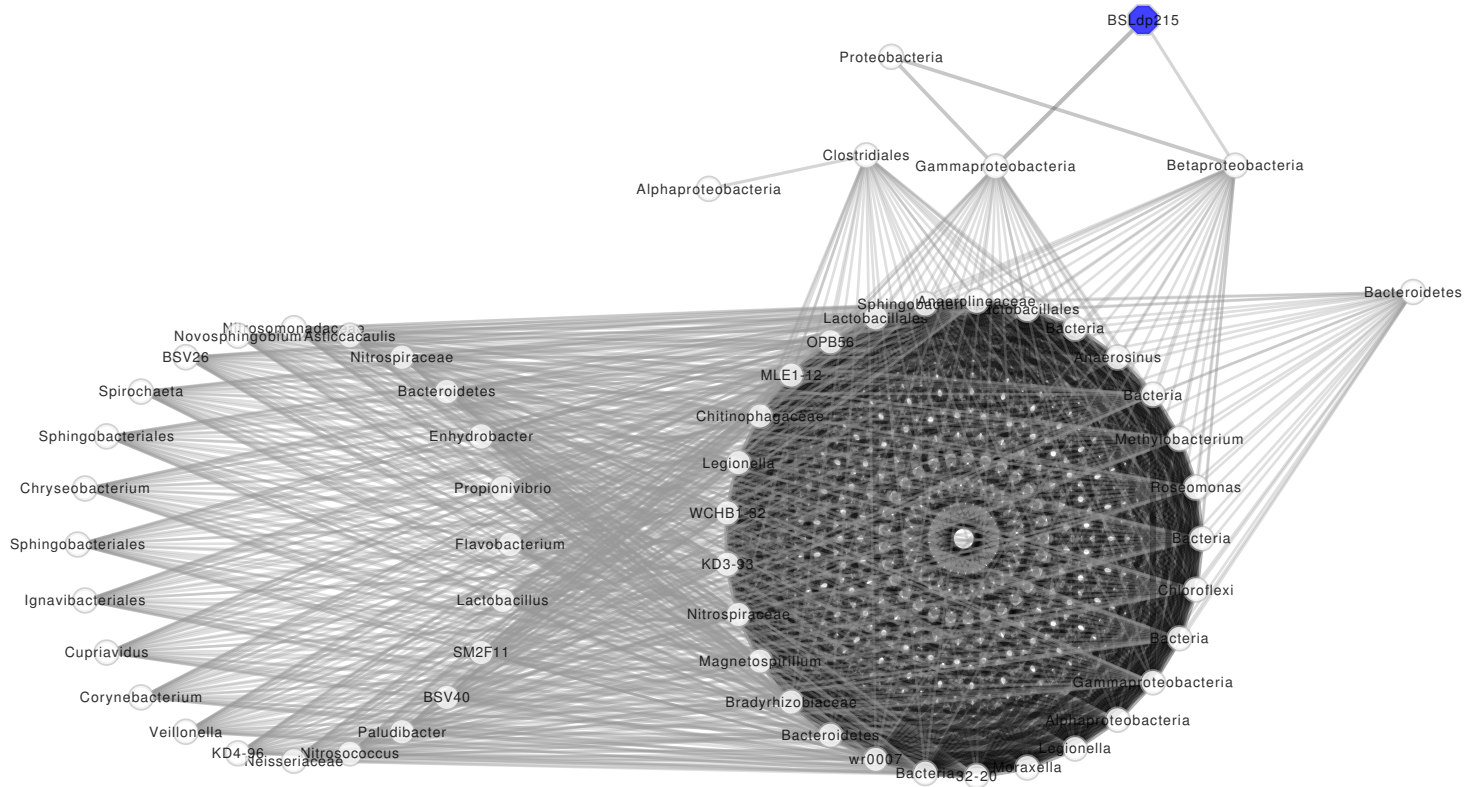
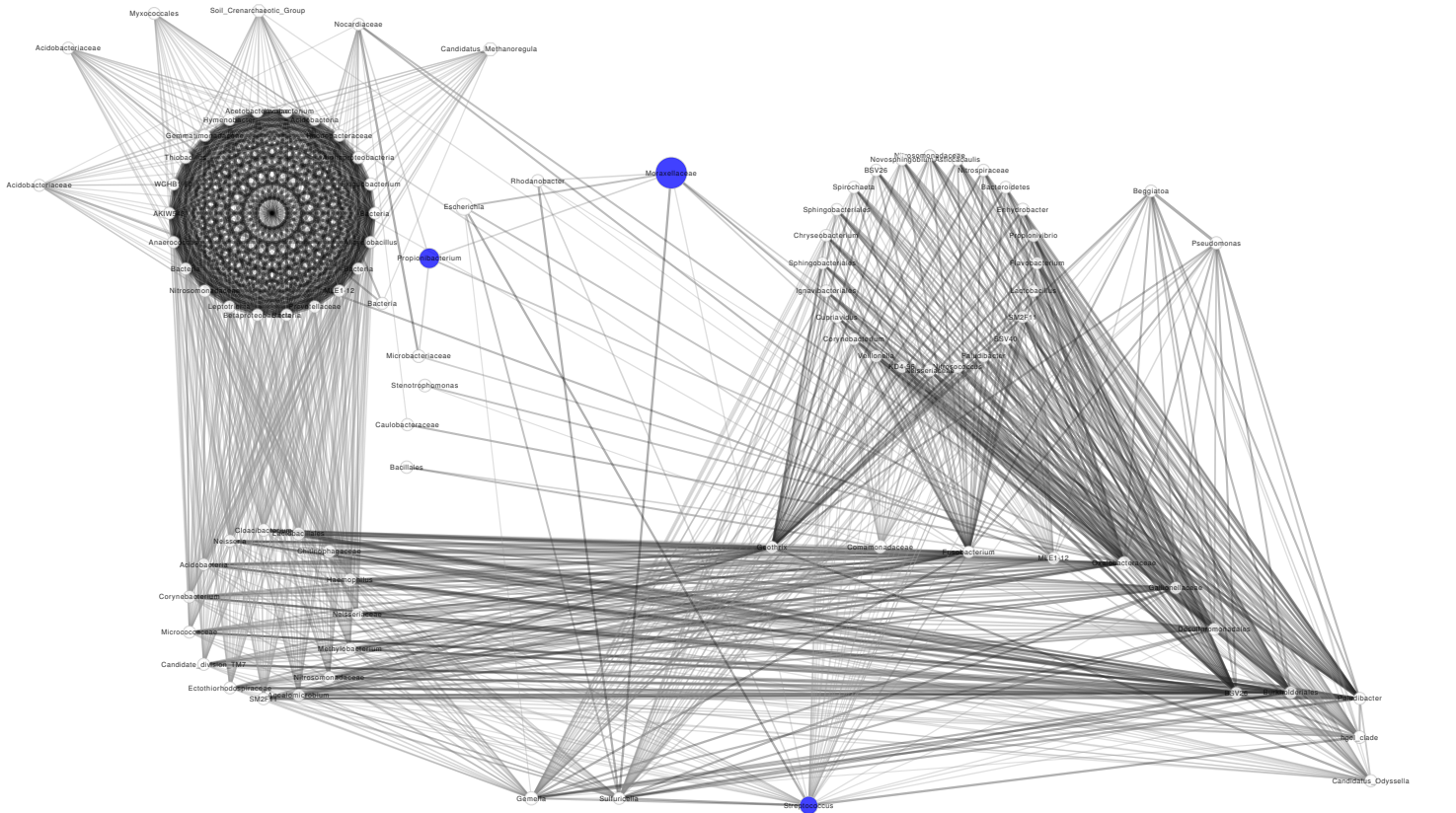


Figure 9

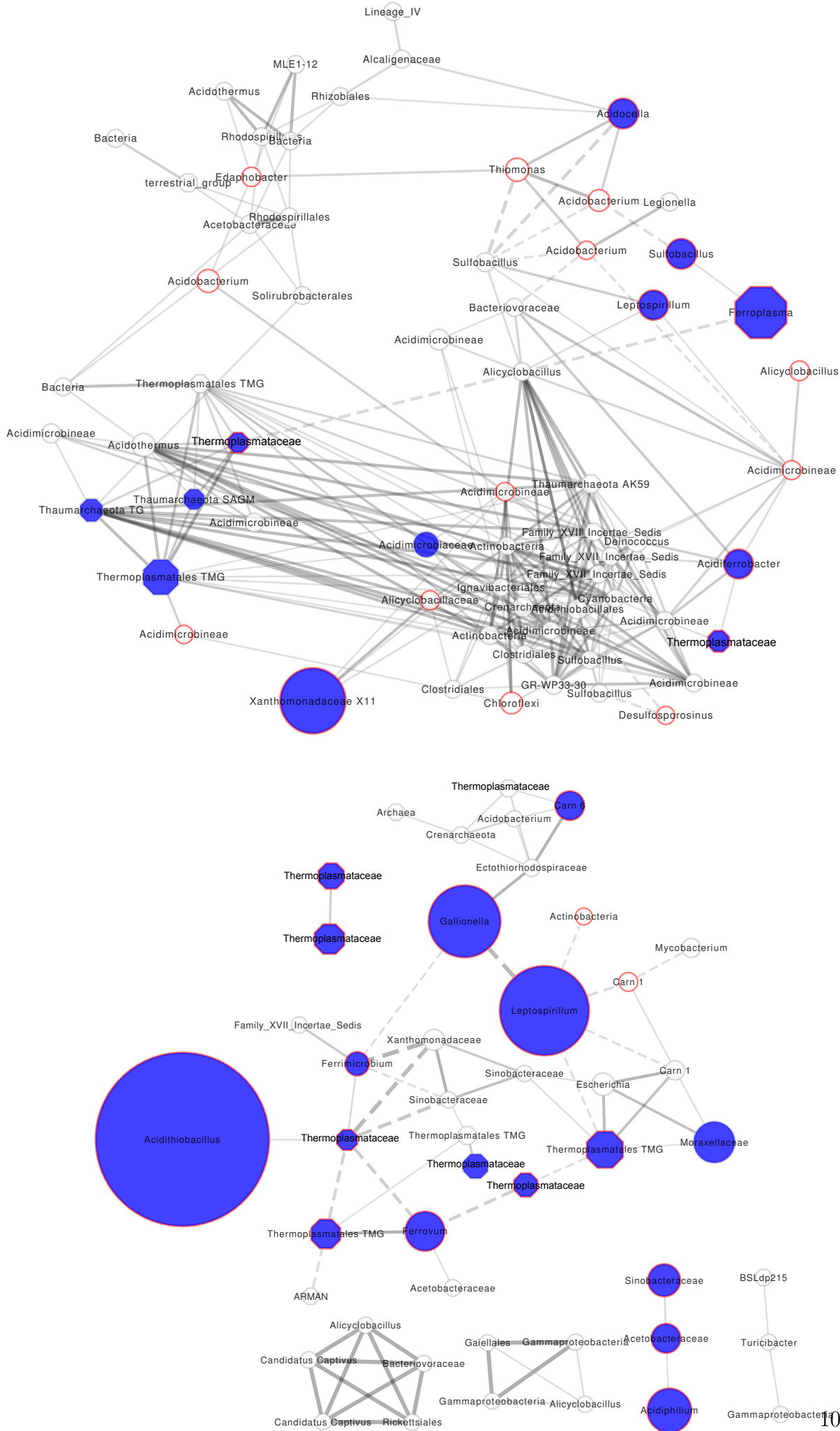


Table 1**A**

| Samples | Full dataset | | | | | Resampled dataset | | | | |
|---------------|------------------|-------------|-----------------|----------------|-------------------|-------------------|-------------|-----------------|----------------|-------------------|
| | No. of sequences | No. of OTUs | Good's coverage | Chao1 richness | Shannon diversity | No. of sequences | No. of OTUs | Good's coverage | Chao1 richness | Shannon diversity |
| BOL_pot | 10224 | 101 | 0.997 | 146.09 | 1.63 | 4363 | 73 | 0.995 | 101.11 | 1.59 |
| BOL_mil | 11450 | 72 | 0.999 | 77.57 | 1.95 | 4363 | 55 | 0.997 | 60.00 | 1.95 |
| BOL_oru0 | 6646 | 291 | 0.990 | 343.13 | 3.72 | 4363 | 253 | 0.983 | 303.12 | 3.67 |
| BOL_oru1 | 6979 | 195 | 0.996 | 207.50 | 3.23 | 4363 | 181 | 0.993 | 203.89 | 3.21 |
| FRA_carn_cowg | 6797 | 79 | 0.995 | 116.71 | 1.93 | 4363 | 64 | 0.994 | 99.10 | 1.93 |
| FRA_carn_s1 | 8477 | 50 | 0.998 | 61.14 | 1.42 | 4363 | 37 | 0.997 | 44.86 | 1.39 |
| ESP_odi_st | 4363 | 62 | 0.997 | 69.86 | 2.78 | 4363 | 62 | 0.997 | 69.86 | 2.78 |
| ESP_odi_th | 4588 | 45 | 0.997 | 62.14 | 1.63 | 4363 | 45 | 0.996 | 67.67 | 1.63 |
| ESP_R.T_pt1 | 6289 | 15 | 1.000 | 15.25 | 0.98 | 4363 | 15 | 0.999 | 18.00 | 0.99 |
| ESP_R.T_pt2 | 6692 | 100 | 0.995 | 141.33 | 2.36 | 4363 | 91 | 0.992 | 134.15 | 2.37 |
| ESP_R.T_ri | 6183 | 41 | 0.999 | 42.20 | 1.76 | 4363 | 39 | 0.999 | 40.00 | 1.77 |
| PRT_sd | 4826 | 56 | 0.998 | 62.00 | 2.61 | 4363 | 56 | 0.998 | 61.14 | 2.61 |
| Total | 83514 | 533 | 0.998 | 621.12 | 3.22 | | | | | |

B

| Samples | Full dataset | | | | | Resampled dataset | | | | |
|---------------|------------------|-------------|-----------------|----------------|-------------------|-------------------|-------------|-----------------|----------------|-------------------|
| | No. of sequences | No. of OTUs | Good's coverage | Chao1 richness | Shannon diversity | No. of sequences | No. of OTUs | Good's coverage | Chao1 richness | Shannon diversity |
| BOL_pot | 2623 | 7 | 1.00 | 7.00 | 1.43 | 1204 | 7 | 1.00 | 7.00 | 1.42 |
| BOL_mil | 1564 | 8 | 1.00 | 8.00 | 1.69 | 1204 | 8 | 1.00 | 8.00 | 1.67 |
| BOL_oru0 | 5009 | 17 | 1.00 | 18.00 | 2.06 | 1204 | 15 | 1.00 | 15.00 | 2.08 |
| BOL_oru1 | 1509 | 6 | 1.00 | 6.00 | 1.35 | 1204 | 6 | 1.00 | 6.00 | 1.36 |
| FRA_carn_cowg | 3116 | 32 | 1.00 | 44.00 | 1.68 | 1204 | 24 | 0.99 | 31.00 | 1.66 |
| FRA_carn_s1 | 3337 | 36 | 1.00 | 45.00 | 1.37 | 1204 | 27 | 0.99 | 36.00 | 1.32 |
| ESP_odi_st | 4542 | 30 | 1.00 | 32.14 | 1.76 | 1204 | 21 | 0.99 | 42.00 | 1.74 |
| ESP_odi_th | 1204 | 14 | 1.00 | 14.00 | 1.51 | 1204 | 14 | 1.00 | 14.00 | 1.51 |
| ESP_R.T_pt1 | 3303 | 12 | 1.00 | 15.00 | 1.68 | 1204 | 9 | 1.00 | 9.00 | 1.65 |
| ESP_R.T_pt2 | 4214 | 27 | 1.00 | 30.75 | 1.73 | 1204 | 22 | 1.00 | 27.00 | 1.72 |
| ESP_R.T_ri | 1693 | 8 | 1.00 | 8.50 | 0.56 | 1204 | 7 | 1.00 | 7.50 | 0.56 |
| PRT_sd | 4860 | 26 | 1.00 | 26.75 | 2.12 | 1204 | 21 | 1.00 | 21.50 | 2.10 |
| Total | 36974 | 81 | 0.999 | 131.14 | 2.71 | | | | | |

Table 2

| OTU | Nb. of sequences | % | Phylum/Class | Genus | BOL_pot | BOL_mil | BOL_oru0 | BOL_oru1 | FRA_cam_cowG | FRA_cam_st | ESP_odi_st | ESP_odi_th | ESP_R_T_pt1 | ESP_R_T_pt2 | ESP_R_T_r1 | PRT_sd |
|----------------|------------------|-------|---------------------|---------------------------------|---------|---------|----------|----------|--------------|------------|------------|------------|-------------|-------------|------------|--------|
| AOtu23 | 458 | 1,24 | Euryarchaeota | Thermoplasmatales BSLdp215 | + | - | - | + | - | + | - | - | - | - | - | - |
| AOtu53 | 1980 | 5,36 | Euryarchaeota | Unclassified Thermoplasmataceae | - | + | + | - | + | + | - | - | - | + | - | + |
| AOtu60 | 547 | 1,48 | Euryarchaeota | Thermoplasmatales E-plasma | - | + | + | - | + | + | + | - | + | + | + | + |
| AOtu67 | 5115 | 13,83 | Euryarchaeota | Thermoplasmatales TMEG | - | - | - | - | + | - | + | - | + | + | - | + |
| AOtu69 | 3812 | 10,31 | Euryarchaeota | Thermoplasmatales TMEG | + | + | + | + | + | + | + | + | - | + | - | + |
| AOtu70 | 824 | 2,23 | Euryarchaeota | Unclassified Thermoplasmataceae | + | - | + | - | + | + | + | + | - | + | - | + |
| AOtu73 | 3127 | 8,46 | Euryarchaeota | Thermoplasmatales TMEG | - | + | + | + | + | + | + | + | + | + | + | + |
| AOtu74 | 1469 | 3,97 | Euryarchaeota | Unclassified Thermoplasmataceae | - | - | + | - | + | + | + | + | + | + | + | + |
| AOtu75 | 4583 | 12,40 | Euryarchaeota | Thermoplasmatales BSLdp215 | - | - | + | + | + | + | + | + | - | + | - | + |
| AOtu76 | 2813 | 7,61 | Euryarchaeota | Unclassified Thermoplasmataceae | - | + | + | - | + | + | + | + | + | + | - | + |
| AOtu77 | 831 | 2,25 | Euryarchaeota | Thermoplasmatales Gpl | + | - | - | - | + | + | + | + | + | - | + | + |
| AOtu78 | 690 | 1,87 | Euryarchaeota | Unclassified Thermoplasmataceae | - | - | - | - | + | + | + | + | + | + | - | + |
| AOtu80 | 1304 | 3,53 | Euryarchaeota | Thermoplasmatales A-plasma | - | - | + | - | + | + | + | + | + | - | - | + |
| AOtu81 | 4962 | 13,42 | Euryarchaeota | Ferroplasma | + | + | + | + | + | + | - | + | + | + | + | + |
| AOtu62 | 1567 | 4,24 | Thaumarchaeota | Terrestrial group | - | - | - | - | - | - | + | - | - | + | - | + |
| AOtu65 | 908 | 2,46 | Thaumarchaeota | SAGMCG_1 | - | - | - | - | + | - | + | - | - | + | - | + |
| Total Archaea | 34990 | 94,63 | | | | | | | | | | | | | | |
| Otu007 | 3188 | 3,82 | Alphaproteobacteria | Acidiphilium | + | + | + | + | + | - | + | + | + | + | - | + |
| Otu009 | 1617 | 1,94 | Alphaproteobacteria | Acidocella | + | + | + | + | + | + | - | + | - | + | - | - |
| Otu012 | 1554 | 1,86 | Alphaproteobacteria | Unclassified Acetobacteraceae | + | + | + | + | + | + | + | + | - | + | + | + |
| Otu005 | 4004 | 4,79 | Betaproteobacteria | Ferrovum | + | + | + | + | + | + | + | + | - | + | - | + |
| Otu008 | 1832 | 2,19 | Betaproteobacteria | unclassified | + | + | + | + | + | + | + | - | + | - | - | - |
| Otu003 | 8197 | 9,82 | Betaproteobacteria | Gallionella | + | + | + | + | + | + | + | + | - | + | - | - |
| Otu001 | 19875 | 23,80 | Gammaproteobacteria | Acidithiobacillus | + | + | + | + | + | + | + | + | + | + | + | + |
| Otu004 | 7245 | 8,68 | Gammaproteobacteria | Unclassified Xanthomonadaceae | + | + | + | + | + | + | + | + | - | + | - | + |
| Otu010 | 1608 | 1,93 | Gammaproteobacteria | Acidiferrobacter | + | - | + | - | + | + | + | + | - | + | + | + |
| Otu011 | 1578 | 1,89 | Gammaproteobacteria | Unclassified Xanthomonadaceae | + | + | + | + | + | - | + | + | - | + | - | + |
| Otu006 | 3256 | 3,90 | Gammaproteobacteria | Moraxellaceae | + | + | + | + | + | - | + | + | - | + | - | - |
| Otu013 | 1391 | 1,67 | Actinobacteria | Propionibacterium | + | + | + | + | - | - | - | - | - | + | + | - |
| Otu016 | 1075 | 1,29 | Actinobacteria | Unclassified Acidimicrobiaceae | + | + | + | + | - | + | + | - | - | + | - | + |
| Otu019 | 833 | 1,00 | Actinobacteria | Ferrimicrobium | + | - | - | - | + | - | + | + | + | - | + | + |
| Otu002 | 8421 | 10,08 | Nitrospirae | Leptospirillum | + | + | + | + | + | + | + | + | + | + | + | + |
| Otu015 | 1268 | 1,52 | Nitrospirae | Leptospirillum | + | - | + | + | - | - | + | + | + | + | + | + |
| Otu014 | 1334 | 1,60 | Firmicutes | Sulfobacillus | + | + | + | + | - | - | + | + | + | + | + | + |
| Otu018 | 955 | 1,14 | Firmicutes | Streptococcus | + | - | + | + | - | - | + | - | - | + | - | - |
| Total Bacteria | 69231 | 82,90 | | | | | | | | | | | | | | |

Table S1

| Samples | Organic carbon | Dissolved O ₂ | T (°C) | pH | Eh | Fe (total) | Sulfate | As (III) | As (V) | Al | Co | Cu | Cr | Zn | Cd | Mn | V | Sr | Sb | Rb | Conductivity | Longitude | Latitude |
|---------------|----------------|--------------------------|--------|------|-------|------------|---------|----------|--------|---------|-------|--------|-----|---------|------|--------|-----|-------|-------|------|--------------|------------|------------|
| BOL_pot | 0.09 | 6.66 | 10.70 | 3.04 | 382.5 | 1340 | 9843 | 918 | 4086 | 238300 | 456 | 2586 | 13 | 2810000 | 6801 | 29310 | 20 | 127 | 17 | 99 | 6870 | -65.73855 | -19.626817 |
| BOL_mil | 0.3 | 4.63 | 10.30 | 3.49 | 355.6 | 360 | 1827 | 11 | 152 | 4153 | 166 | 357 | 0 | 117000 | 301 | 28250 | 1 | 236 | 0 | 8 | 2053 | -68.148764 | -16.319103 |
| BOL_oru0 | 0.21 | 4.96 | 22.00 | 2.10 | 465.7 | 821 | 5990 | 527 | 4856 | 529400 | 216 | 1552 | 123 | 50770 | 988 | 8420 | 252 | 44350 | 745 | 4921 | 1272 | -67.13395 | -17.963917 |
| BOL_oru1 | 0.19 | 3.77 | 21.20 | 1.31 | 406.0 | 4920 | 29843 | 13910 | 83980 | 1677000 | 1742 | 73030 | 682 | 180400 | 5996 | 16900 | 892 | 10210 | 10300 | 3070 | 9840 | -67.124056 | -17.961358 |
| FRA_carn_cowg | 0.04 | 7.65 | 17.40 | 3.22 | 570.4 | 1149 | 2819 | 127100 | 33260 | 60450 | 338 | 769 | 131 | 22390 | 102 | 8355 | 38 | 220 | 34 | 128 | 4890 | -4.00167 | 44.11694 |
| FRA_carn_sl | 0.04 | 1.29 | 16.70 | 3.21 | 489.5 | 1256 | 3557 | 180500 | 20230 | 64990 | 356 | 833 | 144 | 24060 | 106 | 8865 | 37 | 221 | 59 | 136 | 5190 | -4.00222 | 44.11722 |
| ESP_odi_st | 0.07 | 1.75 | 23.10 | 2.53 | 420.1 | 1972 | 40465 | 1 | 73 | 1520000 | 5843 | 173300 | 715 | 456900 | 956 | 214800 | 64 | 462 | 6 | 38 | 18690 | -6.976 | 37.798694 |
| ESP_R.T_pt2 | 0.03 | 5.63 | 19.30 | 1.79 | 510.0 | 17220 | 208140 | 8 | 8959 | 3079000 | 20800 | 15240 | 314 | 15560 | 99 | 20590 | 157 | 199 | 10 | 47 | 25325 | -6.5515 | 37.722111 |
| ESP_R.T_ri | 0.1 | 8.77 | 14.80 | 2.64 | 489.1 | 238 | 25581 | 1 | 39 | 142800 | 991 | 1637 | 18 | 5220 | 21 | 20700 | 1 | 366 | 1 | 3 | 3520 | -6.551056 | 37.722361 |
| ESP_R.T_pt1 | 0.08 | 4.63 | 22.00 | 1.02 | 488.0 | 20620 | 194186 | 176 | 34785 | 2670000 | 20950 | 109400 | 193 | 99310 | 270 | 11060 | 655 | 795 | 67 | 720 | 58800 | -6.554333 | 37.720639 |
| Sao_Domingos | 0.07 | 5.94 | 27.20 | 2.09 | 600.4 | 8432 | 48140 | 12 | 39255 | 923600 | 11190 | 661900 | 506 | 364600 | 579 | 25380 | 893 | 147 | 188 | 11 | 12120 | -7.50475 | 37.649361 |
| ESP_odi_th | 0.08 | 4.25 | 26.80 | 2.51 | 426.4 | 2718 | 38605 | 367 | 16625 | 918600 | 9021 | 125300 | 754 | 450300 | 941 | 135700 | 505 | 275 | 149 | 14 | 14285 | -7.071361 | 37.599806 |

3.3 Conclusion

Notre analyse complète des communautés microbiennes des DMA fournit de précieux renseignements sur les règles qui déterminent la structure des communautés dans de tels environnements. Bien que nous mettons en évidence, avec plusieurs auteurs que les filtres environnementaux, y compris l'arsénite, le sulfate, le carbone et le pH, contribuent à façonner la communauté dans les DMA, nous montrons que la forte spécialisation de niche explique l'assemblage singulier des communautés procaryotes dans chaque DMA, même si ces environnements partagent des caractéristiques physico-chimiques fortes. À l'échelle de notre enquête nous avons également montré que la dispersion ne peut pas être un processus important dans la composition des communautés dans les DMA. La spécialisation des niches est une hypothèse qui peut expliquer la limitation de dispersion apparente des procaryotes dans les DMA, bien que d'autres processus comme les propriétés intrinsèques des espèces et les effets de priorité ne sont peut être pas négligeables. Enfin nous avons étudié les interactions entre espèces et mis en évidence des pistes pour des investigations complémentaires. La conclusion notable de cette étude est que nous démontrons malgré 15 ans d'études de la microbiologie des DMA, que l'impulsion donnée par le développement des techniques récentes permet d'aller encore plus loin avec ces modèles historiques en permettant d'explorer de nombreux processus et tirer des conclusions plus générales en écologie microbienne des communautés.

Chapitre 4

Structure des communautés *Archaea* dans les DMA

4.1 Introduction

Les procaryotes jouent un rôle de premier ordre dans la mise en place et le maintien des DMA. La diversité des procaryotes dans les DMA a été très étudiée, et bien que beaucoup d'informations soient disponibles sur les bactéries, il y en a peu sur les *Archaea* (Hallberg 2010). La difficulté d'isoler les archées est une des causes qui explique ce manque d'information, illustrée par le tableau 1.2. Cependant les quelques souches disponibles nous renseignent sur leurs métabolismes en général. Il a été isolé des archées oxydant le fer et/ou le soufre comme les *Sulfolobus*, *Acidianus*, *Metallosphaera*, *Sulfurisphaera* et *Ferroplasma* (Edwards *et al.* 2000, Golyshina *et al.* 2000, Baker et Banfield 2003). Des études écologiques ont aussi montré qu'elles pouvaient même surpasser les bactéries dans la dissolution de la pyrite dans le drainage de la mine Richmond (États Unis) (Edwards 2000). Bruneel *et al.* (2008) ont aussi montré que *Ferroplasma acidiphilum* et des *Thermoplasmatales* étaient responsables de la majeure partie du lessivage du drainage de Carnoulès. Ce travail a pour but de décrire la diversité des *Archaea* dans les DMA afin d'enrichir nos connaissances sur leur écologie. Nous avons utilisé les données du séquençage intensif des gènes des ARNr 16S des 12 sites de DMA du projet MIGRAMD pour décrire la diversité des archées et la structure des communautés archées dans les DMA. Le séquençage haut débit nous permet d'avoir une vue exhaustive des taxons composant les communautés. Puis la comparaison de communautés issues de DMA

présentant des conditions environnementales différentes –en terme de pH, et de concentrations en arsenic notamment– peuvent nous permettre d’identifier des paramètres structurant ces communautés. Enfin nous avons aussi utilisé les bases de données publiques pour savoir si des archées adaptées à des milieux contraints comme les DMA pouvaient être observées dans d’autres environnements.

4.2 Résultats et discussion

4.2.1 Diversité et structure des communautés dans les DMA

Le séquençage 454 a permis un échantillonnage profond des communautés archées de nos DMA. En effet le taux de couverture de Good calculé pour une définition d’OTU à 95 % d’identité est de 1 pour tous les échantillons. C’est une couverture qui a été conservée après le sous échantillonnage que nous avons effectué pour ramener les échantillons au même nombre de séquences. Cependant les sites présentant la richesse la plus élevée (FRA_carn_cowg/s1, et ESP_odi_st) ont une couverture un peu moins bonne après le sous échantillonnage de 0,99. La richesse observée dans ces sites est de 32 ; 36 et 30 respectivement. La richesse la plus faible est observée dans les sites BOL_oru1/pot/mil et ESP_R.T_ri avec 6, 7 et 8 OTU. La diversité la plus faible est celle de ESP_R.T_ri et la plus importante celle du site BOL_oru0. Le sous échantillonnage n’a pas eu d’impact sur l’estimation de la diversité archées des sites d’études (Table 4.1, 4.2). La majorité des OTU observés ont été affiliés aux phyla des *Euryarchaeota* et *Thaumarchaeota*. Les *Thaumarchaeota* représentent jusqu’à 30 % des effectifs de séquences dans les sites ESP_R.T_pt2, ESP_odi_st et PRT_sd. Hormis dans le site BOL_oru0, les *Thaumarchaeota* sont marginaux ou n’ont pas été détectés comme dans les sites Bol_pot/mil et ESP_R.T_ri/pt1. Le reste des séquences a été affilié aux *Euryarchaeota*, qui regroupe la grande majorité des archées dans les DMA étudiés (Figure 4.1). Nous avons utilisé la distance Unifrac pour voir si les communautés des différents DMA étaient phylogénétiquement proches ou pas (Figure 4.2). On n’observe pas un patron de regroupement géographique des communautés, sauf pour les deux sites du drainage de Carnoulès. Les autres sites forment trois groupes : BOL_pot, ESP_R.T_pt1, ESP_R.T_ri, ESP_odi_th ; BOL_oru0, BOL_oru1, BOL_mil et PRT_sd, ESP_odi_th, ESP_R.T_pt2. Ce dernier groupe semble se distinguer des autres par la présence de *Thaumarchaeota*. Les

TABLE 4.1 – Description de la richesse et de la diversité des communautés *Archaea* dans les DMA de l'étude pour une définition d'OTU à 95 %.

| Site | Nbr. de séquences | Nbr. d'OTU | couverture de Good | richesse Chao1 | diversité Shannon |
|---------------|-------------------|------------|--------------------|----------------|-------------------|
| BOL_pot | 2623 | 7 | 1,00 | 7,00 | 1,43 |
| BOL_mil | 1564 | 8 | 1,00 | 8,00 | 1,69 |
| BOL_oru0 | 5009 | 17 | 1,00 | 18,00 | 2,06 |
| BOL_oru1 | 1509 | 6 | 1,00 | 6,00 | 1,35 |
| FRA_carn_cowg | 3116 | 32 | 1,00 | 44,00 | 1,68 |
| FRA_carn_s1 | 3337 | 36 | 1,00 | 45,00 | 1,37 |
| ESP_odi_st | 4542 | 30 | 1,00 | 32,14 | 1,76 |
| ESP_odi_th | 1204 | 14 | 1,00 | 14,00 | 1,51 |
| ESP_R.T_pt1 | 3303 | 12 | 1,00 | 15,00 | 1,68 |
| ESP_R.T_pt2 | 4214 | 27 | 1,00 | 30,75 | 1,73 |
| ESP_R.T_ri | 1693 | 8 | 1,00 | 8,50 | 0,56 |
| RT_sd | 4860 | 26 | 1,00 | 26,75 | 2,12 |

TABLE 4.2 – Description de la richesse et de la diversité des communautés *Archaea* dans les DMA de l'étude pour une définition d'OTU à 95 %, après un échantillonnage aléatoire des séquences pour ramener les échantillons au même nombre de séquence.

| Site | Nbr. de séquences | Nbr. d'OTU | couverture de Good | richesse Chao1 | diversité Shannon |
|---------------|-------------------|------------|--------------------|----------------|-------------------|
| BOL_pot | 1204 | 7 | 1,00 | 7,00 | 1,42 |
| BOL_mil | 1204 | 8 | 1,00 | 8,00 | 1,67 |
| BOL_oru0 | 1204 | 15 | 1,00 | 15,00 | 2,08 |
| BOL_oru1 | 1204 | 6 | 1,00 | 6,00 | 1,36 |
| FRA_carn_cowg | 1204 | 24 | 0,99 | 31,00 | 1,66 |
| FRA_carn_s1 | 1204 | 27 | 0,99 | 36,00 | 1,32 |
| ESP_odi_st | 1204 | 21 | 0,99 | 42,00 | 1,74 |
| ESP_odi_th | 1204 | 14 | 1,00 | 14,00 | 1,51 |
| ESP_R.T_pt1 | 1204 | 9 | 1,00 | 9,00 | 1,65 |
| ESP_R.T_pt2 | 1204 | 22 | 1,00 | 27,00 | 1,72 |
| ESP_R.T_ri | 1204 | 7 | 1,00 | 7,50 | 0,56 |
| PRT_sd | 1204 | 21 | 1,00 | 21,50 | 2,10 |

communautés archées de nos 12 sites de DMA semblent se structurer davantage en fonction de conditions environnementales locales, que de leur localisation.

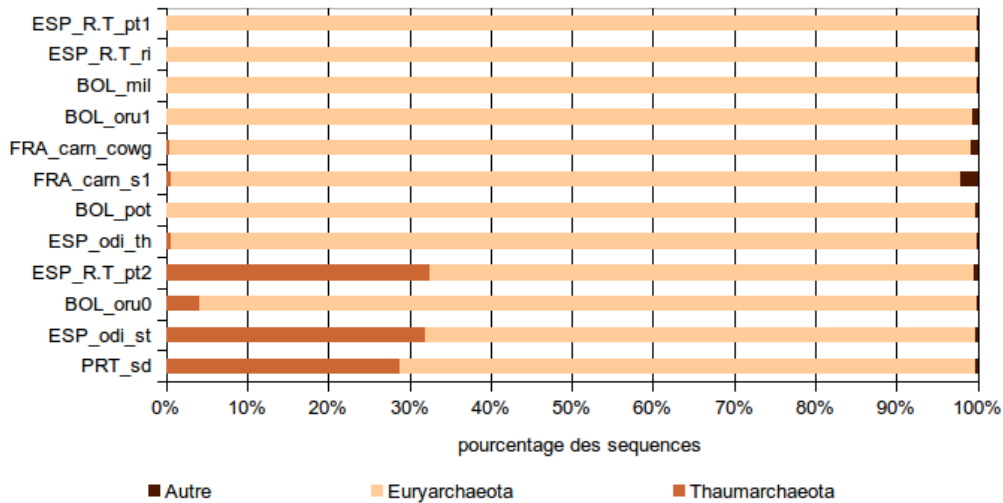


FIGURE 4.1 – Répartition des phyla *Archaea* majoritaires dans les sites d'étude.

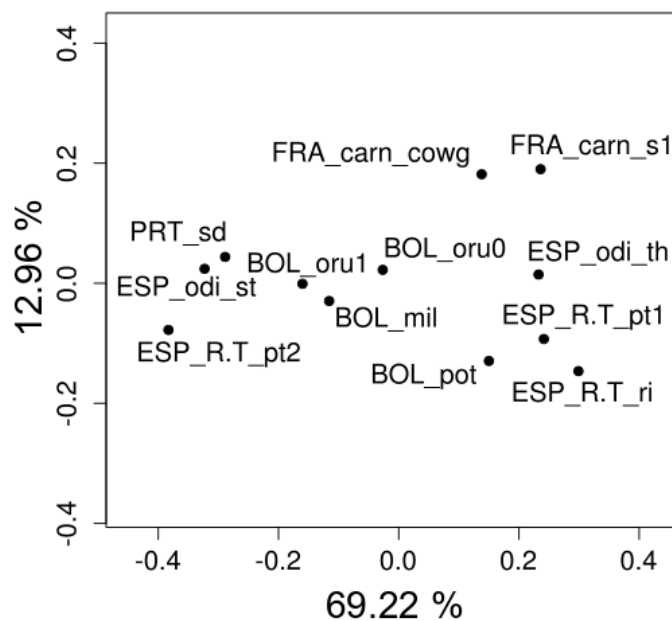


FIGURE 4.2 – Représentation des deux premiers axes (axe 1 abscisse ; axe 2 ordonné) d'une analyse en coordonnées principales des distances Unifrac entre les communautés archées des DMA étudiés.

4.2.2 Taxa *Archaea* retrouvés dans les DMA

Dans la figure 4.1, on observe que la majorité des séquences appartient aux *Euryarchaeota*. L'arbre phylogénétique construit à partir des séquences représentatives des OTU (Figure 4.3) montre que ce phylum possède le plus d'OTU (24) et les plus abondants (5115 séquences pour l'OTU_67). Parmi les *Euryarchaeota*, tous les OTU sauf un sont dans l'ordre des *Thermoplasmatales*, qui englobent 88 % du total des séquences de l'étude. Un OTU est affilié aux *Ferroplasma* (OTU_81), il est abondant (13 % de l'effectif total) et partage 98 % d'identité avec les séquences de référence de *Ferroplasma acidiphilium* et *Ferroplasma acidarmanus*. Toutes les séquences de notre étude qui ont été affiliées aux *Ferroplasma* se regroupent dans le même OTU, qui aurait aussi absorbé les séquences de référence si on les avait intégrées lors du calcul des OTU. Notre définition d'OTU est concordante avec la définition du groupe *Ferroplasma*. Neuf OTU ont été affiliés aux *Thermoplasmataceae*. Aucun n'est proche d'espèces de *Thermoplasma* isolées. En effet, ces séquences partagent 90 % d'identité de séquences avec les autres séquences du groupe. *Thermoplasma acidophilum* et *Thermoplasma volcanium* sont des archées thermophiles (Table 1.2) et les DMA que nous avons échantillonnés sont mésothermaux ce qui expliquerait pourquoi nous n'avons pas amplifié de gènes ARNr 16S de ces micro-organismes. Trois OTU en revanche sont proches de génomes reconstitués (Tyson *et al.* 2004, Dick *et al.* 2009) : E-plasma et OTU_60, A-plasma et OTU_80 et *Thermoplasmatales* Gpl et OTU_77. Yelton *et al.* (2013) ont réalisé des études de génomique comparatives pour avoir plus d'informations sur le métabolisme potentiel de ces procaryotes. Ils se révèlent proches génétiquement des *Ferroplasma*, mais des particularités de membrane et de régulation d'enzymes feraient qu'ils colonisent des niches différentes car ils cohabitent dans les mêmes biofilms. Ces OTU sont potentiellement des genres qui sont à définir dans la famille *Thermoplasmataceae*. Le "Terrestrial Miscellaneous Euryarchaeotal Group" (TEMG), regroupe 8 OTU pour 34 % des effectifs totaux de séquences. Ce groupe des *Thermoplasmatales* est construit à partir de séquences environnementales d'environnements terrestres. Les distances entre les représentants de nos OTU dans ce groupe sont importantes, les séquences les plus éloignées ne partagent que 81 % d'identité. Le groupe BSLdp215 est aussi défini avec des séquences environnementales, mais provenant d'un lac chaud volcanique du "Lassen Volcanic National Park". Ce groupe contient un OTU très abondant (OTU_75) et les deux OTU les plus éloignés partagent 90 % d'identité de séquence. Un OTU (OTU_55)

dont l'effectif n'est pas négligeable est affilié à la famille des *Methanosaetaceae*. Ce sont des archées méthanogènes acétotrophes. Elles transforment l'acétate en méthane et dioxyde de carbone. Ce sont des micro-organismes anaérobies stricts, et les DMA que nous étudions sont oxygénés, ce qui amène à se poser des questions sur leur présence. Des archées méthanogènes ont aussi été observées dans le drainage de Carnoulès (Volant *et al.* 2012) ainsi que dans le bassin du Rìo Tinto (Sanz *et al.* 2011). Dans cette dernière étude une activité méthanogène a été mesurée dans des microcosmes de sédiments du Rìo Tinto. La profondeur du sédiment crée des niches, anaérobies, réductrices et avec un pH plus neutre, favorables aux archées méthanogènes. L'OTU 55 n'est pas présent dans le Rìo Tinto mais seulement dans les sites boliviens BOL_oru0 et Bol_mil. Nous n'avons pas échantillonné assez profondément pour atteindre la zone de sédiment anoxique décrite par Sanz *et al.* (2011). Les sédiments boliviens sont plus fins et compacts que ceux du Rìo Tinto et plus riches en matière organique. Il est donc possible que des niches favorables aux méthanogènes soient plus proches de la surface dans les DMA boliviens. D'après la taxonomie SILVA utilisée, 4 OTU appartiennent au groupe environnemental DHVEG-6 et 2 au groupe MEG des *Euryarchaeota*. Ce sont des OTU minoritaires que l'on observe principalement dans les sites de Carnoulès. Cependant ces groupes contiennent les séquences de gènes ARNr 16S des ARMAN-4 *Candidatus Parvarchaeum acidiphilum* et ARMAN-2 *Candidatus Micrarchaeum acidiphilum* (voir chapitre 1, 1.2.4). Une étude de génomique comparative de procaryotes non cultivés provenant de taxa peu documentés, postérieure à la base de donnée SILVA que nous utilisons, propose de réorganiser ces groupes (Rinke *et al.* 2013). Notre groupe d'OTU du clade DHVEG-6 ferait partie d'un nouveau phylum *Parvarchaeota* et les OTU du clade MEG du nouveau phylum *Micrarchaeota*. Ces phyla ainsi que d'autres se branchant profondément avec le phylum *Euryarchaeota* (*Nanoarchaeota* par exemple), forment un groupe monophylétique d'*Archaea* de petites tailles ayant des petits génomes. Le phylum *Thaumarchaeota* est le deuxième phylum dans lequel des OTU de notre étude ont été affiliés. Ils sont moins importants en effectif que les OTU affiliés aux *Euryarchaeota*, mais sont localement abondants. Les OTU les plus abondants sont retrouvés dans le groupe SAGMCG-1 qui regroupe des séquences d'un métagénome de mine sud africaine, dans le "terrestrial group" et le THSCG qui contient des séquences environnementales terrestres de sources chaudes. L'OTU_65 du SAGMCG-1 n'est quasiment observé que dans le site ESP_R.T._pt2. L'OTU_62 du "terrestrial group" est ob-

servé exclusivement dans les sites ESP_R.T._pt2, ESP_odi_st et PRT_sd. Enfin l'OTU_72 du THSCG n'a été détecté que dans les sites ESP_odi_st, PRT_sd et BOL_oru0. Les OTU affiliés aux *Thaumarchaeota* semblent être très sensibles aux pressions environnementales de certains DMA et prospèrent dans des niches particulières.

4.2.3 Spécificité des *Archaea* des DMA

De l'analyse de fouille de données réalisée sur les communautés globales des DMA présentée dans le chapitre 3, nous avons extrait les informations relatives aux OTU *Archaea* (Figure 4.4). La première observation que l'on peut faire est que la diversité des environnements dans lesquels on observe des archées des DMA est moins importante que celle des bactéries. Les OTU communs ont plus de 80 % de leur réponses de BLAST qui sont issues de milieux miniers ou acide, près de 10% de réponses de milieux de type sol et 5 % de type géothermal. Les OTU que l'on retrouve dans moins de la moitié des sites d'études ont une répartition des réponses de BLAST plus équitable entre ces 4 catégories d'environnements, à savoir 33 % pour les DMA, et environ 20 % pour les environnements acides, géothermaux et sol. Il est clair que les archées que l'on a observées dans les DMA sont spécifiques et adaptées aux conditions de ce milieu (acidité, concentration en métaux). Cependant il est possible que des niches d'autres milieux correspondent également à des niches présentes dans les DMA. Les sols dans lesquels ont été retrouvées les séquences très proches de celles des archées des DMA, sont peut-être acides sans que cela ne soit enregistré dans le fichier d'information de la séquence. Les sources géothermales peuvent présenter des conditions proches des DMA (acidité, soufre, métaux) mais avec une température élevée. Les DMA que nous étudions sont mésothermaux, mais il est possible qu'il existe des micro-niches hyperthermales permettant la croissance de tels organismes. Ces résultats sont concordants avec l'arbre phylogénétique qui montre des taxons de *Thaumarchaeota* de sol et de sources hydrothermales ainsi qu'un taxon de sol des *Thermoplasmatales*.

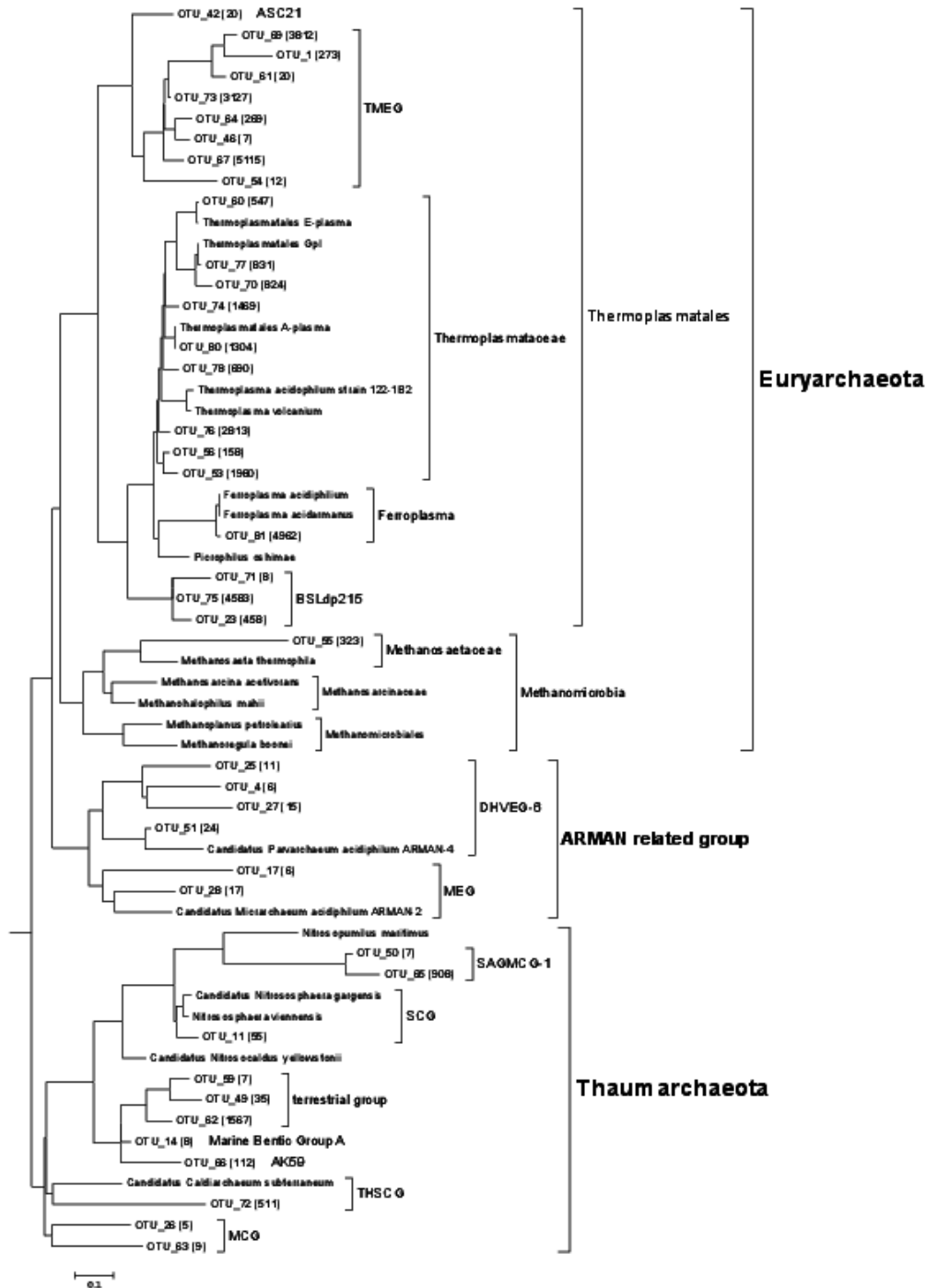


FIGURE 4.3 – Arbre phylogénétique des séquences des représentants des OTU (95 %) *Archaea* ayant plus de 5 séquences. L'arbre a été construit avec le logiciel ARB et la base de données SILVA v.111 datant de Juillet 2012. Les séquences ont été intégrées dans l'arbre SILVA par la méthode du maximum de parcimonie, qui est celle par défaut dans ARB avec le filtre *Archaea*. La nomenclature des taxons et des divisions est celle utilisée dans SILVA (hormis le "ARMAN related group"). L'arbre est enraciné sur le domaine *Bacteria*, l'échelle représente le nombre de substitutions de nucléotides. MCG : Miscellaneous Crenarchaeotic Group; THSCG : Terrestrial Hotspring Crenarchaeotic Group; SCG : Soil Crenarchaeotic Group; SAGMCG-1 : South African Gold Mine Group 1; MEG : Miscellaneous Euryarchaeotic Group; DHVEG-6 : Deep Sea Hydrothermal vent Group 6; TMEG : Terrestrial Miscellaneous Euryarchaeotal Group.

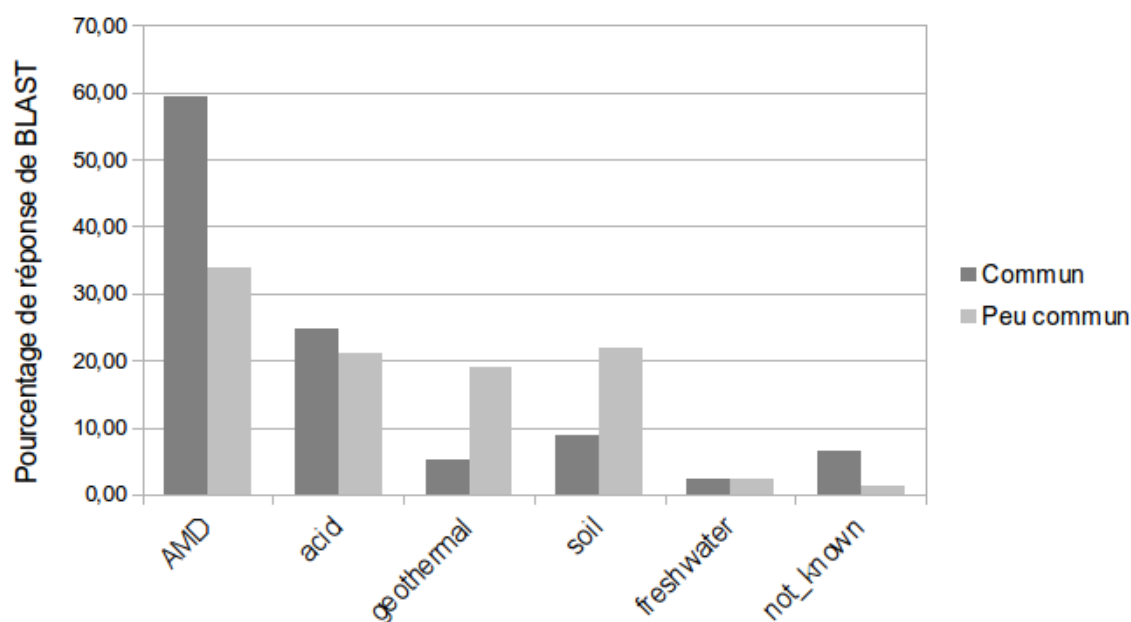


FIGURE 4.4 – Diagramme illustrant les résultats de la fouille de données sur les séquences *Archaea*. Les barres représentent la proportion de réponse BLAST appartenant à un environnement défini. La définition des environnements est disponible dans le chapitre "Matériel et méthodes". Les OTU communs sont ceux qui ont été observés dans plus de la moitié des échantillons. Les peu communs dans moins de la moitié.

TABLE 4.3 – Paramètres physico-chimiques de l'eau des sites d'étude. O₂ dissout, sulfate et Fe en mg.L⁻¹; Eh en mV; conductivité en S; AsIII, AsV, Al, Cu, Zn en µg.L⁻¹; C est le carbone total mesuré par perte au feu exprimé en g de C / g de matière sèche des sédiments.

| Sites | O ₂ dissout | T (°C) | pH | Fe (total) | sulfate | AsIII | AsV | Al | Cu | Zn | C |
|---------------|---------------------------|--------|------|------------|---------|--------|-------|---------|--------|---------|------|
| BOL_pot | 6,66 | 10,7 | 3,04 | 1340 | 9843 | 918 | 4086 | 238300 | 2586 | 2810000 | 0,09 |
| BOL_mil | 4,63 | 10,3 | 3,49 | 360,3 | 1827 | 11 | 152 | 4153 | 357 | 117000 | 0,30 |
| BOL_oru0 | 4,96 | 22,0 | 2,10 | 820,7 | 5990 | 527 | 4856 | 529400 | 1552 | 50770 | 0,21 |
| BOL_oru1 | 3,77 | 21,2 | 1,31 | 4920 | 29843 | 13910 | 83980 | 1677000 | 73030 | 180400 | 0,19 |
| FRA_carn_cowg | 7,65 | 17,4 | 3,22 | 1149 | 2819 | 127100 | 33260 | 60450 | 769 | 22390 | 0,04 |
| FRA_carn_s1 | 1,29 | 16,7 | 3,21 | 1256 | 3557 | 180500 | 20230 | 64990 | 833 | 24060 | 0,04 |
| ESP_odi_st | 1,75 | 23,1 | 2,53 | 1972 | 40465 | 1 | 73 | 1520000 | 173300 | 456900 | 0,07 |
| ESP_R.T_pt1 | 5,63 | 19,3 | 1,79 | 17220 | 208140 | 8 | 8959 | 3079000 | 15240 | 15560 | 0,08 |
| ESP_R.T_pt2 | 8,77 | 14,8 | 2,64 | 238 | 25581 | 1 | 39 | 142800 | 1637 | 5220 | 0,03 |
| ESP_R.T_ri | 4,63 | 22,0 | 1,02 | 20620 | 194186 | 176 | 34785 | 2670000 | 109400 | 99310 | 0,10 |
| Sao_Domingos | 5,94 | 27,2 | 2,09 | 8432 | 48140 | 12 | 39255 | 923600 | 661900 | 364600 | 0,07 |
| ESP_odi_th | 4,25 | 26,8 | 2,51 | 2718 | 38605 | 367 | 16625 | 918600 | 125300 | 450300 | 0,08 |

4.2.4 Communautés *Archaea* structurées par les paramètres environnementaux

De la même façon que pour les communautés procaryotes globales, nous nous sommes demandés si les paramètres environnementaux structuraient spécifiquement les communautés *Archaea* des DMA. En effet, les 12 DMA échantillonnés proposent des conditions environnementales différentes, comme le pH, les concentrations en ions métalliques ou en carbone total (Table 4.3). Nous avons pour cela construit un modèle d'ACC parcimonieux (Figure 4.5) pour identifier ces paramètres. Notre modèle a identifié deux paramètres importants pour la structuration de la communauté *Archaea* : la quantité de carbone total des sédiments et la concentration en As III. Notre modèle explique environ 31% de la variance des OTU au sein des sites étudiés. D'autres facteurs sont donc en cause dans la structuration de ces communautés. Dans notre étude, la concentration en As III sélectionne fortement les OTU 27, 28 et 51 qui se trouvent dans les groupes phylogénétiques des ARMAN. Ce paramètre sélectionne aussi les OTU 56 et 76 des *Thermoplasmataceae* et l'OTU 75 du groupe BSLdp215 des *Thermoplasmales*. Le bassin du *Lassen Volcanic National Park* contient en moyenne 10 mg.L⁻¹ d'arsenic dans ses systèmes fluviaux (Sorey *et al.* 1994). Cela concorde avec la sélection de l'OTU 75 par l'arsenite. Le carbone total a sélectionné l'OTU 55 *Methanosaeta*, l'OTU 11 *Thaumarchaeota*, et les OTU 60 ; 23 ; 69 ; 53 ; 64 de *Thermoplasmatales*. Il semble que des taxons de *Thermoplasmataceae* soient sélectionnés par leur résistance à l'arsenic et leur oligotrophie.

4.3 Conclusion

Cette étude avait pour but de décrire la diversité des *Archaea* dans différents systèmes de DMA afin d'enrichir les connaissances sur leur écologie. Le séquençage haut débit nous a permis d'avoir un échantillonnage suffisant pour avoir une vision exhaustive des taxons composants les communautés archées des DMA. D'après la systématique actuelle, les taxa que nous avons observés dans les DMA appartiennent à deux phyla : *Euryarchaeota* qui regroupe la majorité des séquences (88 % des séquences sont affiliées aux *Thermoplasmatales*, un ordre des *Euryarchaeota*) et *Thaumarchaeota* dont les effectifs sont moins importants globalement mais le sont localement. Les OTU de l'ordre *Thermoplasmatales* sont observés

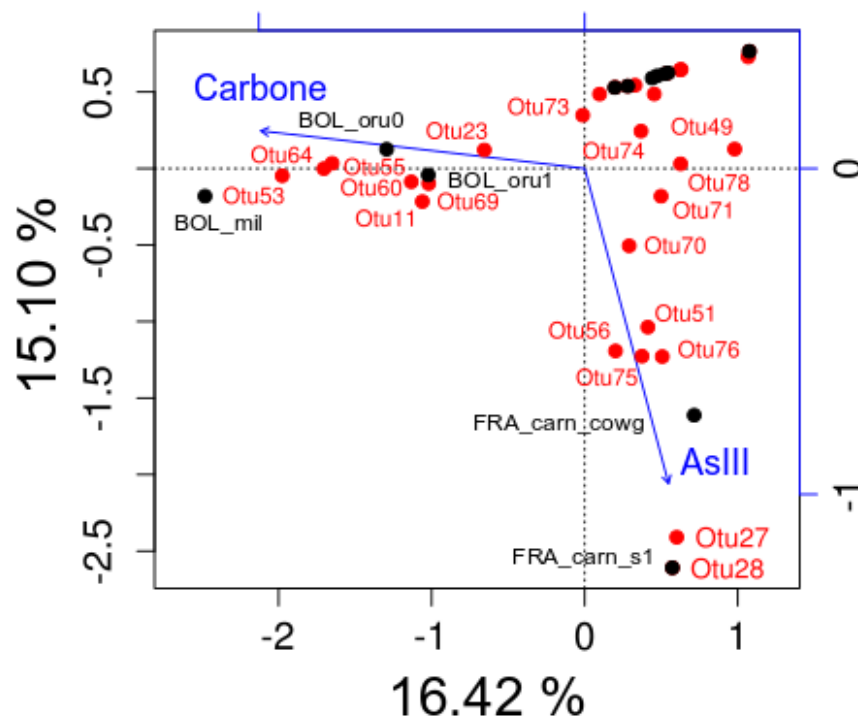


FIGURE 4.5 – ACC sur les OTU *Archaea* ayant plus de 5 séquences, en utilisant les paramètres environnementaux mesurés comme contrainte. Les axes 1 (abscisse) et 2 (ordonnée) sont représentés. Les OTU (rouge) se répartissent le long des vecteurs bleus en fonction de leur maxima écologique pour ce paramètre, les sites (noir) sont des centroïdes des OTU. Seul le nom des OTU et des sites influencés par les paramètres significatifs de l'ACC ont été représentés.

dans tous les DMA, l'OTU_81 est affilié au genre *Ferroplasma* et contient potentiellement des espèces différentes de celles qui ont été isolées. Aucun de nos OTU en revanche n'a été affilié au genre *Thermoplasma* défini par les deux souches types *Thermoplasma acidophilum* et *Thermoplasma volcanium*. Ces souches ne semblent pas représentatives de la diversité des *Thermoplasmataceae* dans les DMA, les mêmes résultats sont observés chez Baker et Banfield (2003), Volant *et al.* (2012). Cependant trois OTU sont très proches des séquences de gène ARNr 16S des génomes reconstitués des *Thermoplasmatales* A-plasma, E-plasma et Gpl (Tyson *et al.* 2004, Dick *et al.* 2009). Les analyses génomiques de Yelton *et al.* (2013) fournissent des informations sur le potentiel métabolique de ces *Archaea* qui sont très similaires entre eux et ainsi que de celui des *Ferroplasma*, malgré leurs différences taxonomiques. Or ces organismes cohabitent dans les mêmes biofilms. Des particularités de structure de membrane et des régulations enzymatiques différentes doivent être responsables de leur adaptation à différentes niches. Une grande partie des OTU des *Thermoplasmatales* sont dans des groupes de séquences environnementales qui fournissent peu d'informations hormis une localisation géographique ou un type d'environnement. La définition de ces groupes évolue avec l'enrichissement des informations sur les archées. Dans cette étude, nous avons l'exemple de nos OTU des groupes DHVEG-6, et MEG qui contiennent les séquences des ARMAN et qui d'après les travaux de Rinke *et al.* (2013) pourraient être deux autres phyla *Parvarchaeota* et *Micrarchaeota*, définis par des petites tailles cellulaires et des génomes compacts. Ces deux exemples montrent comment les approches génomiques peuvent compléter les méthodes culturelles en apportant des informations sur des organismes qui sont réfractaires à la culture en laboratoire. Cependant elle ne permettent l'accès qu'au contenu du génome, et les informations sur les activités et les régulations enzymatiques ne sont pas accessibles. Des études physiologiques et biochimiques (sur des souches isolées ou des consortia) sont donc nécessaires pour confirmer les hypothèses émises par les études génomiques. Il n'y a pas de *Thaumarchaeota* de drainage minier acide isolée, et les OTU de notre étude qui y sont affiliés, sont dans des groupes environnementaux. Même si on ne peut pas caractériser ces archées, on apprend qu'ils sont spécifiques de quelques drainages, et donc probablement sélectionnés par des conditions particulières. Cependant notre modèle ACC n'a pas été capable d'identifier des paramètres permettant de sélectionner des OTU *Thaumarchaeota*. Notre modèle a en revanche identifié deux paramètres importants dans la structure des communautés, sé-

lectionnant des OTU *Euryarchaeota* : la concentration en arsenite et la quantité de carbone total. Ils sélectionnent notamment des OTU non caractérisés des *Thermoplasmatales* par leur résistance à de fortes concentrations en arsenite et des conditions plus ou moins oligotrophes. La concentration en arsénite a aussi sélectionné les OTU ARMAN. On ne peut pas comparer nos conditions avec celles du drainage de Richmond où ces archées ont été isolées car les données sur l'arsenic ne sont pas disponibles. Méndez Garcia (2013) dans ses travaux a séquencé beaucoup de clones ARMAN dans le drainage de Los Ruedos où les concentrations en arsenic sont bien moins élevées qu'à Carnoulès (3,6-6 mg.L⁻¹ contre environ 200 mg.L⁻¹). Dans ces travaux les ARMAN ont été observés dans des biofilms. Notre étude montre qu'on peut les retrouver dans les sédiments aussi. Des études de microscopie électronique suggèrent des relations parasitaires entre des ARMAN et des *Thermoplasmatales* (Comolli *et al.* 2009). Des hôtes compatibles sont peut être présents dans le drainage de Carnoulès. Cependant il est clair que d'autres informations sont nécessaires pour comprendre la répartition des ARMAN dans les DMA. Enfin l'analyse de fouille de données nous a permis de montrer que les OTU fréquemment observés dans les DMA sont spécifiques des milieux miniers et acides. Cependant localement il est possible de retrouver des archées adaptées à des environnements de type sol et géothermaux. Ces résultats indiquent que les conditions environnementales dans les DMA sont importantes dans la structuration des communautés archées.

Chapitre 5

Diversité des transporteurs d'arsenic dans les DMA

5.1 Introduction

Le drainage minier acide est un milieu contraint par des concentrations élevées en métaux et métalloïdes, dont l'arsenic. Des teneurs élevées en arsenic sont mortelles, les procaryotes habitant les DMA ont donc développé des stratégies de résistance à cet élément. Parmi ces stratégies de résistance il y a la mise en place de transporteurs pour expulser l'arsenic hors de la cellule. Chez les procaryotes les transporteurs les plus communs sont de type ArsB et Acr3. Ce sont des transporteurs uniports qui expulsent l'arsenite hors de la cellule (voir chapitre 1.2.3). L'origine de ces transporteurs semble très ancienne, et serait issue d'une convergence évolutive (Silver et Phung 2005, Cai *et al.* 2009, Li *et al.* 2014). À savoir que les transporteurs ArsB et Acr3 ne semblent pas avoir d'origine commune, et sont tous les deux une réponse évolutive créée indépendamment pour s'adapter à la contrainte de la présence d'arsenic. La famille des transporteurs Acr3 est divisée en deux groupes Acr3.1 et Acr3.2 et un couple d'amorces spécifiques de chaque groupe a été dessiné par Achour *et al.* (2007). La famille des transporteurs Acr3 est présente dans le domaine eucaryote, notamment chez les champignons, par la sous-famille Acr3.1. D'après les données actuelles la famille ArsB n'est présente que chez les procaryotes. La diversité des gènes codant pour les transporteurs d'arsenic a été étudiée dans les sols (Achour *et al.* 2007, Cai *et al.* 2009) dans le drainage de Carnoulès (Fahy *et al.* 2014) et dans une étude de génomique comparative de

Burkholderiales (Li *et al.* 2014). Elles ont révélé une diversité importante des transporteurs, ainsi que des indices suggérant des transferts horizontaux et des arrangements génomiques des gènes des opérons *ars*. En effet l'étude phylogénétique des transporteurs ArsB de souches cultivées par Cai *et al.* (2009) a montré des divergences dans le phylum des protéobactéries entre l'arbre des gènes ARNr 16S et l'arbre des séquences du gène *arsB*. Dans l'étude de Li *et al.* (2014) il a été montré qu'il existe des structures différentes des opérons *ars* chez les *Burkholderiales* qui sont issues de ré-arrangements entre différentes copies de l'opéron dans le génome ou les plasmides ou d'acquisitions extérieures. Ces travaux de thèse permettent l'accès à l'information génétique de sites miniers variées à travers le monde. Ainsi nous pouvons étudier la diversité des transporteurs exprimée par des communautés différentes. Nous avons construit des banques de clones pour accéder à cette diversité et réalisé une étude phylogénétique des séquences d'acides aminés des transporteurs observés. Une telle étude n'a pas encore été réalisée à une échelle aussi globale. Nous nous sommes aussi demandés si des séquences de transporteur étaient partagées entre les différents DMA étudiés, indiquant des échanges de communautés entre ces sites ou une origine commune de telles communautés.

5.2 Résultats et discussion

Les gènes codant les transporteurs d'arsenite n'ont pas pu être amplifiés dans tous les sites d'étude. Nous n'avons pu amplifier aucun des trois gènes dans les sites BOL_oru0/oru1 et ESP_R.T_ri. De plus le gène *arsB* n'a pas été amplifié dans les sites ESP_R.T_pt1 et BOL_mil alors que le gène *acr3.2* n'a pas été amplifié dans les sites BOL_pot, ESP_odi_th, PRT_sd (Table 5.1). Pour définir les OTU nous avons testé s'il existait une corrélation entre les distances des ARNr 16S et des séquences ArsB, Acr3.1 et Acr3.2 (nucléotidiques et d'acides aminés) dans des génomes complet de micro-organismes. Si cette corrélation existait cela montrerait une co-évolution des deux gènes et on pourrait définir des seuils de similarité qui correspondraient à des taxons. Or nous n'avons pas mis en évidence de corrélations significatives entre ces distances, la définition d'OTU que nous avons choisi ne correspond donc pas à un niveau taxonomique. C'est un seuil arbitraire qui nous permet de comparer les richesses et la diversité entre les différents sites d'étude, et d'agglomérer les séquences pour rendre les arbres plus lisibles. C'est d'ailleurs à partir de ces arbres que le seuil de 80

TABLE 5.1 – Tableau récapitulatif des statistiques sur les gènes codant les transporteurs d’arsenite ArsB et Acr3. On retrouve la richesse observée en nombre d’OTU défini à 80 % d’identité de séquence d’acides aminés ; le taux de couverture de Good, la richesse totale estimée par le Chao 1 et la diversité estimée par l’indice de Shannon. NA représente l’absence de données lorsque le gène n’a pas été amplifié.

| Sites | Nbr. OTU (80 % id) | | | Couverture de Good | | | Richesse Chao 1 | | | Diversité Shannon | | |
|---------------|--------------------|--------|--------|--------------------|--------|--------|-----------------|--------|--------|-------------------|--------|--------|
| | ArsB | Acr3.1 | Acr3.2 | ArsB | Acr3.1 | Acr3.2 | ArsB | Acr3.1 | Acr3.2 | ArsB | Acr3.1 | Acr3.2 |
| BOL_pot | 2 | 10 | NA | 100 | 93 | NA | 2 | 11 | NA | 0,51 | 1,87 | NA |
| BOL_mil | NA | 3 | 4 | NA | 100 | 93 | NA | 3 | 4,5 | NA | 0,95 | 0,53 |
| ESP_R.T_pt1 | NA | 4 | 4 | NA | 100 | 95 | NA | 4 | 4,5 | NA | 1,11 | 0,44 |
| ESP_R.T_pt2 | 7 | 11 | 2 | 93 | 89 | 100 | 10 | 16 | 2 | 1,36 | 1,87 | 0,66 |
| ESP_odi_th | 4 | 9 | NA | 100 | 90 | NA | 4 | 11 | NA | 0,73 | 1,60 | NA |
| ESP_odi_st | 4 | 5 | 1 | 100 | 100 | 100 | 4 | 5 | 1 | 1,27 | 0,99 | 0 |
| FRA_carn_cowg | 3 | 10 | 6 | 100 | 86 | 91 | 3 | 15 | 9 | 0,98 | 1,83 | 1,19 |
| FRA_carn_s1 | 3 | 9 | 4 | 100 | 88 | 95 | 3 | 19 | 5 | 0,91 | 1,63 | 0,50 |
| PRT_sd | 5 | 2 | NA | 100 | 100 | NA | 5 | 2 | NA | 1,42 | 0,32 | NA |

% d’identité a été choisi. En effet nous voulions que les OTU des séquences de transporteurs des DMA forment des groupes cohérents par rapport aux groupes formés par nos séquences de référence. À titre d’exemple le groupe des *Acidiphilium* de l’arbre des séquences ArsB, les séquences partagent 80 % d’identité de séquence ; dans le groupe *Acidithiobacillus* entre 84 et 88 % d’identité de séquences.

5.2.1 Phylogénie des séquences ArsB des DMA

La première observation que nous pouvons extraire de l’arbre des gènes ArsB (Figure 5.1) est que les séquences d’acides aminés de AsrB d’organismes appartenant aux mêmes phyla ne sont pas monophylétiques. Cependant on retrouve des groupes monophylétiques de séquences appartenant à des micro-organismes partageant des rangs taxonomiques inférieurs. Les groupes observés sont dans les taxons : *Acidiphilium*, *Thiomonas*, *Rhizobiales*, *Xanthomonadales*, *Pseudomonadales* et *Acidithiobacillales*. La plupart des séquences ArsB observées dans les DMA se regroupent avec des séquences de *Thiomonas* et d’*Acidiphilium*. Les séquences ArsB qui se regroupent avec les séquences de *Thiomonas* et *Desulfovibrio* proviennent de tous les sites où nous avons amplifié le gène *arsB*. Cependant notre étude du gène ARNr 16S montre très peu de *Thiomonas*, et ils n’ont pas été détectés dans les sites PRT_sd, ESP_odi_st/th, les *Desulfovibrio* sont quasiment indétectables dans tous l’échantillonnage des DMA (5 séquences affiliées). Le fait que deux séquences identiques soient portées par deux organismes appartenant à deux classes bactériennes différentes, montre que le gène *arsB*

peut être échangé entre des taxons éloignés. Ces transporteurs ArsB pourraient donc être portés par des procaryotes différents que l'on retrouverait fréquemment dans les DMA. Aucun de nos OTU ArsB ne se groupe avec les séquences de *Leptospirillum*. Ceci est surprenant car les *Leptospirillum* sont ubiquistes et abondants dans les DMA étudiés. Il est possible que nos amorces n'amplifient pas ces gènes *arsB*. Les OTU se regroupant avec les *Acidithiobacillales* ont peu de séquences et aucun OTU provenant du site ESP_R.T_pt2 n'est présent dans ce groupe. L'OTU abondant du site ESP_R.T_pt2 est dans le cluster *Acidiphilium*. Ce qui concorde avec nos observations d'ARNr 16S, car beaucoup de séquences affiliées aux *Acidiphilium* étaient présentes dans ce site. Dans ce groupe on remarque que les séquences de transporteurs ArsB des souche *Acidiphilium cryptum* JF-5 et *Acidiphilium multivorum* AIU301 sont identiques, ceci suggère des transferts horizontaux de gènes codants ce transporteur dans le genre *Acidiphilium*. Toujours dans ce groupe deux séquences ESP_odi_st_20 et PRT_sd_19 ont des transporteurs identiques, suggérant un mélange de populations entre ces sites ou un partage de micro-organismes d'origine commune. On retrouve le même patron dans le groupe des séquences de *Thiomonas arsenitoxidans* et *Thiomonas intermedia* avec les séquences ESP_odi_st_26 et PRT_sd_28. Deux OTU observés dans les sites de Carnoulès tombent dans un groupe paraphylétique de séquences de *alpha* et *gammaproteobacteria*. Ce groupe n'est pas appuyé par de fortes valeurs de bootstrap, les deux OTU représentés par les séquences FRA_carn_cowg_8 et FRA_carn_s1_36 sont clairement distincts. Ils pourraient être portés par des procaryotes spécifiques à ces points de prélèvements, dont les micro-niches locales ne permettraient pas les échanges entre elles. Enfin nous avons mis en évidence des séquences annotées comme ArsB dans les bases de données publiques, qui sont très distantes des autres séquences étudiées (Figure 5.1, a)). Dans ce cluster on retrouve des séquences ArsB issues de gènes génomiques ou plasmidiques de *alpha* et *betaproteobacteria*. L'interrogation des bases de données par BLAST semble indiquer que la séquence d'*Acidiphilium cryptum* JF5-2 appartient à la famille des transporteurs YbiR, cependant cette séquence reste encore différente au sein de ce groupe divergent. Dans la base de données "Conserved Domain Database" (CDD) de NCBI ArsB fait parti d'une super-famille de protéine "ArsB_NhaD_permease" dont le code est "cl17221". Les transporteurs YbiR font aussi parti de cette super-famille, et partagent des caractéristiques communes avec ArsB. Ce sont des perméases d'anions qui ont entre 8 et 13 hélices transmembranaires et qui

peuvent fonctionner soit indépendamment comme un transporteur chimiosmotique ou bien comme le canal d'une pompe à anion ATP dépendante. La nature des molécules transportées par les canaux YbiR n'est pas connue. Pourtant les autres séquences de ce groupe externe appartiennent aussi à la famille des transporteurs ArsB d'après les recherches BLAST. C'est l'indice qu'il existe peut être plusieurs sous-familles de ArsB comme c'est le cas pour le transporteur Acr3.

5.2.2 Phylogénie des séquences Acr3.1 des DMA

Le gène *acr3.1* a été amplifié dans neuf sites au total. C'est aussi le transporteur qui présente la plus grande richesse de séquences d'acides aminés (Table 5.1). Cette richesse contraste avec le peu de séquences issues de génomes complets disponibles dans les bases de données pour aider à leur affiliation. Les séquences provenant des génomes d'*Archaea* se mêlent avec les séquences des *Firmicutes* (Figure 5.2). Aucune des séquences Acr3.1 des DMA ne se groupe avec celle de *Saccharomyces cerevisiae*. Cela est cohérent avec le dessin des amorces qui n'a pas été fait pour amplifier les gènes *acr3.1* eucaryotes (Achour *et al.* 2007). On identifie un groupe important de séquences des DMA où se trouve aussi les séquences issues des *Actinobacteria*, ainsi qu'un groupe avec les séquences d'*alphaproteobacteria*. Enfin un troisième cluster se détache aussi mais il ne contient pas de séquences de références. Un résultat notable est que nous avons mis en évidence une corrélation linéaire faible (coefficient de Pearson = 0,48) mais significatif ($p < 0.01$ pour 1000 permutations) entre les distances Unifrac basées sur les ARNr 16S (définition d'OTU à 95 % d'identité) et celles basées sur les séquences d'acides aminés Acr3.1 (définition d'OTU à 80 % d'identité). Cette corrélation indique que plus les communautés microbiennes des DMA sont proches phylogénétiquement, plus les séquences Acr3.1 qu'elles vont posséder seront proches. Cela peut indiquer l'acquisition du transporteur par l'ancêtre commun du taxon, mais aussi des transferts horizontaux qui devraient se faire entre micro-organismes de taxons proches. Si on considère que les transporteurs Acr3.1 ayant des séquences proches appartiennent à des taxons proches, on peut faire des hypothèses sur des échanges de populations entre les DMA. Les séquences représentantes d'OTU ESP_odi_st_8 et PRT_sd_6 ainsi que ESP_odi_st_31 et PRT_sd_37 sont presque identiques. Cela suggère un partage de population entre ces deux régions avec des organismes ayant une origine géographique commune. Un autre patron que l'on observe

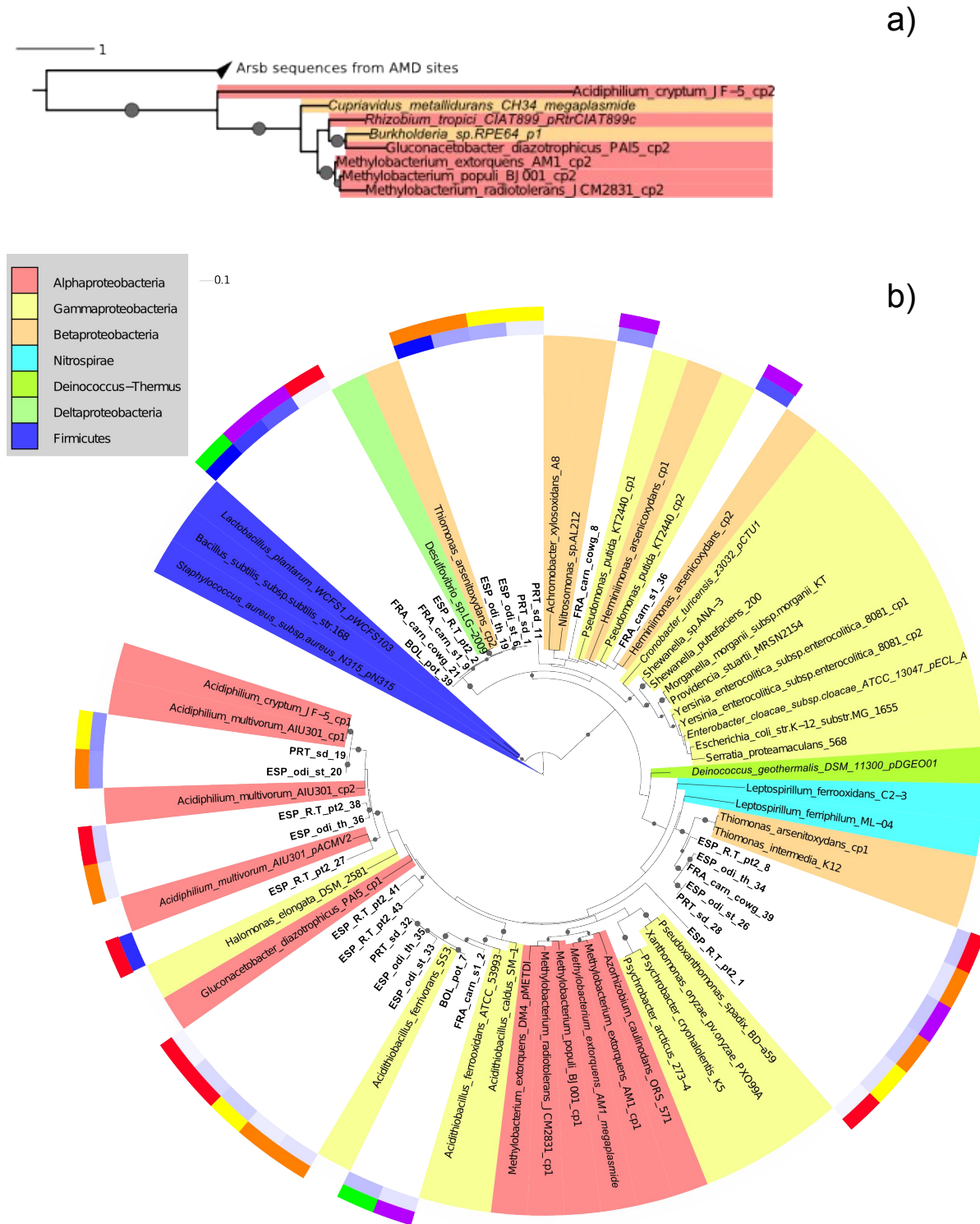


FIGURE 5.1 – Arbre phylogénétique inféré des séquences d'acides aminés des représentants des OTU ArsB et des séquences de ArsB des génomes et des plasmides séquencés. Nous avons éliminé les régions des séquences trop peu conservées avec Gblocks. L'arbre a été calculé avec l'algorithme de maximum de vraisemblance et 1000 répétitions de bootstrap, qui utilise un modèle LG avec une correction γ (4 catégories de site et un paramètre α estimé). **a)** groupe de séquences affiliées à ArsB mais distantes des autres séquences de l'étude. **b)** séquences des DMA et du reste des génomes et plasmides. Les cercles sur les nœuds sont proportionnels aux valeurs de bootstrap supérieure à 50. L'intensité de la couronne bleue interne est proportionnelle au nombre de séquences dans l'OTU. Les couleurs de la couronne extérieure représentent les régions d'échantillonnage. Violet Carnoulès; Rouge Rio Tinto; Orange Odiel; Jaune le district de São Domingo et Vert la Bolivie. Les représentants des OTU sont écrits en gras et les plasmides en italique. Les différentes copies du gène dans le même organisme sont indiquées par "cp". Les échelles indiquent un nombre de substitution d'acide aminé.

est celui des séquences FRA_carn_s1_36 et FRA_carn_cowg_17 qui sont très proche et spécifique du bassin de Carnoulès. Elles sont très proches de la séquence ESP_odi_th_38 du bassin Odiel, mais séparées par un nœud consolidé par un bootstrap fort, appuyant leurs différences. Cela pourrait illustrer une population commune entre ces deux sites, qui se serait ensuite adaptée aux conditions différentes qui y règnent et auraient évolué de façon divergente. Cette même observation peut être faite entre les séquences ESP_R.T_pt2_21 et ESP_odi_th_4 du bassin espagnol et les séquences BOL_mil_5 et BOL_pot_11 du bassin bolivien.

5.2.3 Phylogénie des séquences Acr3.2 des DMA

La phylogénie des séquences d'acides aminés Acr3.2 présente également des paraphylies dans les phyla bactériens. La séquence provenant de la δ -protéobactérie *Desulfobacula toluolica* est proche des cyanobactéries, et la séquence de *Methylacidiphilum infernovorum* (*Verrucomicrobia*) est au milieu d'un cluster de *alpha*, *beta*, *gammaproteobacteria* (Figure 5.3). L'histoire évolutive de ce gène semble être marquée par des acquisitions indépendantes du gène *acr3.2* par les bactéries. Les séquences des *gamma-proteobacteria* *Shigella flexneri* et *Azotobacter vinelandii* se regroupent avec des *beta-proteobacteria*, ce qui suggère une acquisition de ces gènes après la séparation de ces trois classes. La majorité des transporteurs Acr3.2 de nos sites d'étude sont proches de ceux retrouvés dans des *beta-proteobacteria*. BOL_mil_10, FRA_carn_cowg_5 et ESP_odi_st_6, ESP_R.T_pt2_13 sont dans un groupe terminal sans séquence de référence proche. Ces deux couples de représentants d'OTU sont très proches mais séparés par des nœuds avec de fortes valeurs de bootstrap. Il y aurait des populations communes à BOL_mil_10 et FRA_carn_cowg_5 ainsi qu'à ESP_odi_st_6 et ESP_R.T_pt2_13, qui retrouveraient dans ces différents sites des niches favorables. Ces deux populations qui ont une origine commune et qui ont ensuite divergées en fonction des conditions locales. Les deux OTUs majoritaires de Carnoulès (FRA_carn_s1_35 et FRA_carn_cowg_14) sont très similaires de la séquence de *Albidiferax ferrireducens* indiquant des partages de micro-organismes entre les deux sites de Carnoulès. Les séquences de l'OTU majoritaire de ESP_R.T_pt1 se placent dans un groupe près de *Cupriavidus metallidurans* contenant aussi des séquences des sites ESP_R.T_pt2, et de Carnoulès. Des micro-organismes peuvent donc être partagés par le drainage de Carnoulès et celui du Rìo

Tinto. Ces deux espèces appartiennent à l'ordre des *Burkholderiales*. Comme ce que l'on observe dans l'arbre ArsB les OTUs majoritaires s'associent avec des *Burkholderiales*. Cependant les procaryotes appartenant à cet ordre sont très minoritaires dans les drainages étudiés (maximum : 2,76 % des séquences d'ARNr 16S dans le site Rio_Tinto_Pt2), et aucune séquence affiliée aux *Burkholderiales* n'a été observée dans le site Rio_Tinto_Pt1. Il est possible que notre vision sur la diversité des transporteurs Acr3.2 soit biaisée par des amorces PCR amplifiants préférentiellement certaines séquences. Cependant le fait que nous n'ayons pas observé de séquences affiliées aux *Burkholderiales* dans le site Rio_Tinto_Pt1 et que des séquences Acr3.2 appartenant à des micro-organismes différents peuvent être proches, il est probable que les séquences de l'OTU Rio_Tinto_Pt1_10 appartiennent à d'autres organismes que des *Cupriavidus*.

5.3 Conclusion

La méthode moléculaire choisie, le clonage puis séquençage Sanger, pour l'étude de la diversité des transporteurs ArsB et Acr3 dans les DMA a permis une profondeur d'échantillonnage satisfaisante pour la définition d'OTU que nous avons choisie qui est 80 % d'identité de séquence d'acide aminé. Beaucoup des OTU ArsB et Acr3.2 majoritaires sont proches de séquences issues de *Burkholderiales*, pourtant minoritaires dans les DMA étudiés. Alors que nous n'avons pas observé de séquences ArsB proches des *Leptospirillum* qui sont pourtant très abondants dans les DMA, et peu de séquences ArsB proches des *Acidithiobacillus*, aussi abondants dans les DMA. Une étude récente de 188 génomes de *Burkholderiales* (Li *et al.* 2014) montrent que les transporteurs d'arsenic sont très répandus chez ces procaryotes avec une prédominance du type Acr3. Il y a peut être un effet loupe sur ces séquences à cause d'une amplification préférentielle des gènes *arsB* et *acr3.2* de *Burkholderiales* par nos amorces. Un biais qui serait plus important pour les amorces ciblant le gène *acr3.2*. Les amorces que nous avons utilisées ont été dessinées par Achour *et al.* (2007) pour identifier les transporteurs d'arsenic dans des souches isolées de sols contaminés à l'arsenic, avec 75 séquences. Elles ont ensuite été retravaillées par Fahy *et al.* (2014) pour les utiliser dans les DMA. Il serait opportun de réévaluer l'universalité de ces amorces suite à l'enrichissement des bases de données de séquences depuis 2007. C'est cependant un travail laborieux car

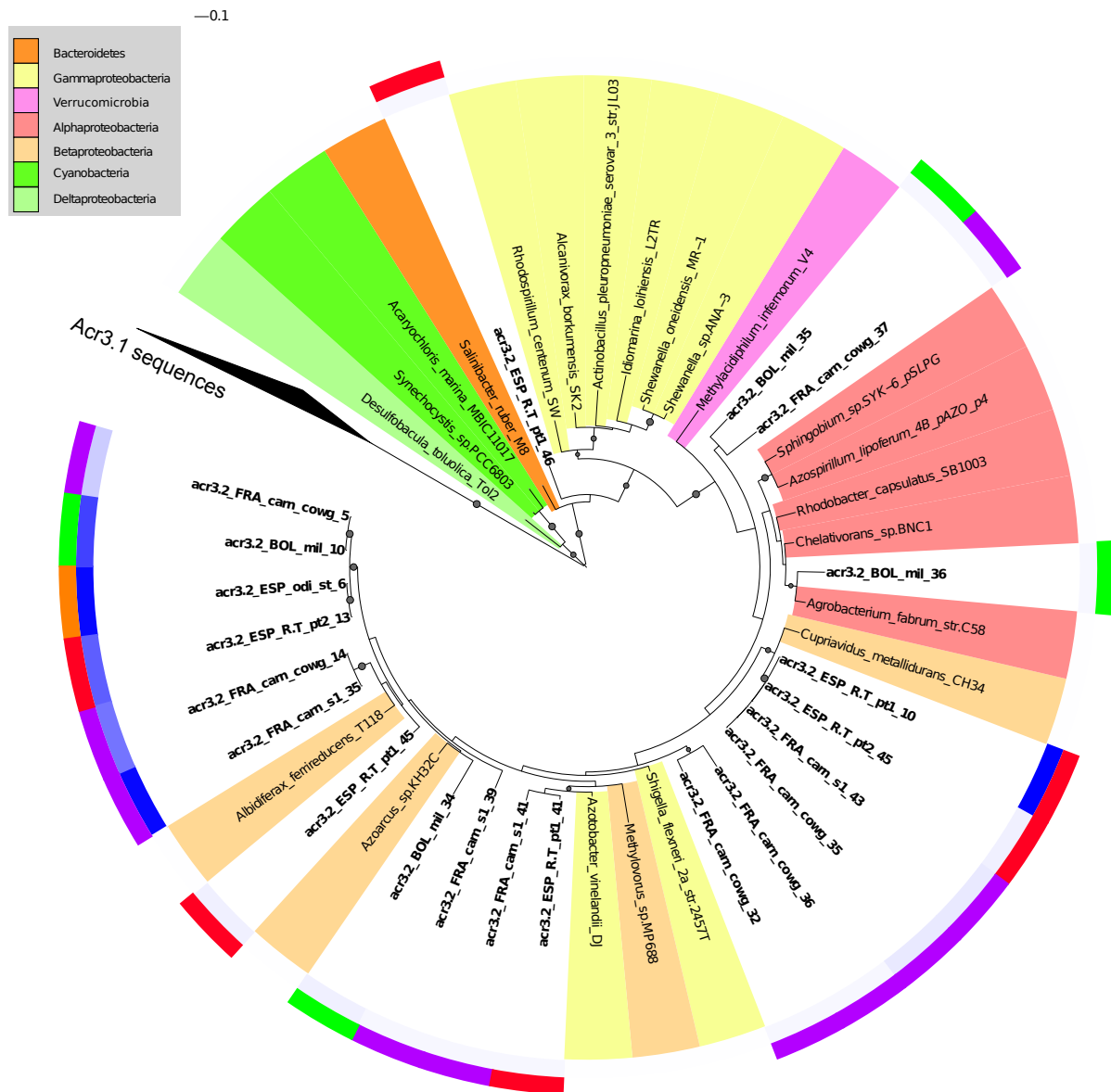


FIGURE 5.3 – Arbre phylogénétique inféré des séquences d'acides aminés des représentants des OTU Acr3.2 et des séquences de Acr3.2 des génomes et des plasmides séquencés. Nous avons éliminé les régions des séquences trop peu conservées avec Gblocks. L'arbre a été calculé avec l'algorithme de maximum de vraisemblance et 1000 répétitions de bootstrap, qui utilise un modèle LG avec une correction γ (4 catégories de site et un paramètre α estimé). Les cercles sur les noeuds sont proportionnels aux valeurs de bootstrap supérieure à 50. L'intensité de la couronne bleue interne est proportionnelle au nombre de séquences dans l'OTU. Les couleurs de la couronne extérieure représentent les régions d'échantillonnage. Violet Carnoules; Rouge Rio Tinto; Orange Odriel; Jaune le district de São Domingo et Vert la Bolivie. Les représentants des OTU sont écrits en gras et les plasmides en italique. Les différentes copies du gène dans le même organisme sont indiquées par "cp". Les échelles indiquent un nombre de substitution d'acide aminé.

on apprend de l'étude de Li *et al.* (2014) que beaucoup de séquences sont mal annotées et nécessitent une réévaluation pour confirmer leur annotation. Nous avons mis en évidence une corrélation entre les communautés procaryotes des DMA et les séquences d'acides aminés Acr3.1 observées dans les sites d'études. Les communautés présentant des taxons proches auront tendance à posséder des séquences Acr3.1 plus proches que des communautés ayant des taxons éloignés phylogénétiquement. De telles corrélations n'ont pas été mises en évidence pour les séquences ArsB et Acr3.2. C'est un résultat cohérent avec le biais d'amplification exprimé précédemment car une observation biaisée de la diversité des transporteurs ne peut pas être comparée à une diversité des gènes ARNr 16S beaucoup plus exhaustive. L'utilisation des transporteurs d'arsenic comme marqueurs phylogénétiques a permis de mettre en évidence des patrons de répartition géographique des procaryotes dans les DMA à différentes échelles : à l'échelle locale comme dans le drainage de Carnoulès où l'on a observé des OTU spécifiques des points de prélèvement (transporteur ArsB), ou bien partagés par les deux sites (transporteur Acr3.2); à une échelle supérieure on observe des OTU se partager dans différents sites du bassin de pyrite ibérique, ainsi qu'entre ce bassin et le drainage de Carnoulès; enfin à une échelle encore plus globale on observe des partages de transporteurs entre les sites européens et boliviens.

Nous ne pouvons cependant pas identifier les espèces impliquées à cause du manque de séquences de références disponibles dans les bases de données publiques, mais aussi des échanges horizontaux. Quant aux mécanismes qui régissent la dispersion des micro-organismes dans le monde, ils sont nombreux et très complexes à tester *in vivo*. Que l'on considère des mécanismes stochastiques de colonisation par une méta-population régionale (Hubbell 2001), ainsi que des mécanismes déterministes de filtre de l'habitat ou de compétition, les combinaisons possibles sont nombreuses. Dans une étude récente Székely *et al.* (2013) ont essayé de montrer *in vitro* avec des communautés simples les mécanismes qui pouvaient régir la dispersion des procaryotes. Ils ont montré que le filtre de l'habitat, la plasticité fonctionnelle et la compétition étaient des mécanismes importants. Enfin des études ont suggéré des échanges entre organismes des transporteurs d'arsenic (Achour *et al.* 2007, Cai *et al.* 2009) notamment au sein des opérons ars (Cai *et al.* 2009, Li *et al.* 2014). Nos arbres phylogénétiques vont dans ce sens en exprimant des groupes de séquences de transporteurs paraphylétiques appartenant aux mêmes phyla. On pourrait émettre l'hypothèse que les transporteurs portés

par des plasmides pourraient être une source de ces échanges. Cependant nos analyses ne vont pas dans le sens de cette hypothèse car les séquences plasmidiques se regroupent avec les séquences génomiques des organismes dont ils sont issus ou avec des taxons proches.

Conclusion générale et perspectives

Comprendre les mécanismes qui structurent les communautés est essentiel pour comprendre le fonctionnement des écosystèmes. Bien que les processus structurant les communautés des macro-organismes aient été étudiés depuis des décennies, c'est un sujet qui continue d'être discuté. L'étude des communautés de micro-organismes s'est heurtée aux difficultés de dresser un inventaire exhaustif des individus composant les communautés. Les méthodes d'identifications culturales des micro-organismes ne permettent l'identification que d'une infime partie des micro-organismes (Amann *et al.* 1995), biaisant fréquemment les résultats vers l'idée que les espèces microbiennes sont ubiquistes (Green et Bohannan 2006). Les avancées des méthodes moléculaires ont permis de se libérer de la mise en culture pour l'identification des micro-organismes, et la puissance des technologies de séquençage et de traitement informatique autorise des inventaires profonds des communautés microbiennes. Cependant la confirmation ou l'infirmité de la première partie du postulat de Baas Becking « everything is everywhere »¹ reste difficile car chaque méthode possède une limite de détection. De plus, l'universalité des amorces de PCR est relative car elles sont dessinées à partir des séquences d'ADN connues. Les séquences trop divergentes ne sont donc pas détectées (Baker *et al.* 2003). Cette idée fait aussi écho aux processus stochastiques de structuration des communautés proposés par Hubbell (2001) à savoir des événements de colonisation par un pool régional et des événements de naissance et de mort à une échelle locale. La deuxième partie du postulat de Baas Becking « but the environment selects » renvoie à des processus déterministes de structuration des communautés, comme le filtre de l'habitat, la théorie des niches ou les interactions entre les micro-organismes.

Les travaux réalisés lors de cette thèse dans le cadre du projet **MIGRAMD** renseignent sur les processus qui structurent les communautés dans les DMA. Cela a été d'abord possible par la profondeur d'échantillonnage des communautés procaryotes des DMA qu'autorise la technologie de séquençage haut-débit 454. En effet nous avons été capables d'observer plus de 98 % de la richesse des genres bactériens et archées, donnant un inventaire quasi exhaustif des communautés procaryotes des DMA. La richesse bactérienne observée correspond à celle qui a été observée par les méthodes culturales (Table 1.2) ou bien par les banques de clones (Hallberg 2010). Cependant il est intéressant de remarquer que malgré leur facilité de mise en

1. Baas-Becking, Lourens G.M. (1934), *Geobiologie of inleiding tot de milieukunde*, The Hague, the Netherlands : W.P. Van Stockum & Zoon

culture, les espèces *Acidithiobacillus* peuvent parfois représenter une part insignifiante de la communauté (environ 0,1 % des séquences dans BOL_mil et ESP_R.T_pt2). Sa prédisposition à la culture a peut-être surévalué leur incidence dans les DMA. La richesse des archées est moins importante que celle des bactéries dans les DMA. Cependant elle est beaucoup moins bien caractérisée. Beaucoup d'OTU se groupent dans des clades environnementaux, et les quelques souches de références disponibles sont bien différentes de la plupart des OTU que nous avons observés. Notamment dans l'ordre des *Thermoplasmataceae* où les deux souches de référence partagent 90 % d'identité de séquence avec les OTU des DMA. Cela suggère une diversité physiologique et taxonomique importante au sein de cet ordre qui reste à découvrir. Dans cette tâche la génomique peut être utile, comme le montrent les travaux de Yelton *et al.* (2013) où la reconstruction de génomes à partir d'ADN environnemental a permis d'enrichir la caractérisation des *Thermoplasmataceae*. Nous avons aussi détecté des ARMAN principalement dans les sites de Carnoulès, malgré l'utilisation d'amorces classiques (Baker *et al.* 2006). Ce sont des OTU marginaux mais intéressants d'un point de vue physiologique car on suppose que ce sont des parasites. Bien que l'on sache peu de chose sur leur physiologie, leur découverte a permis d'enrichir la taxonomie des *Archaea* (Rinke *et al.* 2013). Quelques OTU, dont certains ont un effectif important, sont affiliés au phylum *Thaumarchaeota*. Ils font partie de clades environnementaux non caractérisés, et sont dans notre étude, spécifiques à quelques drainages. Certaines conditions particulières doivent sélectionner ces archées, cependant nous n'avons pas été en mesure de les identifier. La caractérisation taxonomique des procaryotes des 12 DMA étudiés dans cette thèse, permet aussi d'affiner des représentations conceptuelles des chaînes trophiques des procaryotes des DMA, comme a pu le proposer Baker et Banfield (2003) pour les biofilms de la mine Richmond. D'après nos données il est clair que les principaux métabolismes, relatifs au fer, au soufre et à la matière organiques, dans les DMA ne sont pas réalisés par les mêmes organismes dans la totalité de ceux-ci (Fig 5.4). Ce schéma se base sur les connaissances physiologiques acquises par l'étude d'espèces cultivables proches des taxons identifiés dans les DMA. La justesse de cette représentation doit être encore confirmée. Pour se faire le séquençage du métagénome des drainages peut confirmer/infirmar la présence des gènes nécessaires à ces métabolismes. De même un isolement des espèces de procaryotes permettrait une observation de leurs phénotypes pour les comparer à nos hypothèses.

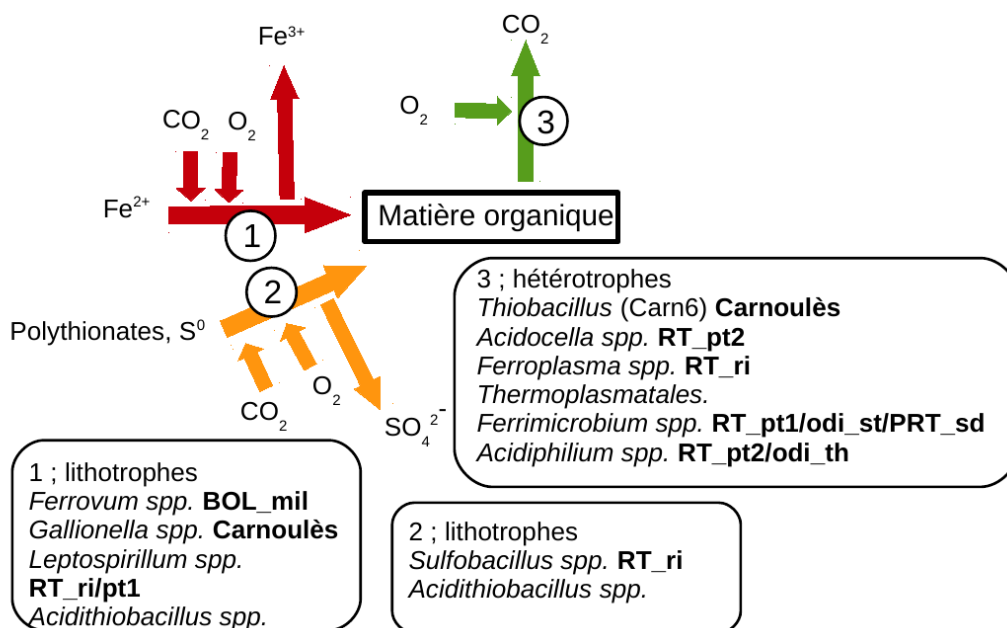


FIGURE 5.4 – Schémas conceptuel des interactions trophiques entre les procaryotes lithotrophes et hétérotrophes des DMA étudiés. En face des noms des organismes sont indiqués les sites dans lesquelles ils ont été observés. Ce schéma a été construit à partir de la caractérisation taxonomique des DMA ainsi que des données physiologiques disponibles sur ces organismes.

Les bases de données publiques de séquences, sont une source d'informations abondantes sur des communautés microbiennes d'environnements très variés. Nous avons utilisé cette manne d'informations pour savoir si les procaryotes observés dans les DMA sont aussi présents dans d'autres environnements. Nos résultats montrent que la majorité des microorganismes observés dans les DMA sont spécifiques aux milieux acides et miniers. Il existe aussi une population que l'on observe dans des environnements volcaniques et géothermaux. Ces milieux présentent des conditions similaires aux DMA par leur acidité et la présence de sulfures, mais avec des températures élevées. Il est donc possible que ces milieux partagent des niches écologiques avec les DMA. Une partie non négligeable des bactéries et des archées que nous avons observées dans nos drainages sont aussi retrouvées dans des sols. Ces organismes sont spécifiques à quelques sites où il peut exister des micro-niches moins extrêmes, que l'on peut retrouver dans des sols. Le filtre de l'habitat est donc important pour structurer les communautés dans les DMA. On a aussi observé dans les sites boliviens des organismes de la flore commensale des animaux, indiquant l'anthropisation de ces DMA. Cependant on ne sait pas si cette population se maintient dans le drainage, ni si elle modifie

les communautés naturelles du drainage. En effet il serait intéressant d'affiner cette analyse en précisant la fraction active des procaryotes dans les DMA. Il faudrait pour cela séquencer les ARN ribosomiaux en plus des gènes codant pour ceux-ci. On pourrait ainsi définir plus précisément les niches de viabilité de ces organismes. L'approche de fouille de données est puissante mais est bridée par un manque de standardisation. En effet il n'est pas obligatoire de fournir l'information sur le lieu d'isolement des séquences lors de la soumission des séquences aux bases de données, environ 10 % des séquences n'étaient pas renseignées. L'autre limitation est que les informations sur la localisation des séquences ne sont pas standardisées. Il est donc nécessaire de reprendre les résultats manuellement pour regrouper les données en catégories pour les interpréter, ce qui empêche la réalisation de telles analyses à très grande échelle. Ces observations amènent à nous poser des questions sur la colonisation des DMA par les procaryotes et notamment leur origine. On observe qu'en plus des environnements de type minier et acide, on retrouve des procaryotes des DMA dans les sols et les milieux d'eau douce. Il est donc possible que les communautés des DMA soient enrichies par des apports locaux des sols et des eaux douces.

La comparaison des communautés procaryotes observées dans nos DMA avec des communautés générées par un modèle nul, nous indique que les processus stochastiques ont peu de place dans la structuration des communautés dans les DMA. Les processus déterministes et notamment le filtre de l'habitat sont importants. Le paramètre le plus fort que nous avons identifié est le pH qui avait déjà été proposé avec la température du drainage (Hallberg 2010). Nos DMA sont tous mésothermaux donc la température n'était pas une pression de sélection très forte dans notre échantillonnage. Cependant nous avons mis en évidence d'autres paramètres importants. La quantité de carbone total est déterminante, ceci est cohérent avec le métabolisme des souches cultivées dont beaucoup sont oligotrophes (Chapitre 1). Parmi tous les éléments métalloïdes que nous avons mesurés dans nos sites l'arsenite (AsIII) est celui qui impose la pression de sélection la plus forte. Ce résultat met en avant la nécessité de prendre en compte les différentes formes d'oxydo/réduction des ions métalliques car ils n'ont pas systématiquement la même réactivité (Tamaki et Frankenberger 1992). Le dernier paramètre important pour structurer les communautés est la concentration en sulfate. Une concentration élevée en sulfate indique une dissolution importante des minéraux de pyrite (équations 1.1 et 1.3). De plus, parmi les OTU sélectionnés par cette condition, deux sont du genre

Leptospirillum et un du genre *Ferroplasma*, deux genres connus pour leur efficacité dans la dissolution des minéraux sulfurés (Johnson et Hallberg 2008). Ce dernier paramètre n'est donc peut être pas une cause mais plutôt une conséquence d'une communauté particulière.

L'analyse des séquences d'acides aminés des transporteurs d'arsenic ArsB et Acr3 a été faite pour évaluer les échanges de populations au sein ou entre les DMA. Nous avons observé différents patrons, le premier suggère des populations spécifiques des sites échantillonnés. Le deuxième des populations partagées à différentes échelles, entre site du même drainage, entre drainages au sein d'une même région ou entre régions. Cela indiquerait des échanges de populations entre les DMA, ou bien une colonisation par un réservoir commun. Si on considère les conclusions sur l'importance du filtre de l'habitat dans la structuration des communautés procaryotes des DMA, il faudrait que des niches soient partagées entre les DMA. Enfin nous avons observé une corrélation entre la proximité taxonomique des communautés et la proximité des transporteurs Acr3.1. On peut penser que ce gène a été acquis par un ancêtre commun et transmis à la descendance. Si on considère un échange horizontal de ce gène il aura plutôt tendance à se faire entre organismes de taxons proches.

Enfin dans le but d'avoir une étude la plus complète possible sur les processus structurant les communautés, nous avons essayé de prendre en compte les relations entre les organismes au sein des communautés. Ces relations sont nombreuses et complexes et leur prise en compte est récente (Faust et Raes 2012). Pourtant ce sont des processus déterministes qui sont potentiellement importants, bien que difficiles à étudier avec des organismes réfractaires à la culture. Cependant l'étude de réseaux de cooccurrences peut identifier des relations (d'exclusions ou de d'agrégations) entre les micro-organismes. Dans notre étude nous avons mis en évidence l'existence de groupes d'organismes qui partagent plus de liens entre eux qu'avec des organismes hors du groupe mais les outils actuels ne permettent pas de conclure sur la nature des relations. C'est aussi une méthodologie jeune qui commence juste à être évaluée (Berry et Widder 2014). Cette analyse permet de faire ressortir des relations potentielles qui peuvent ensuite être enrichies par des analyses de génomes et d'identifier des métabolismes compatibles/antagonistes et/ou confirmées par des approches culturelles si cela est possible. Cette analyse est aussi un canevas sur lequel peut être superposé d'autres informations. On peut penser aux informations de viabilité des organismes évoqué précédemment. Il est aussi possible de rajouter des données fonctionnelles. Des outils tels que PICRUSt (Langille *et al.*

2013) permettent d'estimer le potentiel métabolique de taxons procaryotes. Cette information associée aux cliques identifiées, peut nous indiquer si leur formation est cohérente avec des fonctions métaboliques et d'identifier des redondances fonctionnelles. Cette redondance étant importante dans les processus de résilience des communautés.

Bibliographie

- ACHOUR, A. R., BAUDA, P. et BILLARD, P. (2007). Diversity of arsenite transporter genes from arsenic-resistant soil bacteria. *Research in microbiology*, 158(2):128–37.
- AJEES, A., YANG, J. et ROSEN, B. (2011). The ArsD As(III) metallochaperone. *Biometals*, 24:391–399.
- ALAZARD, D., JOSEPH, M., BATTAGLIA-BRUNET, F., CAYOL, J.-L. et OLLIVIER, B. (2010). Desulfosporosinus acidiphilus sp. nov. : a moderately acidophilic sulfate-reducing bacterium isolated from acid mining drainage sediments : New taxa : Firmicutes (Class Clostridia, Order Clostridiales, Family Peptococcaceae). *Extremophiles : life under extreme conditions*, 14(3):305–12.
- ALBER, B., OLINGER, M., RIEDER, A., KOCKELKORN, D., JOBST, B., HÜGLER, M. et FUCHS, G. (2006). Malonyl-coenzyme A reductase in the modified 3-hydroxypropionate cycle for autotrophic carbon fixation in archaeal Metallosphaera and Sulfolobus spp. *Journal of bacteriology*, 188(24):8551–9.
- ALBERS, S.-V., van de VOSSENBURG, J., DRIESSEN, A. et KONINGS, W. (2001). Bioenergetics and solute uptake under extrem conditions. *Extremophiles*, 5:285–294.
- ALTSCHUL, S., GISH, W., MILLER, W., MEYERS, E. et LIPMAN, D. (1990). Basic Local Alignment Search Tool. *Journal of Molecular Biology*, 215(3):403–410.
- AMANN, R. I., LUDWIG, W. et SCHLEIFER, K. H. (1995). Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiological reviews*, 59(1):143–69.
- AMARAL-ZETTLER, L. a., ZETTLER, E. R., THEROUX, S. M., PALACIOS, C., AGUILERA, A. et AMILS, R. (2011). Microbial community structure across the tree of life in the extreme Río Tinto. *The ISME journal*, 5(1):42–50.
- ANTRANIKIAN, G., HERZBERG, C. et GOTTSCHALK, G. (1982). Characterisation of ATP citrate lyase from *Chlorobium limicola*. *Journal of Bacteriology*, 152:1284–1287.
- BAAS-BECKING, L. G. M. (1934). *Geobiologie ; of inleiding tot de milieukunde*. Den Haag [Netherlands] : W.P. Van Stockum & Zoon N.V. Includes index.
- BAKER, B. J. et BANFIELD, J. F. (2003). Microbial communities in acid mine drainage. *FEMS microbiology ecology*, 44(2):139–52.
- BAKER, B. J., COMOLLI, L. R., DICK, G. J., HAUSER, L. J., HYATT, D., DILL, B. D., LAND, M. L., VERBERKMOES, N. C., HETTICH, R. L., BANFIELD, J. F. et MOUNTAIN, I. (2010a). Enigmatic, ultrasmall, uncultivated Archaea. *pnas*, 107(19):8806–8811.
- BAKER, B. J., HUGENHOLTZ, P., DAWSON, S. C. et BANFIELD, J. F. (2010b). Extremely Acidophilic Protists from Acid Mine Drainage Host Rickettsiales-Lineage Endosymbionts That Have Intervening Sequences in Their 16S rRNA Genes. *Applied and Environmental Microbiology*, 69(9):5512–5518.
- BAKER, B. J., TYSON, G. W., WEBB, R. I., FLANAGAN, J., HUGENHOLTZ, P., ALLEN, E. E. et BANFIELD, J. F. (2006). Lineages of acidophilic archaea revealed by community genomic analysis. *Science (New York, N.Y.)*, 314(5807):1933–5.
- BAKER, G., SMITH, J. et COWAN, D. (2003). Review and re-analysis of domain-specific 16S primers. *Journal of Microbiological Methods*, 55(3):541–555.
- BARBERÁN, A., BATES, S. T., CASAMAYOR, E. O. et FIERER, N. (2012). Using network analysis to explore co-occurrence patterns in soil microbial communities. *The ISME journal*, 6(2):343–51.
- BATHE, S. et NORRIS, P. (2007). Ferrous iron- and sulfur-induced genes in *Sulfolobus metallicus*. *Applied Environmental Microbiology*, 73:2491–2497.
- BEAM, J. P., JAY, Z. J., KOZUBAL, M. a. et INSKEEP, W. P. (2014). Niche specialization of novel Thaumarchaeota to oxic and hypoxic acidic geothermal springs of Yellowstone National Park. *The ISME journal*, 8(4):938–51.

- BEDNAR, A. J., GARBARINO, J. R., RANVILLE, J. F. et WILDEMAN, T. R. (2002). Preserving the Distribution of Inorganic Arsenic Species in Groundwater and Acid Mine Drainage Samples. *Environmental Science & Technology*, 36(10):2213–2218.
- BELL, G. (2001). Neutral macroecology. *Science (New York, N.Y.)*, 293(5539):2413–8.
- BENJAMINI, Y. et HOCHBERG, Y. (1995). Controlling the False Discovery Rate : A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1):289–300.
- BERRY, D. et WIDDER, S. (2014). Deciphering microbial interactions and detecting keystone species with co-occurrence networks. *Frontiers in microbiology*, 5(May):219.
- BLOWES, D., JAMBOR, J. et ALPERS, C. (1994). *The environmental geochemistry of sulfide mine-wastes*. Short course handbook. Mineralogical Association of Canada.
- BORCARD, D., GILLET, F. et LEGENDRE, P. (2011). *Numerical Ecology with R*. Springer New York, New York, NY.
- BRAY, J. R. et CURTIS, J. T. (1957). An Ordination of the Upland Forest Communities of Southern Wisconsin. *Ecological Monographs*, 27(4):325.
- BRIDGE, T. et JOHNSON, D. (2000). Reductive dissolution of ferric iron minerals by *Acidiphilium* SJH. *Geomicrobiology Journal*, 17:193–206.
- BROCHIER-ARMANET, C., BOUSSAU, B., GRIBALDO, S. et FORTERRE, P. (2008). Mesophilic Crenarchaeota : proposal for a third archaeal phylum, the Thaumarchaeota. *Nature reviews. Microbiology*, 6(3):245–52.
- BROCHIER-ARMANET, C., GRIBALDO, S. et FORTERRE, P. (2012). Spotlight on the Thaumarchaeota. *The ISME journal*, 6(2):227–30.
- BRUNEEL, O., PASCAULT, N., EGAL, M., BANCON-MONTIGNY, C., Goñi URRIZA, M. S., ELBAZ-POULICHET, F., PERSONNÉ, J.-C. et DURAN, R. (2008). Archaeal diversity in a Fe-As rich acid mine drainage at Carnoulès (France). *Extremophiles : life under extreme conditions*, 12(4):563–71.
- BRUNEEL, O., VOLANT, A., GALLIEN, S., CHAUMANDE, B., CASIOT, C., CARAPITO, C., BARDIL, A., MORIN, G., BROWN, G. E., PERSONNÉ, C. J., LE PASLIER, D., SCHAEFFER, C., VAN DORSSELAER, A., BERTIN, P. N., ELBAZ-POULICHET, F. et ARSÈNE-PLOETZE, F. (2011). Characterization of the active bacterial community involved in natural attenuation processes in arsenic-rich creek sediments. *Microbial ecology*, 61(4):793–810.
- CAI, L., LIU, G., RENSING, C. et WANG, G. (2009). Genes involved in arsenic transformation and resistance associated with different levels of arsenic-contaminated soils. *BMC microbiology*, 9(1):4.
- CALDWELL, P. E., MACLEAN, M. R. et NORRIS, P. R. (2007). Ribulose biphosphate carboxylase activity and a Calvin cycle gene cluster in *Sulfobacillus* species. *Microbiology (Reading, England)*, 153(Pt 7):2231–40.
- CALVIN, M. (1962). The path of carbon in photosynthesis. *Science*, 135:879–899.
- CASIOT, C., LEBRUN, S., MORIN, G., BRUNEEL, O., PERSONNÉ, J. C. et ELBAZ-POULICHET, F. (2005). Sorption and redox processes controlling arsenic fate and transport in a stream impacted by acid mine drainage. *The Science of the total environment*, 347(1-3):122–30.
- CHAFFRON, S., REHRAUER, H. et PERNTHALER, J. (2010). A global network of coexisting microbes from environmental and whole-genome sequence data. *Genome Research*, 20:947–959.
- CHAO, A. (1984). Non-parametric estimation of the number of classes in a population. *Scandinavian Journal of Statistics*, 11:265–270.
- CHAO, A. et LEE, S.-M. (1992). Estimating the Number of Classes via Sample Coverage. *Journal of the American Statistical Association*, 87(417):210–217.
- COLE, J. R., WANG, Q., CARDENAS, E., FISH, J., CHAI, B., FARRIS, R. J., KULAM-SYED-MOHIDEEN, A. S., MCGARRELL, D. M., MARSH, T., GARRITY, G. M. et TIEDJE, J. M. (2009). The Ribosomal Database Project : improved alignments and new tools for rRNA analysis. *Nucleic Acids Research*, 37(suppl 1):D141–D145.
- COMOLLI, L. R., BAKER, B. J., DOWNING, K. H., SIEGERIST, C. E. et BANFIELD, J. F. (2009). Three-dimensional analysis of the structure and ecology of a novel, ultra-small archaeon. *The ISME journal*, 3(2):159–67.
- CONNOR, E. F. et SIMBERLOFF, D. (1979). The Assembly of Species Communities : Chance or Competition ? *Ecology*, 60(6):1132–1140.
- COUPLAND, K. et JOHNSON, D. (2008). Evidence that the potential for dissimilatory ferric iron reduction is widespread among acidophilic heterotrophic bacteria. *FEMS Microbiology Letters*, 279:30–35.

- CULLEN, W. et REIMER, K. (1989). Arsenic speciation in the environment. *Chemical Reviews*, 89(124):713–764.
- DAI, Z., GUO, X., YIN, H., LIANG, Y., CONG, J. et LIU, X. (2014). Identification of nitrogen-fixing genes and gene clusters from metagenomic library of acid mine drainage. *PLoS one*, 9(2):e87976.
- de WIT, R. et BOUVIER, T. (2006). 'Everything is everywhere, but, the environment selects'; what did Baas Becking and Beijerinck really say? *Environmental microbiology*, 8(4):755–8.
- DESANTIS, T. Z., HUGENHOLTZ, P., KELLER, K., BRODIE, E. L., LARSEN, N., PICENO, Y. M., PHAN, R. et ANDERSEN, G. L. (2006). NAST : a multiple sequence alignment server for comparative analysis of 16S rRNA genes. *Nucleic Acids Research*, 34(suppl 2):W394–W399.
- DIAMOND, J. M. (1975). In CODY, M. L. et DIAMOND, J. M., éditeurs : *Ecology and evolution of communities*, pages 342–444. Belknap Press of Harvard University Press, Cambridge, MA.
- DICK, G. J., ANDERSSON, A. F., BAKER, B. J., SIMMONS, S. L., THOMAS, B. C., YELTON, A. P. et BANFIELD, J. F. (2009). Community-wide analysis of microbial genome sequence signatures. *Genome biology*, 10(8):R85.
- DIREITO, S. O. L., MAREES, A. et RÖLING, W. F. M. (2012). Sensitive life detection strategies for low-biomass environments : optimizing extraction of nucleic acids adsorbing to terrestrial and Mars analogue minerals. *FEMS microbiology ecology*, 81(1):111–123.
- DOMBROWSKI, P., LONG, W., FARLEY, K., MAHONY, J., CAPITANI, J. et DI TORO, D. (2007). Thermodynamic analysis of arsenic methylation. *Environmental Science & Technology*, 39:2169–2176.
- DOPSON, M., BAKER-AUSTIN, C. et BOND, P. (2005). Analysis of differential protein expression during growth states of *Ferroplasma* strains and insights into electrons transport for iron oxydation. *Microbiology*, 151:4127–4137.
- DOPSON, M., BAKER-AUSTIN, C., HIND, A., BOWMAN, J. P. et BOND, P. L. (2004). Characterization of *Ferroplasma* Isolates and *Ferroplasma acidarmanus* sp. nov., Extreme Acidophiles from Acid Mine Drainage and Industrial Bioleaching Environments. *Applied and Environmental Microbiology*, 70(4):2079–2088.
- EDGAR, R. C. (2004). MUSCLE : multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research*, 32(5):1792–7.
- EDGAR, R. C., HAAS, B. J., CLEMENTE, J. C., QUINCE, C. et KNIGHT, R. (2011). Uchime improves sensitivity and speed of chimera detection. *Bioinformatics*, 27(16):2194–2200.
- EDWARDS, K., BOND, P. et BANFIELD, J. (2000). Characteristics of attachment and growth of *Thiobacillus caldus* on sulphide minerals : a chemotactic response to sulphur minerals? *Environmental microbiology*, 2:324–332.
- EDWARDS, K., GOEBEL, B., RODGERS, T., SCHRENK, M., GIH-RING, T., CARDONA, M., HU, B., MCGUIRE, M., HAMERS, R., PACE, N. et BANFIELD, J. (1999). Geomicrobiology of pyrite (FeS₂) dissolution : Case Study at Iron Mountain, California. *Geomicrobiology Journal*, 16:155–179.
- EDWARDS, K., HU, B., HAMERS, R. et BANFIELD, J. (2001). A new look at microbial leaching patterns on sulfide minerals. *FEMS microbiology ecology*, 34:197–206.
- EDWARDS, K., SCHRENK, M., HAMERS, R. et BANFIELD, J. F. (1998). Microbial oxidation of pyrite : experiments using microorganisms from an extreme acidic environment. *American Mineralogist*, 83:1444–1453.
- EDWARDS, K. J. (2000). An Archaeal Iron-Oxidizing Extreme Acidophile Important in Acid Mine Drainage. *Science*, 287(5459):1796–1799.
- EHRlich, H. (2002). Bacterial oxidation of As(III) compounds. In FRANKENBERGER, W., éditeur : *Environmental Chemistry of Arsenic*, pages 313–328. New York, USA : Marcel Dekker.
- EILER, A., HEINRICH, F. et BERTILSSON, S. (2012). Coherent dynamics and association networks among lake bacterioplankton taxa. *The ISME journal*, 6(2):330–42.
- ELTON, C. (1927). *Animal Ecology*. Biology series. University of Chicago Press.
- FAHY, A., GILOTEAUX, L., BERTIN, P., LE PALER, D., MÉDIGUE, C., WEISSENBACH, J., DURAN, R. et LAUGA, B. (2014). 16S rRNA and As-related functional diversity : contrasting fingerprints in sediments from an acid mine drainage. (accepté). *Microbial Ecology*.
- FAUST, K. et RAES, J. (2012). Microbial interactions : from networks to models. *Nature Reviews Microbiology*, 10(8):538–550.
- FISH, J. A., CHAI, B., WANG, Q., SUN, Y., BROWN, C. T., TIEDJE, J. M. et COLE, J. R. (2013). FunGene : the Functional Gene Pipeline and Repository. *Frontiers in Microbiology*, 4:291.

- FOWLER, T., HOLMES, P. et CRUNDWELL, F. (1999). Mechanism of pyrite dissolution in the presence of *Thiobacillus ferrooxidans*. *Applied and Environmental Microbiology*, 65:2987–2993.
- FU, H.-L., MENG, Y., ORDÓÑEZ, E., VILLADANGOS, A. F., BHATTACHARJEE, H., GIL, J. A., MATEOS, L. M. et ROSEN, B. P. (2009). Properties of arsenite efflux permeases (Acr3) from *Alkaliphilus metalliredigens* and *Corynebacterium glutamicum*. *The Journal of biological chemistry*, 284(30):19887–95.
- FUHRMAN, J. a. (2009). Microbial community structure and its functional implications. *Nature*, 459(7244):193–9.
- FYSON, A. (2000). Angiosperm in acidic waters at pH 3 and below. *Hydrobiologia*, 433:129–135.
- GALE, N. et BECK, J. (1967). Evidence for the Calvin cycle and hexose monophosphate pathway in *Thiobacillus ferrooxidans*. *Journal of Bacteriology*, 94:1052–1059.
- GEOLOGY, E. (2003). REVIEWS. *Economic Geology*, 98(6):1253–1254.
- GLENN, D. (2002). Multivariate Regression Trees : A New Technique for Modeling Species-Environment Relationships. *Ecology*, 83(4):1105–1117.
- GOLYSHINA, O. V., PIVOVAROVA, T. A., KARAVAİKO, G. I., KONDRAT'ĖVA, T. F., MOORE, E., ABRAHAM, W. R., LUNSDORF, H., TIMMIS, K. N., YAKIMOV, M. M. et GOLYSHIN, P. N. (2000). *Ferroplasma acidophilum* gen. nov., sp. nov., an acidophilic, autotrophic, ferrous-iron-oxidizing, cell-wall-lacking, mesophilic member of the Ferroplasmaeaceae fam. nov., comprising a distinct lineage of the Archaea. *International Journal of Systematic and Evolutionary Microbiology*, 50(3):997–1006.
- GÓMEZ, F., AMARAL, L. A., ZETTLER, F., ZETTLER, E., KEENAN, B. G., AMILS, R. et SOGIN, M. L. (2002). Eukaryotic diversity in Spain's River of Fire. *Nature*.
- GOOD, I. J. (1953). The Population Frequencies of Species and the Estimation of Population Parameters. *Biometrika*, 40(3/4): 237–264.
- GOTELLI, N. J. (2000). Null model analysis of species co-occurrence patterns. *Ecology*, 81(9):2606–2621.
- GOUGH, H. L. et STAHL, D. A. (2011). Microbial community structures in anoxic freshwater lake sediment along a metal contamination gradient. *ISME J*, 5(3):543–558.
- GOYER, R. et CLARKSON, T. (2001). Toxic effect of metals. In KLAASSEN, C., éditeur : *Casarett & Doull's Toxicology : The Basic Science of Poisons*, pages 811–868. New York, USA : McGraw Hill.
- GREEN, J. et BOHANNAN, B. J. M. (2006). Spatial scaling of microbial biodiversity. *Trends in ecology & evolution*, 21(9):501–7.
- GUAY, R. et SILVER, M. (1975). *Thiobacillus acidophilus*, sp. nov. ; isolation and some physiological characteristics . *Canadian Journal of Microbiology*, 21:281–285.
- GUBRY-RANGIN, C., HAI, B., QUINCE, C., ENGEL, M., THOMSON, B. C., JAMES, P., SCHLOTTER, M., GRIFFITHS, R. I., PROSSER, J. I. et NICOL, G. W. (2011). Niche specialization of terrestrial archaeal ammonia oxidizers. *Proceedings of the National Academy of Sciences*, 108(52):21206–21211.
- GUINDON, S. et GASCUEL, O. (2003). A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by Maximum Likelihood. *Systematic Biology*, 52(5):696–704.
- HALL, T. A. (1999). BioEdit : a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, 41:95–98.
- HALLBERG, K. (2010). New perspectives in acid mine drainage microbiology. *Hydrometallurgy*, 104(3-4):448–453.
- HALTER, D., CORDI, A., GRIBALDO, S., GALLIEN, S., GOULHEN-CHOLLET, F., HEINRICH-SALMERON, A., CARAPITO, C., PAGNOUT, C., MONTAUT, D., SEBY, F., VAN DORSSELAER, A., SCHAEFFER, C., BERTIN, P. N., BAUDA, P. et ARSÈNE-PLOETZE, F. (2011). Taxonomic and functional prokaryote diversity in mildly arsenic-contaminated sediments. *Research in microbiology*, 162(9):877–87.
- HAMADY, M. et KNIGHT, R. (2009). Microbial community profiling for human microbiome projects : Tools, techniques, and challenges. *Genome research*, 19(7):1141–52.
- HARRISON, A. P. (1984). The acidophilic thiobacilli and other acidophilic bacteria that share their habitat. *Annual review of microbiology*, 38:265–92.
- HIRAISHI, A., MATSUZAWA, Y., KANBE, T. et WAKAO, N. (2000). *Acidisphaera rubrifaciens* gen. nov., sp. nov., an aerobic bacteriochlorophyll-containing bacterium isolated from acidic environments. *International Journal of Systematic and Evolutionary Microbiology*, 50(4):1539–1546.

- HORNER-DEVINE, M. C. et BOHANNAN, B. J. M. (2006). Phylogenetic clustering and overdispersion in bacterial communities. *Ecology*, 87(7 Suppl):S100–8.
- HORNER-DEVINE, M. C., SILVER, J. M., LEIBOLD, M. a., BOHANNAN, B. J. M., COLWELL, R. K., FUHRMAN, J. a., GREEN, J. L., KUSKE, C. R., MARTINY, J. B. H., MUYZER, G., OVREÅ S, L., REYSENBACH, A.-L. et SMITH, V. H. (2007). A comparison of taxon co-occurrence patterns for macro- and microorganisms. *Ecology*, 88(6):1345–53.
- HUBBELL, S. P. (2001). *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton University Press.
- HUBER, H., HOHN, M. J., RACHEL, R., FUCHS, T., WIMMER, V. C. et STETTER, K. O. (2002). A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. *Nature*, 417(6884):63–7.
- HUGHES, J. B., HELLMANN, J. J., RICKETTS, T. H. et BOHANNAN, B. J. M. (2001). Counting the Uncountable : Statistical Approaches to Estimating Microbial Diversity. *Applied and Environmental Microbiology*, 67(10):4399–4406.
- HUSE, S. M., WELCH, D. M., MORRISON, H. G. et SOGIN, M. L. (2010). Ironing out the wrinkles in the rare biosphere through improved OTU clustering. *Environmental Microbiology*, 12(7):1889–1898.
- HÜGLER, M., HUBER, H., MOLYNEAUX, S., VETRIANI, C. et SIEVERT, S. (2003). Characterization of acetyl-CoA/propionyl-CoA carboxylase in *Metallosphaera sedula*. *European Journal of Biochemistry*, 270:736–744.
- ISHII, M., MIYAKE, T., SATOH, T., SUGIYAMA, H., OSHIMA, Y., KODAMA, T. et IGARASHI, Y. (1997). Community structure and metabolism through reconstruction of microbial genomes from the environment. *Archives of Microbiology*, 166:368–371.
- IZZO, V., NOTOMISTA, E., PICARDI, A., PENNACCHIO, F. et DI DONATO, A. (2005). The thermophilic archaeon *Sulfolobus solfataricus* is able to grow on phenol. *Research in microbiology*, 156(5-6):677–89.
- JOHNSON, D. B. (2006). Biohydrometallurgy and the environment : Intimate and important interplay. *Hydrometallurgy*, 83(1-4):153–166.
- JOHNSON, D. B., BACELAR-NICOLAU, P., OKIBE, N., THOMAS, A. et HALLBERG, K. B. (2009). *Ferrimicrobium acidiphilum* gen. nov., sp. nov. and *Ferritrix thermotolerans* gen. nov., sp. nov. : heterotrophic, iron-oxidizing, extremely acidophilic actinobacteria. *International journal of systematic and evolutionary microbiology*, 59(Pt 5):1082–9.
- JOHNSON, D. B. et HALLBERG, K. B. (2003). The microbiology of acidic mine waters. *Research in microbiology*, 154(7):466–73.
- JOHNSON, D. B. et HALLBERG, K. B. (2008). Carbon, Iron and Sulfur Metabolism in Acidophilic Micro-Organisms. volume 54 de *Advances in Microbial Physiology*, pages 201–255. Academic Press.
- JOHNSON, D. B., HALLBERG, K. B. et HEDRICH, S. (2014). Uncovering a microbial enigma : isolation and characterization of the streamer-generating, iron-oxidizing, acidophilic bacterium "Ferrovum myxofaciens". *Applied and environmental microbiology*, 80(2):672–80.
- JOHNSON, D. B. et RANG, L. (1993). Effects of acidophilic protozoa on populations of metal-mobilizing bacteria during the leaching of pyritic coal. *Journal of General Microbiology*, 139:1417–1423.
- JOHNSON, D. B., SEN, A. M., KIMURA, S., ROWE, O. F. et HALLBERG, K. B. (2006). Novel biosulfidogenic system for selective recovery of metals from acidic leach liquors and waste streams. *Mineral Processing and Extractive Metallurgy*, 115(1):19–24.
- JONES, R. M., HEDRICH, S. et JOHNSON, D. B. (2013). *Acidocella aromatica* sp. nov. : an acidophilic heterotrophic alphaproteobacterium with unusual phenotypic traits. *Extremophiles : life under extreme conditions*, 17(5):841–50.
- KIM, M., MORRISON, M. et YU, Z. (2011). Evaluation of different partial 16S rRNA gene sequence regions for phylogenetic analysis of microbiomes. *Journal of microbiological methods*, 84(1):81–7.
- KIMURA, S., HALLBERG, K. et JOHNSON, D. (2006). Sulfidogenesis in Low pH (3.8–4.2) Media by a Mixed Population of Acidophilic Bacteria. *Biodegradation*, 17(2):57–65.
- KINDT, R. et COE, R. (2005). *Tree diversity analysis. A manual and software for common statistical methods for ecological and biodiversity studies*. World Agroforestry Centre (ICRAF), Nairobi.
- KNIGHTS, D., COSTELLO, E. K. et KNIGHT, R. (2011). Supervised classification of human microbiota. *FEMS microbiology reviews*, 35(2):343–59.
- KUROSAWA, N., ITOH, Y. H., IWAI, T., SUGAI, A., UDA, I., KIMURA, N., HORIUCHI, T. et ITOH, T. (1998). *Sulfurisphaera ohwakuensis* gen. nov., sp. nov., a novel extremely thermophilic acidophile of the order Sulfolobales. *International Journal of Systematic Bacteriology*, 48(2):451–456.
- LANE, D. J., PACE, B., OLSEN, G. J., STAHL, D. A., SOGIN, M. L. et PACE, N. R. (1985). Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. *Proc Natl Acad Sci U S A.*, 83(13):6955–6959.

- LANGILLE, M. G. I., ZANEVELD, J., CAPORASO, J. G., McDONALD, D., KNIGHTS, D., REYES, J. A., CLEMENTE, J. C., BURKEPILE, D. E., VEGA THURBER, R. L., KNIGHT, R., BEIKO, R. G. et HUTTENHOWER, C. (2013). Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat Biotech*, 31(9):814–821.
- LARRSON, L., GUNNEL, O., HOLST, O. et KARLSSON, H. (1993). Oxidation of pyrite by *Acidianus brierleyi* : Importance of close contact between the pyrite and the microorganisms. *Biotechnology Letters*, 15:99–104.
- LEKBERG, Y., SCHNOOR, T., KJØLLER, R., GIBBONS, S. M., HANSEN, L. H., AL-SOUD, W. A., SØRENSEN, S. r. J. et ROSENDAHL, S. r. (2012). 454-sequencing reveals stochastic local reassembly and high disturbance tolerance within arbuscular mycorrhizal fungal communities. *Journal of Ecology*, 100(1):151–160.
- LETUNIC, I. et BORK, P. (2007). Interactive Tree Of Life (iTOL) : an online tool for phylogenetic tree display and annotation. *Bioinformatics (Oxford, England)*, 23(1):127–8.
- LI, X., ZHANG, L. et WANG, G. (2014). Genomic evidence reveals the extreme diversity and wide distribution of the arsenic-related genes in Burkholderiales. *PLoS one*, 9(3):e92236.
- LIU, S., ATHAR, M., LIPPAI, I., WALDREN, C. et HEI, T. (2001). Induction of oxyradicals by arsenic : implication for mechanism of genotoxicity. *Proceedings of the National Academy of Sciences*, 98:1643–1648.
- LOBOS, J. H., CHISOLM, T. E., BOPP, L. H. et HOLMES, D. S. (1986). *Acidiphilium organovorum* sp. nov., an Acidophilic Heterotroph Isolated from a *Thiobacillus ferrooxidans* Culture. *International Journal of Systematic Bacteriology*, 36(2):139–144.
- LORTIE, C. J., BROOKER, R. W., CHOLER, P., KIKVIDZE, Z., MICHALET, R., PUGNAIRE, F. I. et CALLAWAY, R. M. (2004). Rethinking plant community theory. *Oikos*, 107(2):433–438.
- LOWSON, R. T. (1982). Aqueous oxidation of pyrite by molecular oxygen. *Chemical Reviews*, 82(5):461–497.
- LOZUPONE, C. et KNIGHT, R. (2005). UniFrac : a new phylogenetic method for comparing microbial communities. *Applied and environmental microbiology*, 71(12):8228–35.
- LU, S., GISCHKAT, S., REICHE, M., AKOB, D. M., HALLBERG, K. B. et KÜSEL, K. (2010). Ecophysiology of Fe-cycling bacteria in acidic sediments. *Applied and environmental microbiology*, 76(24):8174–83.
- LUDWIG, W., STRUNK, O., WESTRAM, R., RICHTER, L., MEIER, H., YADHUKUMAR, BUCHNER, A., LAI, T., STEPPI, S., JOBB, G., FÖRSTER, W., BRETTSCHE, I., GERBER, S., GINHART, A. W., GROSS, O., GRUMANN, S., HERMANN, S., JOST, R., KÖNIG, A., LISS, T., LÜSSMANN, R., MAY, M., NONHOFF, B., REICHEL, B., STREHLOW, R., STAMATAKIS, A., STUCKMANN, N., VILBIG, A., LENKE, M., LUDWIG, T., BODE, A. et SCHLEIFER, K.-H. (2004). ARB : a software environment for sequence data. *Nucleic acids research*, 32(4):1363–71.
- MANSOUR, N., SAWHNEY, M., TAMANG, D., VOGL, C. et SAIER, M. (2007). The bile/arsenite/riboflavin transporter (BART) superfamily. *FEBS Journal*, 274:612–629.
- MARGULIES, M., EGHOLM, M., ALTMAN, W. E., ATTIYA, S., BADER, J. S., BEMBEN, L. A., BERKA, J., BRAVERMAN, M. S., CHEN, Y.-J., CHEN, Z., DEWELL, S. B., DU, L., FIERRO, J. M., GOMES, X. V., GODWIN, B. C., HE, W., HELGESEN, S., HO, C. H., HO, C. H., IRZYK, G. P., JANDO, S. C., ALENQUER, M. L. I., JARVIE, T. P., JIRAGE, K. B., KIM, J.-B., KNIGHT, J. R., LANZA, J. R., LEAMON, J. H., LEFKOWITZ, S. M., LEI, M., LI, J., LOHMAN, K. L., LU, H., MAKHIJANI, V. B., MCDADE, K. E., MCKENNA, M. P., MYERS, E. W., NICKERSON, E., NOBILE, J. R., PLANT, R., PUC, B. P., RONAN, M. T., ROTH, G. T., SARKIS, G. J., SIMONS, J. F., SIMPSON, J. W., SRINIVASAN, M., TARTARO, K. R., TOMASZ, A., VOGT, K. A., VOLKMER, G. A., WANG, S. H., WANG, Y., WEINER, M. P., YU, P., BEGLEY, R. F. et ROTHBERG, J. M. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, 437(7057):376–80.
- MASON, J. et KELLY, D. (1988). Mixotrophic and autotrophique growth of *Thiobacillus acidophilus* on tetrathionate. *Archives of Microbiology*, 149:317–323.
- MC GINNESS, S. et JOHNSON, D. B. (1992). Grazing of acidophilic bacteria by a flagellated protozoan. *Microbial Ecology*, 23:75–86.
- MÉNDEZ GARCIA, C. (2013). *A systems microbiology-based exploration of the community structure and metabolism of acid streamres and mats thriving at an abandoned mercury mine drainage system*. Thèse de doctorat, Universidad de Oviedo.
- MENENDEZ, C., BAUER, Z., HUBER, H., GAD'ON, N., STETTER, K. et FUCHS, G. (1999). Presence of acetyl coenzyme A (CoA) carboxylase and propionyl-CoA carboxylase in autotrophic *Crenarchaeota* and indication for operation of a 3-hydroxypropionate cycle in autotrophic carbon fixation. *Journal of Bacteriology*, 181:1088–1098.
- MENG, Y.-L., LIU, Z. et ROSEN, B. P. (2004). As(III) and Sb(III) uptake by GlpF and efflux by ArsB in *Escherichia coli*. *The Journal of biological chemistry*, 279(18):18334–41.

- MEYER, M., STENZEL, U. et HOFREITER, M. (2008). Parallel tagged sequencing on the 454 platform. *Nature protocols*, 3(2):267–78.
- MI, S., SONG, J., LIN, J., CHE, Y., ZHENG, H. et LIN, J. (2011). Complete genome of *Leptospirillum ferriphilum* ML-04 provides insight into its physiology and environmental adaptation. *Journal of microbiology (Seoul, Korea)*, 49(6):890–901.
- MORIN, G. et CALAS, G. (2006). Arsenic in Soils, Mine Tailings, and Former Industrial Sites. *Elements*, 2(2):97–101.
- MOROZOVA, O. et MARRA, M. A. (2008). Applications of next-generation sequencing technologies in functional genomics. *Genomics*, 92(5):255–64.
- NEEDLEMAN, S. B. et WUNSCH, C. D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology*, 48(3):443–453.
- NORRIS, P. (1992). Acidophile diversity in mineral sulfide oxidation. In RAWLINGS, D. et JOHNSON, D., éditeurs : *Biomining*, pages 199–216. Springer-Verlag, Heidelberg.
- NORRIS, P. et INGLEDEW, W. (1992). Acidophilic bacteria : adaptations and applications. In HEBERT, R. et R.J., S., éditeurs : *Molecular Biology and Biochemistry of Extremophiles*, pages 121–131. Royal Society for Chemistry, Cambridge.
- OKSANEN, A. J., BLANCHET, F. G., KINDT, R., MINCHIN, P. R., SIMPSON, G. L., SOLY, P. et WAGNER, H. (2012). *Package 'vegan'*.
- OREMLAND, R., SALTIKOV, C., WOLFE-SIMON, F. et STOLZ, J. (2009). Arsenic in the evolution of earth and extraterrestrial ecosystems. *Geomicrobiology Journal*, 26:522–536.
- OREMLAND, R. et STOLZ, J. (2003). The ecology of arsenic. *Science*, 300:939–944.
- PAEZ-ESPINO, D., TAMAMES, J., de LORENZO, V. et CANOVAS, D. (2009). Microbial responses to environmental arsenic. *Annual Review of Microbiology*, 22:117–130.
- PEZACKA, E. et WOOD, H. (1986). The autotrophic pathway of acetogenic bacteria : role of CO dehydrogenase disulfide reductase. *Journal of Biological Chemistry*, 261:1609–1615.
- PINTO, A. J. et RASKIN, L. (2012). PCR biases distort bacterial and archaeal community structure in pyrosequencing datasets. *PloS one*, 7(8):e43093.
- PLUMLEE, G., LOGSDON, M. et FILIPEK, L. (1999). *The environmental geochemistry of mineral deposits*. Reviews in economic geology. Society of Economic Geologists.
- PONTARP, M., CANBÄCK, B., TUNLID, A. et LUNDBERG, P. (2012). Phylogenetic analysis suggests that habitat filtering is structuring marine bacterial communities across the globe. *Microbial ecology*, 64(1):8–17.
- QIN, J., ROSEN, B., ZHANG, Y., WANG, G., FRANKE, S. et RENSING, C. (2006). Arsenic detoxification and evolution of trimethylarsine gas by a microbial arsenite S-adenosylmethionine methyltransferase. *PNAS*, 103:2075–2080.
- QUAST, C., PRUESSE, E., YILMAZ, P., GERKEN, J., SCHWEER, T., YARZA, P., PEPLIES, J. et GLÖCKNER, F. O. (2013). The SILVA ribosomal RNA gene database project : improved data processing and web-based tools. *Nucleic acids research*, 41(Database issue):D590–6.
- QUATRINI, R., APPIA-AYME, C., DENIS, Y., JEDLICKI, E., HOLMES, D. S. et BONNEFOY, V. (2009). Extending the models for iron and sulfur oxidation in the extreme acidophile *Acidithiobacillus ferrooxidans*. *BMC genomics*, 10:394.
- QUINCE, C., LANZEN, A., DAVENPORT, R. et TURNBAUGH, P. (2011). Removing Noise From Pyrosequenced Amplicons. *BMC Bioinformatics*, 12(1):38.
- RAINEY, F. A., WARD-RAINEY, N. L., JANSSEN, P. H., HIPPE, H. et STACKEBRANDT, E. (1996). *Clostridium paradoxum* DSM 7308T contains multiple 16S rRNA genes with heterogeneous intervening sequences. *Microbiology (Reading, England)*, 142 (Pt 8:2087–95.
- RAM, R. J., VERBERKMOES, N. C., THELEN, M. P., TYSON, G. W., BAKER, B. J., BLAKE, R. C., SHAH, M., HETTICH, R. L. et BANFIELD, J. F. (2005). Community proteomics of a natural microbial biofilm. *Science (New York, N. Y.)*, 308(5730):1915–20.
- RASKIN, L., POULSEN, L. K., NOGUERA, D. R., RITTMANN, B. E. et STAHL, D. A. (1994). Quantification of methanogenic groups in anaerobic biological reactors by oligonucleotide probe hybridization. *Applied and Environmental Microbiology*, 60(4):1241–1248.
- RINKE, C., SCHWIENSTEK, P., SCZYRBA, A., IVANOVA, N. N., ANDERSON, I. J., CHENG, J.-F., DARLING, A., MALFATTI, S., SWAN, B. K., GIES, E. a., DODSWORTH, J. a., HEDLUND, B. P., TSIAMIS, G., SIEVERT, S. M., LIU, W.-T., EISEN, J. a., HALLAM, S. J., KYRPIDES, N. C., STEPANAUSKAS, R., RUBIN, E. M., HUGENHOLTZ, P. et WOYKE, T. (2013). Insights into the phylogeny and coding potential of microbial dark matter. *Nature*, 499(7459):431–7.

- RITCHIE, A. (1994). The waste-rock environment. In JAMBOR, J. L. et BLOWES, D. W., éditeurs : *The environmental geochemistry of sulfide mine-wastes*, chapitre 22, pages 133–161. Mineralogical Association of Canada.
- RODIER, J., BROUTIN, J. P., CHAMBON, P., CHAMPSAUR, H. et RODI, L. (1996). *L'Analyse des Eaux*. Dunod, Paris.
- RONAGHI, M., KARAMOHAMED, S., PETTERSSON, B., UHLÉN, M. et NYRÉN, P. (1996). Real-time DNA sequencing using detection of pyrophosphate release. *Analytical biochemistry*, 242(1):84–9.
- ROSE, A. W. et CRAVOTTA, C. A. (1999). Geochemistry of coal mine drainage. In *Coal mine drainage prediction and pollution prevention in Pennsylvania*. The Pennsylvania Department of Environmental Protection.
- ROSEN, B. et LIU, Z. (2009). Transport pathways for arsenic and selenium : a minireview. *Environment International*, 35:512–515.
- ROSINDELL, J., HUBBELL, S. P. et ETIENNE, R. S. (2011). The unified neutral theory of biodiversity and biogeography at age ten. *Trends in ecology & evolution*, 26(7):340–8.
- ROSSELLÓ-MORA, R. (2003). Opinion : the species problem, can we achieve a universal concept? *Systematic and applied microbiology*, 26(3):323–6.
- ROTHBERG, J. M. et LEAMON, J. H. (2008). The development and impact of 454 sequencing. *Nature biotechnology*, 26(10):1117–24.
- SANGER, F., NICKLEN, S. et COULSON, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*, 74(12):5463–5467.
- SANTELLI, C., WELCH, S., WESTRICH, H. et BANFIELD, J. (2001). The effect of Fe-oxidizing bacteria on Fe-silicate mineral dissolution. *Chemical Geology*, 180:99–115.
- SANZ, J. L., RODRÍGUEZ, N., DÍAZ, E. E. et AMILS, R. (2011). Methanogenesis in the sediments of Rio Tinto, an extreme acidic river. *Environmental microbiology*, 13(8):2336–41.
- SCHLOSS, P. D. (2012). Secondary structure improves OTU assignments of 16S rRNA gene sequences. *The ISME journal*, pages 1–4.
- SCHLOSS, P. D., GEVERS, D. et WESTCOTT, S. L. (2011). Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. *PLoS one*, 6(12):e27310.
- SCHLOSS, P. D. et HANDELSMAN, J. (2004). Status of the Microbial Census. *Microbiology and Molecular Biology Reviews*, 68(4):686–691.
- SCHLOSS, P. D., WESTCOTT, S. L., RYABIN, T., HALL, J. R., HARTMANN, M., HOLLISTER, E. B., LESNIEWSKI, R. A., OAKLEY, B. B., PARKS, D. H., ROBINSON, C. J., SAHL, J. W., STRES, B., THALLINGER, G. G., VAN HORN, D. J. et WEBER, C. F. (2009). Introducing mothur : Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. *Applied and Environmental Microbiology*, 75(23):7537–7541.
- SHANNON, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423 and 623–656.
- SHANNON, P., MARKIEL, A., OZIER, O., BALIGA, N. S., WANG, J. T., RAMAGE, D., AMIN, N., SCHWIKOWSKI, B. et IDEKER, T. (2003). Cytoscape : A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research*, 13:2498–2504.
- SHAW, A. K., HALPERN, A. L., BEESON, K., TRAN, B., VENTER, J. C. et MARTINY, J. B. H. (2008). It's all relative : ranking the diversity of aquatic bacterial communities. *Environmental microbiology*, 10(9):2200–10.
- SHEN, H., CHENG, X., CAI, K. et HU, M.-b. (2009). Detect overlapping and hierarchical community structure in networks. *Physica A : Statistical Mechanics and its Applications*, 388(8):1706 – 1712.
- SHIVELY, J., van KEULEN, G. et MEIJER, W. (1998). Something from almost nothing : carbon dioxide in chemoautotrophs. *Annual Review of Microbiology*, 52:191–230.
- SHUTTLEWORTH, K. L., UNZ, R. F. et WICHLACZ, P. L. (1985). Glucose catabolism in strains of acidophilic, heterotrophic bacteria. *Applied and environmental microbiology*, 50(3):573–9.
- SILVER, S. et PHUNG, L. (2005). Genes and enzymes involved in bacterial oxidation and reduction of inorganic arsenic. *Applied and Environmental Microbiology*, 71:599–608.
- SIMPSON, E. H. (1949). Measurement of Diversity. *Nature*, 163(4148):688–688.
- SINGER, P. C. et STUMM, W. (1970). Acidic mine drainage : The rate-determining step. *Science*, 167(3921):1121–1123.

- SLOAN, W. T., LUNN, M., WOODCOCK, S., HEAD, I. M., NEE, S. et CURTIS, T. P. (2006). Quantifying the roles of immigration and chance in shaping prokaryote community structure. *Environmental microbiology*, 8(4):732–40.
- SLYEMI, D. et BONNEFOY, V. (2012). How prokaryotes deal with arsenic †. *Environmental microbiology reports*, 4(6):571–586.
- SOREY, M. L., COLVARD, E. M. et INGBRITSEN, S. (1994). Measurements of thermal-water discharge outside Lassen Volcanic National Park, California, 1983-94. Rapport technique, U.S. DEPARTMENT OF THE INTERIOR, Menlo Park, California.
- STACKERBRANDT, E. et GOEBEL, B. M. (1994). Taxonomic Note : A Place for DNA-DNA Reassociation and 16S rRNA Sequence Analysis in the Present Species Definition in Bacteriology. *International Journal of Systematic Bacteriology*, 44(4):846–849.
- STACKERBRANDT, E. et RAINEY, F. (1995). Partial and complete 16S rDNA sequences, their use in generation of 16S rDNA phylogenetic trees and their implications in molecular ecological studies. *Molecular Microbial Ecology*, 4:27–34.
- STAHL, D. et AMANN, R. (1991). Development and application of nucleic acid probes in bacterial systematics. In JOHN WILEY AND SONS LTD., éditeur : *Nucleic acid techniques in bacterial systematics.*, pages 205–248. Chichester, UK.
- STOLZ, J., BASU, P., SANTINI, J. et OREMLAND, R. (2006). Arsenic and selenium in microbial metabolism. *Annual Review of Microbiology*, 60:107–130.
- STONE, L. et ROBERTS, A. (1990). The checkerboard score and species distributions. *Oecologia*, 85:74–79.
- STRAUSS, G. et FUCHS, G. (1993). Enzymes of a novel autotrophic CO₂ fixation pathway in the phototrophic bacterium *Chloroflexus aurantiacus*, the 3-hydroxypropionate cycle. *European Journal of Biochemistry*, 215:633–643.
- STUMM, W. et MORGAN, J. J. (1981). *Aquatic Chemistry*. John Wiley & Sons, New York, 2 édition.
- SUGIO, T., KATAGIRI, T., INAGAKI, K. et TANO, T. (1989). Actual substrate for elemental sulfur oxidation by sulfur :ferric ion oxidoreductase purified from Thiobacillus ferrooxidans. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 973(2):250–256.
- SUGIO, T., MIZUNASHI, W., INAGAKI, K. et TANO, T. (1987). Purification and some properties of sulfur :ferric ion oxidoreductase from Thiobacillus ferrooxidans. *Journal of bacteriology*, 169(11):4916–22.
- SUZUKI, I. et WERKMAN, C. (1958). Chemoautotrophic carbon dioxide fixation by extracts of *Thiobacillus thiooxidans*. *Archives of Biochemistry and Biophysics*, 76:103–111.
- SZÉKELY, A. J., BERGA, M. et LANGENHEDER, S. (2013). Mechanisms determining the fate of dispersed bacterial communities in new environments. *The ISME journal*, 7(1):61–71.
- TALAVERA, G. et CASTRESANA, J. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic biology*, 56(4):564–77.
- TAMAKI, S. et FRANKENBERGER, W. (1992). Environmental biochemistry of arsenic. *Reviews of Environmental Contamination and Toxicology*, 2(124):79–110.
- TESKE, A., WAWER, C., MUYZER, G. et RAMSING, N. B. (1996). Distribution of sulfate-reducing bacteria in a stratified fjord (Mariager Fjord, Denmark) as evaluated by most-probable-number counts and denaturing gradient gel electrophoresis of PCR-amplified ribosomal DNA fragments. *Applied and Environmental Microbiology*, 62(4):1405–1415.
- TYSON, G. W., CHAPMAN, J., HUGENHOLTZ, P., ALLEN, E. E., RAM, R. J., RICHARDSON, P. M., SOLOVYEV, V. V., RUBIN, E. M., ROKHSAR, D. S. et BANFIELD, J. F. (2004). Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature*, 428(6978):37–43.
- VALDÈS, J., PEDROSO, I., QUATRINI, R., HALLBERG, K., VLENZUELA, P. et HOLMES, D. (2007). Insights into the metabolism and ecophysiology of three acidithiobacilli by comparative genome analysis. *Advanced Materials Research*, 20-21:439–442.
- VAUGHAN, D. (2006). Arsenic. *Elements*, 2(2):71–75.
- VOLANT, A. (2012). *Etude des communautés microbiennes (Bactérie, Archaea et Eucaryotes) et de leurs variations spatio-temporelles dans la mine de Carnoulès fortement contaminée en arsenic*. Thèse de doctorat, Université Montpellier II.
- VOLANT, a., DESOEUVRE, a., CASIOT, C., LAUGA, B., DELPOUX, S., MORIN, G., PERSONNÉ, J. C., HÉRY, M., ELBAZ-POULICHET, F., BERTIN, P. N. et BRUNEEL, O. (2012). Archaeal diversity : temporal variation in the arsenic-rich creek sediments of Carnoulès Mine, France. *Extremophiles : life under extreme conditions*, 16(4):645–57.
- WAKAO, N., NAGASAWA, N., MATSUURA, T., MATSUKURA, H., MATSUMOTO, T., HIRAISHI, A., SAKURAI, Y. et SHIOTA, H. (1994). *Acidiphilium multivorum* sp. nov., an acidophilic chemoorganotrophic bacterium from pyritic acid mine drainage. *The Journal of General and Applied Microbiology*, 40(2):143–159.

- WANG, X., CAI, Y., SUN, Y., KNIGHT, R. et MAI, V. (2012). Secondary structure information does not improve OTU assignment for partial 16s rRNA sequences. *The ISME journal*, 6(7):1277–80.
- WARD, D. M. (1998). A natural species concept for prokaryotes. *Current Opinion in Microbiology*, 1(3):271–277.
- WATERHOUSE, A. M., PROCTER, J. B., MARTIN, D. M. a., CLAMP, M. et BARTON, G. J. (2009). Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics (Oxford, England)*, 25(9):1189–91.
- WEBB, C. O., ACKERLY, D. D., MCPEEK, M. a. et DONOGHUE, M. J. (2002). Phylogenies and Community Ecology. *Annual Review of Ecology and Systematics*, 33(1):475–505.
- WICHLACZ, P. L., UNZ, R. F. et LANGWORTHY, T. A. (1986). *Acidiphilium angustum* sp. nov., *Acidiphilium facilis* sp. nov., and *Acidiphilium rubrum* sp. nov. : Acidophilic Heterotrophic Bacteria Isolated from Acidic Coal Mine Drainage. *International Journal of Systematic Bacteriology*, 36(2):197–201.
- WILMES, P. et BOND, P. L. (2006). Metaproteomics : studying functional gene expression in microbial ecosystems. *Trends in microbiology*, 14(2):92–7.
- WOESE, C. R. et FOX, G. E. (1977). Phylogenetic structure of the prokaryotic domain : The primary kingdoms. *Proceedings of the National Academy of Sciences*, 74(11):5088–5090.
- WOOD, A. et KELLY, D. (1983). Autotrophic and mixotrophic growth of three thermoacidophilic iron-oxidizing bacteria. *FEMS Microbiology Letters*, 20:107–112.
- WOOD, A. et KELLY, D. (1984). Growth and sugar metabolism of a thermoacidophilic iron-oxidizing mixotrophic bacterium. *Journal of General Microbiology*, 130:1337–1349.
- WOODCOCK, S., van der GAST, C. J., BELL, T., LUNN, M., CURTIS, T. P., HEAD, I. M. et SLOAN, W. T. (2007). Neutral assembly of bacterial communities. *FEMS microbiology ecology*, 62(2):171–80.
- WOOTTON, J. T. (2005). Field parameterization and experimental test of the neutral theory of biodiversity. *Nature*, 433(7023):309–12.
- WYSOCKI, R., CLEMENS, S., AUGUSTYNIAK, D., GOLIK, P., MACIASZCZYK, E., TAMÁS, M. J. et DZIADKOWIEC, D. (2003). Metalloid tolerance based on phytochelatins is not functionally equivalent to the arsenite transporter *Act3p*. *Biochemical and Biophysical Research Communications*, 304(2):293–300.
- YELTON, A. P., COMOLLI, L. R., JUSTICE, N. B., CASTELLE, C., DENEFF, V. J., THOMAS, B. C. et BANFIELD, J. F. (2013). Comparative genomics in acid mine drainage biofilm communities reveals metabolic and structural differentiation of co-occurring archaea. *BMC genomics*, 14(1):485.
- YUAN, C., LU, X., QIN, J., ROSEN, B. et LE, X. (2008). Volatile arsenic species released from *Escherichia coli* expressing the *AsIII* S-adenosylmethionine methyltransferase gene. *Environmental Science & Technology*, 42:3201–3206.
- ZARGAR, K., CONRAD, A., BERNICK, D. L., LOWE, T. M., STOLC, V., HOEFT, S., OREMLAND, R. S., STOLZ, J. et SALTIKOV, C. W. (2012). *ArxA*, a new clade of arsenite oxidase within the DMSO reductase family of molybdenum oxidoreductases. *Environmental microbiology*.
- ZHANG, G. et FANG, Z. (2005). The contribution of direct and indirect actions in bioleaching of pentlandite. *Hydrometallurgy*, 80:59–66.

Table des figures

| | | |
|-----|--|-----|
| 1.1 | Schémas représentant le cycle de Calvin. | 18 |
| 1.2 | Cycle du 3-hydroxypropionate de <i>Chloroflexus aurantiacus</i> | 19 |
| 1.3 | Modèle de l'oxydation du Fe(II) chez <i>At. ferrooxidans</i> ATCC 23270 | 24 |
| 1.4 | Modèle de l'oxydation des sulfures chez <i>Acidithiobacillus</i> | 26 |
| 1.5 | Représentation schématique des différentes transformations de l'arsenic par les procaryotes | 29 |
| 1.6 | Résumé des relations écologiques entre les membres d'espèces différentes . . . | 33 |
| 1.7 | Schémas des groupes microbiens impliqués dans le cycle du fer dans le lac minier acide Lake 77 | 34 |
| 1.8 | Relation syntrophique proposée entre <i>Desulfosporosinus</i> M1 et <i>Acidocella</i> PFBC | 35 |
| 1.9 | Représentation schématique des rôles et des interactions entre les différents groupes physiologiques des procaryotes acidophiles | 36 |
| 2.1 | Planisphère où sont représentées les localisations des 12 sites d'études. | 48 |
| 2.2 | Schémas représentant les amorces fusions utilisées pour le séquençage des gènes codant pour l'ARNr 16S. | 50 |
| 4.1 | Répartition des phyla <i>Archaea</i> majoritaires dans les sites d'étude. | 112 |
| 4.2 | ACoP des distances Unifrac entre les communautés archées des DMA étudiés | 112 |
| 4.3 | Arbre phylogénétique des représentants des OTU <i>Archaea</i> | 116 |
| 4.4 | Diagramme illustrant les résultats de la fouille de données sur les séquences <i>Archaea</i> | 117 |
| 4.5 | ACC sur les OTU <i>Archaea</i> ayant plus de 5 séquences | 119 |
| 5.1 | Arbre phylogénétique inféré des séquences d'acides aminés des représentants des OTU ArsB | 128 |
| 5.2 | Arbre phylogénétique inféré des séquences d'acides aminés des représentants des OTU Acr3.1 | 130 |
| 5.3 | Arbre phylogénétique inféré des séquences d'acides aminés des représentants des OTU Acr3.2 | 132 |
| 5.4 | Schémas conceptuel des interactions trophiques entre les procaryotes lithotrophes et hétérotrophes des DMA étudiés. | 139 |

Liste des tableaux

| | | |
|-----|--|-----|
| 1.1 | Principaux sulfures métalliques retrouvés dans les résidus miniers classés par ordre croissant de réactivité | 10 |
| 1.2 | Procaryotes observées dans les DMA et leurs principales caractéristiques physiologiques | 17 |
| 1.3 | Pourcentage de la fraction cultivable de la diversité microbienne pour différents écosystèmes | 39 |
| 2.1 | Coordonnées GPS et dates des campagnes de prélèvement des 12 sites d'étude. | 47 |
| 2.2 | Liste des amorces utilisées pour les amplifications PCR. | 51 |
| 2.3 | Liste des identifiants multiplex utilisés pour les amplifications PCR des gènes ARNr 16S. | 51 |
| 4.1 | Description de la richesse et de la diversité <i>Archaea</i> : jeu de données complet | 111 |
| 4.2 | Description de la richesse et de la diversité <i>Archaea</i> : jeu de données rééchantillonné | 111 |
| 4.3 | Description des paramètres physico-chimiques de l'eau des sites d'étude . . . | 117 |
| 5.1 | Statistiques sur les gènes codant les transporteurs d'arsenite ArsB et Acr3. . | 125 |

Liste des abréviations

| | |
|----------------|--|
| ACC | Analyse Canonique des Correspondances |
| ACP | Analyse en Composante Principale |
| ADN | Acide DésoxyriboNucléique |
| ARMAN | Archaeal Richmond Mine Nanoorganisms |
| ARN | Acide RiboNucléique |
| ATP | Adénosine 5'-TriPhosphate |
| BLAST | Basic Local Alignment Search Tool |
| BSR | Bactéries Sulfato-Réductrices |
| CBB | cycle de Calvin-Benson-Bassham |
| COD | Composés Organiques Dissous |
| DAR | Drainage Acide de Roche |
| DMA | Drainage Mimier Acide |
| EDTA | EthyleneDiamineTetraacetic Acid |
| HDR | l'HétéroDisulfure Réductase |
| ICP-MS | Inductively Coupled Plasma Mass Spectrometry |
| IPTG | IsoPropyl β -D-1-ThioGalactopyranoside |
| LAH | List Abbreviations Here |
| LB | Lysogeny Broth |
| MIGRAMD | MIcrobial biogeoGRaphy of Acid Mine Drainage |
| MRT | Multivariate Regression Tree |
| NAD(P) | Nicotinamide Adénine Dinucléotide (Phosphate) |
| NAD(P)H | forme réduite du NAD(P) |
| NCBI | The Natinoal Center for Bioinformatic Information |
| NRI | Net Relatedness Index |
| NTI | Nearest Taxa Index |
| OTU | Operational Taxonomic Unit |
| PCR | Polymerase Chain Reaction |
| Pit | Phosphate Inorganic Transport |
| QH2 | pool de quinone |
| RISC | composés soufrés inorganiques réduits (Reduced Inorganic Sulfur Compounds) |
| RUBISCO | Ribulose-1,5-Bisphosphate Carboxylase/Oxygénase |
| SQR | la Sulfure Quinone Réductase |
| TQR | la Thiosulfate Quinone Réductase |
| X-Gal | 5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside |

Liste des symboles

| | |
|--|-----------------------------------|
| °C | degré Celsius |
| As | arsenic |
| As(III) | arsenite |
| As(V) | arsenate |
| BaCl ₂ | chlorure de baryum |
| BaSO ₄ | sulfate de baryum |
| Cd | cadmium |
| Co | cobalt |
| CO ₂ | dioxyde de carbone |
| CuFeS ₂ | chalcopyrite |
| e ⁻ | électron |
| E ⁰ | potentiel red/ox |
| Fe | fer |
| Fe(II) ; Fe ²⁺ | fer ferreux |
| Fe(III) ; Fe ³⁺ | fer ferrique |
| Fe(OH) ₃ | hydroxyde ferrique |
| FeAsS | arsénopyrite |
| FeOOH | oxyhydroxyde ferrique |
| FeS ₂ | pyrite |
| H ₂ O | eau |
| H ₂ S | sulfure d'hydrogène |
| H ⁺ ; H ₃ O ⁺ | proton ; ion hydronium |
| HCO ₃ ⁻ | bicarbonate ou hydrogénocarbonate |
| Hg | mercure |
| HS ⁻ | ion hydrogénosulfure |
| K _m | constante de Michaelis |
| K ⁺ | ion potassium |
| Mn | manganèse |
| Na ₂ HPO ₄ | hydrogénophosphate de sodium |
| NaH ₂ PO ₄ | dihydrogénophosphate de sodium |
| Ni | nickel |
| O ₂ | dioxygène |
| S | soufre |
| SO ₄ ²⁻ | ion sulfate |
| ZnS | sphalérite |

Annexe A

Production scientifique

Diversity and spatiotemporal dynamics of bacterial communities: physicochemical and other drivers along an acid mine drainage

Aurélie Volant¹, Odile Bruneel¹, Angélique Desoeuvre¹, Marina Héry¹, Corinne Casiot¹, Noëlle Bru², Sophie Delpoux¹, Anne Fahy³, Fabien Javerliat³, Olivier Bouchez⁴, Robert Duran³, Philippe N. Bertin⁵, Françoise Elbaz-Poulichet¹ & Béatrice Lauga³

¹Laboratoire HydroSciences Montpellier, HSM, UMR 5569 (IRD, CNRS, Universités Montpellier 1 et 2), Université Montpellier 2, Montpellier, France; ²Laboratoire de Mathématiques et de leurs Applications, UMR 5142 (CNRS), Université de Pau et des Pays de l'Adour, Pau, France; ³Équipe Environnement et Microbiologie, EEM, UMR 5254 (IPREM, CNRS), Université de Pau et des Pays de l'Adour, Pau, France; ⁴INRA Auzeville, Plateforme Génomique Chemin de Borde Rouge, Castanet-Tolosan, France; and ⁵Département Microorganismes, Génomes, Environnement, Laboratoire de Génétique Moléculaire, Génomique, Microbiologie, GMGM, UMR 7156 (Université de Strasbourg, CNRS), Strasbourg, France

Correspondence: Odile Bruneel, Laboratoire HydroSciences Montpellier, UMR5569, Université Montpellier 2, Place E. Bataillon, CC MSE, 34095 Montpellier, France.
Tel.: (+33)4 67 14 36 59;
fax: (+33)4 67 14 47 74;
e-mail: odile.bruneel@msem.univ-montp2.fr

Received 16 April 2014; revised 10 July 2014; accepted 16 July 2014. Final version published online 09 September 2014.

DOI: 10.1111/1574-6941.12394

Editor: Tillmann Lueders

Keywords

spatial and temporal dynamics; bacterial diversity; acid mine drainage; arsenic.

Abstract

Deciphering the biotic and abiotic factors that control microbial community structure over time and along an environmental gradient is a pivotal question in microbial ecology. Carnoulès mine (France), which is characterized by acid waters and very high concentrations of arsenic, iron, and sulfate, provides an excellent opportunity to study these factors along the pollution gradient of Reigous Creek. To this end, biodiversity and spatiotemporal distribution of bacterial communities were characterized using T-RFLP fingerprinting and high-throughput sequencing. Patterns of spatial and temporal variations in bacterial community composition linked to changes in the physicochemical conditions suggested that species-sorting processes were at work in the acid mine drainage. Arsenic, temperature, and sulfate appeared to be the most important factors that drove the composition of bacterial communities along this continuum. Time series investigation along the pollution gradient also highlighted habitat specialization for some major members of the community (*Acidithiobacillus* and *Thiomonas*), dispersal for *Acidithiobacillus*, and evidence of extinction/re-thriving processes for *Gallionella*. Finally, pyrosequencing revealed a broader phylogenetic range of taxa than previous clone library-based diversity. Overall, our findings suggest that in addition to environmental filtering processes, additional forces (dispersal, birth/death events) could operate in AMD community.

Introduction

Acid mine drainage (AMD) is one of the most pernicious forms of pollution in the world and is widely recognized as having costly environmental and socioeconomic impacts (Hallberg, 2010). AMD occurs when waste from the extraction and processing of sulfide ore comes into contact with oxygenated water. Drainages are typically acidic and usually contain high concentrations of sulfate, metals and metalloids including arsenic. Although perceived as extreme environments hostile to life, a variety of microorganisms are able to thrive in it. For some of them, their role in the oxidation of sulfide minerals,

which leads to bioleaching, is well known, as is their role in natural attenuation of such polluted waters (Edwards *et al.*, 2000; Hallberg, 2010; Johnson, 2012). Despite the central role of microorganisms in such ecosystem functioning, our understanding of the mechanisms shaping microbial community structure and diversity in AMD remains limited. As pointed out by Miller *et al.* (2009), deciphering how microbial communities are patterned along environmental gradients is a pivotal question in microbial ecology. Although AMD is characterized by changing conditions over time and space, few studies were interested in comparing the microbial communities along such environmental gradients. The AMD of

Carnoulès in southern France provides an excellent opportunity to investigate these fundamental questions of microbial ecology. This site is characterized by an acid pH (2–4) and high levels of metal and metalloids, in particular As (up to 10 g L^{-1} in the tailings stock pore water, and $100\text{--}350 \text{ mg L}^{-1}$ at the source of Reigous Creek), and natural attenuation processes result in a strong spatial pollution gradient along the drainage. Indeed, nearly 95% of the arsenic in solution is removed between the source of Reigous Creek, which emerges from the mine tailings, and its confluence with the Amous River, 1.5 km downstream. To our knowledge, this AMD is one of the most As-rich AMDs reported to date (Morin & Calas, 2006). It is an outstanding example of adaptation to life in an extreme environment.

In this study, we used a combination of molecular approaches to investigate the spatial and temporal dynamics of bacterial communities in relation to the physicochemical parameters in Carnoulès acid mine drainage. Using 16S rRNA gene pyrosequencing and terminal restriction fragment length polymorphism (T-RFLP), our aim was (1) to characterize the spatial dynamics of the structure and composition of the bacterial communities along an environmental gradient, (2) to evaluate the temporal changes in the composition of the communities, and (3) to determine whether their dynamics could be linked to variations of environmental conditions.

Materials and methods

Description of the study site

The Pb–Zn Carnoulès mine, located in southern France, produced 1.2 Mt of solid wastes that are stored behind a dam and contain 0.7% Pb, 10% Fe, and 0.2% As. The aquifer is not fed by vertical percolation of rainwater through the tailings, but rather originates from natural springs that were buried under the tailings (Koffi *et al.*, 2003). The water table is 1–10 m below the surface of the tailings stock, depending on the season and location. With the exception of temperature, which is almost constant with average values around $15 \text{ }^\circ\text{C}$, the physicochemical parameters of the groundwater vary as a function of the hydrological conditions (Casiot *et al.*, 2003b). In 2001, the groundwater below the tailings contained extremely high levels of arsenic: up to $10\,000 \text{ mg L}^{-1}$ (Casiot *et al.*, 2003b). The water emerges at the bottom of the dam, forming the source of the Reigous Creek. This AMD is acid ($\text{pH} \leq 3$), with high concentrations of sulfate ($2000\text{--}7700 \text{ mg L}^{-1}$), iron ($500\text{--}1000 \text{ mg L}^{-1}$), and arsenic ($50\text{--}350 \text{ mg L}^{-1}$). Iron and arsenic are mainly present in their reduced forms Fe(II) and As(III) (Casiot *et al.*, 2003a). The natural attenuation of As is the

result of microbiologically mediated As–Fe coprecipitation (Morin *et al.*, 2003; Bruneel *et al.*, 2006). 10–47% of Fe, and 20–60% of As are removed from the aqueous phase within the first 30 m of the creek. Beyond this point (COWG sampling site, located 30 m downstream from the spring, Fig. 1), the Reigous receives water from quarries and mine galleries, especially after rainfall events, which strongly influence its acidity and metal content (Egal *et al.*, 2010).

Sampling procedure and measurement of physicochemical properties

Six sampling campaigns were carried out in November 2007, February 2008, October 2008, March 2009, November 2009, and January 2010 at five sampling sites, resulting in a set of 30 samples. Groundwater was collected from a borehole (S5, between 10 and 12 m deep) located within the tailings. Water samples were also taken at four sites along Reigous Creek (collecting downstream seepage water from the surroundings) at the spring (S1), 30 m downstream from the spring (COWG), 150 m downstream (GAL), and 1500 m downstream (CONF), just before the confluence between Reigous Creek and the Amous River (Fig. 1). Water samples (300 mL) were immediately filtered through sterile $0.22 \text{ }\mu\text{m}$ Nuclepore filters, which were transferred to a collection tube (Nunc), frozen in liquid nitrogen, and stored at $-80 \text{ }^\circ\text{C}$ until DNA extraction. This sampling was carried out in triplicate. Measurements of water conductivity, temperature, redox potential, pH, and dissolved oxygen concentration were carried out as previously described (Bruneel *et al.*, 2011). For chemical analyses, 500 mL water samples were immediately filtered through $0.22 \text{ }\mu\text{m}$ 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 $4 \text{ }^\circ\text{C}$ in polyethylene bottles until analysis. A $20 \text{ }\mu\text{L}$ aliquot of the filtered water sample was added either to a mixture of acetic acid and EDTA (Samanta & Clifford, 2005) for As speciation or to a mixture of 0.5 mL acetate buffer (pH 4.5) and 1 mL of 1,10-phenanthroline solution (Rodier *et al.*, 1996) for Fe(II) determination. The vials were filled to 10 mL with deionized water. Samples destined for Fe and As speciation and sulfate determination were stored in the dark and analyzed within 24 h. Chemical analyses were carried out as previously described (Bruneel *et al.*, 2011).

DNA isolation

Genomic DNA was extracted in triplicate from filtered water using the UltraClean Soil DNA Isolation Kit

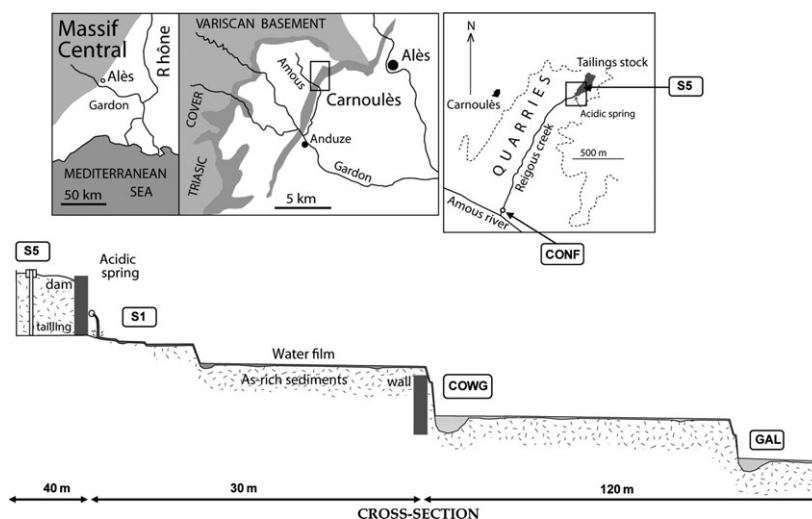


Fig. 1. Map of Carnoulès mine and location of sampling sites.

(MoBio Laboratories Inc., Carlsbad, CA) according to the manufacturer's recommendations. These triplicate extractions were pooled before PCR amplification. All genomic DNA extracts were stored at -20°C until further analysis.

Terminal restriction fragment length polymorphism

The 16S rRNA genes were amplified by PCR, and the bacterial community structure was identified by T-RFLP. The fluorescent labeled primers HEX 357F (5'-hexachloro-fluorescein-phosphoramidite-CCTACGGGAGGCA GCAG-3') (Lane, 1991) and 926R (5'-CCGTCAATTCMT TTRAGT-3') (Muyzer & Ramsing, 1995), described as universal within the bacterial domain, were used. Triplicate PCR amplifications were performed on each sample. The reaction mixture contained 1 μL of DNA template, 1 μL of both primers (10 μM), and 12.5 μL of PCR Master Mix Ampli Taq Gold 360 (Applied Biosystems, Foster City, CA). Sterile distilled water was added to obtain a final volume of 25 μL . PCR conditions were as follows: one cycle at 95°C for 10 min, 35 cycles at 95°C for 45 s, 55°C for 45 s, and 72°C for 45 s, followed by 10 min at 72°C . The 90 PCR products were purified with Illustra GFXTM PCR DNA and the Gel Band Purification Kit (GE Healthcare, Munich, Germany). The concentration of PCR product was determined by comparison with molecular markers (Smartlader, Eurogentec) after migration on agarose gel. Approximately 100 ng of purified amplicon was digested in 10 μL reaction with 0.3 U of enzyme HpaII or AluI (New England Biolabs Inc., Ipswich, MA) at 37°C for 3 h. Terminal restriction frag-

ment (T-RF) profiles were obtained from the digested amplicons by suspending 1 μL aliquots in 8.75 μL formamide with 0.25 μL of Genescan ROX 500 size standard (Applied Biosystems). T-RFs were separated on an ABI PRISM 3130xl Genetic Analyser (Applied Biosystems). Data were collected and analyzed using GENEMAPPER software (version 1.4, Applied Biosystem). To increase stringency for the T-RF profiles of 16S rRNA genes, T-RFs outside the range of the size standard (35–500 bp) were discarded, and the background noise level was set at 30 fluorescence units. T-ALIGN software (Smith *et al.*, 2005) was used to compare replicate profiles and to generate consensus profiles containing only T-RFs that occurred in replicate reactions. Consensus profiles were then aligned on the basis of the length of the T-RFs and individual peak areas as previously described by Smith *et al.* (2005) with the confidence interval set at 0.5, resulting in the generation of data sets of aligned T-RFs that gave individual relative peak areas as a percentage of the overall profile. T-RFs were included in the subsequent analysis if they represented $> 1\%$ of the cumulative peak height for the sample.

Construction of the libraries, 454-pyrosequencing, and sequence quality control

The 16S rRNA genes were also amplified by PCR for multiplex pyrosequencing using barcoded primers. The primer pairs used, targeting the V3 to V5 variable regions of the 16S rRNA gene, were 357F (5'-AxxxCTTACGGGAGGCAGCAG-3') and 926R (5'-BxxxCCGTCAATTCMTTTRAGT-3'). A and B represent the two FLX Titanium adapters (A adapter sequence: 5'-CGTATCGCCTC CCTCGGCCATCAG-3'; B adapter sequence: 5'-CTAT

GCGCCTTGCCAGCCCGCTCAG-3'), and xxx represent the sample-specific barcode sequence. PCR was performed using 30–35 cycles under conditions identical to those described above for T-RFLP. The number of cycles was varied with the samples to obtain a strong band with a minimum number of cycles to respect the initial abundances of bacterial communities. The 90 PCR products with a proximal length of 569 bp were excised from 1% agarose gel and purified with the QIAquick Gel Extraction Kit (QIAGEN Inc., Valencia, CA). To minimize random PCR bias, triplicates were pooled in equimolar ratios prior to pyrosequencing. Pyrosequencing of the 30 amplicon libraries was performed on a GS-FLX-Titanium sequencer (Roche 454 Life Sciences) at the GenoToul genomic platform in Toulouse (France) using four separate 1/8 region of a plate.

Processing of pyrosequencing data and taxonomic classification

Preliminary quality checks, sorting, and trimming of the 454 reads were carried out using the NG6 pipeline (<http://vm-bioinfo.toulouse.inra.fr/ng6/>). Tags were extracted from the 454 reads using the sff file (Roche software), and three kinds of analysis were performed as described by Ueno *et al.* (2010): (1) BLAST search for *E. coli*, phage, and yeast contaminants, (2) read quality analysis, and (3) removal of sequences that were too long or too short (sequences with more or less than two standard deviations from the mean), sequences containing more than 4% of N, low-complexity sequences and duplicated reads, using Pyrocleaner. The sequences were then analyzed with the software Mothur version 1.30 (Schloss *et al.*, 2009). Preprocessing of unaligned sequences included removing sequences < 450 bp, all sequences containing ambiguous characters, and sequences with more than eight homopolymers. We also removed sequences that did not align over the same span of nucleotide positions. Identical sequences were grouped, and representative sequences were aligned against the SILVA bacterial and archaeal reference database using the Needleman–Wunsch algorithm (Needleman & Wunsch, 1970). Chimeric sequences were detected and removed using the implementation of Chimera Uchime. A further screening step (precluster) was carried out to reduce sequencing noise by clustering reads differing by only one base every 100 bases (Huse *et al.*, 2010). The remaining high-quality reads were used to generate a distance matrix and were clustered into operational taxonomic units (OTUs) defined at 97% cut-off using the average neighbor algorithm. Next, the OTUs were phylogenetically classified to genus level using the naive Bayesian classifier (80% confidence

threshold) trained on the RDP taxonomic outline implemented in Mothur and a modified bacterial database. *In silico* T-RF prediction of the 16S rRNA gene sequences obtained in this study was performed using the program TRiFLE (Junier *et al.*, 2008), and predicted T-RFs were linked to measured T-RFs from the microbial community profiles.

Estimation of diversity and statistical analysis

Diversity indices

Nonparametric Chao1 and Shannon alpha diversity estimates, as well as coverage and rarefaction curves, were calculated with MOTHUR v.1.30 for each sample. Analysis of variance (ANOVA) was performed with Tukey's tests to identify differences between sampling sites.

Cluster analysis

To compare community composition based on T-RFLP and 454-pyrosequencing data, normalized OTUs abundances were square-root-transformed and pairwise dissimilarities among samples were calculated using the relative abundance-based Bray–Curtis index (BC). Non-metric multidimensional scaling (nMDS) analysis was performed on the dissimilarity matrices to visualize patterns of community composition. Using the 454-pyrosequencing data, we carried out a random sampling procedure to make equal the number of sequences per sample (486 sequences) and we removed singleton OTUs (sequences that only occurred once) to reduce the influence of rare OTUs. One-way analysis of similarity (ANOSIM) and multiple pairwise comparisons were used to test whether there were significant differences in community composition in space. *R*-values > 0.75 are commonly interpreted as well-separated bacterial compositions; *R* > 0.5 as overlapping, but clearly different; and *R* < 0.25 as practically not separable.

CCA

Canonical correspondence analyses (CCAs) were used to explore variations in the bacterial communities under the constraint of our set of environmental variables. Explanatory variables were $\log(x + 1)$ -transformed where necessary to approximate normal distribution. This model was tested with Monte Carlo permutation tests (499 randomized runs) to determine significance, and each environmental parameter was tested by stepwise analysis to detect significant predictors. All statistical analyses were performed with R 3.0.1 (R Development Core Team, 2012) including the VEGAN package.

Results

Spatial and temporal analyses of the environmental data set

The physicochemical characteristics of the water samples were determined for six sampling dates and at five different sites in a borehole and along Reigous Creek (Fig. 1; Supporting Information, Table S1). With the exception of pH, the environmental variables measured differed significantly between sites (ANOVA, $P < 0.05$). A significant decrease (Tukey's test, $P < 0.01$) in temperature was observed in Reigous Creek, where the influence of air temperature causes a larger range of values (Fig. 2a) in contrast to the temperature of the water in the borehole (S5) and at the source of the Reigous (S1), which did not differ significantly (13.3 ± 1.4 °C and 13.6 ± 1 °C, respectively, and Tukey's test, $P = 0.99$). All samples were characterized by low pH (≤ 3.7). No significant variations in pH (ANOVA, $P = 0.49$) were observed between the sampling sites located downstream from the source (COWG, GAL and CONF; Fig. 2b). Dissolved oxygen (DO) concentrations in the water at the upstream sites presented a mean of 1.0 ± 1.0 mg L⁻¹ for S5 and 0.8 ± 0.6 mg L⁻¹ for S1, denoting generally suboxic conditions at these sites. DO increased sharply between S1 and COWG (mean of 6.1 ± 1.8 mg L⁻¹) (Fig. 2c) and continued to increase slightly all along the creek, to reach a mean of 10.5 ± 1.3 mg L⁻¹ at CONF. The redox potential (Eh) showed average values of 558 ± 85 mV at S5. Eh increased along Reigous Creek from 512 ± 39 mV at S1 to 635 ± 99 mV at CONF (Fig. 2d). In contrast, average conductivity decreased from 7588 ± 5203 μ S cm⁻¹ at S5 to 5121 ± 631 μ S cm⁻¹ at S1 and reached minimum (1612 ± 97 μ S cm⁻¹) at CONF (Fig. 2e). Sulfate (SO₄²⁻) concentrations were maximum in the groundwater at S5 with a mean of $14\,080 \pm 10\,630$ mg L⁻¹. After a sharp decrease at S1 (average values of 2682 ± 1180 mg L⁻¹), concentrations gradually decreased along Reigous Creek (Fig. 2f). Dissolved Fe concentrations in the groundwater at S5 exhibited average values of 4474 ± 2855 mg L⁻¹. Fe concentrations decreased from the source (S1, average values of 1317 ± 383 mg L⁻¹) to CONF, where Fe remained below 82 mg L⁻¹ (Fig. 2g). The proportion of Fe(III) (difference between Fe(total) and Fe(II)) was generally negligible except at some sampling dates at S5 and CONF (Table S1). At S5, concentrations of dissolved arsenic (As) exhibited an average value of 440 ± 184 mg L⁻¹ (Fig. 2h). Along Reigous Creek, As concentrations decreased with increasing distance from the source (average value of 175 ± 71 mg L⁻¹), to values below 6 mg L⁻¹ at CONF (Fig. 2h), with predominance of the reduced form As(III). An average of 65% of sulfate,

96% of iron, and 99% of arsenic were removed from the aqueous phase between S1 and CONF sampling sites.

Diversity and species richness estimators of bacterial communities

Hex-labeled PCR products were digested separately with two restriction enzymes. HpaII that produced the largest numbers of T-RFs (data not shown) was used to assess the differences in the microbial communities. T-RFLP profiles generated showed a total of 43 different T-RFs for the five sites, and the number of T-RFs detected in each sample varied from 2 to 17 (Fig. 3). Average T-RF richness (number of T-RFs) and average Shannon diversity indices calculated from relative peak intensity data were highest at S1 and COWG ($H = 2.03 \pm 0.3$ and $H = 1.96 \pm 0.3$, respectively), while the lowest values were observed at GAL ($H = 1.26 \pm 0.4$; Fig. 4a). Values at CONF were intermediate ($H = 1.67 \pm 0.7$). Although bacterial community diversity varied among the sampling sites, the differences were not significant (ANOVA, $F = 2.46$, $P = 0.071$). For each site, the bacterial community showed variations over time, but no particular trend could be identified. Some T-RFs were found in the majority of the profiles (e.g. T-RF 150), where they usually accounted for a high proportion of the total T-RFs (Fig. 3). Between one and three site-specific T-RFs were identified in all the sites (in red in Fig. 3).

A total of 99 441 sequence reads were generated in a single run of 454-pyrosequencing from 30 independent 16S rRNA gene libraries. Note that pyrosequencing of two samples taken in February 2008 (S5 and S1) failed and were thus excluded from analysis. After trimming and processing with Mothur, 63 442 reads remained with an average length of 530 bp. Clustering of the remaining sequences led to the identification of 6613 OTUs (including 4510 singletons) defined at 97% identity. Although singletons represented 68% of the total number of OTUs, they only accounted for 7% of the total DNA sequences. The results of rarefaction analysis along with the Chao1 and the Shannon indexes and coverage values are listed in Table 1. In the resampled data set, Good's coverage ranged from 69% to 97% with an average value of 85%, indicating that the majority of bacterial phylotypes were recovered. Species richness (Chao1 index) of the bacterial communities presented significant variations along Reigous Creek (ANOVA, $P = 0.001$, $F = 6.66$) (Fig. 4b). The nonparametric estimators Chao1 ranged between 52 and 495 estimated OTUs for all the sites considered (Table 1). The highest average OTU richness was found at CONF and S1 (Chao1 = 364 ± 145 and 296 ± 32 , respectively), suggesting that an important number of OTUs were not revealed by the analysis of these two sites. Situated

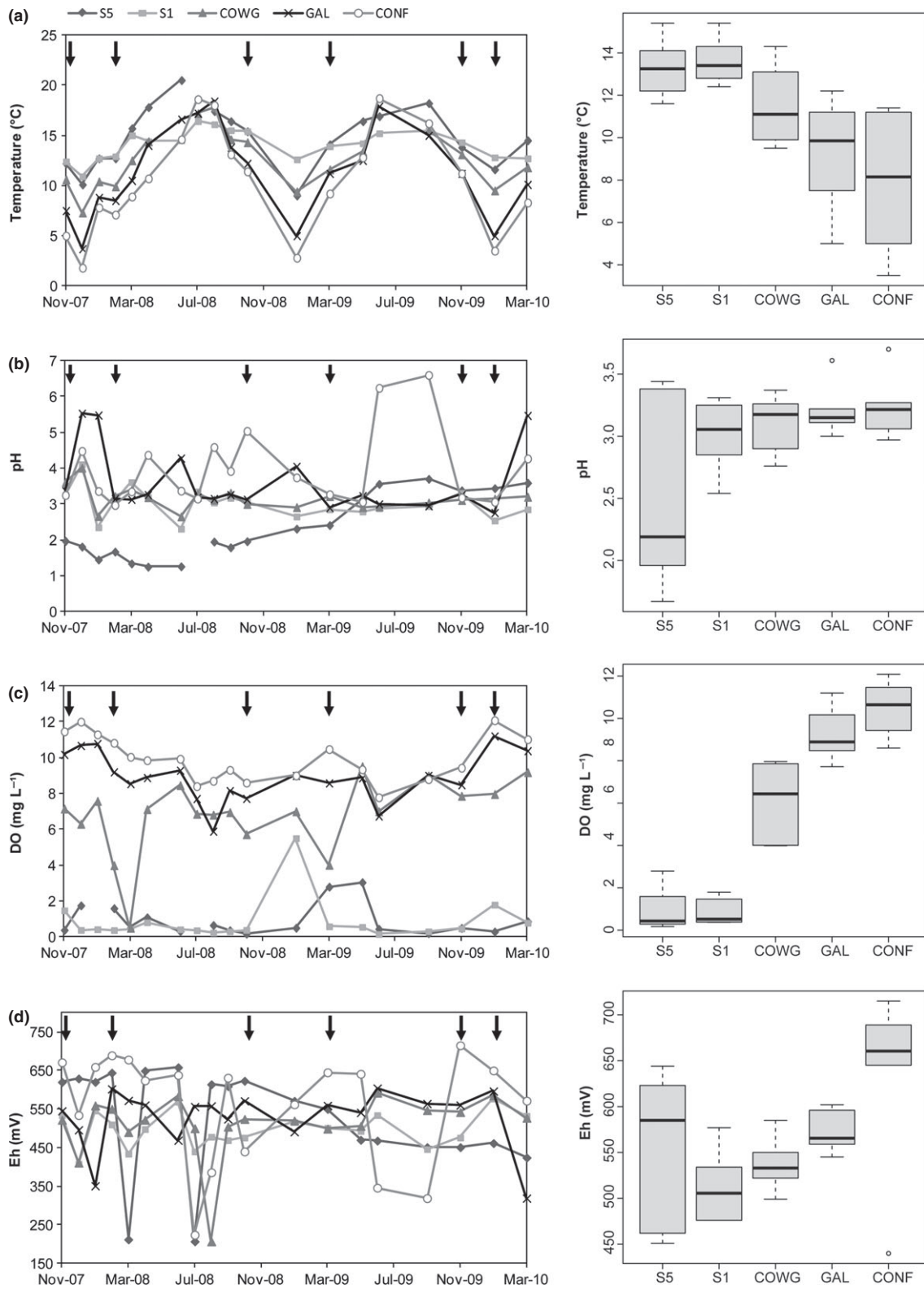


Fig. 2. Variations in the main physicochemical parameters over the course of the study and boxplot of each variable per sampling site. Arrows indicate sampling dates for T-RFLP and pyrosequencing analysis. Note that some data are missing, as shown by the gaps in the curves. DO, dissolved oxygen; Eh, redox potential.

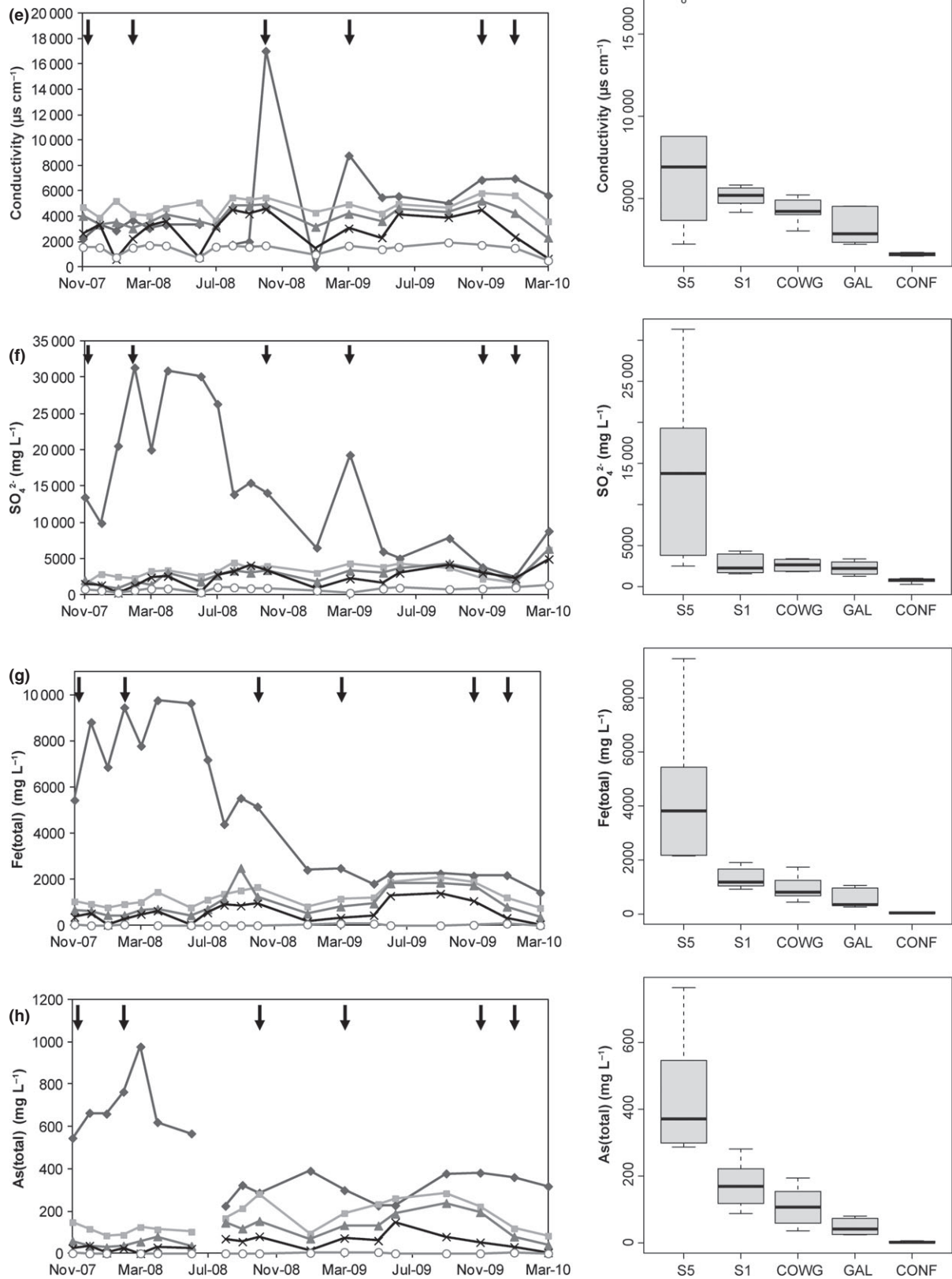


Fig. 2. Continued

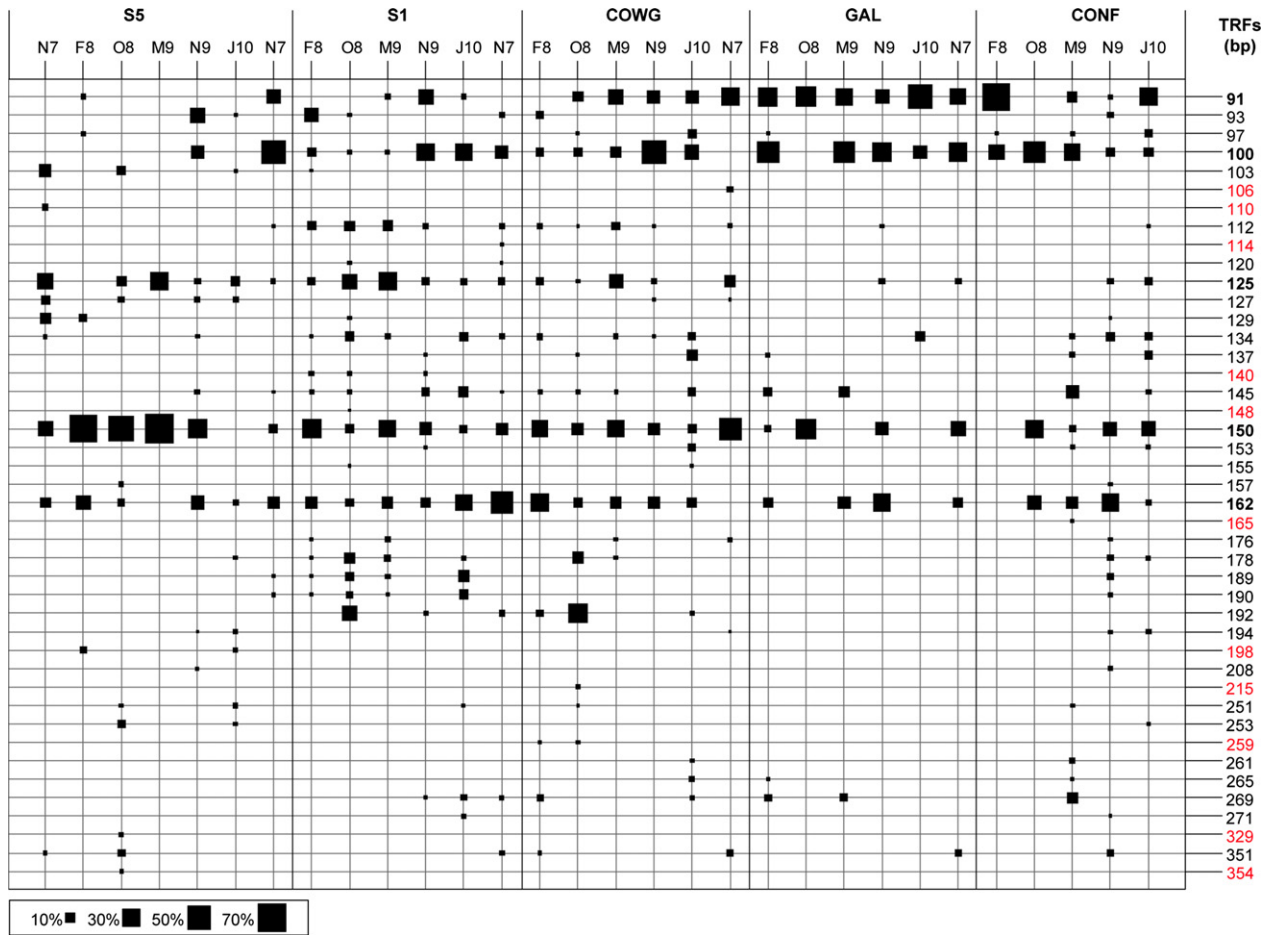


Fig. 3. Relative abundance of terminal restriction fragments (T-RFs) derived from bacterial communities. Single T-RFs per sampling site are in red. Dominant T-RFs are in bold. Taxonomic affiliation of T-RFs was carried out by *in silico* T-RFLP analysis. T-RF 91 and 100 could not be assigned to any phylogenetic group; T-RF 125 represented *Armatimonadetes* gp4 and *Chlorobi*; T-RF 150 was mainly related to *Deinococcus-Thermus*, *Spirochaetes* and *Actinobacteria*; T-RF 162 was mainly related to *A. ferrooxidans* but could be assigned to other proteobacterial phylotypes detected in the AMD. N7: November 2007; O8: October 2008; M9: March 2009; N9: November 2009; J10: January 2010.

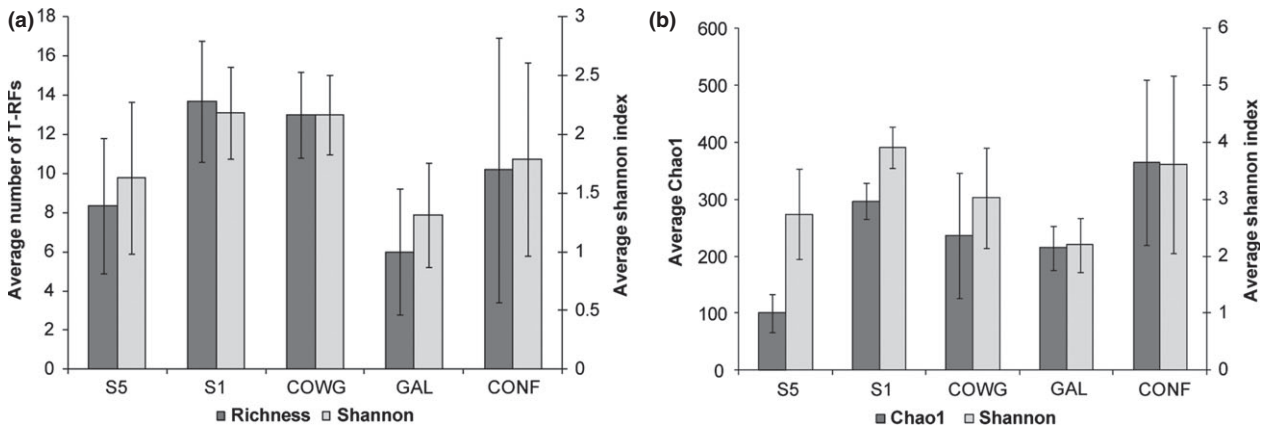


Fig. 4. Average diversity and species richness index per group ± standard deviation calculated based on (a) T-RFLP profiles and (b) 454 pyrosequencing reads of the reduced data set based on the 16S rRNA genes.

Table 1. Estimated OTU richness, diversity indices, and estimated sample coverage for each 16S rRNA gene library. Results are presented for full data set reads (full data set) and for reduced data sets without singletons and randomly resampled to make the sample size equal (reduced data set)

| Sampling sites | Reduced data set | | | | Full data set | | | | | |
|----------------|------------------|------------|----------------|-------------------|---------------|--------------|------------|-------------------|-------------------|-----------|
| | No. of reads | Obs. OTUs* | Chao1 | Shannon† | Coverage‡ | No. of reads | Obs. OTUs* | Chao1 | Shannon† | Coverage‡ |
| S5 | | | | | | | | | | |
| S5N7 | 486 | 79 | 116 (95; 163) | 3.18 (3.04; 3.32) | 92 | 2089 | 216 | 576 (435; 810) | 3.16 (3.07; 3.24) | 93 |
| S5O8 | 486 | 32 | 52 (38; 98) | 1.66 (1.52; 1.80) | 97 | 2523 | 87 | 258 (165; 465) | 1.49 (1.42; 1.56) | 98 |
| S5M9 | 486 | 52 | 89 (67; 146) | 2.14 (1.99; 2.29) | 94 | 2838 | 177 | 381 (298; 520) | 2.20 (2.13; 2.28) | 96 |
| S5N9 | 486 | 92 | 142 (115; 199) | 3.31 (3.17; 3.46) | 91 | 1086 | 159 | 481 (334; 751) | 3.22 (3.11; 3.34) | 91 |
| S5J10 | 486 | 70 | 102 (82; 150) | 3.40 (3.29; 3.51) | 94 | 2354 | 195 | 598 (423; 908) | 3.41 (3.35; 3.48) | 95 |
| S1 | | | | | | | | | | |
| S1N7 | 486 | 138 | 266 (209; 368) | 3.86 (3.71; 4.01) | 83 | 2719 | 436 | 1329 (1051; 1773) | 3.88 (3.79; 3.96) | 90 |
| S1O8 | 486 | 131 | 289 (217; 422) | 3.93 (3.80; 4.05) | 83 | 3422 | 727 | 2396 (1994; 2925) | 4.59 (4.52; 4.66) | 85 |
| S1M9 | 486 | 127 | 270 (204; 392) | 3.82 (3.68; 3.95) | 84 | 2057 | 365 | 1068 (844; 1397) | 3.97 (3.89; 4.06) | 88 |
| S1N9 | 486 | 137 | 315 (237; 455) | 3.48 (3.30; 3.65) | 81 | 2021 | 392 | 1040 (848; 1313) | 3.53 (3.41; 3.64) | 86 |
| S1J10 | 486 | 181 | 342 (277; 449) | 4.48 (4.35; 4.61) | 78 | 4573 | 845 | 2158 (1859; 2544) | 4.81 (4.75; 4.88) | 88 |
| COWG | | | | | | | | | | |
| CGN7 | 486 | 121 | 193 (159; 256) | 3.75 (3.61; 3.89) | 87 | 688 | 194 | 367 (301; 474) | 4.06 (3.93; 4.19) | 82 |
| CGF8 | 486 | 105 | 201 (154; 295) | 3.16 (2.99; 3.33) | 87 | 2160 | 305 | 835 (659; 1099) | 3.21 (3.11; 3.30) | 90 |
| CGO8 | 486 | 98 | 206 (152; 312) | 2.90 (2.72; 3.07) | 87 | 2119 | 255 | 1025 (715; 1544) | 2.82 (2.72; 2.92) | 92 |
| CGM9 | 486 | 146 | 382 (279; 565) | 3.87 (3.72; 4.01) | 79 | 2756 | 511 | 1710 (1364; 2196) | 4.11 (4.02; 4.19) | 87 |
| CGN9 | 486 | 112 | 349 (232; 578) | 3.00 (2.81; 3.18) | 84 | 1317 | 201 | 624 (455; 907) | 2.80 (2.67; 2.93) | 89 |
| CGJ10 | 486 | 50 | 88 (65; 148) | 1.44 (1.26; 1.62) | 94 | 1827 | 136 | 472 (311; 781) | 1.51 (1.41; 1.62) | 95 |
| GAL | | | | | | | | | | |
| GLN7 | 486 | 91 | 144 (117; 200) | 2.28 (2.08; 2.49) | 90 | 1638 | 223 | 543 (423; 734) | 2.23 (2.10; 2.35) | 91 |
| GLF8 | 486 | 65 | 224 (135; 430) | 1.33 (1.14; 1.52) | 90 | 1679 | 177 | 665 (456; 1030) | 1.56 (1.44; 1.68) | 92 |
| GLO8 | 486 | 93 | 253 (172; 418) | 2.57 (2.39; 2.75) | 87 | 2093 | 236 | 614 (473; 839) | 2.64 (2.54; 2.74) | 93 |
| GLM9 | 486 | 107 | 239 (178; 354) | 2.36 (2.14; 2.57) | 85 | 2489 | 356 | 1070 (843; 1403) | 2.49 (2.37; 2.60) | 90 |
| GLN9 | 486 | 110 | 227 (171; 334) | 2.60 (2.39; 2.82) | 86 | 1682 | 285 | 715 (570; 934) | 2.65 (2.51; 2.78) | 88 |
| GLJ10 | 486 | 83 | 201 (140; 331) | 2.05 (1.85; 2.25) | 88 | 2246 | 271 | 862 (654; 1183) | 2.12 (2.01; 2.23) | 91 |
| CONF | | | | | | | | | | |
| CFN7 | 486 | 239 | 492 (400; 635) | 5.07 (4.97; 5.18) | 70 | 3256 | 933 | 1995 (1764; 2290) | 5.92 (5.87; 5.98) | 84 |
| CFF8 | 486 | 62 | 186 (115; 352) | 1.46 (1.27; 1.65) | 91 | 1768 | 240 | 972 (692; 1428) | 1.94 (1.82; 2.06) | 89 |
| CFO8 | 486 | 191 | 335 (279; 426) | 4.40 (4.25; 4.55) | 77 | 1731 | 548 | 1163 (1004; 1377) | 5.02 (4.93; 5.12) | 80 |
| CFM9 | 486 | 84 | 202 (142; 323) | 1.87 (1.66; 2.07) | 88 | 527 | 125 | 462 (306; 751) | 2.28 (2.07; 2.50) | 81 |
| CFN9 | 486 | 233 | 477 (390; 612) | 4.82 (4.69; 4.95) | 69 | 1994 | 699 | 1708 (1467; 2026) | 5.45 (5.37; 5.53) | 77 |
| CFJ10 | 486 | 180 | 495 (367; 711) | 4.06 (3.89; 4.23) | 74 | 5718 | 1343 | 3763 (3330; 4289) | 4.79 (4.71; 4.86) | 84 |

*OTUs were defined at 97% cutoff.

†Takes into account the number and evenness of species.

‡Coverage: sum of probabilities of observed classes calculated as $(1 - (n/N)^M)$, where n is the number of singleton sequences and N is the total number of sequences. Values in brackets are 95% confidence intervals.

between these two sites, COWG and GAL exhibited intermediate richness estimates (Chao1 = 236 ± 110 and 215 ± 39 , respectively). The lowest richness was observed in the tailing groundwater at S5 (Chao1 = 100 ± 33). Bacterial OTU diversity, estimated by the Shannon index, also differed significantly between sites (ANOVA, $F = 3.01$, $P = 0.039$), with values ranging from 1.33 to 5.07 (Table 1). In agreement with T-RFLP data analysis (Fig. 4a), the highest average diversity value was found at S1 ($H = 3.91 \pm 0.36$) and the lowest value at GAL ($H = 2.20 \pm 0.47$) (Fig. 4b). As predictable, average T-RF diversity is lower than OTU diversity (*c.* 50%); indeed, taxon-specific resolution of pyrosequencing is much higher than fingerprinting (Pilloni *et al.*, 2012). Again, no seasonal trend was observed. The same richness and diversity patterns were observed in both the full and resampled data sets, although the richness estimator and Shannon index were higher in the full data set, due to the larger number of sequences (data not shown).

Taxonomic assignment of bacterial pyrosequencing reads and T-RFs

At a confidence threshold of 80%, we were able to assign 56 426 of 63 442 qualified reads (that is, 89%) to a known phylum (Table S2) and 76% to a known order (Supporting Information, Fig. S1). Most of the unclassified reads (55% representing 9.6% to 37% of the qualified reads of each sample) were associated with samples collected at S1. Altogether, 23 bacterial phyla were recovered from our samples, with 4–8 different phyla found in samples collected at S5, 12–13 at S1, 9–14 at COWG, 10–13 at GAL, and 9–20 at CONF (Table S2). Most of the bacterial sequences (86%) belonged to phyla that are most often encountered in acid mine drainages worldwide (*Proteobacteria*, *Actinobacteria*, *Firmicutes*, *Acidobacteria*, *Bacteroidetes*, and *Nitrospirae*). In addition, microorganisms representing 0.5% of the total sequences were related to CARN1, 'Candidatus Fodinabacter communicans'. *Proteobacteria* was the most abundant phylum in all the samples, accounting for 69.6% of all sequences retrieved. This phylum was represented by bacteria belonging to the *Alphaproteobacteria*, *Betaproteobacteria*, *Gammaproteobacteria*, *Deltaproteobacteria*, and *Epsilonproteobacteria*. The most abundant classes in nearly all the samples were *Betaproteobacteria* and *Gammaproteobacteria* (average values of 63.4% and 30.4% of the pyrosequencing reads, respectively). There were dominated by *Gallionellales* and *Acidithiobacillales*, respectively (Fig. S1a). Three other phyla, *Actinobacteria* represented mainly by *Acidimicrobiales* and *Actinomycetales*, *Firmicutes* (principally *Clostridiales* and *Bacilliales*), and *Acidobacteria*, were also abundant but their proportion varied depending on the

sample analyzed (Fig. S1b–d). Most of the sequences associated with *Acidobacteria* could not be classified to the order level except for *Acidobacteriales* and *Holophagales*. As can be seen in Fig. S2, a relatively small number of OTUs dominated at all sites (> 1% in total abundance per sample). The most abundant OTUs were phylogenetically related to *Gallionella ferruginea* (*Gallionellales*), *Acidithiobacillus ferrooxidans* (*Acidithiobacillales*), and *Thiobacillus* sp. (*Hydrogenophilales*), collectively accounting for 41% of all the sequences.

When possible, T-RFs were assigned to a taxon or a group of taxon by *in silico* restriction of 16S rRNA gene sequences. Among the five dominant T-RFs (91, 100, 125, 150, and 162 bp in size), T-RFs 91 and 100 could not be assigned to any specific phylogenetic group. *Armatimonadetes* gp4 and *Chlorobi* were represented by T-RF 125, and T-RF 150 was mainly related to *Deinococcus-Thermus*, *Spirochaetes*, and *Actinobacteria*. While T-RF 162 was mainly related to *A. ferrooxidans*, it could be assigned to other proteobacterial phylotype detected in this AMD.

Spatial and temporal variations in bacterial community structure

Spatiotemporal dynamics of bacterial populations were identified by T-RFLP analysis and 454-pyrosequencing of 16S rRNA genes (Fig. S3).

Although samples formed overlapping clusters on the nMDS plot of the T-RFLP profiles, weak but significantly different bacterial communities at the five sites were revealed (ANOSIM Global $R = 0.2819$, $P < 0.001$). S5 differed significantly from the other sites, with some overlapping communities (pairwise tests: r -values ranging from 0.45 to 0.61, $P < 0.05$). These results highlighted changes in the structure of the bacterial communities between the tailing groundwater (S5) and the water in Reigous Creek. The high dissimilarity observed within each site revealed variations in community structure over time, especially at S5, GAL, and CONF (Fig. S3a). These variations may have masked a spatial pattern.

nMDS analyses of 454-pyrosequencing data also showed that the composition of the bacterial communities differed significantly along the spatial gradient from the sterile (S5) to the confluence (CONF) (Fig. S3b). Furthermore, an ANOSIM test corroborated the nMDS plot data, revealing significantly different bacterial compositions in water as a function of the spatial location (Global $R = 0.6192$, $P < 0.001$), except at GAL and COWG which did not differ significantly (ANOSIM pairwise comparison $r = 0.206$, $P = 0.37$). Higher temporal variation at CONF was highlighted by the large cluster on the nMDS plot. The marked temporal variations in the bacterial community at S5 and CONF highlighted by the two data sets

may be linked to the stronger seasonal fluctuation of some physicochemical parameters at these sites, particularly temperature at CONF (Fig. 2a) and pH, Eh, sulfate, Fe, and As at S5 (Fig. 2b, d, f, g and h).

We investigated the four most abundant phyla to get a global view of the variations of bacterial communities along the creek (Fig. S1). *Proteobacteria* distribution varied between samples, with a relative predominance of *Acidithiobacillales* in samples from S5 and S1, followed by the dominance of *Gallionellales* in the majority of other samples (Fig. S1a). A general increase in *Betaproteobacteria* was observed in the downstream direction of Reigous Creek. Among *Actinobacteria*, the *Acidimicrobiales* were present in all samples except those collected at CONF where almost equal proportions of *Acidimicrobiales* and *Actinomycetales* were retrieved (Fig. S1b). The *Firmicutes* phylum revealed the dominance of *Clostridiales* in samples from S5, whereas *Bacillales* were dominant at the other sites, again except CONF. Different orders were dominant at CONF over time, including *Lactobacillales* and *Selenomonadales* (Fig. S1c). No *Acidobacteria* were retrieved at S5 in October 2008 (Fig. S1d).

We also assessed the dynamics of the dominant genera (> 5% in total abundance per sample) (Fig. 5). The relative abundance of genera at each site varied over the sampling period. While the relative abundance of *Gallionella* was almost constant in COWG and GAL, there was an important temporal variation in the other sites. This was evident in S5 where this genus was extinct and re-thrived over time (Fig. 5a). Although *Gallionella* was present in almost all sites for a sampling date, there was no clear relationship between sites. Indeed, GAL and COWG exhibited a relatively high proportion of *Gallionella* at almost all the sampling dates (as much as 85% of all pyrosequencing reads at GAL) without any link with the upstream sites S5 and S1 (Fig. 5b). In contrast, *Acidithiobacillus* represented a minor fraction of the bacterial community, except at S5 where this OTU was dominant (24–72% of the pyrosequencing reads). The relative abundance of *Acidithiobacillus* showed a decreasing trend along the continuum for each sampling date (Fig. 5b). This genus also exhibited an increase in October 2008 and March 2009 for S5 and S1 (Fig. 5a). Members of the *Thiobacillus* genus showed higher proportion in COWG at each sampling date independently of the other sites (Fig. 5b). The temporal variation of this genus was minor except an important increase at GAL in October 2008 (Fig. 5a).

T-RFLP profiles from the 30 samples were investigated to assess the dynamics of T-RFs (Fig. 3). Among the five dominant T-RFs, T-RFs 91 and 100 represented a large proportion of the T-RFLP profiles in almost all samples except those from S5. T-RF 125 related to *Armatimonadetes* gp4 and *Chlorobi* was more abundant in samples from

S5 and S1 than in downstream samples. T-RF 150 (*Deinococcus-Thermus*, *Spirochaetes*, and *Actinobacteria* – related) was the most abundant phylotype in the tailings groundwater (S5), accounting for up to 70% of the T-RFLP profiles. It was relatively abundant along the creek especially at GAL. T-RF 162 (related to *A. ferrooxidans* but also to other proteobacterial phylotype) was not dominant at S5 and represented a minor fraction of the bacterial community along the creek.

Linking bacterial community structure to environmental variables

Canonical correspondence analysis (CCA) was performed to elucidate the main relationships between physicochemical variables and bacterial community structure and composition (Fig. 6). Samples were plotted in different areas of the diagram depending on their environmental characteristics. The resolution of 454-pyrosequencing allowed to account for more variation than T-RFLP (36.4% and 20.5%, respectively) in the species–environment relationship across the first two canonical axes. CCA axis 1 based on T-RFLP data only separated the samples into two clusters, one containing the tailings site (S5) and the other grouping the sites along the creek (S1, COWG, GAL, and CONF), with sulfate, DO, and pH being the strongest determinants of bacterial community structure (Fig. 6a). In contrast, a higher resolution was observed with CCA axis 1 based on 454-pyrosequencing data, which was most closely correlated with iron, arsenic, and conductivity and separated the sampling sites into three clusters (CONF, GAL+COWG, and S1+S5) as a function of the pollution gradient (Fig. 6b). The upstream site S5 was highly polluted and little oxygenated, whereas the downstream site CONF was less polluted and characterized by a higher redox potential (Eh). CCA axis 2 separated samples according to water temperature. After the Monte Carlo permutation test, the environmental variables significantly correlated with the canonical axes based on 454-pyrosequencing data were arsenic (F -ratio = 1.9, P = 0.01), temperature (F -ratio = 1.4, P = 0.01), and sulfate (F -ratio = 1.4, P = 0.01). The differences between the two data sets were probably due to the power of 454-pyrosequencing over T-RFLP for taxon resolution. Focusing on 454-pyrosequencing data, the influence of environmental variables on dominant OTUs (> 5% of total abundance per sample) was also investigated (Fig. S4). Nineteen of the 23 dominant OTUs showed a strong correlation with the physicochemical parameters. Five OTUs (15, 16, 28, 32, and 52) were strongly correlated with elevated DO and Eh and 14 with high temperatures and high concentrations of As, Fe, and sulfate. As indicated by the position of OTUs 1, 3, 4, and 11 on the graph, near the origin of the

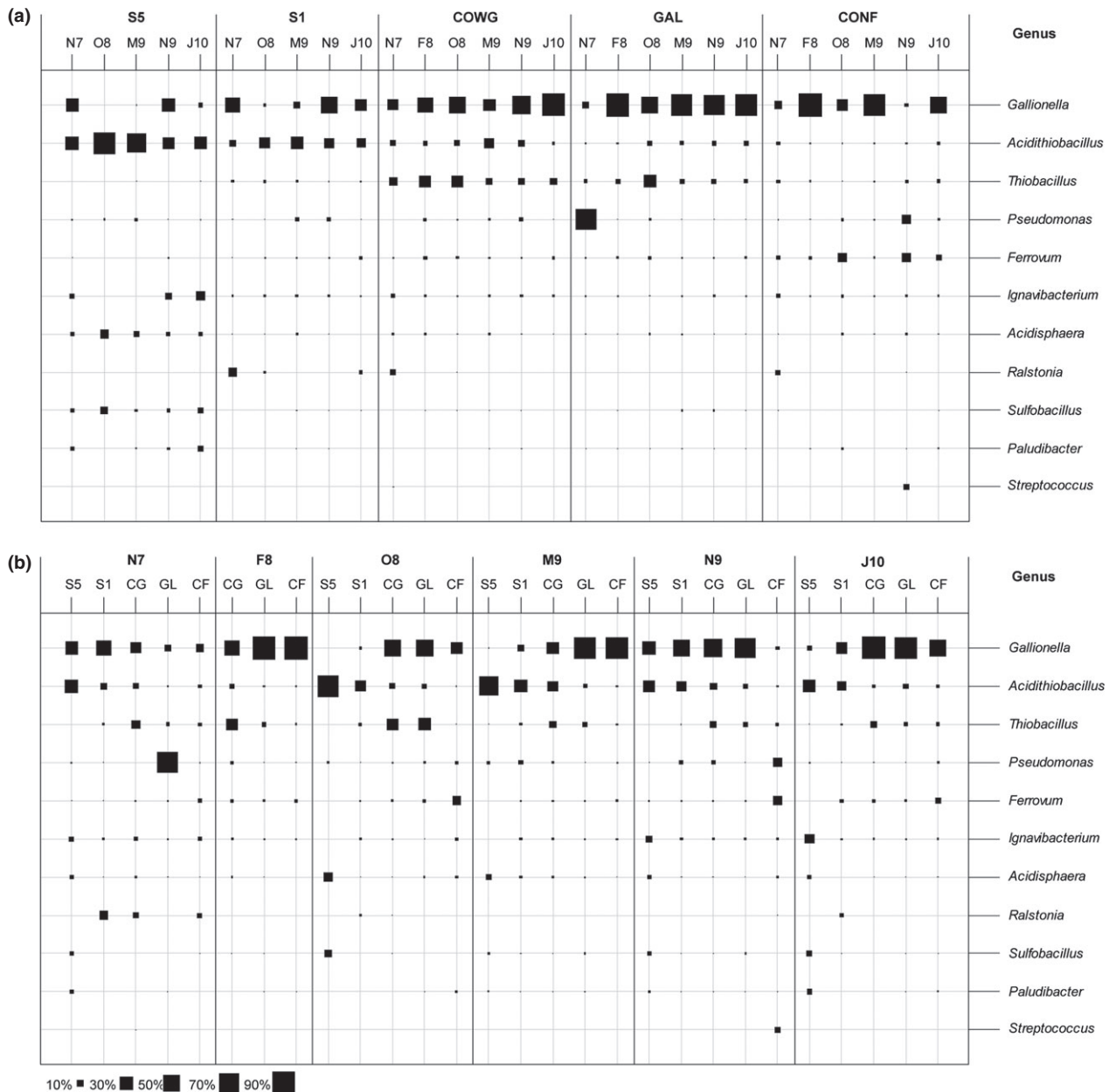


Fig. 5. Relative abundance of the dominant genera (> 5% in total abundance per sample) presented by (a) sampling sites and (b) sampling date. N7: November 2007; O8: October 2008; M9: March 2009; N9: November 2009; J10: January 2010.

axes, none of the environmental variables measured in the study could explain their distribution and thus their niche. At the least polluted site (CONF), *Gallionella*, *Ferrovum*, and *Acidiferrobacter* were the main genera detected, whereas, at the most polluted sites (S5 and S1), a higher number of genera were codominant (*Acidithiobacillus*, *Ignavibacterium*, *Ralstonia*, *Leptospirillum*, *Gallionella*, *Ferrovum*, etc.).

Discussion

This study combined a classical fingerprinting method (T-RFLP) and a high-throughput barcoded pyrosequencing of 16S rRNA genes to investigate the diversity, spatial distribution, and seasonal variation of bacterial communities in Carnoulès AMD (France), which is heavily contaminated with As.

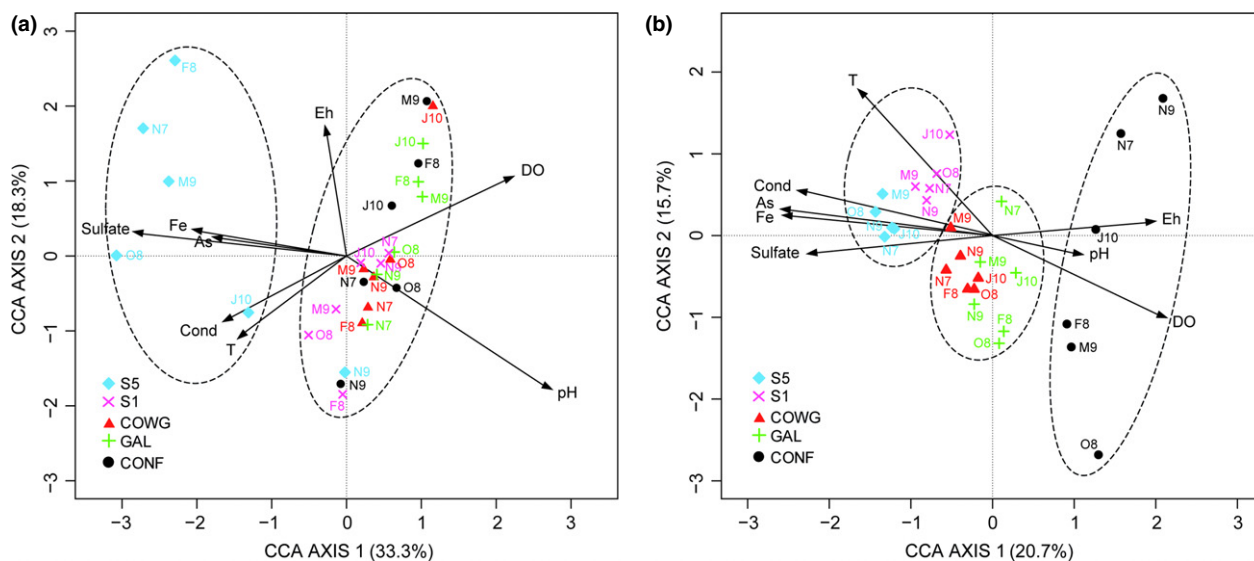


Fig. 6. Canonical correspondence analysis (CCA) correlating the bacterial community structure at each sampling site with arsenic (As), iron (Fe), conductivity (Cond), temperature (T), dissolved oxygen (DO), redox (Eh), pH and sulfate. The bacterial community structures correspond to OTU abundances from (a) T-RFLP data and (b) pyrosequencing data. The main clusters are highlighted.

Spatial and temporal variations in the environmental data set and in the bacterial community

Monitoring the physicochemical parameters of Reigous Creek confirmed previous results (Casiot *et al.*, 2003a; Egal *et al.*, 2010), showing a significant decrease in concentrations of dissolved As, Fe, and sulfate with increasing distance from the source: 72% of sulfate, 96% of iron, and 99% of arsenic had been removed by the time Reigous Creek flowed into the River Amous (Table S1). In addition, the concentrations of As and Fe in the water from the tailings stock were much lower in 2007–2010 (average values of $440 \pm 184 \text{ mg L}^{-1}$ and $4474 \pm 2855 \text{ mg L}^{-1}$, respectively), than those measured in 2001 (up to 10 g L^{-1} for As and around 20 g L^{-1} for Fe, Casiot *et al.*, 2003b), although these concentrations are still very high compared to other AMDs.

Both molecular methods highlighted a higher bacterial diversity than expected in this extreme habitat. T-RFLP profiles showed for the five sites a total of 43 T-RFs ranging from 2 to 17 T-RFs per sampling site (Fig. 3). For pyrosequencing data, a total of 63 442 reads led to the identification of 6613 OTUs, including 4510 singletons representing 68% of the total number of OTUs. As expected, a larger number of phylotypes were identified using the pyrosequencing method leading to a significant increase in resolution. Average Good's coverage was over 89%, suggesting that the 16S rRNA gene sequences into each sample represented the majority of the bacterial

phylotypes present. Nonetheless, additional sequencing effort would be required to exhaustively characterize the bacterial community, particularly for samples from the least polluted site CONF, as shown by the lower coverage values and the lack of asymptote in the rarefaction curves (data not shown).

nMDS analyses revealed significant differences in the composition of the bacterial communities in the five sites along the AMD (Fig. S3). However, different clustering patterns were obtained based on T-RFLP or pyrosequencing data. With pyrosequencing, individual sequences can be classified at the genus level. In contrast, one T-RF can correspond to several different bacterial phylotypes (belonging to different genera or even different higher taxonomic levels). Such differences in the resolution of the two methods may explain the differences obtained in the cluster analyses (Hwang *et al.*, 2012). Nevertheless, changes in bacterial community structure between the tailings groundwater (S5) and the Reigous Creek were revealed by the two sets of data, reflecting important differences in ecological conditions between the two habitats. According to both methods, S1 was the most diverse bacterial community, while GAL was the least diverse. Therefore, bacterial diversity varied independently of the sampling site, suggesting that globally upstream communities do not influence downstream communities. The apparent minimum effect of immigration suggests that species-sorting processes best describe bacterial community structure in these connected environments, with local environmental factors driving the composition of the

bacterial community at each site. Temporal variations of the bacterial communities could also be observed at each site although no particular trend could be identified. However, the important temporal variation of the bacterial community observed at S5 and CONF may be due to a higher seasonal fluctuation in physicochemical parameters at these two sites, especially temperature at CONF, and pH, Eh, sulfate, Fe, and As at S5.

However, our fine-scale investigation at the genus level of the bacterial communities along the Reigous Creek over time provided some important data and allowed to establish some hypothesis about community composition. Indeed, *Gallionella* in contrast to *Acidithiobacillus* do not seem to benefit from the seed bank provided by the most upstream sites (S5 and S1). This suggests that *Gallionella*, under a process that still need to be elucidated, extinct/re-thrived at each site over time. In contrast, *Acidithiobacillus* that is preferentially encountered upstream of the Reigous Creek or *Thiobacillus* that thrived at COWG could be found at these sites, under conditions that reflect their preferential habitats. The presence of these organisms downstream of the sites would instead reflect dispersal from upstream sites.

Physicochemical parameters shape the composition of the bacterial community

This work highlighted a spatial gradient of physicochemical conditions linked to a significant shift in bacterial community composition along the continuum. Indeed, canonical correspondence analysis of the whole pyrosequencing data set indicated that arsenic, temperature, and sulfate were the factors that most influence the composition of the bacterial communities (Fig. 6b). The level of pollution affects also some dominant bacterial populations (> 5% of relative abundance). *Gallionella*, *Ferroplasma*, and *Acidiferrobacter* were the dominant genera in water sampled at the least polluted site (CONF) and were correlated with high DO and Eh, whereas in water from the most polluted sites (S5 and S1), a larger number of dominant genera were detected (*Acidithiobacillus*, *Ignavibacterium*, *Ralstonia*, *Leptospirillum*, *Gallionella*, and *Ferroplasma*) whose relative abundance was correlated with higher temperature and high concentrations of As, Fe, and sulfate (Fig. S4). Thus, different members of a given genus such as *Gallionella* (OTUs 19, 15 and 1) or *Ferroplasma* (OTUs 25 and 28) were correlated with different environmental parameters, suggesting that these OTUs correspond to bacterial phylotypes with some specificity explaining these different behaviors. Furthermore, the high abundance of *Gallionella*-related sequences in these acidic ecosystems characterized by contrasted levels of pollution is consistent with results of a previous study suggesting that

Gallionella-like organisms may be more tolerant to acid and metal than currently thought (Fabisch *et al.*, 2013). In accordance with our results, temperature has been previously suggested as a primary factor controlling the structure and dynamics of microbial communities in AMD (Edwards *et al.*, 1999) and in various natural environments like hot springs (Ward *et al.*, 1998; Miller *et al.*, 2009) or marine environments (Fuhrman *et al.*, 2008). Nevertheless, sulfate and arsenic concentrations have not previously been shown to be significantly correlated with bacterial diversity in AMDs. Earlier studies identified different environmental predictors of microbial populations in AMD including conductivity and rainfall (Edwards *et al.*, 1999), pH (Lear *et al.*, 2009), oxygen gradient (González-Toril *et al.*, 2011), and season (Streten-Joyce *et al.*, 2013), which may result in site-specific physicochemical and geochemical characteristics (Kuang *et al.*, 2012). Furthermore, while many studies highlighted pH as the most important factor structuring AMD communities (Kuang *et al.*, 2012; Chen *et al.*, 2013), our study produced no evidence of the influence of this parameter, probably due to the limited variation in pH among our samples (average values of 2.5 ± 0.8 – 3.2 ± 0.3).

Composition of the bacterial communities

In this study, we were able to identify a wider phylogenetic range of taxa than in any previous clone library-based diversity survey of the Carnoulès AMD, including sequences of several previously undetected taxa. These new taxa include members of the *Bacteroidetes*, *Chlorobi*, *Chloroflexi*, *Elusimicrobia*, *Chlamydiae*, *Cyanobacteria*, *Dinococcus-Thermus*, *Spirochaetes*, *Fibrobacteres*, *Fusobacteria*, *Gemmatimonadetes*, *Plantctomycetes*, *Verrucomicrobia*, and of the uncultured ODI-PO11-TM7 clade. The majority of phyla that were not previously detected on clone libraries accounted only for < 1% of the pyrosequencing data, explaining why they were missed with the clone library approach. The high rate of low-abundance populations (68% of singletons) increased the phylogenetic bacterial diversity. However, despite the preponderance of this rare biosphere in most studies, its ecological and functional roles remain largely unexplored (Galand *et al.*, 2009). Recent studies indicated that such organisms may be at a dormant or a spore stage, but in favorable conditions they may become active and even dominant (Delavat *et al.*, 2012). Thus, these taxa may play an important role in extreme habitats like AMD, buffering the effects of important environmental shifts (Sogin *et al.*, 2006; Monchy *et al.*, 2011). However, further investigations will be needed to determine whether they play a role in this ecosystem and/or whether they reflect allochthonous input from surrounding environments. Moreover, the high-

throughput sequencing also questions the accuracy of OTU richness estimates, as sequencing errors and inadequate clustering algorithms can lead to overestimates of community richness (Huse *et al.*, 2010). The majority of the most abundant taxa detected in this study were related to orders commonly encountered in AMD, most of which are known to be involved in Fe, As, and S cycles: namely *Gallionellales* (*Betaproteobacteria*), *Acidithiobacilliales* (*Gammaproteobacteria*), *Acidimicrobiales* (*Actinobacteria*), *Hydrogenophilales* (*Betaproteobacteria*), *Burkholderiales* (*Betaproteobacteria*), *Nitrospirales* (*Nitrospirae*), *Desulfomonadales* (*Deltaproteobacteria*), and *Desulfobacterales* (*Deltaproteobacteria*). The ecological role of previously detected taxa has been widely characterized (Bruneel *et al.*, 2005, 2006, 2011; Bertin *et al.*, 2011) in this ecosystem. A relatively small number of OTUs dominated at each sampling site (Fig. S2) and the majority of them were phylogenetically related to taxa previously found in AMD (*Gallionella ferruginea*, *Acidithiobacillus ferrooxidans*, and *Thiobacillus* sp.), as well in Carnoulès revealing their persistence in such ecosystems (Baker & Banfield, 2003; Bruneel *et al.*, 2006, 2011; Hallberg *et al.*, 2006; Heinzl *et al.*, 2009; Hallberg, 2010). These three genera varied in their relative abundance over the sampling period. *Gallionella* was present in high proportions in almost all samples, mainly at GAL and COWG. In contrast, except at S5 where this genus was dominant, *Acidithiobacillus* accounted for a minor fraction of the bacterial community (Fig. 5). Furthermore, our study confirmed the presence of relatives of a novel bacterial phylum, 'Candidatus Fodinabacter communificans' detected by a recent metagenomic investigation of Carnoulès AMD and prominent in the active COWG community (Bertin *et al.*, 2011; Fahy *et al.*, unpublished data). The relatively high number of unclassified bacteria per sample (0.3–37%) supports the fact that many bacteria remain to be cultured. These results thus corroborate the main observations made in previous studies, except for the predominance of organisms related to sulfate reducing bacteria (SRB) identified in water from the tailings by Bruneel *et al.* (2005) using a cloning–sequencing approach. Instead, our results revealed the dominance of *Acidithiobacillales* over SRB in these samples. The very low proportion of SRB populations in our study (on average 0.1% of total abundance per sample) could be partly due to differences in the physicochemical variables of the water, to the choice of a stringent similarity cutoff but also to the different primers used for PCR amplification. Furthermore, relatives of *Thiomonas* belonging to the *Burkholderiales* order were retrieved and accounted for < 1% of the total sequences (Fig. S1a). Despite their low abundance, several strains of *Thiomonas* sp. have been previously isolated and shown to be active in the

oxidation of As (Bruneel *et al.* 2003). A metaproteomic approach also showed that *Gallionella*, *Thiomonas*, and *A. ferrooxidans* actively express proteins *in situ*, thus probably playing a functional role in this AMD (Bruneel *et al.*, 2011). These populations could play an important role in the efficient remediation process observed along this creek by favoring the oxidation of Fe(II) and the co-precipitation of As (Casiot *et al.*, 2003a; Bruneel *et al.*, 2011).

This work has increased our knowledge of bacterial diversity and dynamics in acid mine drainage. Bacterial diversity in Carnoulès AMD was revealed to be much higher than previously evidenced using clone library techniques (Bruneel *et al.*, 2011), as suggested by culture-dependent methods (Delavat *et al.*, 2012). Our study revealed complex patterns of spatial and temporal variations in bacterial community composition, suggesting that community composition reflects changes in physicochemical conditions. This investigation provided a first step to the study of spatial and temporal structure of bacterial communities and the factors that control it. To improve our understanding of the functioning of this ecosystem, future efforts should be oriented toward active communities and how they fluctuate in response to environmental changes. Such knowledge will help to determine their roles in the functioning of the AMD ecosystems and explain important assembly processes in microbial ecology.

Acknowledgements

This study was financed by the FRB (*Fondation pour la recherche sur la Biodiversité*) program blanc AAP-IN-2009-039, the « *Observatoire de Recherche Méditerranéen de l'Environnement* » (OSU-OREME). A.V. was supported by a grant from the French Ministry of Education and Research and F.J. by a grant from the Direction Générale de l'Armement (DGA). This work was performed within the framework of the *Groupement de recherche: Métabolisme de l'Arsenic chez les Microorganismes* (GDR2909-CNRS).

References

- Baker BJ & Banfield JF (2003) Microbial communities in acid mine drainage. *FEMS Microbiol Ecol* **44**: 139–152.
- Bertin PN, Heinrich-Salmeron A, Pelletier E *et al.* (2011) Metabolic diversity among main microorganisms inside an arsenic-rich ecosystem revealed by meta- and proteo-genomics. *ISME J* **5**: 1735–1747.
- Bruneel O, Personné JC, Casiot C, Leblanc M, Elbaz-Poulichet F, Mahler BJ, Le Flèche A & Grimont P AD (2003) Mediation of arsenic oxidation by *Thiomonas* sp. in acid-mine drainage (Carnoulès, France). *J Appl Microbiol* **95**: 492–499.

- Bruneel O, Duran R, Koffi K, Casiot C, Fourçans A, Elbaz-Poulichet F & Personné JC (2005) Microbial diversity in a pyrite-rich tailings impoundment (Carnoulès, France). *Geomicrobiol J* **22**: 249–257.
- Bruneel O, Duran R, Casiot C, Elbaz-Poulichet F & Personné JC (2006) Diversity of microorganisms in Fe-As-rich acid mine drainage waters of Carnoulès, France. *Appl Environ Microbiol* **72**: 551–556.
- Bruneel O, Volant A, Gallien S *et al.* (2011) Characterization of the active bacterial community involved in natural attenuation processes in arsenic-rich Creek sediments. *Microb Ecol* **61**: 793–810.
- Casiot C, Morin G, Juillot F, Bruneel O, Personné JC, Leblanc M, Duquesne K, Bonnefoy V & Elbaz-Poulichet F (2003a) Bacterial immobilization and oxidation of arsenic in acid mine drainage (Carnoulès creek, France). *Water Res* **37**: 2929–2936.
- Casiot C, Leblanc M, Bruneel O, Personné JC, Koffi K & Elbaz-Poulichet F (2003b) Geochemical processes controlling the formation of As-rich waters within a tailings impoundment (Carnoulès, France). *Aquat Geochem* **9**: 273–290.
- Chen LX, Li JT, Chen YT, Huang LN, Hua ZS, Hu M & Shu WS (2013) Shifts in microbial community composition and function in the acidification of a lead/zinc mine tailings. *Environ Microbiol* **15**: 2431–2444.
- Delavat F, Lett MC & Lievreumont D (2012) Novel and unexpected bacterial diversity in an arsenic-rich ecosystem revealed by culture-dependent approaches. *Biol Direct* **7**: 28.
- Edwards KJ, Gihring TM & Banfield JF (1999) Seasonal variations in microbial populations and environmental conditions in an extreme acid mine drainage environment. *Appl Environ Microbiol* **65**: 3627–3632.
- Edwards KJ, Bond PL, Druschel GK, McGuire MM, Hamers RJ & Banfield JF (2000) Geochemical and biological aspects of sulfide mineral dissolution: lessons from Iron Mountain, California. *Chem Geol* **169**: 383–397.
- Egal M, Casiot C, Morin G, Elbaz-Poulichet F, Cordier MA & Bruneel O (2010) An updated insight into the natural attenuation of As concentrations in Reigous Creek (southern France). *Appl Geochem* **25**: 1949–1957.
- Fabisch M, Beulig F, Akob DM & Küsel K (2013) Surprising abundance of *Gallionella*-related iron oxidizers in creek sediments at pH 4.4 or at high heavy metal concentrations. *Front Microbiol* **4**: 390.
- Fuhrman JA, Steele JA, Hewson I, Schwalbach MS, Brown MV, Green JL & Brown JH (2008) A latitudinal diversity gradient in planktonic marine bacteria. *P Natl Acad Sci USA* **105**: 7774–7778.
- Galand PE, Casamayor EO, Kirchman DL & Lovejoy C (2009) Ecology of the rare microbial biosphere of the Arctic Ocean. *P Natl Acad Sci USA* **106**: 22427–22432.
- González-Toril E, Aguilera A, Souza-Egipsy V, López Pamo E, Sánchez España J & Amils R (2011) Geomicrobiology of La Zarza-Perrunal acid mine effluent (Iberian Pyritic Belt, Spain). *Appl Environ Microbiol* **77**: 2685–2694.
- Hallberg KB (2010) New perspectives in acid mine drainage microbiology. *Hydrometallurgy* **104**: 448–453.
- Hallberg KB, Coupland K, Kimura S & Johnson DB (2006) Macroscopic streamer growths in acidic, metal-rich mine waters in North Wales consist of novel and remarkably simple bacterial communities. *Appl Environ Microbiol* **72**: 2022–2030.
- Heinzel E, Hedrich S, Janneck E, Glombitza F, Seifert J & Schlömann M (2009) Bacterial diversity in a mine water treatment plant. *Appl Environ Microbiol* **75**: 858–861.
- Huse SM, Welch DM, Morrison HG & Sogin ML (2010) Ironing out the wrinkles in the rare biosphere through improved OTU clustering. *Environ Microbiol* **12**: 1889–1898.
- Hwang C, Ling F, Andersen GL, LeChevallier MW & Liu WT (2012) Microbial community dynamics of an urban drinking water distribution system subjected to phases of chloramination and chlorination treatments. *Appl Environ Microbiol* **78**: 7856–7865.
- Johnson DB (2012) Geomicrobiology of extremely acidic subsurface environments. *FEMS Microbiol Ecol* **81**: 2–12.
- Junier P, Junier T & Witzel KP (2008) TRiFLE, a program for *in silico* terminal restriction fragment length polymorphism analysis with user-defined sequence sets. *Appl Environ Microbiol* **74**: 6452–6456.
- Koffi K, Leblanc M, Jourde H, Casiot C, Pistre S, Gouze P & Elbaz-Poulichet F (2003) Reverse oxidation zoning in mine tailings generating arsenic-rich acidic waters (Carnoulès, France). *Mine Water Environ* **22**: 7–14.
- Kuang JL, Huang LN, Chen LX, Hua ZS, Li SJ, Hu M, Li JT & Shu WS (2012) Contemporary environmental variation determines microbial diversity patterns in acid mine drainage. *ISME J* **7**: 1038–1050.
- Lane DJ (1991) 16S/23S rRNA sequencing. *Nucleic Acid Techniques in Bacterial Systematics* (Stackebrandt E & Goodfellow M, eds), pp. 115–175. Wiley, New York.
- Lear G, Niyogi D, Harding J, Dong Y & Lewis G (2009) Biofilm bacterial community structure in streams affected by acid mine drainage. *Appl Environ Microbiol* **75**: 3455–3460.
- Miller SR, Strong AL, Jones KL & Ungerer MC (2009) Barcoded pyrosequencing reveals shared bacterial community properties along the temperature gradients of two alkaline hot springs in Yellowstone National Park. *Appl Environ Microbiol* **75**: 4565–4572.
- Monchy S, Sancier G, Jobard M *et al.* (2011) Exploring and quantifying fungal diversity in freshwater lake ecosystems using rDNA cloning/sequencing and SSU tag pyrosequencing. *Environ Microbiol* **13**: 1433–1453.
- Morin G & Calas G (2006) Arsenic in soils, mine tailings, and former industrial sites. *Elements* **2**: 97–101.
- Morin G, Juillot F, Casiot C, Bruneel O, Personné JC, Elbaz-Poulichet F, Leblanc M, Ildefonse P & Calas G (2003) Bacterial formation of tooeite 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.

- Muyzer G & Ramsing NB (1995) Molecular methods to study the organization of microbial communities. *Water Sci Technol* **32**: 1–9.
- Needleman SB & Wunsch CD (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol* **48**: 443–453.
- Pilloni G, Granitsiotis MS, Engel M & Lueders T (2012) Testing the limits of 454 pyrotag sequencing: reproducibility, quantitative assessment and comparison to T-Rflp fingerprinting of aquifer microbes. *PLoS ONE* **7**: e40467.
- R Development Core Team (2012) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org>.
- Rodier J, Broutin JP, Chambon P, Champsaur H & Rodi L (1996) *L'Analyse des Eaux*. Dunod, Paris, p. 1383.
- Samanta G & Clifford DA (2005) Preservation of inorganic arsenic species in groundwater. *Environ Sci Technol* **39**: 8877–8882.
- Schloss PD, Westcott SL, Ryabin T *et al.* (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* **75**: 7537–7541.
- Smith CJ, Danilowicz BS, Clear AK, Costello FJ, Wilson B & Meijer WG (2005) T-Align, a web-based tool for comparison of multiple terminal restriction fragment length polymorphism profiles. *FEMS Microbiol Ecol* **54**: 375–380.
- Sogin ML, Morrison HG, Huber JA, Mark Welch D, Huse SM, Neal PR, Arrieta JM & Herndl GJ (2006) Microbial diversity in the deep sea and the underexplored “rare biosphere”. *P Natl Acad Sci USA* **103**: 12115–12120.
- Streten-Joyce C, Manning J, Gibb KS, Neilan BA & Parry DL (2013) The chemical composition and bacteria communities in acid and metalliferous drainage from the wet-dry tropics are dependent on season. *Sci Total Environ* **443**: 65–79.
- Ueno S, Le Provost G, Leger V *et al.* (2010) Bioinformatic analysis of ESTs collected by Sanger and pyrosequencing methods for a keystone forest tree species: oak. *BMC Genomics* **11**: 650.
- Ward DM, Ferris MJ, Nold SC & Bateson MM (1998) A natural view of microbial biodiversity within hot spring cyanobacterial mat communities. *Microbiol Mol Biol Rev* **62**: 1353–1370.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Fig. S1. Composition of different phyla based on classification of 16S rRNA gene sequences of bacteria from each sample using RDP Classifier: (a) *Proteobacteria* orders, (b) *Actinobacteria* orders, (c) *Firmicutes* orders, and (d) *Acidobacteria* orders.

Fig. S2. Histogram of the relative abundance of dominant OTUs at the Carnoulès sampling sites (*G. ferruginea* subsp. *capsiferriformans* ES-2: NC_014394; *A. ferrooxidans* strain HL1: JF815535; *Thiobacillus* sp. ML2-16: DQ145970; *Pseudomonas migulae*: AY605698; *A. ferrivorans* SS3: NR_074660; *Actinobacterium* BGR 105: GU168008; *Acidobacteriaceae bacterium* CH1: DQ355184; *Ferrimicrobium* sp. Py-F2: KC208496; *Metallibacterium* sp. ×11: HE858262; *Alicyclobacillaceae bacterium* Feo-D4-16-CH: FN870323; *Acidisphaera* sp. nju-AMDS1: FJ915153; *Betaproteobacterium* OYT1: AB720115).

Fig. S3. Nonmetric multidimensional scaling analysis of the composition of the bacterial community estimated by (a) T-RFLP and (b) 454 pyrosequencing based on 16S rRNA genes.

Fig. S4. (a) Ordination plot of CCA based on pyrosequencing data showing OTUs with relative abundance >5%. (b) Abundant OTUs and their correlation with environmental variables and phylogenetic affiliation determined by BLAST search.

Table S1. Physicochemical characteristics of the water at each sampling site and sampling date.

Table S2. Relative abundance (in %) of total sequences of bacterial 16S rRNA genes from each sample assigned to different phyla.

Annexe B

Script commenté du traitement des séquences d'ARNr 16S

#Commandes mothur utilisée pour le nettoyage des séquences et le calcul de matrice.

#1) extract the fasta, qual, and flow data from the sff file:

```
mothur > sffinfo(sff="fichier".sff, flow=T)
```

#2) Séparation des flowgrams en fonction des barcodes et primers
(options minflows et maxflows spécifique pour Titanium)

```
mothur > trim.flows(flow="fichier".flow, oligos="fichier".oligos, pdiffs=2,  
bdiffs=1, processors=2,minflows=360, maxflows=720)
```

#3) PyroNoise implementation (très gourmand en ressources PC)

```
mothur > shhh.flows(file="fichier".flow.files, processors=4)
```

#4) On enlève des séquences les barcodes et les primers, puis on filtre sur la taille (400 pb), les homopolymères et les erreurs dans le code barre et les primers.(flip => calcul du complément inverse)

```
mothur > trim.seqs(fasta="fichier".fasta, name="fichier".shhh.names,  
oligos="fichier".oligos, pdiffs=2, bdiffs=1, maxhomop=8, minlength=400, flip=T,  
processors=2)
```

#5) alignement (premier alignement nécessaire pour rechercher les chimères et contaminants), alignement de référence silva.

```
mothur > align.seqs(fasta="fichier".fasta, reference=silva.bacteria.fasta,  
processors=2)
```

#6) preclusterisation, éliminer les erreurs de séquençage.

```
mothur > pre.cluster(fasta="fichier".fasta, name="fichier".names,
```

```
group="fichier".groups, diffs=4)
```

#7) Recherche et suppression de Chimères:

```
mothur > chimera.uchime(fasta="fichier".fasta, name="fichier".names,  
group="fichier".groups, processors=2)
```

```
mothur > remove.seqs(accnos="fichier".accnos, fasta="fichier".fasta,  
name="fichier".names, group="fichier".groups)
```

#8) Classification et suppression des contaminants mitochondriaux et chloroplastiques (archaea ou bacteria en fonction de l'analyse):

```
classify.seqs(fasta="fichier".fasta, name="fichier".names, group="fichier".groups,  
template=trainset9_032012.pds.fasta, taxonomy=trainset9_032012.pds.tax,  
cutoff=80, processors=2, iters=1000)
```

```
mothur > remove.lineage(fasta="fichier".fasta, name="fichier".names,  
group="fichier".groups, taxonomy="fichier".taxonomy,  
taxon=Mitochondria-Cyanobacteria_Chloroplast-Eukarya-unknown-Archaea)
```

#9) Re-calcul de l'alignement pour le calcul de la matrice de distance:

#utilisation de l'aligner infernal de RDP pyro : <http://pyro.cme.msu.edu/>

#10) Calcul de la matrice de distance:

#Calculée sur le site RDP pyro car le format de l'alignement n'est pas reconnu par mothur, cutoff=0.15 : <http://pyro.cme.msu.edu/>

Annexe C

Scripts de soumission sur les clusters de calcul

Script de soumission Genotoul

```
#!/bin/bash
#
# ecrit les erreur dans le fichier de sortie standard
#$ -j y
# shell que l'on veut utiliser
#$ -S /bin/bash
# indiquer son email pour suivre l'execution :
#$ -M fabien.javerliat@univ-pau.fr
# obtenir un message au demarrage (b) , a la fin (e), en cas d'abandon (a)
#$ -m bea
# utiliser 4 coeurs (permet de reverser un noeud en entier
et d'avoir toute la mémoire disponible 32G)
#$ -pe parallel_smp 4
# la queue que l'on veut utiliser :
#$ -q longq

# mettre votre ligne de commande
mothur "#"
```

Script de soumission Pyrene pour des calculs parallèles

```
#!/bin/sh

## DEBUT DIRECTIVES SGE (commencant par '$$')

# Nom du job
#$ -N mothur_test #nom dans la file d'attente

# Queue de travail
#$ -q tlong.q

# Execution du job dans le repertoire courant
```

```

#$ -cwd

# Choix de l'environnement parallele puis du nombre de slots
#$ -pe uppampi 4

# NB: pour connaitre la liste ds environnements paralleles d'une queue faire
# qconf -sq <nom_queue> | grep pe_list

# Initialisation et nettoyage des modules
. /etc/profile.d/modules.sh
module purge
# Chargement des modules
module load mothur/1.28.0

# Pour info : affichage de noeud ou le job est envoye
hostname
# Lancement de l'application (mettre le même nombre de processeurs (np ...)
que dans l'option précédente, pour mothur ne pas utiliser l'option processor)
mpirun -np 4 mothur "#"

```

Script de soumission Pyrene pour des calculs séquentiels

```

#!/bin/sh

## DEBUT DIRECTIVES SGE (commencant par '$$')

# Nom du job
#$ -N traitement_blast

# Queue de travail
#$ -q tlong.q

# Execution du job dans le repertoire courant
#$ -cwd

# NB: pour connaitre la liste ds environnements paralleles d'une queue faire
# qconf -sq <nom_queue> | grep pe_list

# Initialisation et nettoyage des modules
. /etc/profile.d/modules.sh
module purge
# Chargement des modules (indiquer nom du module)

# Pour info : affichage de noeud ou le job est envoye
hostname
# Lancement de l'application (commande à lancer)

```

Annexe D

Scripts Biopython pour la fouille de données

```
#!/usr/bin/python3
# -*- coding: utf-8 -*-

#####
#Auteur: Fabien Javerliat
#date: 26/02/2013
#description: traite des résultats de Blast au format xml pour en récupérer
les numéros d'accessions, les fichiers GB correspondant sur NCBI et
extrait des ces fichiers le champs 'isolation_source'.
#V.1.4.0
#modifications: 26/02/2013 crée un tableau récapitulatif des résultats de BLAST
et des environnements d'isolements
#####

from Bio.Blast import NCBIXML
from Bio import Entrez
from Bio import SeqIO
import sys
import os

####-Recherche les numéros d'accession GI des résultats de Blast#####

def find_access(blast, accession) :

    blast_records=NCBIXML.parse(open(blast))    #lecture du fichier Blast
    accession_file=open(accession,"w")        #création du fichier de sauvegarde

    for blast_record in blast_records:
        for alignment in blast_record.alignments:
            accession_file.write('%s\n' % (alignment.accession))
            #écriture des numéros d'accession dans le fichier de sauvegarde
    accession_file.close()
```

```

###-Recherche les fichier GB correspondant aux numéros d'accension###

def find_gb(accession, genebank) :

    fs= open(accession,"r")
    #ouvrir le fichier contenant les numéro GI des résultats de Blast
    genebank_file_text=open(genebank,"w") #fichier où seront sauver les GenBank
    i=0
    for gi in fs:
        Entrez.email='fabien.javerliat@univ-pau.fr'
        #Adresse mail pour que NCBI sache qui fait la requête
        handle=Entrez.efetch(db='nuccore', id=gi, rettype='gb', retmode='text')
        #téléchargement du fichier GenBank
        text=handle.read() #lecture des données téléchargées
        genebank_file_text.write(text)
        #écriture du GenBank dans le fichier de sauvegarde
        i+=1
        print(i)
    genebank_file_text.close() #fermeture du fichier de sauvegarde
    fs.close() #fermeture du fichier contenant les GI

```

```

####-Recherche le champ 'isolation_source' dans le fichier GB-####
# @author Marc Odunlami
# @date 2012
# Modifié par Fabien Javerliat
##

```

```

def findSource(genebank, isosrc) :

    source=[]

    i = 0
    j = 0
    fout = open("%s_isosrc.gb" % (isosrc), "w")
    #fichier de sortie contenant les isolation source
    fnout = open("%s_N0isosrc.gb" % (isosrc), "w")
    #fichier indiquant les séquences n'ayant pas de champs isolation source
    for seq_record in SeqIO.parse(genebank, "gb") :
        #fichier genbank au format text
        prop = False
        for feat in seq_record.features :
            if "isolation_source" in feat.qualifiers :
                prop = True

```

```

        i += 1
        SeqIO.write(seq_record, fout, "gb")
        aux = feat.qualifiers["isolation_source"][0]
        source.append(aux)
        #ajouter 'if aux not in source :' si on ne veu pas enregistrer
        une source déjà enregistrée
        break
    if not prop :
        SeqIO.write(seq_record, fnout, "gb")
        #écriture dans le fichier fnout si le champ 'isolation_source'
        n'existe pas
        j += 1
fout.close()
fnout.close()

fsrc = open(isosrc, "w")
fsrc.write("\n".join(source))
fsrc.close()

recap=open('%s.recap' % (isosrc,"w"))
recap.write("Nombre de sequences isolation_source :
%s \n Nombre de sequences non isolation_source
: %s \n Nombre total de sequences : %s" %(i,j,i+j))
print ("Nombre de sequences isolation_source : " , i)
print ("Nombre de sequences non isolation_source : " , j)
print ("Nombre total de sequences : ", i + j)

#####
#####tableau récapitulatif des résultats#####
#####

def tableau_recap(blast,genbank, table):
    query=[]
    accession=[]
    source=[]

    save_file=open(table,'w')
    for blast_record in NCBIXML.parse(open(blast)):
        for alignment in blast_record.alignments:
            query.append(blast_record.query)
            #ajout des noms des séquences 'query' dans la liste query
            accession.append(alignment.accession)
            #ajout des numéros d'accèsion des résultats de blast
            dans la liste accession
    for seq_record in SeqIO.parse(genbank, "gb"):

```



```

prop=False
for feat in seq_record.features :
    if "isolation_source" in feat.qualifiers :
        prop=True
        source.append(feat.qualifiers["isolation_source"][0])
        #ajout des lieux d'isolement dans le liste source, si le champ
        n'est pas renseigné 'none' sera ajouté.
        break
    if not prop :
        source.append('none')

a=0
while a < len(query):
    #boucle qui écrit le contenu des listes dans un fichier tabulé
    save_file.write('%s\t%s\t%s\n' % (query[a],accession[a],source[a]))
    a+=1
save_file.close()

```

#####

```

os.chdir('/home/ibeas/Bureau/OTU_abund_rare/Blast_results/results')
files=os.listdir('/home/ibeas/Bureau/OTU_abund_rare/Blast_results/results')
for f in files:
    blast= f
    accession= '%s.access' % (f)
    genebank='%s.gb' % (f)
    isosrc='%s.isosrc' % (f)
    table='%s.csv' % (f)
    print(f)
    find_access(blast, accession)
    #remplacer 'blast' par le nom du fichier contenant les résultats de
    Blast (format xml),et 'accession' par le nom du fichier ou sera sauvé
    les numéros d'accession des résultats de BLAST.
    print("résultats de blast traités")
    find_gb(accession, genebank)
    #remplacer 'genebank' par le nom du fichier de sauvegarde des GB téléchargés
    print('fichiers Gb acquis')
    findSource(genebank, isosrc)
    print('sources trouvées')
    #remplacer 'iosrc' par le nom du fichier de sauvegarde des lieux d'isolement
    tableau_recap(blast,genebank, table)
    #remplacer 'tableau recap' par le nom du fichier de sauvegarde du tableau
    print('%s traitement terminé' %(f))
print('FINI')

```

Script pour compter l'occurrence des mots clefs après la fouille de données

```
#!/usr/bin/python3
```

```
# -*- coding: utf-8 -*-
#####
#Auteur: Fabien Javerliat
#####

fs=open('megablast_bact_isosrc.list','r')
sources=[]
list_resultat_source=[]

for source in fs:
    sources.append(source)

fr=open('fichier_resultat.txt', 'w')
for source in sources:
    if source not in list_resultat_source:
        i = sources.count(source)
        list_resultat_source.append(source)
        source2=source.translate(str.maketrans(' ','','\n'))
        fr.write('%s\t%d\n' %(source2, i))
fr.close()
fs.close()
```


Annexe E

Scripts R utilisés pour l'analyse des données

Calcul des ACC

```
#CANONICAL CORRESPONDANCE ANALYSIS
#####
#CCA of the subsampled bacterial/archaea species data, constrained by all
the selected variables after CA.
#environmental variables in env2
spe.cca<- cca (otu ~ ., env2)
spe.cca
summary(spe.cca) #Scaling 2 (default)

#CCA triplots (using lc site scores)
#####
#scaling1: species scores scaled to relative eigenvalues,
#sites are weighted averages of the species
plot(spe.cca, scaling=1, display=c("sp","lc","cn"), main="Triplot.CCA otu ~ env2
-scaling 1")
#Default scaling 2: site scores scaled to relative eigenvalues,
#species are weighted averages of the sites
plot(spe.cca, display=c("sp","lc","cn"), main="Triplot.CCA otu ~ env2 - scaling 2")

#CCA scaling 1 biplot without species (using lc site scores)
#####
plot(spe.cca, scaling=1, display=c("lc","cn"), main="Biplot.CCA otu ~ env2
- scaling 1")
#CCA scaling 2 biplot without sites
#####
plot(spe.cca, scaling=2, display=c("sp","cn"), main="Biplot.CCA otu ~ env2
- scaling 2")

#Permutation tests of CCA results
#####
#Permutation test of the overall analysis
```

```

anova(spe.cca, step=1000)
#Permutation test of each axis
anova (spe.cca, by="axis", step=1000)

#CCA-based forward selection using vegan's ordistep()
#####
#This function allows the use of factors in env2
cca.step.forward<- ordistep(cca(otu ~ 1, data=env2),scope=formula(spe.cca),
direction="forward", pstep=1000)

#Parsimonious CCA based on variables defined by forward selection
(spe.cca.pars<- cca (otu ~ parameters, data=env2))
anova.cca(spe.cca.pars, step=1000)
anova.cca(spe.cca.pars, step=1000, by="axis")

vif.cca(spe.cca)
vif.cca(spe.cca.pars)

```

Calcul de dendrogramme à partir d'une matrice de distance

```

tab=read.csv("fichier", row.names=1,sep="\t", dec=".", header=TRUE)
dist=as.dist(tab)
plot(hclust(dist, method="average"), hang=-1)

```

Résumé

L'étude des communautés procaryotes est une part importante de l'écologie microbienne. Dans un contexte de changement climatique global, comprendre comment se structurent les communautés est nécessaire pour estimer comment les communautés sont impactées et comment les mécanismes de résilience peuvent se mettre en place. Les processus qui structurent les communautés sont de deux natures : stochastique par des événements aléatoires de dispersion, de naissance et de mort ; et déterministe par la sélection de l'habitat et les interactions entre les micro-organismes. La puissance des technologies de séquençage et du traitement informatique permet actuellement d'enrichir les connaissances sur les processus qui structurent les communautés microbiennes. Nous avons, dans le cadre du projet MIGRAMD, réalisé une étude comparative de douze sites provenant de drainage minier acide (DMA) de trois régions du monde : le sud-est de la France, le sud-ouest de la péninsule ibérique et la Bolivie. Avec un séquençage profond des gènes de l'ARNr 16S des douze sites et la mesure de paramètres physico-chimiques, nous avons montré l'importance du filtre de l'habitat sur la structuration des communautés procaryotes dont une majorité sont spécifique de cet environnement. Les relations entre les organismes sont importantes au sein des communautés, mais leur étude est compliquée surtout avec des organismes réfractaires à la culture. Les approches récentes d'études de réseaux de cooccurrence peuvent nous renseigner sur ces relations. Nous avons mis en place une telle étude qui suggère des relations privilégiées entre certains micro-organismes. Une grande partie de la richesse *archaea* des DMA n'est pas caractérisée et elle est grandement influencée par les particularités environnementales de chaque site. Enfin nous avons étudié la diversité des transporteurs membranaires d'arsenic ArsB et Acr3 et observé des patrons d'échanges de populations entre les communautés des DMA.