

Analyses moléculaires de la diversité et des fonctions de micro-organismes incultivés des sources hydrothermales profondes

THÈSE

présentée et soutenue publiquement le 8 décembre 2006

pour l'obtention du

Doctorat de l'université de Bretagne Occidentale

Discipline: Microbiologie

par

Hélène MOUSSARD

Composition du jury

- Rapporteurs :* Wafa Achouak, Chargée de Recherche CNRS, CEA Cadarache
Philippe Normand, Directeur de Recherche CNRS, Université Claude Bernard Lyon I
- Examineurs :* Purificación López-García, Chargée de Recherche CNRS, Université de Paris Sud
Georges Barbier, Professeur, Université de Bretagne Occidentale
Christian Jeanthon, Directeur de Recherche CNRS, Station Biologique de Roscoff
Daniel Prieur, Professeur, Université de Bretagne Occidentale

Remerciements

Ce travail de thèse a été réalisé au sein du Laboratoire de Microbiologie des Environnements Extrêmes, à l'Institut Universitaire Européen de la Mer.

Je tiens en premier lieu à exprimer toute ma reconnaissance à mon Directeur de thèse, Monsieur Christian Jeanthon, Directeur de Recherche à la Station Biologique de Roscoff, qui m'a guidé lors de ce travail de thèse. Tes conseils avisés, ton exigence et ton esprit critique lors de la correction des articles et de cette thèse m'ont permis sans aucun doute de valoriser au mieux ce travail. Merci pour ta patience.

Je remercie également tous les autres membres du Jury qui me font l'honneur d'examiner cette thèse :

Merci donc à Wafa ACHOUAK, Chargée de recherche au CEA de Cadarache ainsi qu'à Philippe Normand, Directeur de recherche au CNRS, Université Claude Bernard Lyon I, qui ont accepté spontanément d'être rapporteurs de ce travail ;

Merci à Purificación López-García, Chargée de recherche CNRS à l'Université de Paris Sud et à Georges Barbier, Professeur à l'Université de Bretagne Occidentale, qui ont accepté d'examiner ce travail ;

Merci à Daniel Prieur, Professeur à l'Université de Bretagne Occidentale, d'avoir présidé ce jury. Je souhaite vous exprimer tous mes remerciements pour m'avoir accueillie au sein de votre laboratoire, depuis le DEA jusqu'à la fin de cette thèse et de m'avoir décroché une bourse au Ministère de l'Enseignement Supérieur et de la recherche afin de financer ce travail.

Je tiens à adresser mes plus profonds remerciements à Purificación López-García et à David Moreira qui, à deux reprises, m'ont accueillie dans leur laboratoire. J'ai appris beaucoup de choses en votre compagnie, de la construction d'une banque métagénomique à la phylogénie. J'ai particulièrement apprécié les discussions que nous avons pu avoir ensemble, merci sincèrement, à vous deux.

Je tiens à exprimer ma profonde sympathie à toutes les personnes qui m'ont aidé techniquement ou moralement au cours de ce travail :

Bien sur, j'adresse un grand merci à tous les membres de l'équipe IUEM du LM2E, qui m'ont permis de passer de bons moments en leur compagnie :

Claire Geslin, Soizick Lucas, Edmond Jolivet, Laurent Toffin et Olivier Necessian qui terminaient leur thèse à mon arrivée au LM2E, et avec qui j'ai partagé de très bons moments que ce soit au laboratoire, ou en dehors du labo ;

Gaël de Erauso, mon responsable de DEA : merci à toi de m'avoir toujours soutenu, conseillé,

encouragé et surtout cru en moi... Merci pour tout Gaël... Bon vent à Marseille ;

Stéphane L'Haridon, merci à toi de m'avoir donné tant de conseils avisés notamment sur la culture des micro-organismes anaérobies et les différents métabolismes microbiens que nous sommes susceptibles de trouver dans l'écosystème hydrothermal profond ? Merci d'avoir toujours été disponible pour m'aider ;

Je voudrais également remercier deux étudiants que j'ai encadrés : Adélaïde Nieguitsila qui m'a aidé à caractériser la souche *Thermodesulfator indicus*, et Vincent Jouffe qui a travaillé sur l'annotation génomique des fosmidés. Vincent, j'espère que tu trouveras prochainement un travail dans le domaine que tu souhaites. Merci à vous deux pour votre aide précieuse ;

Et puis, je tiens également à remercier les doctorants qui sont arrivés après moi pour les bons moments passés en leur compagnie : les premiers furent Mélusine Gaillard et Mathieu Gonnet. Mélusine, je te souhaite le meilleur qu'il puisse être pour ton avenir professionnel, quant à toi Mathieu, le pro de l'informatique et de la bioinformatique : allez, au travail ! Tu es le prochain à soutenir ta thèse, n'est-ce pas ? Vite vite ! Bon courage à toi. Il y a également Nathalie, Raja et Erwan : je vous souhaite à tous de terminer du mieux possible votre thèse. Merci pour votre sympathie ;

Merci également à tous les autres membres du labo : Charlotte, Nadège, Jean-Louis, Joëlle, Marc (merci à toi d'avoir à plusieurs reprises discuté avec moi dans cette très longue dernière ligne droite) et aux étudiants : Anne-Laure, Aurore, Olivier, Valentin et Saïd : bon stage à vous au sein du LM2E. Très bonne idée ce petit calendrier des gourmands !

Je remercie également tous les membres de l'équipe Ifremer du LM2E. Tout d'abord, je remercie Joël Quérellou, d'avoir toujours porté de l'intérêt à ce que je faisais. J'adresse des remerciements tous particuliers à Marie-Anne Cambon-Bonavita (merci pour tes précieux conseils en phylogénie et les discussions que j'ai pu avoir avec toi), à Gislaine Henneke pour sa collaboration très efficace sur la manip de l'ADN polymérase, merci à Didier Flament d'avoir aidé Vincent lors de l'annotation des fosmidés et merci à Christine tout simplement pour ta gentillesse portée à mon égard et ta bonne humeur permanente.

Merci à vous Berry et Martha pour votre amitié : que de bons souvenirs j'ai en pensant à vous ! Rappelez-vous, les petits plats que nous nous faisons, les sorties et ballades du WE, le séjour au ski avec toi Martha. J'espère qu'il y aura d'autres séjours à venir d'ailleurs ! Elise, tu es dans la dernière ligne droite, bon courage à toi. Je te souhaite que du bon pour la suite ! Merci aux deux roscovites Isabelle et Arnaud pour leur complicité. Sabrina, je ne t'oublie pas ! Accroche toi pour obtenir un poste que tu mérites tant !

Enfin, merci à Monique Briand et à Robert Marc Marc pour leur aide dans la confection des posters scientifiques et des illustrations présentées dans ce manuscrit de thèse et dans mes

articles. Je tiens également à remercier Emmanuel Taboré d'avoir passé tant de temps devant mon PC et celui de Vincent pour la mise en place de logiciels informatiques.

Bien sur, je remercie également tous mes amis du Club d'Aviron de Mer de Plougonvelin qui m'ont encouragé et soutenu notamment pendant les moments les plus difficiles : merci à vous les filles Chantal M, Chantal N, Chantal P, Claudine, Eloise, Estelle, Sophie d'avoir cru en notre équipage qui en a étonné plus d'un ! Merci pour tous les moments de bonheur que j'ai passé en votre compagnie. Merci à toi Joël pour ton enthousiasme à la présidence ce club, merci au "coach" Xav, fournisseur officiel en tout genre de l'équipage 1 féminin, merci au concepteur du circuit muscu qui nous a fait tant souffrir mais nous a certainement bien préparé pour le championnat de France, merci à GrandFred, Franck, Hervé pour votre amitié et les bons moments passés en votre compagnie.

Enfin, je voudrais remercier mes parents, mon frère, mes grands parents et Frédéric. Je vous dédie cette thèse. Encore merci à toi Frédéric d'avoir su m'encourager, de m'avoir soutenue, de m'avoir épaulée, d'avoir pris le temps de relire ce manuscrit qui pourtant parlait de choses qui t'étaient totalement inconnues.

Sommaire

Liste des abréviations	vii
Liste des figures	ix
Liste des tableaux	xiii
Introduction	1
Chapitre 1	
La microbiologie, de ses débuts à l'ère de la génomique	3
Chapitre 2	
L'ère de la métagénomique	29
Chapitre 3	
L'écosystème hydrothermal marin	69
Chapitre 4	
Objectifs de la thèse	91
Matériels et Méthodes	97
Construction et criblage de la banque métagénomique	99
Résultats	111
Étude 1	
Novel uncultured <i>Epsilonproteobacteria</i> dominate a filamentous sulphur mat from the 13°N hydrothermal vent field, East Pacific Rise	113

Étude 2	
Characterization of a 40 kbp uncultivated euryarchaeotal fosmid from a microbial community associated to the hydrothermal vent polychaete annelid, <i>Alvinella pompejana</i>	139
Étude 3	
Thermophilic lifestyle for an uncultured archaeon from hydrothermal vents : evidence from environmental genomics	167
Discussion générale	205
Conclusion et perspectives	221
Annexes	225
Annexe A	
Banques de données de séquences nucléiques et protéiques	227
Annexe B	
Composés chimiques des environnements hydrothermaux marins profonds	231
Annexe C	
Publication de l'article : “ <i>Thermodesulfatator indicus</i> gen. nov., sp. nov., a novel thermophilic chemolithoautotrophic sulfate-reducing bacterium isolated from the Central Indian Ridge”	235
Bibliographie	245

Liste des abréviations

ADN : acide désoxyribonucléique	DHVE : deep-sea hydrothermal vent <i>Euryar-</i> <i>chaeta</i>
ADN polB : ADN polymérase de la famille B	EDS : energy dispersive spectroscopy
AOM : anaerobic oxidation of methane	EPR : East Pacific Rise (dorsale orientale de l'océan Pacifique)
AMO : ammonium monooxygenase	FISH : fluorescence <i>in situ</i> hybridization
anammox : anaerobic ammonium oxidation	Hsp : heat shock protein
ARN : acide ribonucléique	ITS : intergenic spacer
ARNm : ARN messenger	kpb : kilo paires de bases
ARNr : ARN ribosomal	MEB : microscopie électronique à balayage
ARNr 5S : sous-unité 5 S de l'ARNr ribosomal	MAR : Mid-Atlantic Ridge
ARNr 16S : sous-unité 16S de l'ARN ribosomal	MCR : méthyl-coenzyme-M réductase
ARNr 18S : sous-unité 18S de l'ARN ribosomal	Mpb : mega paires de bases
ARNr 23S : sous-unité 23S de l'ARN ribosomal	ORF : open reading frame (cadre ouvert de lec- ture)
ARNt : ARN de transfert	pb : paire de bases
ATP : adénosine tri-phosphate	PCB : polychlorinated biphenyl
BAC : bacterial artificial chromosome	PCR : polymerase chain reaction
BrdU : bromodéoxyuridine	RBS : ribosome binding site
BET : bromure d'éthidium	RFLP : restriction fragment length polymor- phism
BRE : TFIIB recognition element	RuBisCO : D-ribulose-1,5-biphosphate carboxy-
CDS : coding DNA sequence, désigne une région codante située entre une méthionine et un codon stop	
CIR : Central Indian Ridge	

Liste des abréviations

lase/oxygénase

rTCA cycle : reverse tricarboxylic acid cycle

SIP : stable isotope probing

SRB : sulfate-reducing bacteria

UFC : unité formant colonie

UV : ultra violet

VBNC : viable but not cultivable

Liste des figures

Introduction

Chapitre 1

Fig. 1	Principe de fonctionnement d'une colonne de Winogradsky.	6
Fig. 2	Arbre universel du vivant basé sur l'analyse des gènes codant les ARNr (16S et 18S) des petites sous-unités des ribosomes.	14
Fig. 3	Structure secondaire de la molécule d'ARNr 16S basée sur la structure d' <i>Escherichia coli</i>	15
Fig. 4	Phylogénie illustrant l'évolution des connaissances des phyla bactériens depuis Woese en 1987.	18
Fig. 5	Diversité des phyla archaéens.	19
Fig. 6	Structure des gènes chez les <i>Bacteria</i>	23
Fig. 7	Consensus des promoteurs de la transcription chez différents groupes archaéens.	24
Fig. 8	Représentation schématique du complexe de pré-initiation et des promoteurs dans les trois domaines de la vie.	25

Chapitre 2

Fig. 1	Principe de la construction et du criblage d'une banque métagénomique.	32
Fig. 2	Représentation schématique des deux approches utilisées pour extraire de l'ADN génomique contenu dans un échantillon solide.	37
Fig. 3	Enrichissement d'un échantillon en ADN génomique d'intérêt, par la technique du BrdU ou par la technique SIP.	40
Fig. 4	Assemblage de l'ébauche génomique après séquençage aléatoire global.	49

Fig. 5	Fixation autotrophe du CO ₂ via le cycle inverse de l'acide tricarboxylique (rTCA).	56
Fig. 6	Affiliation phylogénétique des séquences provenant de l'analyse métagénomique de la mer des Sargasses.	58
Fig. 7	Arbre phylogénétique illustrant les différentes rhodopsines identifiées dans la mer des Sargasses.	60
Fig. 8	Photographie du biofilm de la mine de Richmond et abondance relative des micro-organismes présents dans ce biofilm	61
Fig. 9	Schéma représentant les cycles biogéochimiques déduits de l'analyse métagénomique du biofilm microbien prélevé dans le drainage acide de la mine de Richmond	63
Fig. 10	Modèle hypothétique de la méthanogénèse inverse chez les archaea de la lignée ANME-1.	65
Fig. 11	Cycle de l'azote.	66

Chapitre 3

Fig. 1	Principe général de la tectonique des plaques.	72
Fig. 2	Les trois types de dorsales océaniques.	73
Fig. 3	Localisation des dorsales océaniques et des sites hydrothermaux marins étudiés par les biologistes.	74
Fig. 4	Principe de la circulation de l'eau de mer dans la croûte océanique, au niveau des zones d'expansion.	75
Fig. 5	Quelques représentants de la faune hydrothermale marine profonde.	79
Fig. 6	Localisation des principaux habitats microbiens des sources hydrothermales marines.	82
Fig. 7	Arbre phylogénétique illustrant les nouvelles lignées d' <i>Archaea</i> (en gras) mises en évidence dans les sources hydrothermales marines profondes.	89

Résultats

Étude 1

Fig. 1	Quelques photographies de mattes microbiennes observées sur des sites hydrothermaux profonds.	116
Fig. 2	Formation de biofilms hydrothermaux contenant des filaments de soufre d'origine bactérienne.	117

Étude 3

- Fig. 1 Distribution des distances entre les boîtes TATA et les codons d'initiation de la traduction prédits. 200
- Fig. 2 Alignement de séquences d'ADN polB de différentes *Euryarchaeota*. 203

Discussion générale

- Fig. 1 Caractéristiques physiologiques des ε -*Proteobacteria* isolées d'habitats hydrothermaux marins profonds et de quelques autres environnements sélectionnés. 209
- Fig. 2 Etudes publiées relatant la présence de DHVE2 dans les banques de clones d'*Archaea*. 219
- Fig. 3 Arbre phylogénétique illustrant la souche "*Aciduliprofundum boonei*", certains représentants des DHVE2 dont Alv-FOS4 et des *Euryarchaeota*. . . . 220

Liste des tableaux

Introduction

Chapitre 1

Tableau 1	Exemples de processus physiologiques catalysés par les micro-organismes. . .	8
Tableau 2	Pourcentage de micro-organismes cultivables provenant de différents habitats.	9
Tableau 3	Niveaux d'analyse informatique des séquences génomiques.	22

Chapitre 2

Tableau 1	Nouvelles enzymes et biocatalyseurs découverts suite au criblage fonctionnel de banques métagénomiques.	35
Tableau 2	Exemples de criblages de banques métagénomiques basés sur la recherche d'activités antimicrobiennes.	35
Tableau 3	Les avantages et les inconvénients d'une banque métagénomique à petits inserts ou à grands inserts.	42
Tableau 4	Comparaison de stratégies : criblage basé sur l'activité versus criblage basé sur la séquence.	45
Tableau 5	Analyses métagénomiques impliquant des <i>Archaea</i> incultivées.	52

Chapitre 3

Tableau 1	Types de métabolismes microbiens pouvant être observés au niveau des sources hydrothermales océaniques profondes.	84
-----------	---	----

Matériels et Méthodes

Tableau 1	Liste des amorces utilisées pour l'amplification par PCR de fragments d'ADNr 16S archaéens et de fragments d'ADNr 18S eucaryotes.	109
-----------	---	-----

Résultats

Étude 1

Tableau 1	Résultats obtenus après amplification des gènes <i>aclB</i> , <i>oorA</i> et des gènes codants pour les ARNr 16S	136
-----------	--	-----

Étude 2

Tableau 1	Affiliations phylogénétiques des séquences d'ADNr 16S bactériennes de la communauté microbienne associée aux tubes d' <i>A. pompejana</i>	164
-----------	---	-----

Étude 3

Tableau 1	Nombre de gènes prédits chez Alv-FOS1 et Alv-FOS4 par les différents programmes utilisés.	199
-----------	---	-----

Introduction

Chapitre 1	
La microbiologie, de ses débuts à l'ère de la génomique	3
Chapitre 2	
L'ère de la métagénomique	29
Chapitre 3	
L'écosystème hydrothermal marin	69
Chapitre 4	
Objectifs de la thèse	91

LA MICROBIOLOGIE, DE SES DÉBUTS À L'ÈRE DE LA GÉNOMIQUE

1	Découverte des micro-organismes : observations et cultures . . .	5
2	L'ère de la microbiologie moderne	7
2.1	Les habitats microbiens	7
2.2	Rôle des micro-organismes dans les cycles biogéochimiques de la Terre	7
2.3	Une minorité de micro-organismes est cultivable par les techniques classiques de culture	9
2.4	Cultiver l'incultivé : quelques exemples de techniques de culture récemment mises au point	11
3	De la classification vers la phylogénie des micro-organismes . . .	12
3.1	Premières classifications de micro-organismes	12
3.2	Principe de l'horloge moléculaire - L'aube de la phylogénie moléculaire	12
3.3	L'arbre universel du vivant	13
4	Ecologie microbienne moléculaire	16
4.1	Premières utilisations de l'ARNr 16S en écologie microbienne	16
4.2	Les apports de l'approche moléculaire en écologie microbienne	16
5	L'ère de la génomique	20
5.1	Caractéristiques des génomes procaryotes	20
5.2	Annotation de génomes procaryotes	21
5.2.1	Analyse structurale	21
5.2.2	Analyse fonctionnelle	26
5.3	L'utilisation de la bioinformatique pour l'analyse des génomes : banques de données et plates-formes d'annotation	27

5.3.1	Les banques de données	27
5.3.2	Serveurs et plateformes de bioinformatiques	27

1 Découverte des micro-organismes : observations et cultures

L'histoire des micro-organismes en biologie commence il y a un peu plus de 300 ans, avec la mise au point du microscope par Antonie van Leewenhoek (1632-1723). Cet instrument lui permis d'observer pour la première fois des micro-organismes appelés alors animalcules. Ce n'est qu'à la fin du XVIIIème siècle, que l'existence de ces animalcules fut confirmée par l'allemand Müller.

Au XIX siècle, les travaux de Louis Pasteur (1822-1895) révolutionnèrent la bactériologie. Il démontra le rôle des micro-organismes à différents niveaux biologiques, notamment comme agents infectieux et comme agents responsables de la fermentation. Il mit également au point plusieurs milieux de culture, inventa l'autoclave, la pasteurisation et le principe de vaccination. La naissance de la microbiologie comme étant "la science qui étudie les micro-organismes invisibles à l'œil nu", est généralement fixée à 1857, avec la publication du premier article de Pasteur.

Parallèlement aux découvertes de Pasteur, le médecin allemand Robert Koch (1843-1910) et ses collaborateurs développèrent des techniques d'isolement et de culture des bactéries. Ils mirent au point la culture des micro-organismes sur milieu solide, en cultivant par exemple des bactéries sur des tranches de pomme de terre cuite, ou en solidifiant des milieux liquides avec de l'agar. Koch a également introduit le concept d'isolement des micro-organismes en culture pure pour leur étude.

En étudiant les communautés microbiennes du sol et de l'eau, les microbiologistes hollandais Martinus Beijerinck (1851-1931) et le russe Sergei Winogradsky (1856-1953) initièrent les premiers travaux de microbiologie environnementale. Tous deux étaient convaincus que les micro-organismes jouaient un rôle important dans le cycle de la matière sur Terre. Conscients de la diversité métabolique des micro-organismes, ils comprirent que chacun d'eux exigeaient des nutriments essentiels à leur croissance et des conditions de culture qui leur sont propres. Ils mirent ainsi au point les premiers protocoles de culture permettant d'enrichir spécifiquement un micro-organisme donné.

Le principe d'une culture dite d'enrichissement est de cultiver en laboratoire un type de micro-organisme donné, ceci en lui fournissant les conditions essentielles à sa croissance (milieu de culture ayant une composition chimique particulière et définie) et en reproduisant le plus fidèlement possible les conditions présentes dans son environnement naturel. Dans ces conditions, la croissance spécifique du type de micro-organisme visé par les conditions de culture d'enrichissement est favorisée. Finalement, après quelques repiquages ces micro-organismes sont plus facilement isolés en culture pure, même si dans l'échantillon de départ ils étaient minoritaires.

Les techniques d'enrichissements mises au point par Winogradsky et Beijerinck ont permis

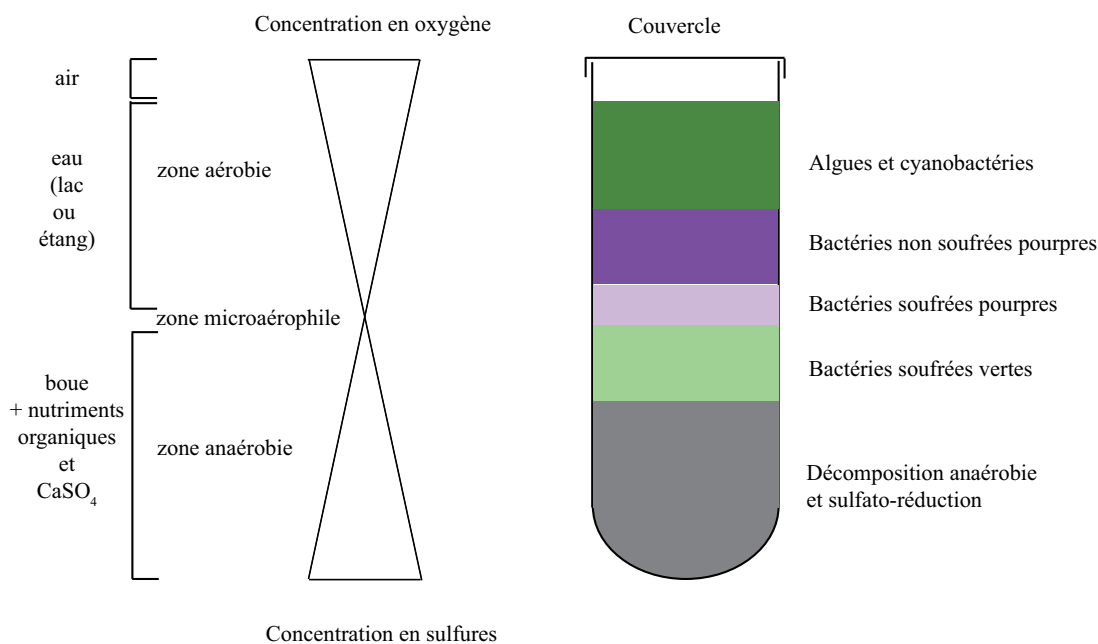


Fig. 1 – Principe de fonctionnement d'une colonne de Winogradsky. La phase supérieure reproduit les conditions retrouvées dans le milieu liquide (luminosité, aération) alors que la phase inférieure reproduit les conditions des premiers centimètres de sédiments (obscurité, augmentation du potentiel réducteur). Au cours du temps, l'établissement de la flore microbienne s'effectue en fonction de leur préférendum au sein des gradients (d'après Madigan, 1996).

d'isoler de nouveaux groupes de micro-organismes, révélant ainsi la pluralité métabolique des micro-organismes. En instaurant des contraintes sélectives de conditions et de milieux de culture sur une population microbienne du sol, Beijerinck réussit à isoler les premières bactéries aérobies fixatrices d'azote (*Azotobacter* sp.). Quelques années plus tard, il prouva que la réduction des sulfates en hydrogène sulfuré (H_2S) était un processus bactérien et isola l'organisme sulfato-réducteur responsable (*Desulfibrio desulfuricans*).

Tout comme Beijerinck, Winogradsky travailla également sur le cycle du soufre. Il mit au point une technique d'enrichissement dite 'colonne de Winogradsky' (Fig. 1). Cette colonne permet de modéliser le cycle du soufre bactérien tel qu'il peut exister au sein d'un environnement naturel. Grâce à cette expérience, il montra la croissance successive et les interactions qui peuvent exister entre différents types de micro-organismes qu'ils soient aérobies ou non, utilisateurs de soufre ou non, phototrophes ou chimiotrophes. Winogradsky fut le premier à proposer le concept de chimiolithoautotrophie. Ce métabolisme est basé sur la fixation du CO_2 couplé à l'oxydation d'un composé inorganique en absence de lumière et de chlorophylle.

Beijerinck et Winogradsky initièrent ainsi les premiers travaux d'écologie microbienne, en étudiant les relations existantes entre les différents types de micro-organismes retrouvés au sein d'une communauté. Plus récemment, Hungate décrivit lui aussi plusieurs techniques d'enrichissement et de culture lui permettant entre autre d'isoler des bactéries cellulolytiques anaérobies

(Hungate, 1950).

2 L'ère de la microbiologie moderne

2.1 Les habitats microbiens

Grâce à la mise au point des techniques de cultures, de nombreux micro-organismes ont été retrouvés aussi bien dans des habitats aquatiques (eau douce ou eau de mer) que terrestres, souterrains et aériens, en associations avec des plantes ou des animaux (Whitman *et al.*, 1998). La Terre offre un vaste panel de conditions environnementales qui fluctuent en permanence dans le temps et dans l'espace, et ceci sous la pression de facteurs abiotiques (chimiques et physiques) et biotiques. Toutes ces conditions environnementales favorisent la formation d'une multitude de niches écologiques qui a priori peuvent toutes être exploitées par les micro-organismes grâce à leur pluralité métabolique et à leur faculté d'adaptation. En effet, les micro-organismes peuvent se développer en absence ou en présence d'oxygène, exposés à la lumière ou non, dans une gamme de températures allant de quelques degrés Celsius à plus de 110°C, et une gamme de pH allant de l'acidité à l'alcalinité extrême, dans de l'eau douce à de l'eau de mer saturée en sels, à pression atmosphérique ou à hautes pressions hydrostatiques, et peuvent être exposés à de fortes radiations ionisantes, à des concentrations élevées en métaux ou à la dessiccation. Les micro-organismes sont donc susceptibles d'être présents dans tous les écosystèmes terrestres. Si l'on considère comme normal un environnement où les températures sont comprises entre 4°C et 40°C, les pH entre 5 et 8.5 et où les salinités varient entre l'eau douce et l'eau de mer, on peut qualifier d'environnements extrêmes ceux qui sont hors de cette gamme.

2.2 Rôle des micro-organismes dans les cycles biogéochimiques de la Terre

On a longtemps sous-estimé l'importance des micro-organismes dans l'équilibre écologique de notre planète. Et pourtant, la répartition des micro-organismes sur Terre est très ubiquiste. De plus, ils représentent à eux seuls une grande proportion de la biosphère terrestre. Whitman et ses collaborateurs (1998) ont estimé que le nombre total de micro-organismes sur Terre était de 4 à 6 $\times 10^{30}$ cellules (estimation fondée sur les teneurs en carbone organique) avec : 1.2×10^{29} cellules dans l'océan, 2.6×10^{29} cellules dans le sol et 0.25 à 2.4×10^{30} de cellules pour la biosphère souterraine. Ces mêmes auteurs suggérèrent que la quantité de carbone contenue dans les cellules procaryotes représentait 60 à 100% du carbone total contenu dans les plantes (la biomasse des plantes était considérée jusque là comme étant majoritaire sur Terre). Les cellules procaryotes qui contiennent 10 fois plus d'azote et de phosphore que les plantes, constituent ainsi le plus grand pool de ces éléments contenu dans des cellules vivantes. Par leur abondance, mais aussi par la pluralité de leurs voies métaboliques et leur ubiquité sur notre planète, les micro-organismes sont des agents spécifiques et primordiaux intervenant dans les cycles de la matière. Ils jouent un rôle

essentiel dans le fonctionnement des processus biogéochimiques sur Terre et sont à la base de nombreux processus écologiques. La biosphère dépend totalement de l'activité microbienne (Tableau 1). Les micro-organismes interviennent entre autre dans la décomposition du sol, dans la genèse et la régulation de la composition de l'atmosphère terrestre, dans la fixation de l'azote et dans la photosynthèse. Ils participent également à la détoxification de nombreux composés organiques et inorganiques (Madsen, 2005). Les micro-organismes catalysent de nombreuses réactions d'oxydo-réduction qui sont au centre de cycles biogéochimiques d'éléments majeurs tels que le carbone, l'azote, le soufre, le phosphore et les métaux. Les micro-organismes sont les premiers responsables des changements géochimiques de la planète, cela d'autant plus que la biomasse globale des procaryotes est approximativement équivalente à celle de toutes les autres formes de vies (eucaryotes) (Whitman *et al.*, 1998).

Processus	Nature du processus	Types d'habitat
Cycle du carbone		
Photosynthèse	Fixation du CO ₂	Em, Ed, EdS, EmS
Respiration	Oxydation de carbone organique en CO ₂	Tous
Décomposition de la cellulose	Dépolymérisation, respiration	Sol
Méthanogénèse	Production de CH ₄	Sw, EmS, EdS
Oxydation aérobie du CH ₄	CH ₄ devient CO ₂	Tous
Oxydation anaérobie du CH ₄	CH ₄ devient CO ₂	EmS
Biodégradation		
Composés organiques synthétiques	Décomposition, formation de CO ₂	Tous
Hydrocarbures - pétrole	Décomposition, formation de CO ₂	Tous
Additifs du fuel (MTBE)	Décomposition, formation de CO ₂	Sol, Sw, Est
Nitroaromatiques	Décomposition, formation de CO ₂	Sol, Sw, Est
Produits pharmaceutiques	Décomposition	Sol, Sw, Est
Solvants chlorés	Composés déchlorinés par respiration dans les habitats anaérobies	Sol, Sw, Est
Cycle de l'azote		
Fixation de l'azote	N ₂ devient NH ₃	Sol, Em
Oxydation de NH ₄ ⁺	NH ₃ devient NO ₂ ⁻ , NO ₃ ⁻	Sol, Sw
Oxydation anaérobie de NH ₄ ⁺	NO ₂ ⁻ et NH ₃ deviennent N ₂	EmS, Sw
Dénitrification	NO ₃ ⁻ accepteur d'électrons devient N ₂	Sol, Sw
Cycle du soufre		
Oxydation des composés soufrés	S ²⁻ et S ⁰ deviennent SO ₄ ²⁻	EmS, EdS
Réduction des sulfates	SO ₄ ²⁻ accepteur d'électrons, devient S ⁰ et S ²⁻	EmS, Sw, Est
Autres éléments		
Oxydation de H ₂	H ₂ est oxydé en H ⁺ , les électrons réduisent d'autres composés	Sw, sol, EmS, EdS
Méthylation et réduction de Hg	Hg organique formé et Hg ²⁺ est converti en Hg	EdS, EmS
Réduction de (per)chlorate	Les oxydants sont convertis en chlorides	Est
Réduction de l'uranium	L'oxyanion U est utilisé comme accepteur d'électron	Est
Réduction de l'arsénique	L'oxyanion As est utilisé comme accepteur d'électron	EdS, Est
Oxydation du fer	Oxydation des minerais de FeS	EdS, Est

Abréviations des habitats: Ed, eau douce; EdS, sédiments d'eau douce, Est, eau souterraine; EmS, sédiments océaniques; Em, eau de mer; Sw, eau d'égouts; Est, eau souterraine
MTBE, methyl tertiary butyl ether

Tableau 1 – Exemples de processus physiologiques catalysés par les micro-organismes (d'après Madsen, 2005).

2.3 Une minorité de micro-organismes est cultivable par les techniques classiques de culture

Malgré le succès remporté par les techniques d'enrichissement et d'isolement en culture pure, il est reconnu que pour la majorité des environnements, on ne sait cultiver qu'une très faible partie des micro-organismes présents dans un échantillon. Le constat est frappant : la fraction cultivable des micro-organismes ne représenterait qu'une très faible proportion (0.05 à 10%) de la diversité totale estimée des procaryotes dans la biosphère (pour revue, Amann *et al.*, 1995 ; Pace, 1997) (Tableau 2).

Habitats	Micro-organismes cultivables (%) ^a
Eau de mer	0.001-0.1
Eau douce	0.25
Eau lacustre mésotrophique	0.1-1
Eau estuarienne non polluée	0.1-3
Boues activées	1-15
Sédiments	0.25
Sol	0.3

Tableau 2 – Pourcentage de micro-organismes cultivables provenant de différents habitats (d'après Amann *et al.*, 1995).

^a le pourcentage des micro-organismes cultivables est calculé de la façon suivante :

$$100 \times \frac{\text{nombre d'unités formant colonies (UFC)}}{\text{nombre total de micro-organismes dénombrés par microscopie dans l'échantillon}}$$

Plusieurs explications à cela sont possibles :

- La première explication réside dans le fait que la plupart des micro-organismes présentent des difficultés à croître dans un milieu de culture artificiel (Jannasch & Jones, 1959).
- Si la majorité des micro-organismes de l'environnement sont viables, ils ne forment pas obligatoirement des colonies visibles lorsqu'on les cultive sur milieu solide. La technique de culture sur milieu gélosé sélectionne les micro-organismes qui ont des croissances rapides, ceux qui peuvent pousser en atteignant de hautes densités cellulaires, (Jannasch & Jones, 1959 ; Keller & Zengler, 2004). Or il faut en moyenne obtenir la croissance d'au moins 10⁵ cellules pour qu'une colonie soit visible à l'œil nu (Keller & Zengler, 2004). L'accumulation de produits toxiques fabriqués par les cellules, mais aussi une carence en éléments nutritifs essentiels lors de leur multiplication, ou bien une infection virale peuvent également être responsables de la difficulté à cultiver ces micro-organismes.
- Reproduire fidèlement en laboratoire les propriétés physico-chimiques d'un environnement donné reste un véritable challenge. Les milieux de culture classiquement utilisés pour le recensement des micro-organismes du sol ou marins sont souvent trop riches en nutriments (Eilers *et al.*, 2000 ; Keller & Zengler, 2004). En supposant que de nombreux micro-

organismes incultivés du sol avaient leur croissance inhibée par des concentrations trop élevées en nutriments dans les milieux de culture couramment utilisés en microbiologie, Janssen et ses collaborateurs (2002) tentèrent de les cultiver en utilisant des milieux de culture dilués. Afin de favoriser la croissance des micro-organismes à taux de croissance lent, ils augmentèrent les temps d'incubation jusqu'à 10 semaines. Par cette approche, ils réussirent à isoler et purifier de nouveaux organismes présents dans le sol. Button qui mit au point une méthode de culture par dilution pour cultiver les micro-organismes présents dans les environnements aquatiques oligotrophes montra que l'addition de nutriments dans les milieux de culture pouvaient stimuler la croissance de quelques micro-organismes, mais que cela inhibait la croissance de la majorité d'entre eux (Button *et al.*, 1993).

- La difficulté de cultiver les micro-organismes sur milieu gélosé peut aussi venir du fait que les micro-organismes vivent rarement sous forme libre, mais plutôt sous forme de biofilms attachés à des surfaces (Davey & O'Toole, 2000). Les biofilms microbiens favorisent les interactions inter-bactériennes en facilitant la diffusion de molécules signales synthétisées par les bactéries douées de quorum sensing. Le quorum sensing, initialement découvert chez des bactéries bioluminescentes a maintenant été identifié chez de nombreuses bactéries à Gram négatif et positif (Whitehead *et al.*, 2001 ; Waters & Bassler, 2005). Ce mécanisme de communication inter-bactérienne dépend de la densité bactérienne. Les techniques classiques d'enrichissement ne favorisent pas la formation de biofilms, limitant ainsi la communication inter-bactérienne. Bruns et ses collaborateurs (2003) ont montré que l'ajout dans un milieu de culture de petites molécules signales telles que de l'AMPc et des acylhomosérines lactones (synthétisées par les bactéries capables de quorum sensing) augmentait le succès de la culture de certains organismes.
- De nombreux micro-organismes développent des relations particulières de symbiose avec d'autres organismes (ces symbioses peuvent être obligatoires ou facultatives, de l'ordre du mutualisme ou du parasitisme, impliquant des ecto- ou des endosymbioses), et ne peuvent être cultivés indépendamment de leur hôte (Amann *et al.*, 1995).
- Certaines bactéries entrent dans un état de dormance dit état viable mais non cultivable (VBNC) lorsqu'elles sont confrontées à des conditions inadéquates à leur physiologie et métabolisme : par exemple une exposition à de l'eau salée, à de l'eau douce, à des milieux oligotrophes, à une pression osmotique ou un pH qui ne leur convient pas ou à de faibles températures (Byrd *et al.*, 1991 ; Colwell, 2000). Une étude menée par Colwell (2000) montra que les souches bactériennes de *Vibrio cholerae* isolées d'environnements aquatiques n'étaient pas directement cultivables sur milieu gélosé. Ces bactéries échappant aux techniques de culture étaient pourtant viables. En effet, après leur passage dans un intestin de souris ou d'humain, ces mêmes bactéries étaient de nouveau cultivables et ré-exprimaient leur pouvoir pathogène. Dans l'état VBNC, ces bactéries qui sont pourtant métaboliquement actives, ne peuvent pas être détectées par les milieux de culture

conventionnels (Pruzzo *et al.*, 2003).

Les méthodes classiques d'enrichissement et d'isolement en culture pure, malgré la diversité des milieux de cultures disponibles, ne sont donc pas adaptées pour cultiver la majorité de la diversité microbienne d'un échantillon donné.

2.4 Cultiver l'incultivé : quelques exemples de techniques de culture récemment mises au point

Nous avons vu que certains micro-organismes doués de quorum-sensing vivaient sous forme de biofilms. Ces biofilms sont essentiels, ils facilitent les communications intercellulaires. Pour favoriser la formation des biofilms, un système de culture en flux continu et sur support solide (le support solide étant une lame de verre ou de polycarbonate) (système de culture nommé "CFSC" pour Continuous-Flow Side Culture) a été récemment mis au point (Stach & Burns, 2002). Stach et Burns (2002) ont réussi par cette méthode d'enrichissement à obtenir à partir d'un même échantillon trois fois plus de diversité génétique que celle obtenue dans des conditions classiques de culture.

Une autre technique d'enrichissement et d'isolement de micro-organismes marins a récemment été mise au point par Kaerberlein et ses collaborateurs (2002). Cette technique a été conçue pour reproduire en laboratoire l'environnement naturel des micro-organismes. Le principe de la technique est le suivant : après dilution d'échantillons marins, des micro-organismes sont pris en masse dans de l'agar puis déposés entre deux membranes. Ces membranes permettent le libre échange de molécules chimiques (substrats, métabolites), mais pas celui des micro-organismes. Le montage est appelé "chambre de diffusion". Ces chambres de diffusion sont placées dans des "aquariums" marins simulant l'environnement naturel des micro-organismes (les chambres de diffusion sont recouvertes d'eau de mer et reposent sur une couche de sédiments marins). Le nombre de micro-organismes récupérés à partir d'un échantillon marin traité de cette façon est 300 fois supérieur à celui observé après culture en boîte de Pétri. Ce système a permis d'obtenir la croissance de micro-organismes jusqu'à présent incultivés.

Zengler a quant à lui récemment mis au point une technique d'enrichissement et d'isolement à haut débit (Zengler *et al.*, 2002). Cette technique consiste à individualiser des micro-organismes dans des microcapsules d'agarose. Ces microcapsules sont alors placées dans différents milieux de culture pour incubation. Les colonies qui se forment dans les microcapsules sont ensuite détectées puis isolées par cytométrie en flux. Cette approche permet le libre échange de composés chimiques entre les micro-organismes d'une même communauté.

3 De la classification vers la phylogénie des micro-organismes

3.1 Premières classifications de micro-organismes

Les techniques classiques et modernes de culture ont permis d'isoler un grand nombre de micro-organismes. Les microbiologistes ont isolé à ce jour de nombreux micro-organismes provenant de divers habitats terrestres, recensant ainsi plus de 5 000 espèces (Garrity & Holt, 2001). Müller dans les années 1770 initia les premiers travaux de classification des micro-organismes. Il décrivit de nombreux procaryotes et les classa en genres et en espèces, suivant les règles de nomenclature binaire que Linné (1707-1778) venait d'éditer (dans cette nomenclature binaire, chaque être vivant est désigné par - un nom de genre qui est commun à plusieurs espèces, et - par un nom d'espèce qui lui est propre). Quelques années plus tard, le botaniste Ferdinand Cohn (1828-1898) établit pour de nombreux micro-organismes leur première classification taxonomique. Cohn élaborait une clé de détermination basée sur l'observation des cellules, lui permettant de définir des genres bactériens sur des critères morphologiques.

Plus tard, les critères morphologiques furent associés à des caractères physiologiques et biochimiques pour pouvoir servir à la classification des micro-organismes (Kluyver & Van Niel, 1936). Cependant, ces méthodes aboutissaient parfois à des classifications contradictoires. En effet, les critères physiologiques et biochimiques n'étaient pas encore suffisamment bien connus pour pouvoir servir à la classification des micro-organismes. De plus ces méthodes étaient limitées dans la mesure où elles ne permettaient pas d'établir avec pertinence des liens de phylogénie entre les micro-organismes. Deux espèces peuvent en effet avoir des caractères morphologiques et/ou physiologiques identiques sans qu'on puisse pour autant leur attribuer une descendance commune. Ainsi, la plupart des relations de parenté qui unissent les grands genres bactériens demeurait à l'époque inconnue.

3.2 Principe de l'horloge moléculaire - L'aube de la phylogénie moléculaire

En 1965, Emile Zuckerkandl et Linus Pauling eurent l'idée d'utiliser pour la première fois des caractères moléculaires pour établir des phylogénies. Ils avancèrent l'hypothèse révolutionnaire selon laquelle les macromolécules pouvaient être des documents de l'histoire évolutive des êtres vivants, ou des horloges moléculaires (Zuckerkandl & Pauling, 1965). Cette hypothèse a pour principe que le taux d'accumulation de mutations chez des organismes différents est du même ordre de grandeur pour des régions homologues (même pression de sélection). Les accumulations sont maximales pour les régions qui ne sont pas soumises à des pressions de sélection naturelles (par exemple les régions ne codant pas pour des gènes), et minimales pour les régions soumises à une forte pression (codant pour des fonctions essentielles à la vie). Si l'on admet cette théorie, et que l'on connaît le taux d'accumulation des mutations en un locus donné, il est possible d'estimer le temps de divergence entre deux espèces en comparant leur séquence (ADN ou protéine).

Les molécules dont la séquence primaire change aléatoirement au cours du temps peuvent être considérées comme un marqueur phylogénétique ou une horloge moléculaire (Woese, 1987). Cependant toute molécule ne peut être utilisée à des fins phylogénétiques. Un bon marqueur moléculaire doit répondre aux critères suivants :

- il doit être universel et ne pas être soumis à des événements de transferts horizontaux de gènes,
- les permutations dans sa séquence doivent être aussi aléatoires que possible,
- sa vitesse d'évolution doit être adaptée à la gamme des distances évolutives devant être mesurée (elle ne doit ni évoluer trop vite sous peine de masquer les changements ancestraux, ni évoluer trop lentement au risque d'occulter les changements récents),
- enfin, il doit contenir une information génétique suffisante.

La formulation de l'hypothèse de l'horloge moléculaire fut un tournant pour la classification des micro-organismes. Grâce à ce concept, on pouvait désormais classer les organismes et les identifier non plus sur des critères phénotypiques (morphologie, physiologie, biochimie...), mais sur des critères génotypiques (informations données par les séquences). En utilisant les séquences primaires d'une protéine, la chaîne β de l'hémoglobine, Emile Zuckerkandl et Linus Pauling (1965) proposèrent la première phylogénie moléculaire des vertébrés.

Parallèlement à premiers travaux de phylogénie moléculaire, l'essor pris par l'informatique a permis l'analyse et l'intégration de plus en plus de données. De plus, si jusqu'au milieu des années 70, les techniques d'obtention des séquences protéiques ou d'acides nucléiques étaient lourdes et coûteuses, Sanger et ses collègues mirent au point en 1977 une technique de séquençage rapide et moins coûteuse, utilisant les didéoxynucléotides terminateurs (Sanger *et al.*, 1977). L'introduction de cette technique de séquençage puis la mise au point par Mullis en 1984, de la technique d'amplification génique dite "Polymerase Chain Reaction" ou PCR provoquèrent une explosion des données de séquences d'ADN disponibles à des fins phylogénétiques.

3.3 L'arbre universel du vivant

La révolution apportée par les phylogénies moléculaires va connaître en 1977 un tournant décisif avec la publication de la première phylogénie moléculaire universelle du vivant par Carl Woese et Fox (1977). Cette phylogénie universelle du vivant, basée sur l'utilisation de la séquence de l'ARN ribosomal des petites sous-unités des ribosomes, a montré que le monde vivant était découpé en trois domaines, à savoir celui des *Eucarya*, des *Bacteria* et des *Archaea* ; ces deux domaines constituant le règne des procaryotes (Fig. 2).

Le principe de la technique d'analyse phylogénétique développée par Woese est le suivant : l'ARNr 16S (18S chez les eucaryotes) de chaque espèce étudiée est digéré par l'ARNase T1. Tous

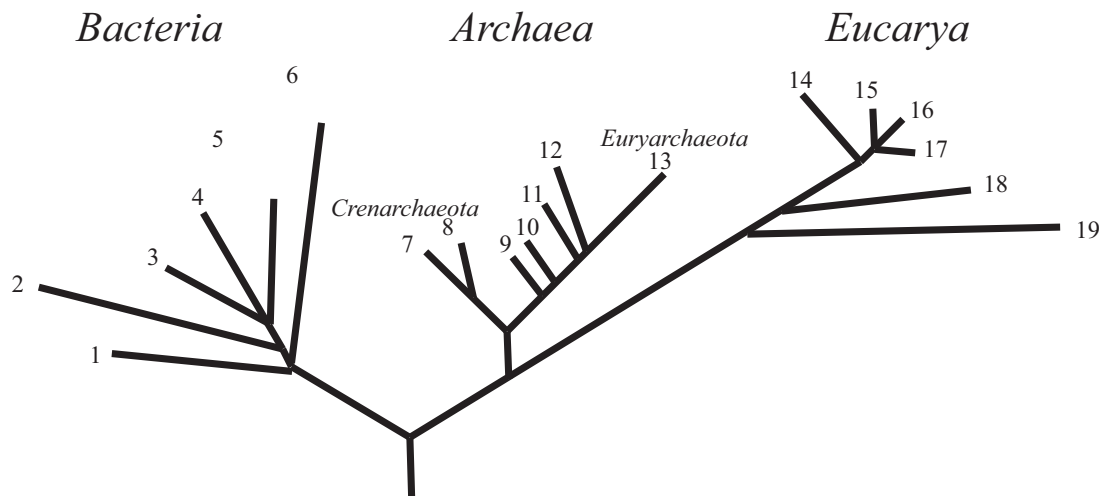


Fig. 2 – Arbre universel du vivant basé sur l'analyse des gènes codant les ARNr (16S et 18S) des petites sous-unités des ribosomes (d'après Woese *et al.*, 1990). L'arbre du vivant montre les 3 domaines du vivant : les *Bacteria*, les *Archaea* et les *Eucarya*. Les nombres à côtés des branches correspondent aux groupes d'organismes suivants : 1, *Thermotogales* ; 2, Flabobactéries ; 3, Cyanobactéries ; 4, Bactéries pourpres ; 5, Bactéries Gram-positif ; 6, Bactéries vertes non sulfureuses ; 7, le genre *Pyrodictium* ; 8, le genre *Thermoproteus* ; 9, *Thermococcales* ; 10, *Methanococcales* ; 11, *Methanobacteriales* ; 12, *Methanomicrobiales* ; 13, Halophiles extrêmes ; 14, Métazoaires ; 15, Ciliés ; 16, Plantes ; 17, Algues ; 18, Flagellés ; 19, Microsporidies

les produits de cette réaction sont ensuite séparés par migration électrophorétique bidimensionnelle. Chaque espèce est alors caractérisée par le profil de migration d'oligonucléotides de son ARNr 16S.

Le choix de l'ARNr 16S/18S par Woese pour établir la phylogénie universelle du vivant n'est pas fortuit. Les ARNr 16S sont d'ailleurs à ce jour les marqueurs moléculaires les plus utilisés pour établir des liens phylogénétiques entre les organismes. Les raisons en sont les suivantes (Woese, 1987) :

- les ARNr sont des constituants des ribosomes, dont la fonction n'a pas changé depuis environ 3.8 milliards d'années. (chez les procaryotes, il existe trois types d'ARNr : les ARNr 5S (120 pb) et ARNr 23S (3000 pb) qui constituent la sous-unité 50S d'un ribosome, et l'ARNr 16S (1500 pb) constituant la sous-unité 30S d'un ribosome),
- les gènes codant les ARNr 16S sont universels, ils sont présents dans toutes les cellules procaryotes,
- leur taille d'environ 1 540 pb contient suffisamment d'information génétique pour permettre d'établir des phylogénies robustes,
- leur séquence primaire composée de positions variables et conservées permet d'établir des liens phylogénétiques entre les organismes appartenant à différents niveaux taxonomiques pouvant aller du domaine à l'espèce (Fig. 3),
- le transfert de gènes horizontaux n'a pas encore été observé chez les procaryotes.

Les analyses phylogénétiques basées sur les séquences de l'ARNr 16S ont ouvert une nou-

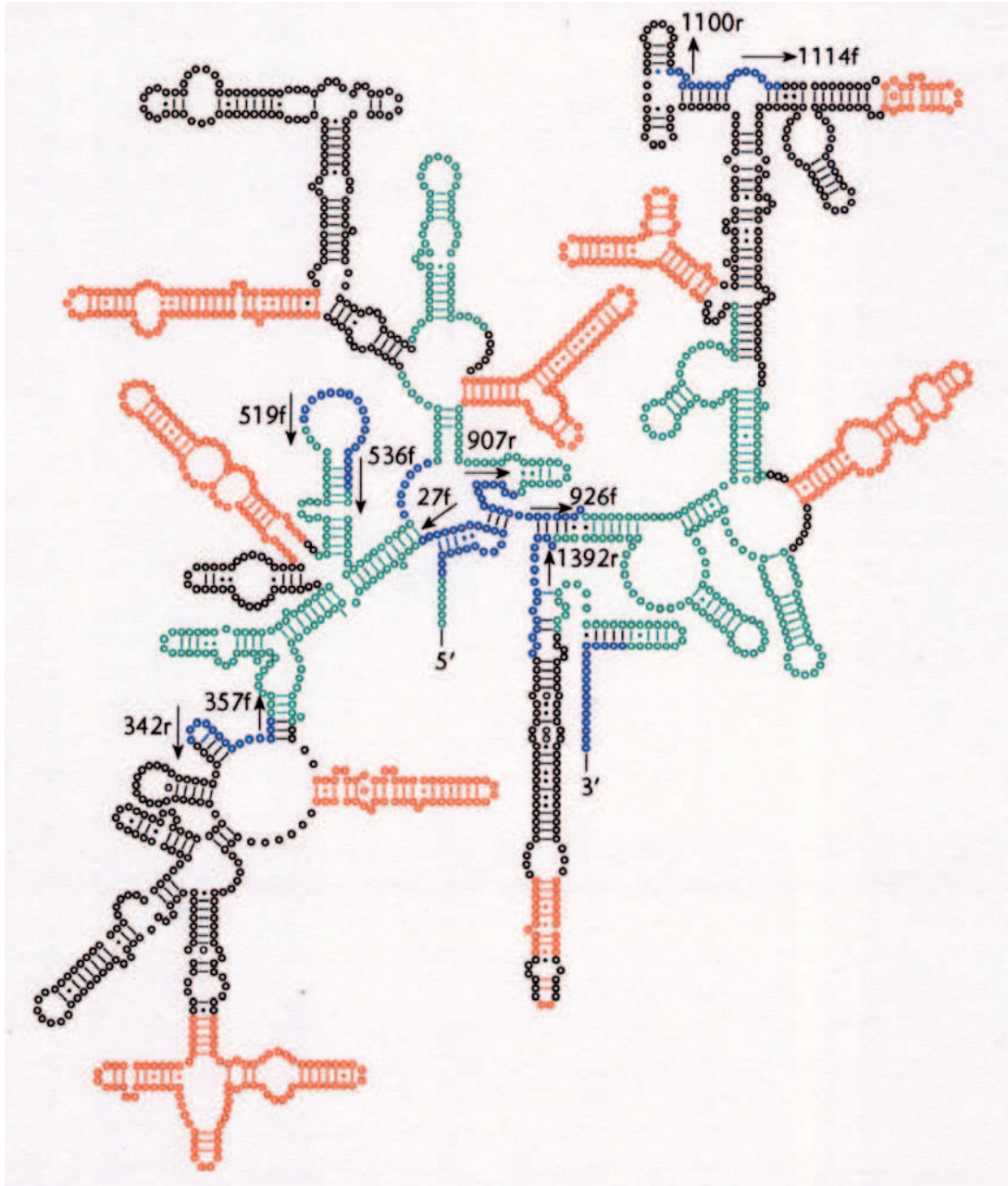


Fig. 3 – Structure secondaire de la molécule d'ARNr 16S basée sur la structure d'*Escherichia coli* (Maidak et al., 1994)35.Yusupov, M. M., Yusupova, G. Zh., Baucom, A., Lieberman, K., Earnest, T. N., Cate, J. H. D., and Noller, H. F. (2001) *Science*, 292, 883-896. . Les régions hypervariables sont en rouge; les régions conservées sont en vert. Les sites d'hybridation des amorces utilisées traditionnellement pour l'amplification génique par PCR sont en bleu, et le sens d'amplification est noté par des flèches. Les autres nucléotides sont en noir.

velle ère pour la biologie évolutive. L'utilisation de ce marqueur moléculaire a été adoptée par la majorité des écologistes moléculaires. Cependant les séquences d'ARNr 16S ne sont pas idéales et les progrès méthodologiques dans le domaine de la reconstruction phylogénétique ont mis en évidence différents artéfacts liés à l'utilisation de ces séquences comme par exemple le biais de composition en G+C chez les organismes hyperthermophiles (Galtier & Lobry, 1997) ou les phénomènes d'attraction des longues branches (Philippe & Germot, 2000).

D'autres marqueurs alternatifs à l'ARNr 16S ont cependant d'ores et déjà été utilisés avec succès pour étudier la phylogénie des procaryotes. C'est le cas des molécules suivantes : l'ARNr 23 S (ARNr des grosses sous-unité des ribosomes), les facteurs d'élongation EF-Tu/-1 α EFG, les ARN polymerases, la sous-unité β de l'ATPase F1/FO, la protéine RecA, les protéines de choc thermique Hsp60 et Hsp70, les ADN polymérasés de la famille B et les ADN gyrases et topoisomérasés (Ludwig & Klenk, 2001). L'utilisation de ces marqueurs alternatifs a permis la construction d'arbres phylogénétiques qui confirment dans la plupart des cas la phylogénie observée avec les ARNr 16S.

4 Ecologie microbienne moléculaire

4.1 Premières utilisations de l'ARNr 16S en écologie microbienne

Suite aux travaux de Carl Woese sur l'ARNr 16S, Norman Pace et ses collaborateurs eurent l'idée d'utiliser une approche moléculaire pour identifier des micro-organismes, sans que ceux-ci ne soient cultivés (Pace *et al.*, 1985 ; Olsen *et al.*, 1986). Cette approche moléculaire consistait à extraire directement les acides nucléiques d'une population microbienne, puis à séquencer des marqueurs moléculaires afin d'établir des phylogénies. On pouvait ainsi soit séquencer les ARNr 16S directement extraits suite à l'extraction des acides nucléiques, soit séquencer des copies (ADN) rétro-transcrites de ces ARNr. Par la suite, les progrès réalisés dans le domaine de la PCR ont permis d'amplifier la quasi totalité du gène codant l'ARNr 16S spécifiquement ciblé grâce à des amorces s'hybridant avec les régions conservées de la séquence (Giovannoni *et al.*, 1990). De plus, grâce à la technique d'hybridation *in situ* (FISH) décrite à l'origine par l'équipe de Norman Pace (Giovannoni *et al.*, 1988 ; DeLong *et al.*, 1989), on peut maintenant en plus de l'identification des micro-organismes les quantifier à l'aide de sondes oligonucléotidiques ciblant des fragments discriminants à l'intérieur de la séquence des ARNr 16S.

4.2 Les apports de l'approche moléculaire en écologie microbienne

Le clonage et le séquençage du gène codant l'ARNr 16S directement extrait à partir d'échantillons environnementaux a permis d'étudier de la biodiversité microbienne sans passer par la très contraignante étape de mise en culture des micro-organismes (Amann *et al.*, 1995). Cette

approche moléculaire a été intensément exploitée ces dix dernières années dans des habitats terrestres et aquatiques.

Quand Carl Woese proposa il y a vingt ans de cela une phylogénie des bactéries basée sur les ARNr 16S, il répertoria 12 phyla bactériens distincts, chacun d'entre eux possédant des représentants cultivés (Woese, 1987). Depuis, quarante et un nouveaux phyla ont été identifiés en vingt ans d'inventaires moléculaires réalisés dans de nombreux environnements, ce qui conduit à un total de 53 phyla bactériens (Rappe & Giovannoni, 2003). Parmi ces derniers, 26 phyla n'ont pas de représentant cultivé et ont été identifiés uniquement grâce à leur signature d'ARNr 16S (Fig. 4). Quant aux *Archaea*, les premières d'entre elles à être identifiées suite à des inventaires moléculaires furent des *Crenarchaeota*. Ces dernières constituent avec les *Euryarchaeota* les deux divisions archéennes initialement décrites par Woese et ses collaborateurs (1990). Ces *Crenarchaeota* ont été identifiées dans des échantillons de plancton marin (DeLong, 1992 ; Fuhrman *et al.*, 1992). Quelques années plus tard, d'autres séquences d'ARNr 16s affiliées aux *Crenarchaeota* furent identifiées dans des environnements terrestres (Bintrim *et al.*, 1997 ; Jurgens *et al.*, 1997, 1999). Cette surprenante découverte a révélé que les *Archaea* ne colonisaient pas uniquement les habitats extrêmes comme on le pensait initialement, mais qu'elles s'étendaient également aux habitats les plus communément répandus sur Terre. Depuis de nouveaux phylotypes d'*Archaea* ont été détectés dans la plupart des environnements. En comparant les séquences d'ARNr 16S, Philip Hugenholtz recensait il y a trois ans de cela, un total de 18 phyla archéens, dont 10 ne contiennent pas encore de représentant cultivé (Hugenholtz, 2002). Depuis, près de 8000 séquences d'ARNr 16S archéennes ont été déposées dans des banques de données, constituant souvent de nouvelles lignées phylogénétiques au sein des *Archaea* (Schleper *et al.*, 2005) (Fig. 5).

L'approche moléculaire a ainsi montré que dans la plupart des environnements, la majorité de la diversité microbienne (plus de 99 % tous environnements confondus) restait inaccessible par des approches culturelles classiques (Amann *et al.*, 1995 ; Torsvik & Ovreas, 2002). Un grand nombre de nouvelles lignées phylogénétiques à la fois chez les *Bacteria* et les *Archaea* mais aussi au sein des *Eucarya* (Diez *et al.*, 2001 ; Lopez-Garcia *et al.*, 2003b) ont été mises en évidence grâce aux inventaires moléculaires. Cette approche a permis d'établir une classification de tous les organismes, cultivés ou non. L'heure est maintenant d'étudier le génome des micro-organismes, cultivés ou non, afin de comprendre leur rôle, leur organisation et leur fonctionnement dans la biosphère.

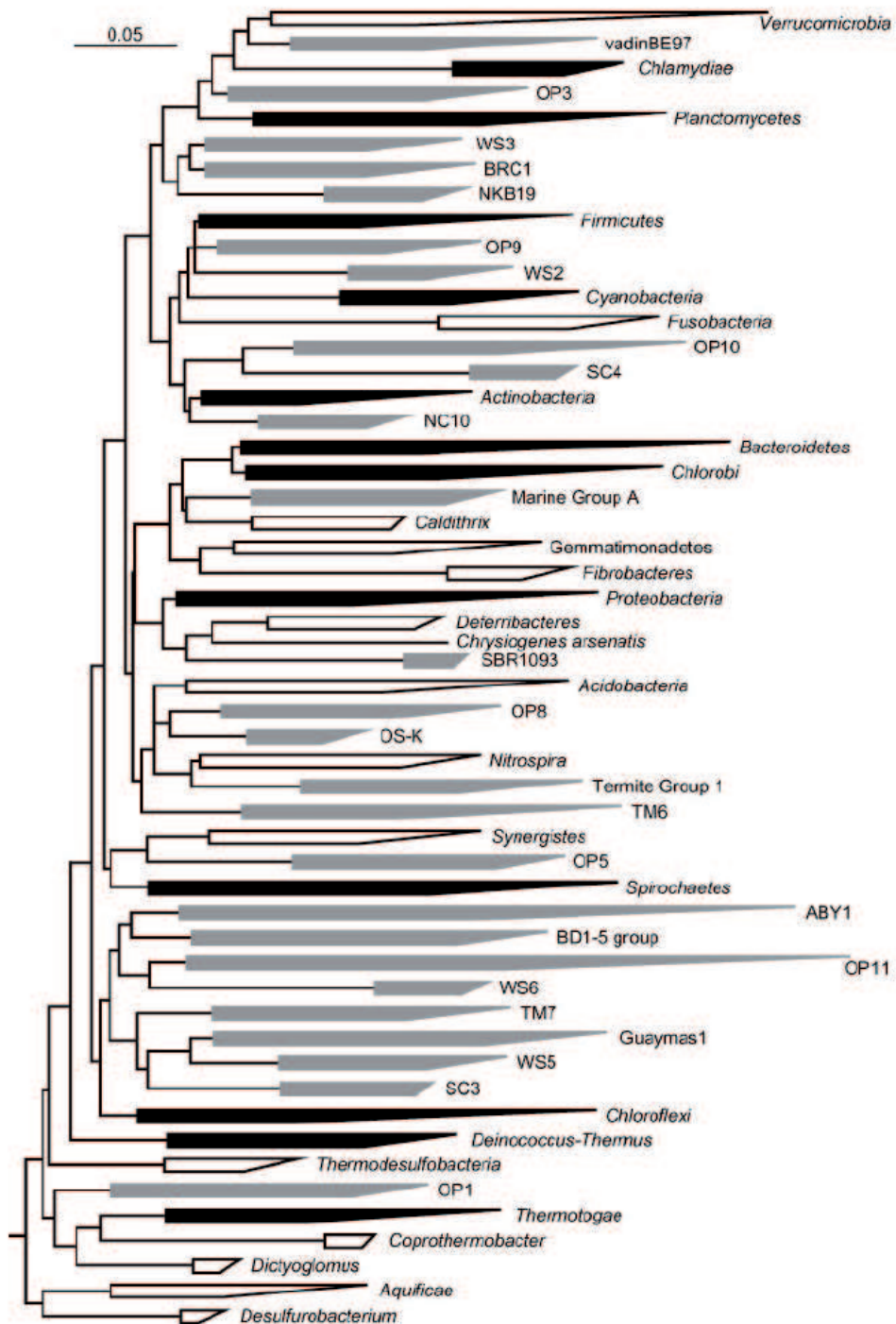


Fig. 4 – Phylogénie illustrant l'évolution des connaissances des phyla bactériens depuis Woese en 1987 (Woese, 1987). Les segments en noir montrent les phyla originaux décrits par Woese en 1987. Les segments blancs représentent les 14 phyla contenant des espèces cultivées, découverts depuis 1987. Les segments en gris indiquent les 26 phyla candidats qui ne contiennent pas d'espèces cultivées (d'après Rappe & Giovannoni, 2003).

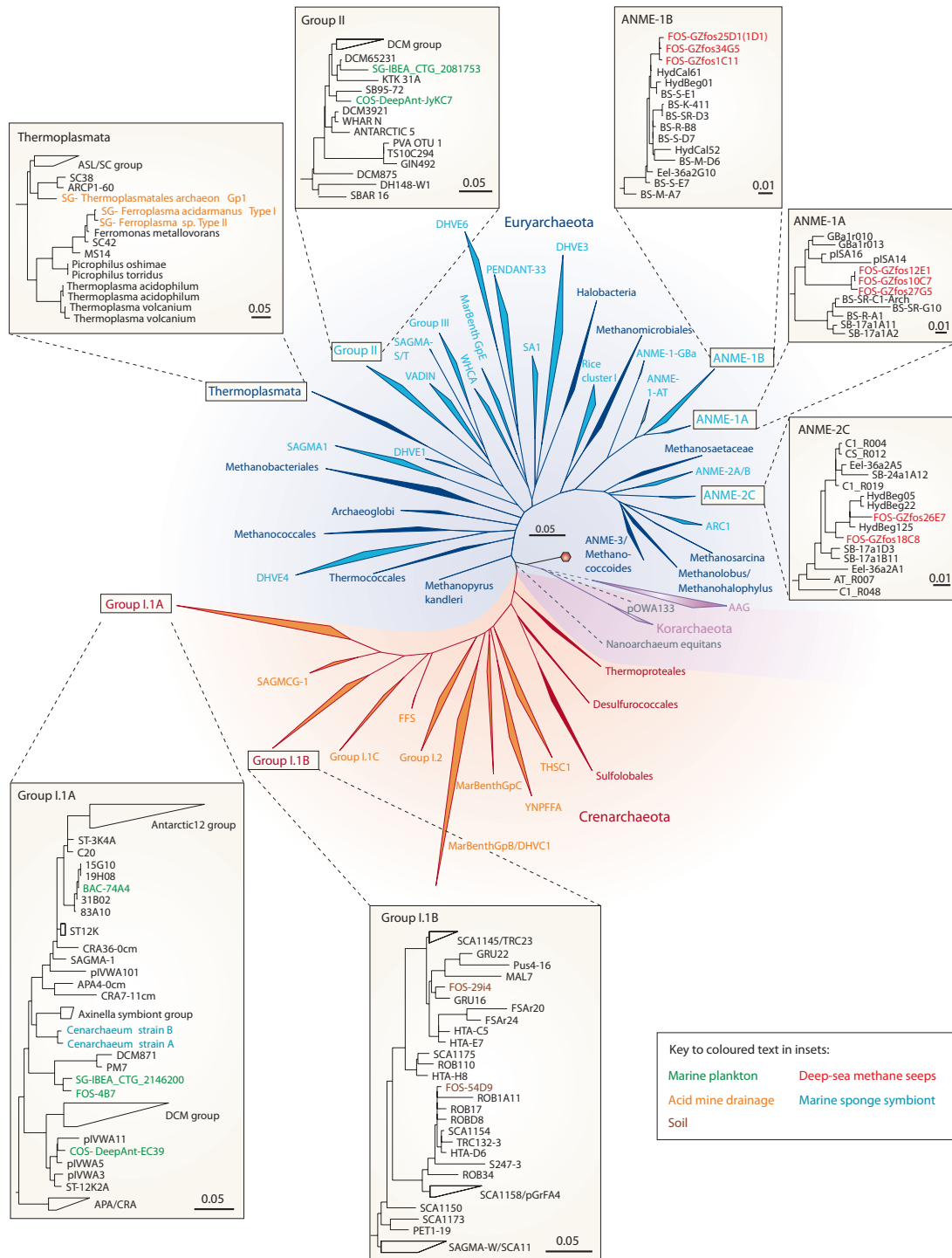


Fig. 5 – Diversité des phyla archaéens. La phylogénie présentée ici a été reconstruite à partir de séquences d'ARNr 16S avec les méthodes de vraisemblance et de maximum de parcimonie. Les espèces non cultivées qui ont fait l'objet d'études génomiques sont détaillées dans les boîtes. Les couleurs claires représentent les branches constituées uniquement d'espèces non cultivées et ceux avec des couleurs sombres représentent les branches où il y a des espèces cultivées. La taille des triangles est proportionnelle à la quantité de séquences analysées (d'après Schleper *et al.*, 2005).

5 L'ère de la génomique

L'origine du terme génomique est récente et fut proposé par Tom Roderick en 1989 pour désigner la science ayant pour sujet l'étude des génomes. Cette nouvelle discipline vise à identifier l'ensemble des gènes d'un organisme vivant. La première séquence d'un génome, celle du bactériophage phiX174 (5386 pb) fut publiée en 1977 (Sanger *et al.*, 1977). Plusieurs génomes de virus, chloroplastes et mitochondries furent ensuite séquencés pendant les 18 années qui suivirent. Ce n'est qu'en 1995, que le premier génome bactérien fut séquencé. Il s'agissait de la bactérie *Haemophilus influenza* dont la séquence (chromosome circulaire de 1830 kpb) fut annotée et publiée par l'équipe américaine du TIGR dirigée par Craig Venter (Fleischmann *et al.*, 1995). Le premier génome d'archée entièrement séquencé fut celui de *Methanocaldococcus jannaschii* (1 chromosome circulaire de 1664 kpb et deux plasmides de 58 kpb et 16 kpb) (Bult *et al.*, 1996). La séquence génomique de *Sacharomyces cerevisiae* (16 chromosomes, 12 Mpb) fut quant à elle la première séquence complète d'un génome eucaryote à être publiée (Goffeau *et al.*, 1996).

Au cours de ces dix dernières années, la génomique a connu un essor extraordinaire. A l'heure actuelle 422 génomes ont été publiés d'après le site GOLD (Genomes OnLine Database : <http://www.genomesonline.org/>). Cette banque de génomes en ligne permet de contrôler l'avancée des projets de séquençage dans le monde. Parmi les 422 génomes publiés, 28 sont ceux d'*Archaea*, 352 de *Bacteria* et 42 d'eucaryotes. De nombreux autres projets de séquençage sont également en voie d'achèvement dans quelques grands centres de séquençage (Sanger, Evry, TIGR, etc.). La connaissance des génomes apporte une information unique pour étudier le fonctionnement et l'évolution de l'organisme (Nelson, 2003). L'analyse génomique d'un organisme permet de comprendre sa physiologie cellulaire, d'appréhender un très grand nombre de ses processus biologiques et de ses activités métaboliques, soit par comparaison avec des résultats expérimentaux, soit par analogie avec un système modèle. Le séquençage de génomes complets peut également mettre en évidence des transferts de gènes (Nelson *et al.*, 1999) et permettre de mieux comprendre les mécanismes de virulence (Tettelin *et al.*, 2001).

5.1 Caractéristiques des génomes procaryotes

A la différence des génomes eucaryotes, les procaryotes ont des génomes plus simples et plus compacts (quelques Mpb). Ces génomes sont constitués d'un chromosome circulaire avec lequel peuvent coexister d'autres structures extra-chromosomiques appelées plasmides. Les génomes procaryotes ne comprennent généralement que 10% de séquences non codantes contre 90% chez les eucaryotes.

Par définition, le gène désigne une séquence nucléotidique (ADN ou ARN) nécessaire à la synthèse d'une protéine ou d'un ARN structural. Les principaux ARN structuraux des organismes

procaryotes sont les ARN de transferts (ARNt) et les ARN ribosomiaux (ARNr). Les ARNr sont des constituants des ribosomes. Les ARNt quant à eux permettent le transfert des acides aminés le long d'une chaîne polypeptidique en cours de formation lors de la traduction.

D'autres critères permettent de distinguer les génomes procaryotes et eucaryotes. Tout d'abord, la séquence de chaque gène est continue dans les génomes procaryotes (il n'y a pas d'introns). Enfin, les gènes des procaryotes sont fréquemment organisés en opéron (ensembles de gènes localisés les uns à la suite des autres, transcrits de façon coordonnée, sous le contrôle d'une protéine régulatrice). Les protéines ainsi produites peuvent intervenir dans une même voie métabolique, ou bien constituer les différentes sous-unités d'une protéine multimérique.

5.2 Annotation de génomes procaryotes

L'annotation d'un génome s'articule autour de deux principaux axes. La première phase connue sous le nom de génomique structurale consiste à cartographier les génomes : à en décrire leur organisation, à identifier les objets génomiques présents sur la séquence comme par exemple les signaux nécessaires à l'expression des gènes et les gènes codant des protéines ou des ARN (Tableau 3). La deuxième phase appelée génomique fonctionnelle consiste à attribuer une fonction biologique à ces objets et/ou à leurs produits et à en définir leurs interactions.

5.2.1 Analyse structurale

Recherche des signaux des gènes microbiens Les signaux nécessaires à l'expression des gènes tels que les codons d'initiation, les codons de terminaison, les sites de liaison du ribosome RBS et les promoteurs sont des motifs particuliers que l'on retrouve le long d'une séquence génomique. Leur localisation permet de délimiter les régions codantes d'une séquence nucléique (Fig. 6).

Les RBS Ils sont en général situés 8 à 10 pb en amont du site du démarrage de la traduction et caractérisés par un motif Shine-Dalgarno (SD), une séquence de 6 pb riche en purine et complémentaire à l'extrémité 3' de l'ARNr 16S. Cette séquence qui s'apparie à l'extrémité 3' de l'ARNr 16S permet au ribosome de repérer le codon start. Ce motif est sujet à peu de variations chez les procaryotes, reflet de la conservation des séquences d'ARNr 16S. Le programme RBSfinder (<http://www.genomics.jhu.edu/RBSfinder/>) permet de localiser les RBS.

Les promoteurs et terminateurs de la transcription Un promoteur est une région d'ADN située en amont des gènes comportant les sites de fixation de l'ARN polymérase ainsi que les sites de fixation de protéines régulatrices de la transcription. Les promoteurs des domaines *Eukarya*, *Archaea* et *Bacteria* contiennent plusieurs blocs de séquences conservées (box) qui servent à l'adressage de facteurs spécifiques de la transcription. Le promoteur des *Bacteria*

	Annotation	Objectifs
Analyse structurale	<p>À l'échelle nucléique</p> <ul style="list-style-type: none"> • Usage du code génétique des régions codantes • Statistiques sur les mots (oligonucléotides / recherche de répétitions) • Recherche de signaux et de régions promotrices • Comparaison avec les banques d'ADNc 	<ul style="list-style-type: none"> → Identification des gènes → Identification des régions de transferts horizontaux
	<p>À l'échelle protéique</p> <ul style="list-style-type: none"> • Comparaison avec les banques de séquences protéiques • Recherche de motifs structuraux • Recherche de régions transmembranaires, de peptides signaux • Prédiction de structures secondaires et tertiaires 	<ul style="list-style-type: none"> → Caractérisation de la fonction biologique des gènes identifiés
Analyse fonctionnelle	<p>Tous les gènes d'un organisme</p> <ul style="list-style-type: none"> • Construction de familles de paralogues • Interférence de voies métaboliques • Définition de classes de gènes : usage des codons dans les gènes, classes fonctionnelles • Reconstruction des réseaux de régulation de l'organisme 	<ul style="list-style-type: none"> → Identification de duplications, mise en évidence de mécanismes de variabilité antigénique chez les organismes pathogènes → Approche de la fonction biologique des gènes par exploration de voisinage (similarité en séquence intra-espèce, proximité chromosomique, métabolique, biais de codage) → Identification de structure en opéron
	<p>Génomique comparative</p> <ul style="list-style-type: none"> • Construction de familles d'orthologues • Analyse de l'organisation génomique en groupe de synténie • Reconstruction des voies métaboliques par comparaison avec des génomes modèles • Analyse «différentielle» des gènes de l'organisme avec des génomes modèles voisins 	<ul style="list-style-type: none"> → Identification des signaux de régulation conservés au voisinage des orthologues → Étude des interactions physiques entre gènes voisins (interactome) → Approche de la fonction biologique des gènes par exploration de voisinage (similarité en séquence inter-espèce, voisinage chromosomique et métabolique) → Identification des fonctions absentes ou spécifiques d'un génome donné → Caractérisations des erreurs d'annotation

Tableau 3 – Niveaux d'analyse informatique des séquences génomiques (d'après Médigue *et al.*, 2002).

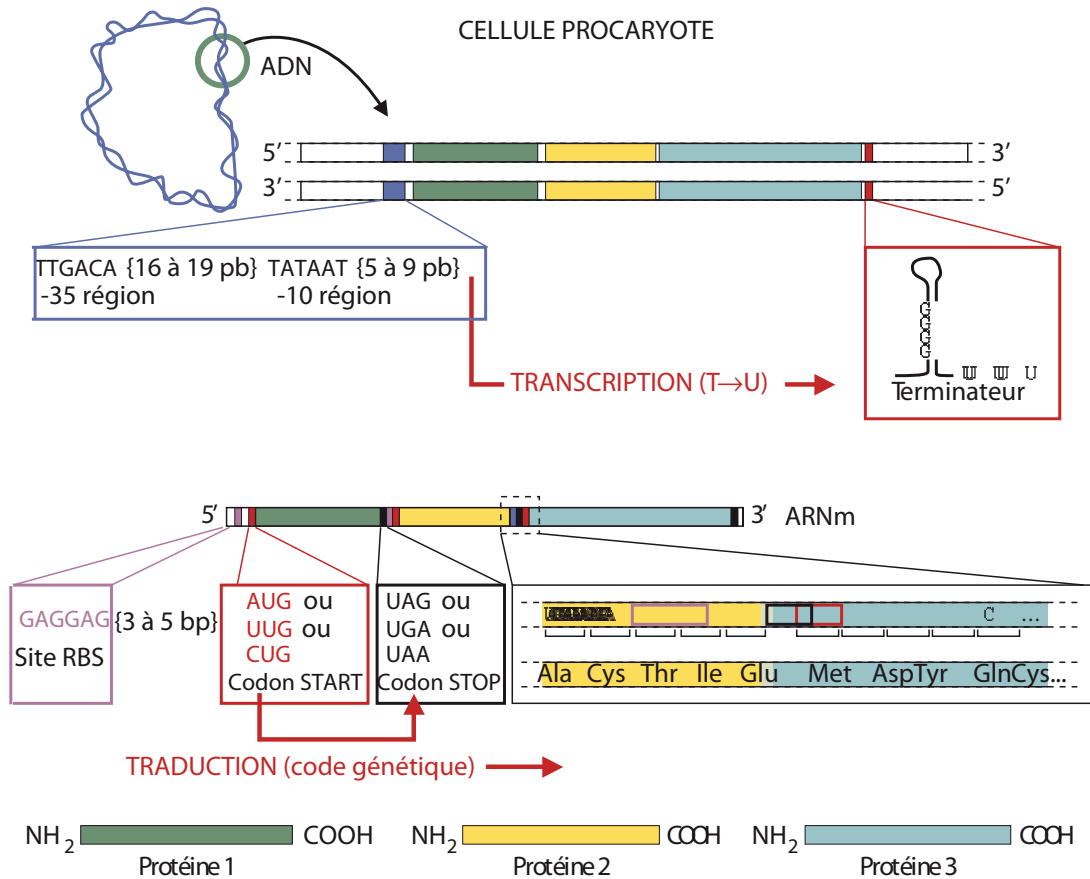


Fig. 6 – Structure des gènes chez les *Bacteria*. Les processus moléculaires associés à la réplication (ADN→ADN), à la transcription (ADN→ARN) et à la traduction (ARN→protéine) se déroulent dans un même compartiment représenté par la cellule de l'organisme. Plusieurs signaux permettent la reconnaissance des zones à transcrire, à savoir la région promotrice sur laquelle se fixe l'ARN polymérase pour déclencher la transcription (boîtes -35 et -10) et une région de terminaison (dite indépendante de rho) qui correspond à une structure secondaire en tige-boucle au niveau de laquelle l'ARN polymérase se décroche (Terminateur). Les gènes sont le plus souvent rassemblés en opéron, c'est-à-dire un groupe de gènes exprimés en même temps sous le contrôle d'une protéine régulatrice. Les mécanismes de transcription et de traduction se produisent de façon simultanée : dès que le ribosome peut, au niveau du site RBS (ribosome binding site), se fixer sur la molécule d'ARN messager en cours de fabrication, la traduction de la protéine est mise en route avant même que la transcription soit achevée. Cette traduction débute au niveau du codon d'initiation formé le plus souvent des trois lettres AUG (et plus rarement CUG ou UUG), et se termine par un des trois condons de terminaison universels, UAA, UAG et UGA (d'après Médigue *et al.*, 2002).

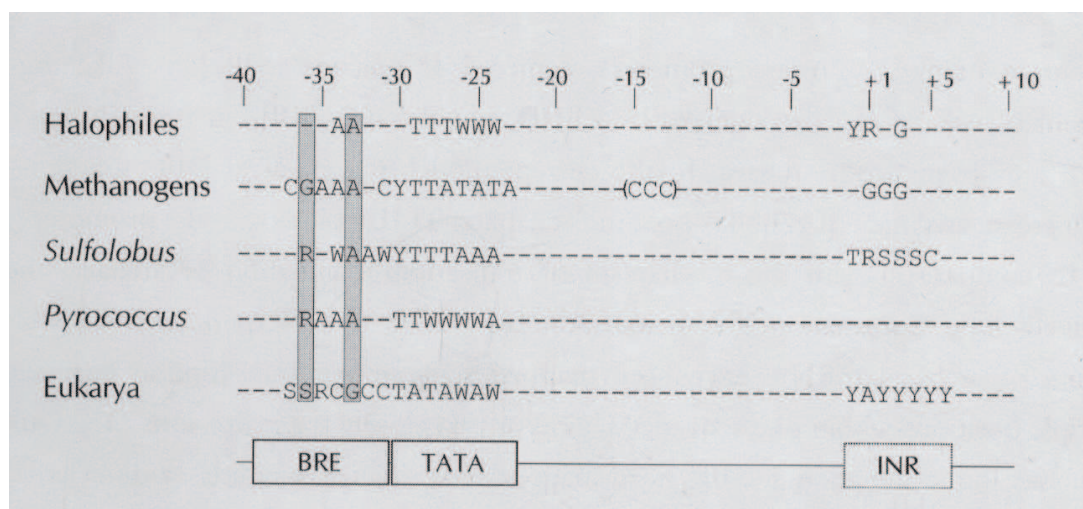


Fig. 7 – Consensus des promoteurs de la transcription chez différents groupes archaéens (d'après Brinkmann, 2002).

comprend trois éléments : la boîte -35, la boîte TATA vers -10 et le site d'initiation de la transcription. L'étude de différents génomes d'*Archaea*, de *Bacteria* et d'*Eukarya*, a montré que la structure du promoteur des *Archaea* correspondait quant à lui, à une version simplifiée de celui des *Eukarya* (Bell & Jackson, 1998). Les promoteurs archaéens ont des séquences de type boîte TATA située environ 30 pb en amont du site d'initiation de la transcription et une boîte BRE (élément de reconnaissance du facteur B) située en amont de la boîte TATA (Brinkmann, 2002) (Fig. 7). Aucun programme de prédiction de promoteur archaéen n'est disponible, une recherche manuelle des motifs consensus est donc effectuée pour les localiser.

Dans le modèle archaéen du complexe d'initiation de la transcription interviennent : une ARN polymérase de type II, la protéine TBP (TATA-box Binding Protein) et les facteurs de transcription TFB (Transcription Factor B) et TFE (Transcription Factor E). La protéine TBP se fixe sur la boîte TATA et permet le recrutement de l'ARN polymérase à l'aide du TFB. Ce dernier se fixe sur le BRE (TFB Recognition Element) (Fig. 8).

La terminaison de la transcription chez les procaryotes peut s'effectuer selon deux systèmes qui se distinguent sur l'intervention ou non du facteur protéique rho. Le système rho-dépendant n'a pas été identifié chez les *Archaea* (Washio *et al.*, 1998 ; Brinkmann, 2002). Le second système dit rho-indépendant a lui été identifié aussi bien chez les bactéries que chez des archées. Chez ces dernières, le système de terminaison repose sur une séquence riche en thymidines. Cette succession de thymidines sur le brin d'ADN n'est pas précédée de séquences pouvant former des structures en tige boucle comme il a été observé dans le système rho-indépendant de terminaison des bactéries (Santangelo & Reeve, 2006).

Codon d'initiation et codon stop de la traduction Trois codons d'initiation de la traduction peuvent être utilisés chez les procaryotes : les codons ATG, TTG et GTG. Le codon

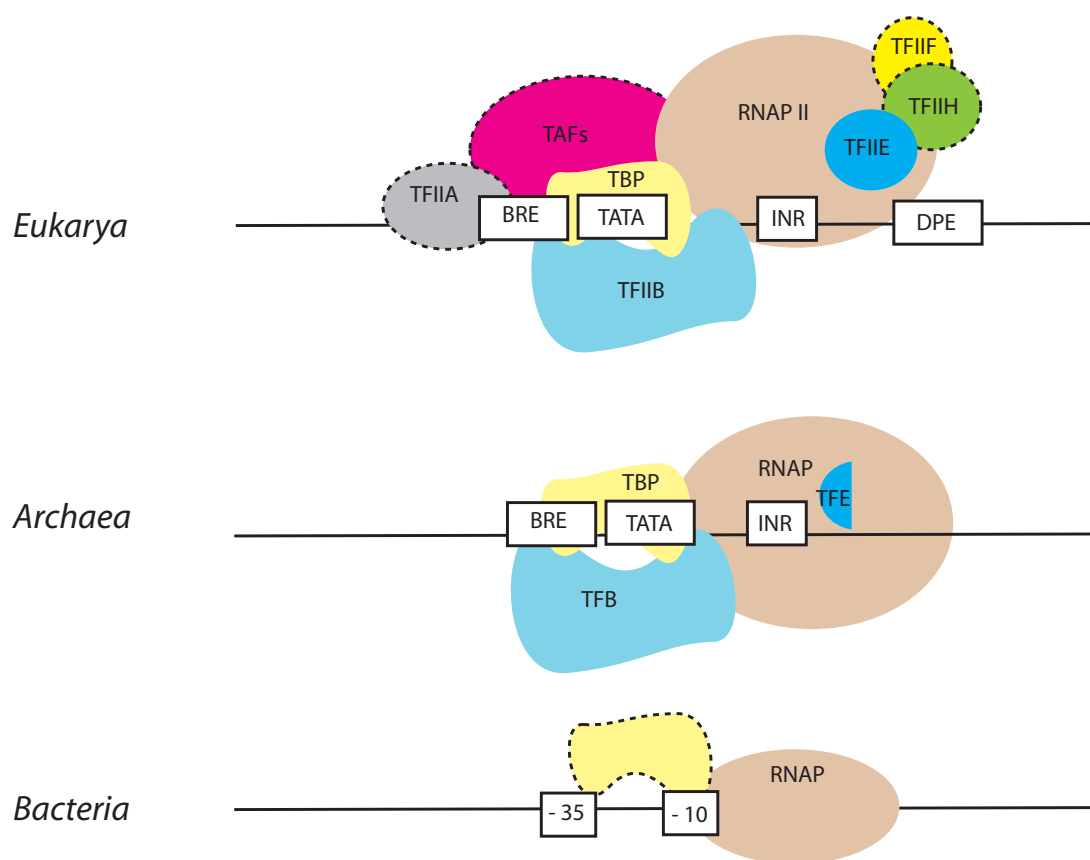


Fig. 8 – Représentation schématique du complexe de pré-initiation et des promoteurs dans les trois domaines de la vie (d'après Brinkmann, 2002). Les facteurs avec un contour noir en pointillé correspondent aux protéines absentes chez les *Archaea*. INR, codon start ; RNAP, ARN polymérase.

méthionine ATG est préférentiellement utilisé. Il peut représenter 70 à 90% des codons d'initiation prédits chez la plupart des génomes archaéens (Torarinsson *et al.*, 2005). Les codons stop sont quant à eux les suivants : TAA, TAG ou TGA.

Prédiction des séquences codantes La prédiction des ARNr sur une séquence génomique est assez simple dans la mesure où leurs séquences sont relativement bien conservées. Il suffit donc de rechercher le long de la séquence génomique des similitudes avec des séquences d'ADNr/ARNr en utilisant des programmes tels que BLAST (<http://www.ncbi.nlm.nih.gov/BLAST>) (Altschul & Lipman, 1990). Les ARNt quant à eux ont la particularité d'adopter une structure secondaire qui leur est propre. Des logiciels tel que tRNAscan-SE (<http://www.genetics.wustl.edu/eddy/tRNAscan-SE>) (Lowe & Eddy, 1997) reposent sur cette propriété des ARNt pour les prédire sur une séquence génomique.

La prédiction des gènes chez les procaryotes peut se faire à l'aide de programmes comme GeneMark.hmm (<http://www.ebi.ac.uk/genemark/>) et le programme Gene Locator and Interpolated Markov Modeler (GLIMMER) (<http://cbbcb.umd.edu/software/glimmer/>). L'analyse statistique du biais des codons, c'est à dire l'usage des codons pour encoder les acides aminés, ainsi que le pourcentage en G+C du génome et la fréquence de la position des bases, fréquence avec laquelle chacune des 4 bases occupe chacune des 3 positions dans le codon, sont autant de signes discriminatoires entre une région codante et une région non codante. Une fois que les séquences codantes ont été prédites par les logiciels, on peut affiner cette prédiction en tenant compte des signaux des gènes décrits précédemment.

5.2.2 Analyse fonctionnelle

Une fois les séquences codantes prédites, il faut ensuite identifier leur structure et prédire leur fonction. Cette seconde étape d'annotation d'une séquence consiste à comparer les gènes hypothétiques avec les séquences de gènes dont la fonction est connue. Cette comparaison de gènes est basée sur l'identification des gènes homologues. Par définition, des gènes homologues, sont des gènes dérivant d'un même ancêtre commun. Il y a deux types d'homologie : l'orthologie, et la paralogie. L'orthologie définit des gènes homologues qui ont divergés suite à un phénomène de spéciation. Des gènes paralogues sont des gènes homologues issus d'un événement de duplication, au sein d'un organisme. Contrairement aux gènes orthologues, des gènes paralogues peuvent diverger suffisamment, pour exprimer des fonctions différentes.

La comparaison des séquences nucléiques se fait couramment à l'aide du logiciel d'alignement BLAST (Altschul *et al.*, 1990). Ce logiciel compare la séquence requête hypothétique à toutes les séquences d'une banque de données. Une des interfaces la plus souvent utilisée se trouve sur le site du NCBI : <http://www.ncbi.nlm.nih.gov/BLAST/>.

5.3 L'utilisation de la bioinformatique pour l'analyse des génomes : banques de données et plates-formes d'annotation

Les progrès réalisés dans les techniques de séquençages (robotisation) et d'assemblages des génomes ainsi que la réduction des coûts, ont permis l'essor de la génomique. L'analyse de séquences génomiques a engendré des quantités énormes de données qui ne pouvaient être traitées manuellement. L'outil informatique est devenu par conséquent un complément essentiel à l'analyse des génomes. De nombreuses banques de données de séquences accessibles sur Internet et des outils de bioinformatiques ont ainsi été développés.

5.3.1 Les banques de données

Afin de trier les quantités énormes de données engendrées par la génomique, des banques de données ont été développées. Les premières banques de données papier ont été compilées dès 1965 (Atlas of Protein Sequences) (Dayhoff, 1965). Puis rapidement, dans les années 1980, des banques de données informatisées ont été créées et mises en ligne sur Internet. Parmi les nombreuses banques qui sont en libre accès sur Internet (719 banques de données de biologie moléculaire recensées en 2005 (Galperin, 2005), deux types de banques sont à distinguer : les premières sont dites généralistes tandis que les secondes sont dites spécialisées. Dans les banques généralistes, les séquences nucléotidiques et/ou protéiques sont collectées et stockées de la façon la plus exhaustive possible, offrant un ensemble plutôt hétérogène d'informations. Les inconvénients de ces banques généralistes sont qu'elles souffrent de leur hétérogénéité dans la nature des séquences (ADN nucléaire, ADN mitochondrial, ARNm, ARNt...), de leur hétérogénéité d'annotation et de leur redondance (il est fréquent de retrouver dans une banque généraliste plusieurs entrées correspondant à un même gène). Devant la croissance des banques généralistes, les banques spécialisées sont apparues (Galperin, 2005). Les informations stockées et diffusées dans ces bases de données spécialisées sont plus homogènes. Elles peuvent par exemple être dédiées à un organisme en particulier, un organisme cellulaire ou à une thématique donnée.

Les principales banques de données nucléiques et protéiques généralistes et spécialisées sont listées dans l'annexe A.

5.3.2 Serveurs et plateformes de bioinformatiques

L'Institut Pasteur (<http://bioweb.pasteur.fr/>) et le Pôle Bioinformatique de Lyon (<http://pbil.univ-lyon1.fr/>) proposent tous deux un ensemble de logiciels dédiés à l'analyse des séquences nucléiques et protéiques. De plus, plusieurs plateformes bioinformatiques ont été mises au point pour faciliter le travail d'annotation et d'exploration des génomes. C'est le cas des plateformes Artemis et Genostar qui ont été utilisées dans le cadre de ce travail et dont voici une présentation succincte :

Artemis est un logiciel gratuit de visualisation et d'annotation manuelle des séquences d'ADN (Rutherford *et al.*, 2000). Ce logiciel développé par le centre Sanger (<http://www.sanger.ac.uk/Software/Artemis/>) permet de voir grâce à une interface graphique conviviale les caractéristiques des séquences et les résultats des annotations dans les six phases de lecture. Artemis est adapté à l'analyse des génomes compacts, tels que ceux des procaryotes et des eucaryotes inférieurs. Il peut lire les entrées complètes des bases de données EMBL et GENBANK, les séquences en format FASTA ou les données brutes. Une fois la séquence génomique chargée dans le logiciel, Artemis permet entre autre de déterminer les propriétés (recherche d'ORFs, contenu en G+C, calcul du biais en G/C, usage des codons, hydrophobicité et hydrophilicité...) des objets génomiques caractérisés (séquence nucléique, CDS, protéines déduites de la séquence nucléique...), et de les sauvegarder au format EMBL, GenBank ou GFF. L'exécution de programmes tels que Blastp et Fasta sur les CDS choisies est réalisée à l'extérieur d'Artemis, mais le résultat peut ensuite être visualisé sur l'interface Artemis.

Genostar (<http://www.genostar.org>) est une plate-forme bioinformatique de génomique exploratoire créée en 1999 par un consortium privé/public composé de GenomeExpress, Hybrigenics, l'INRIA et l'Institut Pasteur. Cette plate-forme est composée de 3 modules, GenoAnnot, GenoLink et GenoBool. Ces trois modules s'articulent autour d'un noyau en charge de la gestion et de la persistance des données, appelé GenoCore. Le module GenoAnnot est destiné à l'annotation de séquences génomiques. GenoLink permet de relier et de confronter des données et des informations de nature très variée, qu'elles soient produites dans Genostar ou importées, comme par exemple la similarité entre gènes, l'appartenance à des familles fonctionnelles, ou encore l'existence de synténies ou d'interactions physiques entre protéines. Enfin, GenoBool aide à la mise en évidence de liens significatifs entre des collections de données hétérogènes. Dans ce but, il offre une interface de manipulation et de codage de ces données ainsi qu'un ensemble de méthodes statistiques d'analyses multi-variées. Les applications concernent, par exemple, l'analyse de l'usage des codons ou de biais liés à la réplication.

L'architecture de la plate-forme Genostar est réalisée sous le modèle d'une base de données (modèle relationnel), permettant aisément le passage d'un module à un autre. L'avantage d'une telle plate-forme par rapport à celle que nous avons précédemment décrite réside dans le fait qu'il s'agit d'une plate-forme modulaire et extensive, dans laquelle on peut facilement ajouter de nouvelles méthodes d'analyses au sein de stratégies personnalisables.

L'ÈRE DE LA MÉTAGÉNOMIQUE

1	La métagénomique : définition, principe et historique	31
2	Métagénomique : intérêts de cette approche	31
2.1	Etudier les communautés microbiennes	31
2.2	Accéder aux génomes des micro-organismes incultivés	33
2.3	Capturer des opérons ou des gènes d'intérêts	34
3	Méthodologie : construction, conservation et caractérisation d'une banque métagénomique	36
3.1	Extraction de l'ADN environnemental	36
3.1.1	Extraction d'ADN d'échantillons aquatiques	36
3.1.2	Extraction d'ADN lié à une matrice solide	36
3.2	Enrichir un échantillon environnemental en ADN génomique d'intérêt	39
3.3	Choix du vecteur de clonage	41
3.3.1	Vecteurs permettant le clonage d'inserts d'ADN de grande taille	41
3.3.2	Les vecteurs d'expression	43
3.3.3	Les vecteurs navettes	44
3.3.4	Les vecteurs inductibles	44
3.4	Choix des cellules hôtes	44
3.5	Caractérisation d'une banque métagénomique	45
3.5.1	Analyse basée sur les séquences	45
3.5.2	Analyse basée sur une activité	47
3.6	Estimer l'effort de séquençage afin de reconstituer les génomes d'une communauté microbienne donnée	48
3.7	Méthode de reconstruction de génomes	48
3.8	Conservation d'une banque métagénomique	49

4	Exemples d'études métagénomiques ayant permis de faire le lien entre l'identité phylogénétique et une fonction spécifique d'un micro-organisme	50
4.1	Découverte d'une nouvelle bactériorhodopsine	50
4.2	Métagénomique des <i>Archaea</i>	50
4.2.1	Les crenarchaeotes du sol	51
4.2.2	Les <i>Archaea</i> marines planctoniques	52
4.2.3	<i>Cernarchaeum symbosium</i> , un organisme archaéen modèle	53
4.3	La symbiose des bactéries sulfo-oxydantes avec le vestimentifère <i>Riftia pachyptila</i>	54
4.4	L'épibiose bactérienne d' <i>Alvinella pompejana</i>	55
5	Exemples d'analyses métagénomiques à l'échelle des communautés microbiennes : reconstruction de génomes et/ou de voies métaboliques	57
5.1	Etude du métagénome de la mer des Sargasses	57
5.1.1	Une production de données considérables	57
5.1.2	Photobiologie dans les océans	58
5.1.3	Biogéochimie des océans	59
5.1.4	Difficultés liées à l'analyse du métagénome de la mer des Sargasses	59
5.2	Etude métagénomique d'un biofilm microbien échantillonné dans le drainage acide d'une mine	61
5.2.1	L'étude de cycles biogéochimiques à l'aide de la métagénomique	61
5.2.2	Caractéristiques de l'eau de drainage de la mine Richmond	61
5.2.3	Assemblage des séquences de la communauté microbienne du biofilm	62
5.2.4	Analyse métabolique	62
5.3	Oxydation anaérobie du méthane dans les sédiments marins profonds	64
5.4	Reconstitution du génome de la bactérie <i>Kuenenia stuttgartiensis</i> , un acteur majeur du cycle de l'azote	66

1 La métagénomique : définition, principe et historique

Depuis les années 1980, de nombreuses études de génomique microbienne ont été réalisées sur des souches isolées en laboratoire. La génomique microbienne a permis d'améliorer profondément notre compréhension de la physiologie, de la biochimie, de la génétique, de l'écologie et de l'évolution microbienne. A la fin des années 1990, une nouvelle approche dérivée de la génomique microbienne a vu le jour grâce aux progrès spectaculaires de la biologie moléculaire, de la bioinformatique et des techniques de séquençage ainsi qu'à la baisse significative du coût du séquençage à haut débit. Cette nouvelle approche connue sous le nom de métagénomique permet désormais d'accéder à l'information génomique des communautés microbiennes de l'environnement, ceci indépendamment de la mise en culture de micro-organismes qui les composent (Fig. 1). D'autres termes sont également employés pour décrire cette même discipline, à savoir : génomique environnementale, écogénomique, génomique des communautés ou génomique des populations (Handelsman, 2004).

Le principe de la métagénomique consiste à cloner dans un vecteur approprié de l'ADN extrait directement d'une communauté microbienne donnée (provenant par exemple d'un échantillon environnemental), puis à en explorer le métagénome (Schloss & Handelsman, 2003). Le métagénome est quant à lui défini comme étant l'ensemble des génomes des micro-organismes présents dans une communauté microbienne donnée. L'exploration du métagénome représente une nouvelle stratégie conduisant à faire l'inventaire complet des gènes présents dans un environnement naturel (Dauga *et al.*, 2005).

Le concept de la métagénomique a été initialement énoncé par Pace (1985). Par la suite, Schmidt et ses collaborateurs (1991) furent les premiers à mettre en application cette idée en clonant directement de l'ADN extrait d'échantillons d'eau de mer dans des phages lambda. La technique fut ensuite reprise et améliorée par DeLong et son équipe qui clonèrent des grands fragments d'ADN planctonique marin (40kpb) dans des fosmidés, marquant le point de départ de la métagénomique (Stein *et al.*, 1996). Ce n'est qu'en 1998, que le terme métagénomique fut introduit pour la première fois (Handelsman *et al.*, 1998).

2 Métagénomique : intérêts de cette approche

2.1 Etudier les communautés microbiennes

Les micro-organismes dans la nature ne vivent pas seuls, mais forment des assemblages complexes dits communautés microbiennes. Ces communautés sont constituées de plusieurs espèces microbiennes qui s'organisent entre elles et co-évoluent ensemble. Chaque espèce au sein de la communauté possède ses propres fonctions ou activités qui se complètent d'ailleurs bien

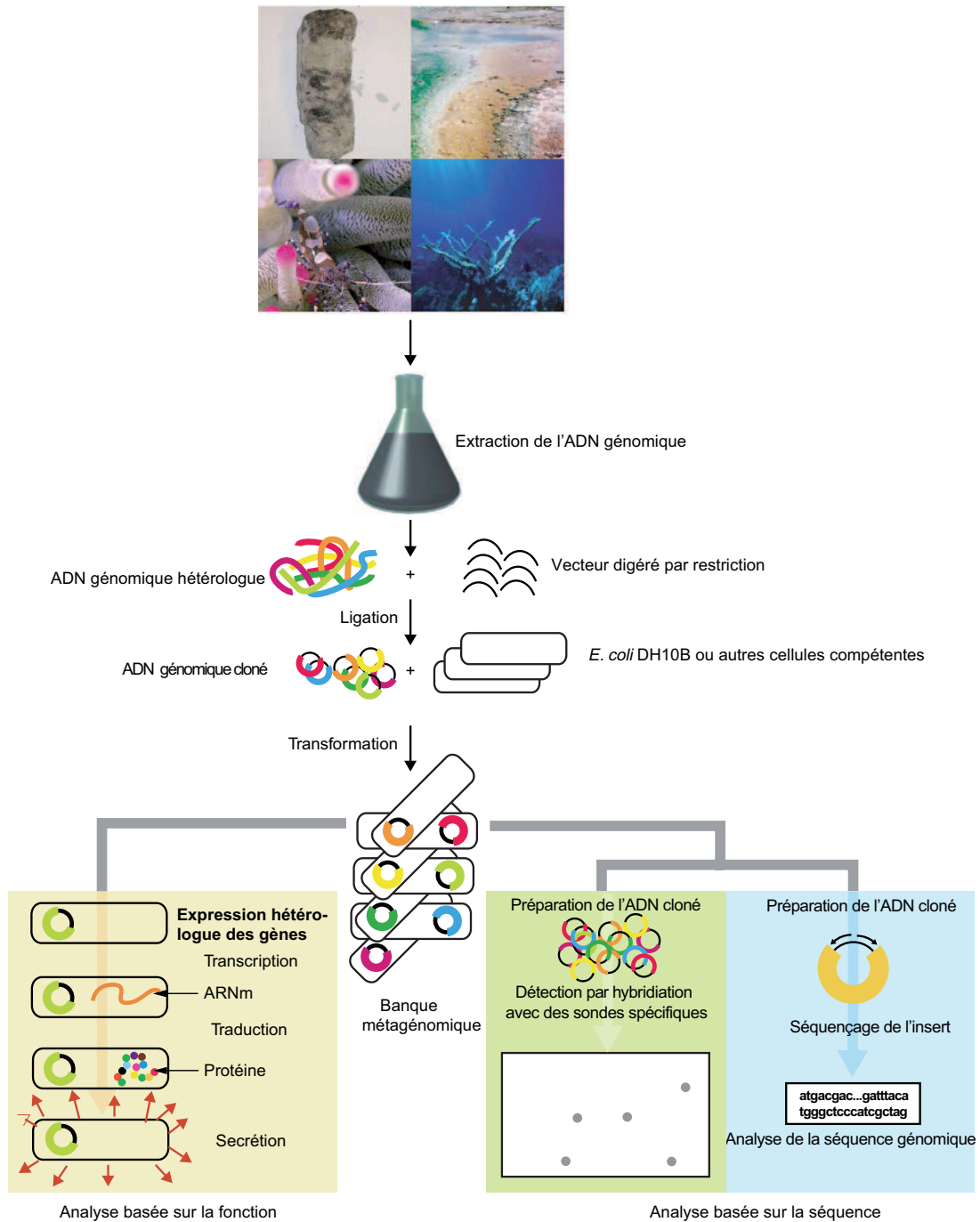


Fig. 1 – Principe de la construction et du criblage d'une banque métagénomique. Les photographies sont de gauche à droite et de haut en bas : un échantillon de sol, des biofilms microbiens du parc de Yellowstone, des mézozoaires de l'océan profond et des éponges (d'après Schloss & Handelsman, 2003).

souvent. Elles interagissent entre elles en échangeant par exemple du matériel génétique, des nutriments et/ou des signaux biochimiques. Si par des techniques d'isolement on individualise ces espèces microbiennes, on les soustrait complètement des forces qui les maintiennent au sein de la communauté. Ces micro-organismes une fois isolés ne peuvent plus entretenir de relations avec les autres espèces microbiennes, comme cela se passait dans leur habitat naturel. Ainsi, l'analyse génomique du pool des gènes d'une communauté microbienne révèle d'avantages de propriétés que si l'on isolait les membres de cette communauté et qu'on en étudiait leur information génomique.

Avant la mise au point de la métagénomique, les membres d'une communauté microbienne devaient bien souvent être isolés pour pouvoir faire l'objet d'une étude génomique. Le principe de la métagénomique étant d'analyser le pool génomique des communautés microbiennes sans avoir à isoler préalablement les micro-organismes, cette approche permet ainsi de mieux comprendre l'organisation des communautés microbiennes et le rôle écologique des micro-organismes qu'elles contiennent. Elle permet également de révéler les processus évolutifs des micro-organismes mettant parfois en évidence des événements de transferts horizontaux de gènes (Lopez-Garcia *et al.*, 2004).

2.2 Accéder aux génomes des micro-organismes incultivés

On a estimé que plus de 99% des micro-organismes observables dans la nature ne pouvaient être cultivés en utilisant les techniques de culture classiques (Amann *et al.*, 1995). Grâce à la métagénomique, on peut capturer de grands fragments d'ADN génomique provenant d'échantillons naturels. L'analyse de ces fragments d'ADN génomique et leur identification à l'aide de marqueurs phylogénétiques, permet d'explorer le potentiel physiologique des micro-organismes incultivables. Ainsi, la physiologie des micro-organismes que l'on croyait difficilement accessible, comme par exemple les micro-organismes pathogènes et les symbiotes obligatoires (qui ne peuvent survivre en dehors de leur hôte) est désormais accessible grâce à l'approche métagénomique. On peut également accéder au génome d'anciens micro-organismes pour lesquels l'ADN fossile est la seule trace encore disponible (Tringe *et al.*, 2005). Par le biais de l'approche métagénomique, le séquençage de l'ADN d'une espèce d'ours éteinte depuis 40 000 ans a récemment été mené (Noonan *et al.*, 2005).

De plus, une analyse métagénomique conduite à grande échelle permet éventuellement de reconstruire des génomes d'organismes incultivés, en assemblant un à un les fragments génomiques se chevauchant.

2.3 Capturer des opérons ou des gènes d'intérêts

Enfin, la métagénomique peut servir de support à des études fonctionnelles. Ces études incluent :

- le contrôle des activités protéiques dans l'environnement (Béjà *et al.*, 2000a),
- la caractérisation biochimique de protéines ayant, après leur expression chez *E. coli*, permis d'inférer à des micro-organismes incultivés une caractéristique biologique donnée (à titre d'exemples : Schleper *et al.*, 1997b ; Béjà *et al.*, 2000a),
- enfin, l'identification de nouvelles enzymes ou métabolites secondaires d'intérêt industriel ou pharmaceutique.

La diversité microbienne encore inexplorée représente potentiellement un immense réservoir de nouvelles classes de gènes ayant des fonctions connues ou non. Ainsi, plus d'un million de nouveaux gènes ont par exemple été identifiés lors de l'analyse du métagénome de la mer des Sargasses (Venter *et al.*, 2004). Cette diversité moléculaire laisse présager de la découverte de nouvelles enzymes, biocatalyseurs, métabolites secondaires¹ ou autres molécules bioactives. On espère ainsi découvrir de nouveaux antibiotiques, antifongiques, antiviraux ou médicaments anticancéreux (Vicente *et al.*, 2003).

L'accès à ces molécules était jusqu'à présent limité par la faible quantité de micro-organismes cultivés. De plus, de nombreux métabolites secondaires connus sont produits soit par des micro-organismes vivant au sein de consortia, soit par des micro-organismes occupant des niches écologiques difficilement reproductibles *in vitro* (Piel *et al.*, 2004). Si plusieurs méthodes ont été à ce jour développées pour contourner les problèmes de culture (Kaeberlein *et al.*, 2002 ; Zengler *et al.*, 2002), la métagénomique semble être l'approche la plus directe et la plus prometteuse pour accéder à ce réservoir encore inexploré de nouveaux gènes.

L'avantage de l'approche métagénomique dans la capture des nouveaux gènes vient du fait que l'on peut directement accéder à la totalité des gènes du métagénome sans qu'aucune connaissance préalable de leur séquence ne soit nécessaire (Schloss & Handelsman, 2003) ; ce qui n'est pas le cas des techniques classiques d'inventaires moléculaires telles que l'amplification par PCR de marqueurs phylogénétiques ou l'hybridation *in situ*. Ces deux techniques nécessitent en effet d'avoir une bonne connaissance de la séquence du marqueur pour dessiner les amorces de PCR ou les sondes d'hybridation.

Plusieurs articles et revues traitant de cette approche sont récemment parues (Handelsman *et al.*, 1998 ; Lorenz & Schleper, 2002 ; Schloss & Handelsman, 2003 ; Daniel, 2004 ; Lorenz & Eck,

¹Un métabolite secondaire est par définition non essentiel à la croissance fondamentale et au développement des organismes qui le produisent. Ces organismes ont acquis la capacité de les produire parce qu'ils leur conféraient un avantage sélectif sur leur environnement, comme par exemple des interactions spécifiques avec des récepteurs présents sur d'autres organismes (Vicente *et al.*, 2003).

Fonction	Habitat	Type de la banque	Taille moyenne des inserts (kpb)	Nombre de clones criblés	Taille de la banque (Mpb)	Substrat	Nombre de clones positifs	Références
Estérase/lipase	Forêt/sol	Plasmide	8	67,000	536	Tributyryne	98	*
Estérase/lipase	Forêt/sol	Fosmide	40	19,968	799	Tributyryne	47	*
Estérase/lipase	Sable	Fosmide	30	29,884	90	Tributyryne	49	*
Estérase/lipase	Sable	Fosmide	40	25,344	1,014	Tributyryne	29	*
Estérase/lipase	Sol	Plasmide	6	286,000	1,716	Tributyryne	3	Henne <i>et al.</i> , 2000
Estérase/lipase	Sol	Plasmide	6	730,000	4,380	Trioléine	1	Henne <i>et al.</i> , 2000
Estérase/lipase	Sol	BAC	27	3,648	100	Bacto-lipide	2	Rondon <i>et al.</i> , 2000
Oxydation des polyols	Sols	Plasmides	3	900,000	2,700	1,2-éthanediol; 1,2-propanediol; 2,3-butanediol	15	Knietsch <i>et al.</i> , 2003a
Alcool oxidoréductase	Sol enrichissement	Plasmide	4	400,000	1,6000	Glycérol/1,2-propanediol	10	Knietsch <i>et al.</i> , 2003b
Amidase	Sol enrichissement	Plasmide	5	193,000	965	D-phenylglycine- L-leucine	7	Gabor <i>et al.</i> , 2004
Amylase	Sol	Plasmide	5	80,000		Amidon	1	Gabor <i>et al.</i> , 2000
Amylase	Sol	BAC	27	3,648	100	Amidon	88	Rondon <i>et al.</i> , 2000
Production de biotine	Sol/excréments enrichissement	Cosmide	35	50,000	1,750	Biotin-déficient medium	7	Entcheva <i>et al.</i> , 2001
Protéase	Sol	Plasmide	10	100,000	1,000	Lait écrémé	1	Gupta <i>et al.</i> , 2002
Cellulase	Sédiment enrichissement	Phage λ	6	310,000	1,860	Carboxyméthyl-cellulose	3	Rees <i>et al.</i> , 2003
Chitinase	Eau de mer	Phage λ	5	825,000	4,125	Methylumbelliferyl-diacetylchitobioside	11	Cottrell <i>et al.</i> , 1999
Déhydratase	Sol/sédiment enrichissement	Plasmides	4	560,000	2,240	Glycérol	2	Knietsch <i>et al.</i> , 2003c
4-Hydroxybutyrate conversion	Sol	Plasmide	6	930,000	5,580	4-Hydroxybutyrate	5	Henne <i>et al.</i> , 1999
-βactamase	Sol	Plasmide	5	80,000	400	Ampicilline	4	Gabor <i>et al.</i> , 2004

E. coli a été utilisé comme cellule hôte pour toutes ces études

* Résultats non publiés, K. Liebeton *et al.*, BRAIN AG.

Tableau 1 – Nouvelles enzymes et biocatalyseurs découverts suite au criblage fonctionnel de banques métagénomiques (d'après Lorenz & Eck, 2005).

Habitat	Type de la banque	Taille moyenne des inserts (kpb)	Nombre de clones criblés	Taille de la banque (Mpb)	Organisme ciblé	Cellule hôte	Nombre de clones positifs	Ref.
Sol	BAC	44,5	24,546	1,092	Several	*	3	Gillespie <i>et al.</i> , 2002
Sol	BAC	63	12,000	756	<i>B. subtilis</i> , <i>S. aureus</i>	<i>E. coli</i>	4	MacNeil <i>et al.</i> , 2001
Sol	BAC	27	3,648	100	<i>B. subtilis</i>	<i>E. coli</i>	1	Rondon <i>et al.</i> , 2000
Sol	BAC	30	13,440	403	<i>S. aureus</i>	<i>E. coli</i>	3	‡
Sol	Cosmide	nd	5,000	nd	<i>B. subtilis</i>	<i>E. coli</i>	1	Courtois <i>et al.</i> , 2003
Sol	Cosmide	30–45	nd	Unknown	<i>B. subtilis</i>	<i>E. coli</i>	10	Brady <i>et al.</i> , 2004

* Colonies identifiées par production d'un pigment. Les pigments isolés ont montré une activité antimicrobienne

‡ Unpublished results, R. Schulze and G. Meurer, BRAIN AG.

B. subtilis, *Bacillus subtilis*; *E. coli*, *Escherichia coli*; nd: non déterminé; *S. aureus*, *Staphylococcus aureus*.

Tableau 2 – Exemples de criblages de banques métagénomiques basés sur la recherche d'activités antimicrobiennes (d'après Lorenz & Eck, 2005).

2005 ; Handelsman, 2005). L'utilisation des micro-organismes incultivés pour le développement de nouveaux produits à haute valeur ajoutée commence à voir le jour. Un nombre croissant d'enzymes (Cottrell *et al.*, 1999 ; Henne *et al.*, 1999, 2000 ; Rondon *et al.*, 2000 ; Richardson *et al.*, 2002 ; Voget *et al.*, 2003 ; Knietsch *et al.*, 2003a,b), d'antibiotiques (Rondon *et al.*, 2000 ; MacNeil *et al.*, 2001 ; Courtois *et al.*, 2003 ; Brady *et al.*, 2004), d'anticancéreux (Piel *et al.*, 2004) ont été ainsi identifiés à partir de métagénomomes provenant de divers environnements tels que le sol, la mer ou des digesteurs anaérobies (Tableau 1 et Tableau 2).

3 Méthodologie : construction, conservation et caractérisation d'une banque métagénomique

3.1 Extraction de l'ADN environnemental

La préparation de l'ADN est une étape cruciale pour la construction d'une banque métagénomique. Pour cette étape, il s'agit d'extraire les acides nucléiques d'un échantillon donné, en quantité suffisante et à la taille désirée. Les acides nucléiques extraits doivent être suffisamment purifiés pour pouvoir être clonés par la suite. Les échantillons naturels contiennent de l'ADN sous différentes formes incluant l'ADN libre, l'ADN viral et l'ADN contenu dans les cellules procaryotes et eucaryotes. Ces ADN peuvent être en suspension dans de l'eau, liés à une matrice solide comme le sol, ou enfermés dans un biofilm ou au sein d'un tissu (Tringe & Rubin, 2005). Les méthodes d'extraction d'ADN doivent être adaptées à l'échantillon et au type d'ADN que l'on veut analyser.

3.1.1 Extraction d'ADN d'échantillons aquatiques

Pour extraire des quantités suffisantes d'ADN à partir d'échantillons aquatiques, ceux-ci doivent être concentrés et éventuellement pré-filtrés pour enlever les débris et les cellules de plus grande taille non souhaitées (Somerville *et al.*, 1989). Des protocoles de filtration ont été mis en place pour sélectionner préférentiellement des particules virales, des cellules procaryotes ou eucaryotes (Breitbart *et al.*, 2002 ; Venter *et al.*, 2004).

3.1.2 Extraction d'ADN lié à une matrice solide

Si extraire de l'ADN pur à partir d'échantillons d'eau de mer ne pose pas de problème, il est techniquement plus difficile d'extraire correctement les acides nucléiques d'échantillons de sol ou de sédiments (Steffan *et al.*, 1988 ; Tsai & Olson, 1992 ; Tebbe & Vahjen, 1993). De nombreux composés présents dans ce type d'échantillons se fixent à l'ADN et inhibent les réactions enzymatiques requises pour le clonage. C'est par exemple le cas des composés polyphéniques (acide humique et acide fulvique) (Lorenz & Schleper, 2002 ; Daniel, 2005).

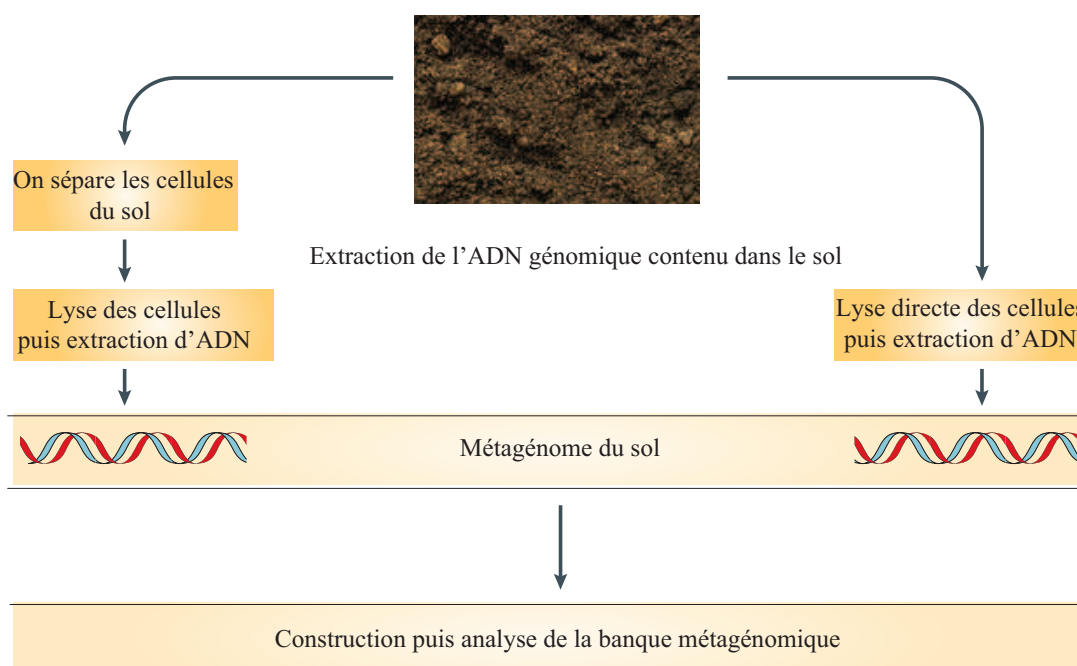


Fig. 2 – Représentation schématique des deux approches utilisées pour extraire de l'ADN génomique contenu dans un échantillon solide (sol ou sédiment), en vue de construire une banque métagénomique (d'après Daniel, 2005).

Plusieurs protocoles d'extraction d'ADN ont été mis au point, et de nombreux kits commerciaux d'extraction d'ADN sont disponibles (Daniel, 2005 ; pour revue). Cependant, ces protocoles d'extraction d'ADN reposent tous sur deux types de méthodes : la méthode de la lyse directe (les cellules sont lysées directement dans l'échantillon), ou indirecte (les cellules sont d'abord extraites de l'échantillon avant leur lyse) (Daniel, 2005) (Fig. 2).

Lyse directe des cellules avant l'extraction d'ADN. Dans la première approche qu'est la lyse directe des cellules (approche initialement expérimentée par Ogram et ses collaborateurs (1987)), l'échantillon est resuspendu dans un tampon de lyse puis est traité successivement avec des détergents et enzymes. S'en suit alors une extraction classique des ADN par l'ajout de phénol/chloroforme. La lyse directe de l'ADN dans l'échantillon permet d'obtenir un bon rendement d'extraction des acides nucléiques. De plus, cette méthode permet de capturer les différents types d'ADN contenu dans cet échantillon (ADN des cellules vivantes, ADN des particules virales, ainsi que l'ADN sous forme libre provenant de cellules mortes).

Certains protocoles prévoient en plus de la lyse chimique des cellules, une lyse mécanique, en utilisant par exemple des billes de verre. Cette approche "drastique" de lyse des cellules permet d'obtenir un bon rendement d'extraction des acides nucléiques ainsi qu'une meilleure représentation de la composition en ADN. Elle permet en effet de s'affranchir des biais d'extraction dus à la

difficulté de lyser certains types d'organismes comme les bactéries à Gram-positif (von Wintzingerode *et al.*, 1997). Cependant, l'ADN extrait par cette méthode est particulièrement fragmenté, rendant moins réalisable la construction de banques métagénomiques à grands inserts. L'ADN extrait sera donc préférentiellement cloné dans des vecteurs de type plasmide ou vecteur lambda portant des promoteurs forts permettant l'expression de gènes.

Lyse indirecte des cellules avant l'extraction d'ADN. Cette technique de lyse indirecte consiste à séparer préalablement les cellules de l'échantillon de sol ou de sédiment, puis à les lyser. Cette méthode d'extraction d'ADN, bien qu'elle soit moins efficace en terme de rendement / de quantité d'ADN obtenu, est moins sévère que la méthode précédente. L'ADN extrait est ainsi moins fragmenté, permettant la construction de banques métagénomiques à grand insert (Courtois *et al.*, 2001). La séparation des micro-organismes de la matrice solide peut se faire par des méthodes mécaniques douces ou chimiques, telles que l'agitation et homogénéisation à l'aide d'un pilon ou l'utilisation de résines échangeuses de cations, suivie par un gradient de densité ou une centrifugation différentielle (Daniel, 2005 ; pour revue).

Une méthode basée sur une technique d'électrophorèse à deux étapes a récemment été mise au point pour permettre une purification rapide d'ADN de haut poids moléculaire à partir d'échantillons du sol (Quaiser *et al.*, 2002). Cette méthode consiste à inclure la biomasse d'un échantillon de sol dans des plugs d'agarose. Au moyen d'une première électrophorèse en champ pulsé contenant 2% de polyvinylpyrrolidone, on élimine les composés polyphénoliques qui ont tendance à être co-purifiés avec l'ADN. Une deuxième électrophorèse en champ pulsé ne contenant plus cette fois-ci de polyvinylpyrrolidone, permet ensuite de sélectionner la taille de l'ADN désirée. Cette méthode permet de concentrer et de purifier suffisamment d'ADN de haut poids moléculaire pour qu'il soit cloné dans des vecteurs acceptant des inserts de grande taille.

Comparaison des deux méthodes de lyse. Des extractions d'ADN réalisées en parallèle avec les deux méthodes d'extraction (lyse directe et indirecte) ont révélé que toutes deux permettaient d'obtenir, à l'exception de la sous-classe des γ -*Proteobacteria*, des représentations phylogénétiques comparables en ce qui concerne le spectre de la diversité bactérienne des échantillons (Courtois *et al.*, 2001). Le choix de telle ou telle méthode pour extraire l'ADN se fera donc de la manière suivante : si l'on veut obtenir de l'ADN de haut poids moléculaire par une méthode rapide et facile on choisira plutôt la méthode d'extraction directe ; par contre, si c'est la qualité et la pureté de l'ADN extrait que l'on privilégie, on choisira alors la méthode d'extraction indirecte. Notons également que si l'on construit une banque métagénomique dans le but d'identifier de nouveaux gènes isolés faisant entre 1 et 2 kpb, la méthode de lyse indirecte sera bien adaptée ; par contre, si on cherche à obtenir des clusters entiers de gènes, la méthode de lyse indirecte sera dans ce

cas plus adaptée. A titre d'exemple, les polyketides synthases² apparaissent dans des clusters de gènes qui peuvent dépasser les 100 kpb d'ADN contigu (Schwecke *et al.*, 1995).

3.2 Enrichir un échantillon environnemental en ADN génomique d'intérêt

Un des inconvénients de l'analyse d'une banque métagénomique tient dans le fait qu'il faille bien souvent analyser beaucoup de clones pour obtenir ceux présentant un intérêt particulier. Pour augmenter la proportion des clones d'intérêts dans une banque, plusieurs stratégies ont été élaborées. Il s'agit d'enrichir l'échantillon environnemental en séquences génomiques d'intérêt, et ceci avant l'étape de clonage de l'ADN.

L'une de ces méthodes d'enrichissement repose sur le contenu en G+C des génomes. Certains micro-organismes ont un contenu génomique en G+C particulièrement élevé. C'est par exemple le cas des *Actinomycètes* (micro-organismes présents dans le sol, réputés pour synthétiser de nombreux antibiotiques) (Waksman & Curtis, 1916) et des *Acidobacteria* (phylum bactérien dont les représentants sont abondamment retrouvés dans le sol, mais dont on connaît peu de choses) (Quaiser *et al.*, 2003). Pour enrichir spécifiquement un échantillon de sol en séquences génomiques de ce type d'organismes, on extrait l'ADN et par ultracentrifugation, on récupère la fraction d'ADN génomique ayant un contenu en G+C élevé (Schloss & Handelsman, 2003). Bien que cette méthode n'apporte pas une séparation complète des génomes, elle permet cependant d'augmenter la représentation de ces génomes à contenu en G+C élevé dans une banque métagénomique.

Une méthode de séparation plus élégante est l'enrichissement au bromodeoxyuridine (BrdU) (Fig. 3). Le BrdU est un analogue de la thymidine. Le principe de cette méthode est le suivant : on incube dans des boîtes de Pétri des échantillons de l'environnement dans lesquels a été ajouté un substrat sélectif et du BrdU. Les micro-organismes de l'échantillon qui métaboliseront ce substrat spécifique, vont de par leur activité métabolique incorporer du BrdU dans leur ADN génomique. L'ADN marqué au BrdU qui peut être récupéré par immuno-rétention, sera ensuite utilisé pour construire la banque métagénomique (Yin *et al.*, 2000 ; Schloss & Handelsman, 2003). On peut ainsi construire des banques métagénomiques enrichies en micro-organismes métabolisant un type de substrat spécifique.

Une autre méthode d'enrichissement est la technique SIP pour Stable-Isotope Probing (Dumont & Murrell, 2005) (Fig. 3). Cette technique permet d'isoler l'ADN des micro-organismes qui sont activement impliqués dans un processus métabolique spécifique. Comme précédemment, on incube dans des boîtes de Pétri des échantillons de l'environnement dans lesquels on a ajouté

²les polyketides constituent un grand groupe de produits naturels qui incluent des antibiotiques, des anticancéreux et des immunosuppresseurs (Schwecke *et al.*, 1995).

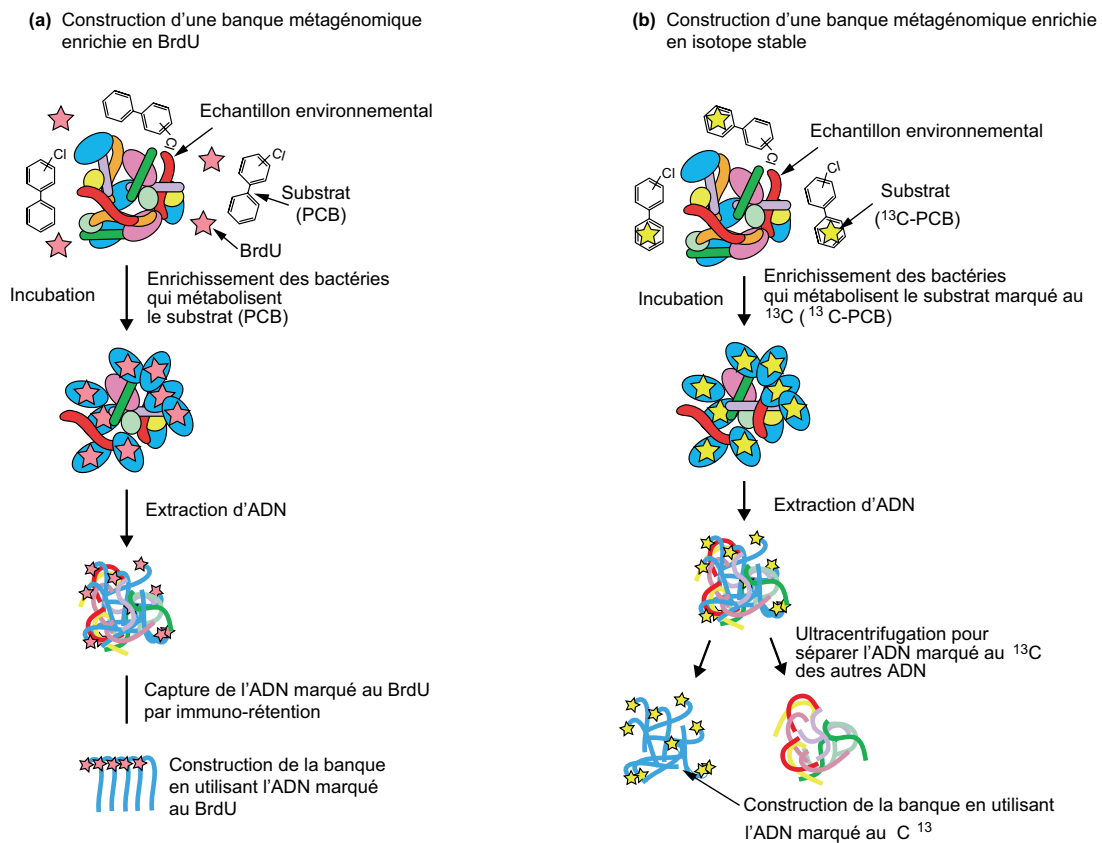


Fig. 3 – Enrichissement d'un échantillon en ADN génomique d'intérêt par la technique du BrdU (a) ou par la technique SIP (b). Dans les deux cas, le PolyChlorinated Biphenyl (PCB) est utilisé comme substrat métabolique (d'après Schloss & Handelsman, 2003).

cette fois ci des substrats marqués au carbone 13 (^{13}C). Les micro-organismes qui utilisent ce substrat vont incorporer du ^{13}C dans leur ADN, et auront un ADN plus dense que le normal qui contient du ^{12}C . Les fragments d'ADN marqués au ^{13}C , seront isolés des autres par centrifugation en gradient de chlorure de césium puis serviront pour construire la banque métagénomique. Cette technique a d'ores et déjà permis de marquer et de séparer des ADN (Radajewski *et al.*, 2002) et des ARN (Manefield *et al.*, 2002a,b).

Notons que le temps ne joue pas en la faveur de ces deux dernières techniques. En effet, plus le temps d'incubation de l'échantillon environnemental avec le substrat métabolique est long, plus grande est la probabilité que le substrat soit recyclé au sein de la communauté microbienne. Dans ce cas, on n'enrichit plus seulement les micro-organismes qui métabolisent spécifiquement ce substrat.

3.3 Choix du vecteur de clonage

La première banque métagénomique fut construite en 1991, en clonant des fragments d'ADN planctonique environnemental (inserts de 10 à 20 kpb) dans le génome du phage Lambda (Schmidt *et al.*, 1991). L'introduction en 1992 de nouveaux vecteurs de clonage, à savoir les fosmides (F1 origin-based cosmid vector) et les BACs (Bacterial Artificial Chromosomes, ou chromosomes bactériens artificiels), a considérablement amélioré les techniques de clonage pour étudier les génomes (Béjà, 2004). Ces vecteurs permettent en effet de cloner des fragments d'ADN bien plus grands que ne le permettent les phages Lambda ou les plasmides traditionnels utilisés en biologie moléculaire. Ils furent respectivement utilisés pour la première fois par Stein (Stein *et al.*, 1996) et par Béjà (Béjà *et al.*, 2000b).

Les techniques de clonage étant maintenant optimisées, le choix du vecteur de clonage pour la construction d'une banque métagénomique dépend maintenant de la taille des inserts que l'on désire cloner mais aussi de l'importance que l'on accorde ou non à exprimer les gènes de ces inserts. Les avantages et les inconvénients de construire une banque à l'aide de vecteurs de clonage acceptant des inserts de petite taille (moins de 15 kpb) ou de grande taille (plus de 40 kpb) sont explicités dans la Tableau 3.

3.3.1 Vecteurs permettant le clonage d'inserts d'ADN de grande taille

Lorsque l'expression des gènes clonés n'est pas requise pour l'analyse de la banque métagénomique, le choix du vecteur repose donc sur la taille des inserts à cloner. Les fragments de plus grandes tailles pourront être clonés dans des vecteurs de type BAC (insert de 25 à 350 kpb), ou dans des cosmides et fosmides (insert de 35 à 45kpb). Les inserts de plus petite taille seront clonés dans des vecteurs de clonage tels que dans le génome du phage lambda (insert de 5 à 24 kpb) ou dans des plasmides (insert de 5 à 8 kpb). Seuls les vecteurs permettant de cloner des

Avantages	Désavantages
Banque à petits inserts (plasmides)	
Pouvoir détecter les gènes exogènes faiblement exprimés grâce aux vecteurs à nombre de copies élevé	Les inserts d'ADN sont de petite taille
Pouvoir exprimer les gènes exogènes grâce aux vecteurs inductibles	Devoir cribler beaucoup de clones pour obtenir ceux d'intérêt
Pouvoir cloner l'ADN de petit PM ou l'ADN du sol contaminé par les substances de la matrice	Ne convient pas au clonage des grands opérons codant des activités métaboliques d'intérêt
Techniquement simple	
Banque à grands inserts (cosmides, fosmides, BACs)	
Inserts de grande taille	Il est plus difficile de détecter les gènes exogènes faiblement exprimés avec les vecteurs à faible nombre de copie
Cribler peu de clones pour obtenir les clones d'intérêt	Expression limitée des gènes exogènes par des vecteurs inductibles
Idéal pour le clonage des grands opérons de gènes codant des voies métaboliques d'intérêt	Nécessite de l'ADN purifié de haut poids moléculaire pour construire la banque
Convient pour la caractérisation partielle de l'ADN génomique des micro-organismes incultivés du sol	Techniquement difficile

BACs: Bacterial Artificial Chromosomes; PM: Poids Moléculaire.

Tableau 3 – Les avantages et les inconvénients d'une banque métagénomique à petits inserts ou à grands inserts (d'après Daniel, 2005).

inserts de grande taille seront présentés ci-dessous :

Les BACs Ils ont été mis au point dans les années 1990, pour la génomique des eucaryotes (Shizuya *et al.*, 1992). Ces vecteurs de clonage se maintiennent de façon stable chez *E. coli*. Ils sont construits sur la base du plasmide appelé "facteur F" qui est un plasmide conjugatif existant naturellement chez *E. coli*. Le facteur F porte des gènes qui lui permettent non seulement de réguler sa propre réplication (gènes *oriS* et *repE*) mais aussi de le maintenir à raison d'une à deux copies par cellule (gènes *parA* et *parB*). Le vecteur de clonage de type BAC reprend les caractéristiques du facteur F. Il contient également un marqueur de sélection (résistance à un antibiotique) ainsi qu'une région à sites multiples de clonage éventuellement située au niveau de la portion *lacZa* du gène β -galactosidase, permettant par test colorimétrique une sélection blanc/bleu des cellules ayant intégré un plasmide recombinant (Shizuya *et al.*, 1992). Ce type de vecteur permet de cloner de grands fragments d'ADN pouvant aller jusqu'à 350 kpb. C'est la raison pour laquelle les vecteurs BACs sont préférés aux autres vecteurs pour cartographier et reconstituer plus rapidement des génomes entiers. Pour la même raison, on a plus de chance d'obtenir avec ce type de vecteur de clonage des voies de biosynthèse complètes, des systèmes de sécrétions ou des îlots de pathogénicité qui sont généralement physiquement regroupés sur les génomes bactériens (Rondon *et al.*, 1999b).

Les cosmides et fosmides Ces vecteurs de clonage sont une alternative aux BACs si l'on veut cloner des fragments de plus petite taille (inserts allant de 35 à 45 kpb). L'efficacité de clonage dans ces vecteurs est généralement 100 à 1000 fois plus performante qu'avec des BACs

(Handelsman *et al.*, 2002). Les cosmides sont des vecteurs qui combinent à la fois les propriétés des bactériophages et des plasmides : ils possèdent une origine de réplication de plasmide, un ou plusieurs marqueurs de sélection, une région de sites multiples de clonage et les séquences cos nécessaires à l'encapsidation d'ADN dans des particules de phage lambda (Collins & Hohn, 1978). Le système d'encapsidation dans les particules phagiques permet de sélectionner les vecteurs ayant intégré des fragments d'ADN de 35 à 45 kpb. Les cosmides recombinants sont injectés dans des cellules d'*E. coli*, après adsorption des particules phagiques sur des récepteurs membranaires d'*E. coli*. Une fois dans la cellule hôte, le cosmide se circularise. Il se réplique alors comme un plasmide, de nombreuses fois, ce qui peut provoquer son instabilité dans la cellule hôte. Pour remédier à ce problème, des vecteurs de clonage à faible nombre de copies, les fosmides, ont été développés (Kim *et al.*, 1992). Ces derniers sont des vecteurs de clonage construits sur la base des cosmides auxquels ont été ajoutés les gènes *oriS*, *repE*, *parA* et *parB* du facteur F. Dans la cellule hôte, la stabilité des fosmides (maintenus à faible nombre de copies) est plus importante qu'avec les vecteurs à grand nombre de copies (Kim *et al.*, 1992). Le système d'encapsidation des fosmides est tout aussi efficace que celui obtenu avec les cosmides, et l'efficacité d'infection n'en est pas amoindrie.

3.3.2 Les vecteurs d'expression

De nombreuses études ont montré qu'il était possible d'exprimer chez *E. coli* divers types de gènes provenant de différents organismes (Ding & Yelton, 1993 ; Ferreyra *et al.*, 1993). Cependant, étant donné que certains gènes requièrent des systèmes de transcription (reconnaissance des promoteurs) et/ou de traduction (signaux d'initiation de la traduction) qui sont absents chez *E. coli*, il n'est donc pas possible, chez cet organisme, d'exprimer tous les gènes d'une banque métagénomique. En effet, si le code génétique est presque universel, chaque organisme voire chaque type de gène dispose de ses propres facteurs de transcription. De surcroît, la traduction est directement affectée par l'usage des codons spécifiques à chaque organisme et est ainsi directement corrélée à la disponibilité de certains ARNt dans la cellule.

Les vecteurs d'expression permettent d'éviter le problème dû à la reconnaissance des promoteurs étrangers à la machinerie transcriptionnelle d'*E. coli* (Newman & Fuqua, 1999). Ces vecteurs d'expression peuvent contenir des promoteurs inductibles afin de limiter les problèmes liés à l'expression constitutive de produits toxiques. Les vecteurs d'expression acceptent généralement des petits fragments d'ADN étranger (<10kpb), et la transcription du ou des gènes clonés ne pourra se faire que si ceux-ci sont correctement orientés par rapport au promoteur du vecteur. L'utilisation de tels vecteurs a permis par exemple d'isoler de nouvelles lipases, estérases et 4-hydroxybutyrate deshydrogénases (Henne *et al.*, 1999, 2000).

3.3.3 Les vecteurs navettes

Si les vecteurs d'expression permettent de limiter les problèmes liés à l'expression des gènes clonés, il peut être judicieux d'utiliser des vecteurs navettes qui sont capables de se répliquer dans au moins deux organismes différents grâce à des origines de réplifications appropriées. Par exemple, il est particulièrement difficile d'exprimer chez *E. coli* des polyketides provenant d'actinomycètes, ceci à cause de différences trop marquées entre les promoteurs et l'usage des codons de ces organismes (Cane *et al.*, 1998 ; Tang *et al.*, 2000 ; Pfeifer *et al.*, 2001). Afin de pallier à ces problèmes, on peut générer une première banque, dans *E. coli* par exemple (avec des cellules ultracompétentes), en utilisant un vecteur navette, puis transférer l'ADN recombinant des clones positifs sélectionnés (identifiés par exemple par similarité de séquence) dans un nouvel organisme hôte permettant cette fois-ci l'expression du gène d'intérêt. Le premier article décrivant l'utilisation d'une cellule hôte autre que *E. coli* pour le criblage de banques métagénomiques est celui de Wang et ses collaborateurs (2000). Grâce à la mise au point d'un vecteur navette *E. coli* - *Streptomyces lividans*, ils ont découvert une nouvelle voie de biosynthèse d'antibiotiques, celle des terragines.

3.3.4 Les vecteurs inductibles

Les vecteurs de clonage à faible nombre de copies sont plus stables que ceux à grand nombre de copies. Le deuxième avantage d'utiliser des vecteurs à faible nombre de copies est le suivant : avec ce type de vecteur, les produits des gènes clonés se retrouvent peu concentrés dans la cellule hôte, ce qui limite la mort cellulaire des cellules hôtes provoquée par la présence de gènes pouvant être toxiques à trop forte concentration. Cependant, il est plus difficile d'extraire ces vecteurs et de les purifier, en vue de séquencer leur insert. De plus, la détection du produit des gènes exogènes présents dans les inserts sera moins aisée qu'avec des vecteurs à grand nombre de copies.

Afin de combiner l'avantage des plasmides à faible et grand nombre de copies, des vecteurs à faible nombre de copies dont la reproduction peut être facilement induite ont été mis au point. A titre d'exemple, on peut citer les vecteur CopyControl™ Fosmid commercialisés par EPICENTRE (pouvant être induit jusqu'à 50 copies par cellules), ou le vecteur superBAC1 (inductible jusqu'à 100 copies par cellules) (Handelsman *et al.*, 2002).

3.4 Choix des cellules hôtes

Le choix de la cellule hôte pour la construction de la banque métagénomique est guidé par l'efficacité de transformation et la stabilité du vecteur (Handelsman *et al.*, 2002). Ce choix est également déterminé par les phénotypes d'intérêts que l'on veut mettre en valeur dans la banque. Pour les criblages fonctionnels ou sélectifs, il est essentiel que la cellule elle-même n'exprime pas la caractéristique d'intérêt.

Avantages	Inconvénients
Méthode de criblage basée sur une activité:	
De nouveaux gènes peuvent être récupérés	Dépend de l'expression (transcription/traduction) des gènes clonés dans la cellule hôte
Sélectionne les gènes clonés dans leur totalité	Nécessite que le produit du gène soit fonctionnel dans la cellule hôte
Sélectionne les produits des gènes fonctionnels	Pouvoir créer une stratégie simple de criblage fonctionnel
Méthode de criblage basée sur la séquence:	
Indépendant des cellules hôtes: pas d'expression des gènes clonés	Les gènes récupérés sont des gènes relativement connus
Une même stratégie de criblage peut être utilisée pour différentes cibles, par ex, hybridation sur colonie et PCR	Les gènes peuvent être partiellement clonés
	Pas de sélection des produits des gènes fonctionnels

Tableau 4 – Comparaison de stratégies : criblage basé sur l'activité versus criblage basé sur la séquence (d'après Daniel, 2005).

3.5 Caractérisation d'une banque métagénomique

De nombreux environnements du sol (Daniel, 2005 ; pour revue) et divers habitats marins (DeLong, 2005, pour revue) ont fait l'objet d'études métagénomiques. Ces études ont permis dans certains cas de décrire la biodiversité des génomes microbiens contenus dans ces environnements ou bien de découvrir de nouveaux gènes à potentiel bioactif ou des molécules à haute valeur ajoutée. En fonction de l'intérêt fondamental ou appliqué de la banque métagénomique, ces banques peuvent être criblées en réalisant soit des analyses basées sur les séquences ou bien des analyses basées sur l'expression d'une activité.

Les avantages et les inconvénients des deux stratégies de criblage d'une banque métagénomique sont détaillés dans la Tableau 4.

3.5.1 Analyse basée sur les séquences

Une première approche consiste à identifier les clones d'intérêt d'une banque, puis à les séquencer. On peut distinguer deux types de "clones d'intérêt" :

- ceux portant des marqueurs phylogénétiques permettant de les identifier (dans ce cas, on espère identifier des fragments génomiques d'organismes incultivés),
- et ceux porteurs de nouveaux gènes d'intérêt industriel ou pharmaceutique.

Les clones d'une banque métagénomique pouvant être phylogénétiquement affiliés à des organismes incultivés, seront entièrement séquencés. Le but de cette approche est d'extraire du fragment génomique séquencé, un maximum d'informations sur la génétique, le métabolisme, la physiologie, le rôle et/ou la fonction écologique des micro-organismes incultivés. La recherche des clones d'intérêt présents dans la banque est généralement conduite par PCR ou hybridation avec des marqueurs/sondes phylogénétiques. Les marqueurs les plus fréquemment utilisés dans le cadre d'études métagénomiques sont ceux ciblant les gènes codants les ARNr de lignées bactériennes ou archéennes (Stein *et al.*, 1996 ; Quaiser *et al.*, 2002 ; Hallam *et al.*, 2003 ; Quaiser *et al.*, 2003 ; Liles *et al.*, 2003 ; Lopez-Garcia *et al.*, 2004 ; Moreira *et al.*, 2004 ; Bèjà *et al.*, 2000b, 2002a) ;

ceux ciblant les homologues des gènes *radA/recA* ont également été utilisés (Sandler *et al.*, 1999).

Plusieurs lignées microbiennes qui étaient auparavant connues uniquement par la signature de leur gène codant l'ARNr 16S, ont pu être ainsi explorées. A titre d'exemple, ce fut le cas des *Acidobacteria*. Plusieurs études métagénomiques ont permis d'identifier des fragments génomiques affiliés aux *Acidobacteria*, permettant ainsi d'émettre des hypothèses sur le rôle écologique de ces bactéries (Quaiser *et al.*, 2002, 2003 ; Liles *et al.*, 2003). D'autres études métagénomiques ciblant plus particulièrement les archaea incultivées du sol ou de la mer ont permis d'élucider dans certains cas la biologie des ces organismes (Stein *et al.*, 1996 ; Quaiser *et al.*, 2002 ; Lopez-Garcia *et al.*, 2004 ; Moreira *et al.*, 2004 ; Hallam *et al.*, 2003 ; Treusch *et al.*, 2004 ; Schleper *et al.*, 1998, 1997b ; Béjà *et al.*, 2002a). Enfin, la découverte de photorécepteurs de type rhodopsine chez des bactéries marines représente un bel exemple de surprise biologique qui a pu être révélée par des études de métagénomique (Béjà *et al.*, 2000a). Plusieurs de ces études seront présentées ultérieurement. Notons cependant que, si cette approche semble très prometteuse, il peut parfois s'avérer très difficile d'attribuer aux organismes non ou difficilement cultivables des caractéristiques métaboliques qui leur sont propres. En effet, il est bien souvent difficile d'attribuer aux gènes prédits une fonction précise (cela peut concerner jusqu'à 60% des gènes) (Tyson & Banfield, 2005). Même dans le cas d'*E. coli*, qui est sans doute la bactérie la plus étudiée, près de 50% des gènes prédits ne ressemblent à aucun gène connu. La plupart de ces gènes peuvent être spécifiques à une lignée particulière, et peuvent jouer un rôle primordial pour l'adaptation à un environnement donné.

Quant aux clones porteurs de nouveaux gènes d'intérêt industriel ou pharmaceutique, eux aussi pourront être identifiés par PCR ou hybridation. Ces techniques de criblage sont soit basées sur des similarités de séquences avec des gènes connus, ou soit basée sur la conservation de motifs. Grâce à ces techniques, il est possible d'identifier de nouvelles variantes de familles ou bien de nouvelles classes fonctionnelles protéiques. Le criblage par PCR d'une banque métagénomique a par exemple permis d'identifier de nouvelles lipases (Bell *et al.*, 2002). La technique d'hybridation de l'ADN a quant à elle permis d'identifier par exemple de nouvelles glycérol - et diol déshydratases (Knietzsch *et al.*, 2003a). Cependant ces deux techniques de criblage présentent le désavantage d'être "conservatrices". Il faut en effet connaître la séquence des gènes ciblés pour pouvoir construire les sondes pour l'hybridation et les amorces pour la PCR. Celles-ci devront en effet être dessinées de sorte à refléter des motifs conservés connus.

Une deuxième approche consiste quant à elle, à séquencer au hasard les clones d'une banque métagénomique, et faire ce que l'on appelle en anglais du "shotgun". Cette approche concerne généralement les banques métagénomiques construites à l'aide de vecteurs de clonage acceptant des petits inserts d'ADN.

Conduit à grande échelle, le séquençage aléatoire de milliers de clones d'une banque métagénomique (whole genome shotgun sequencing) permet théoriquement de reconstruire les génomes présents au sein de la communauté microbienne. Cependant, beaucoup d'environnements contiennent des communautés bien trop complexes pour permettre de séquencer leur métagénome complet. On se contente dans ce cas là de reconstituer partiellement les génomes, comme ce fut le cas pour l'étude du métagénome de la mer des Sargasses (Venter *et al.*, 2004) (voir § 2.5.1). Cette approche donne une bonne idée du contenu et de l'organisation génétique des métagénomes. Dans le cas des communautés microbiennes peu complexes, on peut non seulement faire l'inventaire complet des gènes présents, mais aussi tenter de reconstruire les génomes de micro-organismes incultivés. Les premières études de reconstruction de génome ont été réalisées sur des communautés virales (Breitbart *et al.*, 2002, 2003), puis sur une communauté microbienne peu diversifiée provenant d'un biofilm récolté sur un drainage acide d'une mine (Tyson *et al.*, 2004) (voir § 4.5.2).

3.5.2 Analyse basée sur une activité

La métagénomique a le potentiel de faire découvrir de nouvelles classes de gènes et de nouvelles voies métaboliques (Tableau 3 et Tableau 4). Le criblage fonctionnel des banques métagénomiques basé sur l'expression hétérologue des gènes permettra de découvrir ces nouveaux gènes ou voies métaboliques. Le tri des clones se fait par sélection, lorsqu'un gène donné confère un phénotype particulier à l'hôte. Ce phénotype particulier peut par exemple consister en une résistance à un antibiotique donné. Ce type d'analyse des banques métagénomiques nécessite bien souvent des techniques de criblage à haut débit, des systèmes robotisés semblent conseillés voire obligatoires pour le criblage des banques.

L'analyse des banques métagénomiques basée sur la recherche d'activités est très prometteuse pour la découverte de nouveaux gènes. En effet, il n'est pas nécessaire de connaître la séquence des gènes d'intérêt pour les identifier. Cependant, cette approche ne donne pas toujours les résultats espérés :

- l'expression hétérologue nécessite que la transcription et la traduction du (/ou des gènes d'intérêts s'il s'agit d'un cluster) puissent se faire sans encombre dans la cellule hôte,
- il faut également que le produit du gène d'intérêt puisse être secrété par la cellule hôte si le criblage ou le test le nécessitent. La plupart des études visant à exprimer les gènes d'une banque métagénomique ont cependant été réalisées avec *E. coli* comme cellule hôte étant donné que la manipulation génétique est facile et bien maîtrisée chez cet organisme (Lorenz & Eck, 2005 ; pour revue).

Pour découvrir une gamme plus large de nouveaux produits, on peut améliorer le système d'expression hétérologue des gènes de la banque métagénomique. *E. coli* peut ainsi être génétiquement modifiée afin d'exprimer une gamme plus large de fonctions, en introduisant par exemple

des gènes codant de nouveaux facteurs sigma ou des ARNt rares (Riesenfeld *et al.*, 2004). On peut également construire une première banque métagénomique chez *E. coli* (en utilisant un vecteur navette), puis transférer l'ADN recombinant chez un autre organisme (par exemple chez *Streptomyces*, *Bacillus* ou chez des *Archaea*) qui lui permettra l'expression hétérologue des gènes (Lorenz *et al.*, 2002). Enfin, des techniques astucieuses ont été mises au point pour capturer des gènes par complémentation de mutants auxotrophes avec de l'ADN métagénomique (Henne *et al.*, 1999 ; Majernik *et al.*, 2001).

3.6 Estimer l'effort de séquençage afin de reconstituer les génomes d'une communauté microbienne donnée

Si l'on fabrique une banque métagénomique en vue de reconstruire les génomes présents dans l'échantillon environnemental, il faut estimer la quantité de séquençage à fournir. Pour cela, il faut tenir compte non seulement du taux de couverture des génomes que l'on désire obtenir mais aussi des informations disponibles sur le nombre d'espèces, leur abondance relative et la taille de leur génome (Allen & Banfield, 2005). La richesse de la diversité microbienne d'une communauté donnée peut être évaluée au travers de l'analyse de banques de gènes codant l'ARNr 16S, l'abondance (quantitative) des espèces, elle, pourra être estimée par FISH. Afin de quantifier l'effort de séquençage à fournir, des méthodes statistiques ont été mises au point pour estimer la diversité microbienne des communautés (Hughes *et al.*, 2001). Il est possible d'évaluer la taille d'un génome donné en se référant à celle d'organismes proches déjà séquencés (si disponibles), ou en considérant que la taille moyenne d'un génome de procaryote est de $3.16 \text{ Mpb} \pm 1.79 \text{ Mpb}$ (Allen & Banfield, 2005). Ainsi, à titre d'exemple, si un organisme dont le génome est de 3 Mpb représente 1% de la population d'une communauté donnée, il faudra séquencer 2.4 Gpb pour obtenir un taux de couverture de 8X (presque complet) de son génome (Allen & Banfield, 2005).

3.7 Méthode de reconstruction de génomes

Pour reconstruire des génomes, on assemble les lectures provenant d'un séquençage aléatoire (shotgun sequence). Les séquences sont ordonnées les unes par rapport aux autres par la mise en évidence de recouvrements (extrémités chevauchantes), c'est-à-dire des régions terminales présentant un enchaînement de nucléotides identiques. En assemblant deux ou plusieurs séquences chevauchantes entre elles, on obtient alors une contig. Pour obtenir une séquence génomique contigüe la plus longue possible, il faut ensuite assembler les contigs, c'est-à-dire les relier, les ordonner et les orienter dans le bon sens, le long du chromosome. Petit à petit, on reconstitue l'ossature du génome : on constitue des "scaffolds". Les scaffolds résultent de l'assemblage de contigs qui ne sont pas forcément physiquement rattachés-reliés les uns aux autres (présence de trous/"gaps" dont la taille est connue), mais qui sont ordonnés et orientés dans le bon sens en suivant une procédure de reconstruction de génomes (Fig. 4).

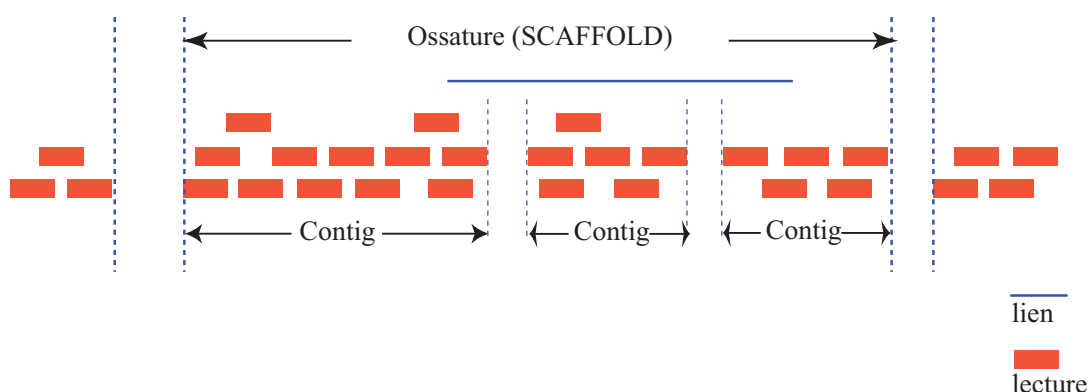


Fig. 4 – Assemblage de l'ébauche génomique après séquençage aléatoire global (whole genome shotgun sequencing). Les lectures ou séquences de chaque clone sont assemblées en contig. On relie, ordonne et oriente ensuite les contigs entre eux, au sein d'ossatures (scaffolds), en comparant par exemple les scaffolds avec des séquences homologues d'organismes pour lesquels le génome complet est connu. source : http://www.cns.fr/externe/Francais/Projets/Projet_AK/AK.html

La procédure de reconstruction des génomes est largement détaillée dans l'article de Tyson et ses collaborateurs Tyson *et al.*, 2004. Elle deviendra certainement le protocole standard utilisé pour les futures études de génomique environnementale (Galperin, 2004). Cette procédure est d'autant plus facile à appliquer que la diversité microbienne de la communauté est faible, et que le contenu en G+C des différents génomes est suffisamment distinct d'un organisme à une autre (Tyson *et al.*, 2004). Le principe de cette procédure consiste à assembler entre eux les contigs pour former des "scaffolds". Les scaffolds sont ensuite classés dans des "cases" ("bin" est le terme anglophone employé par les auteurs). Les "cases" sont ensuite assignées à un type d'organisme donné, en fonction de leur contenu en G+C et de leur taux couverture de séquençage.

D'autres paramètres peuvent également être pris en compte pour assigner les scaffolds au bon organisme. Ces paramètres sont les suivants : la phylogénie de marqueurs conservés (gènes codant l'ARNr 16S, *recA/radA*, *rpoB/C*, HSP60/70, EF-TU, et *aIF2 β*), l'usage des codons, la fréquence des di-, tri-, et tetra-nucléotides (Teeling *et al.*, 2004a,b) et enfin, la comparaison des scaffolds avec des séquences homologues d'organismes pour lesquels le génome complet est connu (Allen & Banfield, 2005).

3.8 Conservation d'une banque métagénomique

Les colonies d'une banque métagénomique, si celle-ci a été faite sur milieu solide, sont généralement repiquées individuellement puis mises en culture pure dans des plaques de 96 puits. Les clones peuvent alors être conservés à -80°C si l'on rajoute au milieu de culture un cryoprotecteur (glycérol ou diméthyl sulfoxyde). Afin d'éviter cette étape supplémentaire, on peut en repiquant les clones dans le milieu dit de "Hogness" cultiver puis conserver directement les cellules après leur croissance (Werner *et al.*, 1997). Ce milieu n'affecte pas la stabilité du vecteur de clonage, ni la viabilité des cellules même après plusieurs cycles de congélation / décongélation.

4 Exemples d'études métagénomiques ayant permis de faire le lien entre l'identité phylogénétique et une fonction spécifique d'un micro-organisme

A titre d'exemple, voici quelques études choisies dans lesquelles il a été possible d'inférer une fonction biologique (physiologique ou métabolique) à des micro-organismes incultivés.

4.1 Découverte d'une nouvelle bactériorhodopsine

La découverte d'un photorécepteur de type rhodopsine chez une bactérie marine non cultivée a été une réelle surprise (Béjà *et al.*, 2000a). Avant cette étude, ces photorécepteurs n'étaient connus que chez les *Archaea* halophiles (initialement découverts en 1970 chez *Halobacterium salinarum*). Les rhodopsines sont présentes en très forte concentration dans la membrane des *Archaea* halophiles. Sous l'effet de la lumière, ces protéines membranaires subissent une série de déformations structurales qui provoquent le passage d'un proton au travers de la membrane et la synthèse d'ATP. Les rhodopsines sont alimentées par l'énergie lumineuse et utilisent une voie autre que celle basée sur la chlorophylle.

Le gène codant la nouvelle bactériorhodopsine a été identifié dans l'insert génomique (130 kpb) d'un BAC (Béjà *et al.*, 2000a). Le clonage et l'expression de la bactériorhodopsine d'origine bactérienne chez *E. coli* a révélé qu'elle avait une activité identique à celle décrite chez *H. salinarum*. Le fragment génomique portant le gène de la bactériorhodopsine a pu être phylogénétiquement identifié. En effet, ce fragment portait entre autre une séquence d'ARNr 16S. Sur la base de cette séquence d'ARNr 16S, cette bactérie marine a été affiliée aux γ -*Proteobacteria*, plus précisément à un groupe de micro-organismes appelé "SAR86" qui n'était jusqu'à présent connu que par ses seules séquences d'ADNr 16S (Béjà *et al.*, 2000a). Notons que le fragment génomique ne contenait aucun autre gène d'origine archaéenne qui aurait pu indiquer un transfert de gène horizontal (Béjà *et al.*, 2000a).

Avant cette étude, le groupe SAR86 avait été détecté à de nombreuses reprises dans des inventaires moléculaires d'habitats marins, mais aucune fonction n'avait pu leur être inférée. L'existence de rhodopsine chez ce groupe de micro-organismes a suggéré que ces derniers se développaient par phototrophie, formant ainsi un nouveau taxon de bactéries marines phototrophes.

4.2 Métagénomique des *Archaea*

Dans les écosystèmes marins et terrestres, les *Archaea* sont si abondantes et diverses qu'elles jouent sans aucun doute un rôle crucial dans tous les cycles de la biosphère terrestre. Si les

Archaea étaient initialement connues pour dominer des habitats qualifiés d'extrêmes tels que les sources chaudes ou acides, les sources hydrothermales marines, les étangs hypersalins et les écosystèmes strictement anoxique, de nombreux inventaires moléculaires ont depuis montré que les *Archaea* étaient abondantes et largement distribuées sur Terre, occupant des biotopes tels que le sol (Bintrim *et al.*, 1997 ; Jurgens *et al.*, 1997 ; Buckley *et al.*, 1998 ; Sandaa *et al.*, 1999 ; Simon *et al.*, 2000 ; Ochsenreiter *et al.*, 2003), les lacs (MacGregor *et al.*, 1997 ; Jurgens *et al.*, 2000 ; Keough *et al.*, 2003 ; Schleper *et al.*, 1997a), la mer (plancton océanique et sédiments marins) (DeLong, 1992 ; Fuhrman *et al.*, 1992 ; Boetius *et al.*, 2000 ; Orphan *et al.*, 2002) ou des environnements terrestres souterrains (Takai *et al.*, 2001b). Ces études montrèrent que les *Archaea* n'étaient pas nécessairement des extrémophiles. D'autre part, malgré la faible diversité métabolique observée chez les *Archaea* cultivées (à titre d'exemple, toutes les *Crenarchaeota* cultivées étaient encore jusqu'à peu, des thermophiles métabolisant principalement le soufre (Burggraf *et al.*, 1997)), ces études laissèrent présager que les *Archaea* pouvaient présenter une gamme de métabolisme très variée.

Parmi le grand nombre de séquences d'*Archaea* mises en évidence par des méthodes moléculaires, seules quelques représentants ont pu être cultivés (Fig. 5 du chapitre 1). Les nombreuses lignées archéennes n'ayant pas de représentant cultivé, nous laisse sans réponse concernant leur métabolisme. L'étude de ces lignées archéennes incultivées par une approche métagénomique (Fig. 5 du chapitre 1) a permis dans certains cas de confirmer certaines hypothèses métaboliques, permettant d'apporter les bases nécessaires pour étudier l'impact écologique de ces espèces au sein de leur communauté microbienne (Tableau 5).

4.2.1 Les crenarchaeotes du sol

La caractérisation des micro-organismes du sol par une approche métagénomique demeure un challenge. Tout d'abord, les sols constituent l'écosystème terrestre le plus diversifié en espèce (la diversité en espèces est 20 fois supérieure à celle rencontrée dans la mer (Curtis *et al.*, 2002)). De plus, les préparations d'ADN à partir d'échantillons du sol sont fortement contaminées en composés polyphénoliques, nécessitant une étape de purification de cet ADN pour pouvoir le cloner (Rondon *et al.*, 1999a ; Quaiser *et al.*, 2002).

Malgré ces difficultés, plusieurs fragments génomiques de grande taille, affiliés à des crenarchaeotes, ont pu être identifiés dans des banques métagénomiques construites avec de l'ADN environnemental provenant d'échantillons du sol (Quaiser *et al.*, 2002 ; Treusch *et al.*, 2004). L'analyse du contenu génique de ces fragments a révélé que des crenarchaeotes non thermophiles étaient capable de nitrification, laissant sous-entendre que ces archaea utilisent l'ammonium comme source principale d'énergie (présence dans ces séquences génomiques d'homologues des gènes codant la petite et la grande sous-unité de l'ammonium monooxygénase (*amoA* et *amoB*))

Environnement	Groupe archaéen concerné	Type	Commentaires	Références
Pacifique nord, plancton marin	Crenarchaeote marine planctonique du groupe I.1A	Fosmides	1er fragment génomique décrit d'un procaryote incultivé	Stein <i>et al.</i> , 1996 Béjà <i>et al.</i> , 2002
Californie côte Pacifique, éponge marine <i>Axinella mexicana</i>	Crenarchaeote du groupe I.1A: <i>Cenarchaeum symbosium</i>	Fosmides	Symbiote archaéen extracellulaire, banque métagénomique enrichie en ADN génomique de cette Archaea	Schleper <i>et al.</i> , 1998 Schleper <i>et al.</i> , 1997
Côte californienne, eau de surface	Euryarchaeote marine du groupe II	BACs	Clone de 60 kpb, identifié par un gène codant l'ARNr 23S	Béjà <i>et al.</i> , 2000
Antarctique, eaux cotières	Crenarchaeote planctonique du groupe I.1A	Fosmides	Analyse comparative de fragments affiliés aux crenarchaeote du groupe I.1A	Béjà <i>et al.</i> , 2002
Echantillon de sol d'une prairie, couche aérobie	Crenarchaeote du sol du groupe I.1B	Fosmides	Identification de fosmides archaéens par des gènes codant l'ARNr 16S ou par séquençage aléatoire	Quaiser <i>et al.</i> , 2002 Treusch <i>et al.</i> , 2004 Schleper <i>et al.</i> , 2004
Rivière Eel, Canyon de Monterey, mattes microbiennes associées aux suintements de méthane	ANME -1 et 2	Fosmides	Méthanotrophes en syntrophie avec des bactéries sulfato-réductrices, gènes pour la méthanogénèse inverse	Hallam <i>et al.</i> , 2003 Hallam <i>et al.</i> , 2004
Nord-ouest de la Mer Noire, mattes microbiennes de suintements de méthane	ANME -1	Fosmides	Caractérisation biochimique d'un nouveau cofacteur au nickel de la méthyl-coenzyme M réductase	Kruger <i>et al.</i> , 2003
Antarctique, eaux des fonds (500m)	Crenarchaeote du groupe I.1A; euryarchaeote du groupe II	Cosmides	Etude de transferts de gènes horizontaux chez les crenarchaeotes	Lopez-Garcia <i>et al.</i> , 2004 Moreira <i>et al.</i> , 2004
Biofilm d'un drainage acide d'une mine	<i>Ferroplasma</i>	Plasmides (petits inserts)	<i>Ferroplasma</i> de type II (génomme presque reconstruit), ainsi que des fragments génomiques du type A et G	Tyson <i>et al.</i> , 2004
Eau de surface de la mer des Sargasses	Crenarchaeotes et euryarchaeotes planctonique	Plasmides (petits inserts)	>1 Gb d'ADN environnemental séquencé	Venter <i>et al.</i> , 2004
Consortium de méthanogènes	Rice cluster I methanogènes	Fosmides	Maintient stable du consortium bactéries/archaea, pendant 3 ans, en anaérobiose, à 50°C	Erkel <i>et al.</i> , 2005

BACs, bacterial artificial chromosomes.

Tableau 5 – Analyses métagénomiques impliquant des *Archaea* incultivées (d'après Schleper *et al.*, 2005).

(Schleper *et al.*, 2005).

4.2.2 Les *Archaea* marines planctoniques

La dominance numérique des crenarchaeotes dans les océans estimée à 10^{28} cellules (Karner *et al.*, 2001), suggère que ces *Archaea* jouent un rôle primordial dans les cycles biogéochimiques sur Terre. Les seules informations dont nous disposions jusqu'à peu, étaient celles provenant de l'analyse isotopique de leurs lipides (Pearson *et al.*, 2001). Ces analyses avaient montré que les *Crenarchaeota* pouvaient fixer le carbone inorganique.

Les premiers travaux de métagénomique portant sur des échantillons d'eau de mer, furent initiés par DeLong et ses collaborateurs (1996). Ces derniers construisirent une banque métagénomique (à l'aide de BACs) à partir d'un échantillon d'ADN planctonique de l'océan Pacifique nord-oriental. Parmi les 3552 clones de leur banque, ils purent identifier puis caractériser un fragment génomique (38.5 kpb) d'une crenarchaeote mésophile incultivée. D'autres fragments génomiques d'*Archaea* marines ont depuis été isolés à partir de banques de BAC, fosmides ou cosmides construites avec des échantillons provenant de l'Antarctique ou du nord de l'océan Pacifique (Lopez-Garcia *et al.*, 2004 ; Moreira *et al.*, 2004 ; Béjà *et al.*, 2000b, 2002a). Dans une plus large mesure, le séquençage aléatoire d'échantillons provenant de la mer des Sargasses a considérablement augmenté la quantité de séquences archéennes marines (Venter *et al.*, 2004). Ainsi, environ 20 scaffolds identifiés dans cette étude (représentant 396 kpb) ont été affiliés à des *Archaea*. Parmi les activités métaboliques identifiées dans cette étude, figurent les voies pour

la biosynthèse du chorismate, des oxoacide-ferrodoxine oxydo-réductases, et les glutamate et phosphoglycérate deshydrogénases.

4.2.3 *Cernarchaeum symbosium*, un organisme archaéen modèle

Parmi la communauté microbienne présente dans les tissus de l'éponge marine *Axinella mexicana*, un seul phylotype archaéen, *C. symbosium*, a été détecté par identification moléculaire (ADNr 16S) (Preston *et al.*, 1996). Cette archée peut être maintenue en co-culture avec son hôte pendant des mois à des températures inférieures de 40°C à 50°C de l'optimum de ses plus proches voisins phylogénétiques, des crenarchaeotes hyperthermophiles. Si *C. symbosium* ne peut être cultivée sans son hôte et les autres bactéries de la communauté, des fractions cellulaires provenant d'enrichissement de cette archée ont permis de construire des banques à large insert (Schleper *et al.*, 1998), conduisant à un projet de séquençage génomique (Schleper *et al.*, 2005). Une ADN polymérase de la famille B, montrant que *C. symbosium* est proche des crenarchaeotes hyperthermophiles cultivées, a été identifiée dans l'un des clones de ces banques génomiques. La caractérisation biochimique de cet ADN polymérase a par la suite révélé que l'archée *C. symbosium* était bien adaptée à des températures basses (Schleper *et al.*, 1997b).

La recherche de synténie³ entre les fragments de génomes de *C. symbosium* et les scaffolds reconstruits lors du séquençage du métagénome de la mer des Sargasses s'est avérée positive à de nombreuses reprises (Schleper *et al.*, 2005). Cette analyse a révélé que *C. symbosium* était phylogénétiquement proche d'archaea planctoniques. Il peut donc s'avérer utile de prendre *C. symbosium* comme organisme modèle pour étudier les crenarchaeotes marines présentes sous forme libre.

Dernièrement, une crenarchaeote marine mésophile a été isolée (Könneke *et al.*, 2005). Il s'agit d'une archaea aérobie chimiolithoautotrophe, qui oxyde l'ammonium en nitrites. En plus d'être la première mésophile isolée du phylum des *Crenarchaeota*, cet isolat constitue également le premier cas de nitrification observé chez une archée. En effet, ce métabolisme n'était jusqu'à présent seulement connu chez des Proteobacteries- β et γ (Bock & Wagner, 2001). Les analyses phylogénétiques portant sur l'ARNr 16S de cet isolat a montré qu'il était proche (>98%) des crenarchaeotes marines du groupe I qui incluent l'archée *C. symbosium* ainsi que des séquences crenarchaeotes identifiées lors de l'analyse du métagénome de la mer des Sargasses (Venter *et al.*, 2004). Le métabolisme autotrophique de cet isolat, ainsi que son affiliation à des séquences environnementales de crenarchaeotes marines, suggèrent que les crenarchaeotes marines nitrifiantes doivent jouer un rôle important dans la régulation globale des cycles du carbone et de l'azote (Könneke *et al.*, 2005).

³On parle de synténie pour décrire, dans le cadre de comparaisons entre génomes, des groupes de gènes co-localisés sur un génome, dont les correspondants, sur un autre génome, sont également co-localisés.

4.3 La symbiose des bactéries sulfo-oxydantes avec le vestimentifère *Riftia pachyptila*

Le vestimentifère *Riftia pachyptila* est un ver tubicole colonisant les parois des cheminées hydrothermales marines profondes. Ce ver est dépourvu de bouche, de tube digestif et d'anus. Il est complètement dépendant de bactéries symbiotiques qui lui assurent son alimentation. Ces dernières sont contenues dans un organe du vestimentifère appelé trophosome (Cary *et al.*, 1993). On a pu identifier ces bactéries grâce à leur séquence d'ADNr 16S (Distel *et al.*, 1988). Elles sont affiliées aux γ -*Proteobacteria*.

Ces bactéries symbiotiques jouent un rôle primordial en faveur de leur hôte puisqu'elles lui fournissent du carbone fixé. Pour cela, elles oxydent l' H_2S présent en grande quantité dans l'environnement hydrothermal. L'énergie produite est suffisante pour fixer le carbone à partir du CO_2 , permettant la transformation de molécules inorganiques en sucres et acides aminés assimilables pour le vestimentifère (Laue & Nelson, 1994 ; Felbeck & Jarchow, 1998). Des études de biochimie ont montré que la fixation du CO_2 se faisait via le cycle de Calvin-Benson (Robinson & Cavanaugh, 1995).

Jusqu'à présent, aucune culture pure de ces bactéries symbiotiques n'a pu être obtenue. Cependant, on peut facilement extraire spécifiquement leur ADN génomique dans la mesure où elles sont les seules à coloniser le trophosome de *R. pachyptila*, et qu'elles atteignent une densité de population très importante. Ces bactéries ont donc pu être étudiées par une approche métagénomique.

Une banque métagénomique a été construite à partir de l'ADN génomique des bactéries symbiotiques (Robinson *et al.*, 1998). Dans cette banque, Robinson et ses collaborateurs ont identifié le gène codant une Ribulose 1,5-biphosphate Carboxylase Oxygénase (RubisCO). Ce gène est l'enzyme clé du cycle de Calvin-Benson. Sa présence confirme donc que la bactérie symbiotique fixe le carbone pour son hôte.

Une autre étude métagénomique a été réalisée pour ces mêmes symbiotes. Cette fois-ci, il s'agissait de comprendre les mécanismes de communication cellulaire qu'entretiennent ces bactéries avec leur hôte (Hughes *et al.*, 1997). Un gène codant un homologue d'une histidine kinase a été identifié dans la banque métagénomique. Ce gène a été cloné puis exprimé chez *E. coli*, montrant que cette histidine kinase s'autophosphorylait en réponse à un signal extérieur et fonctionnait ainsi comme les autres capteurs à histidine kinase déjà caractérisés chez d'autres organismes. Un deuxième gène codant un homologue de régulateurs de réponses a également été identifié dans les fragments génomiques du symbiote. La nature spécifique des signaux reconnus par les symbiotes demeure inexpliquée. Cependant, ces résultats suggèrent que le symbiote dispose d'un

système classique de transduction du signal à deux composants (le capteur à histidine kinase et le régulateur de réponse) qui lui permettent d'interagir avec leur hôte.

Enfin, une dernière analyse métagénomique a permis d'identifier chez ces symbiotes un gène permettant la synthèse de flagelles (Millikan *et al.*, 1999). La présence de ce gène suggère qu'au cours de son cycle de vie, les symbiotes passent par une étape pendant laquelle ils sont présents sous forme libre. Lors de cette étape ils peuvent ainsi coloniser le trophosome des autres vestimentifères.

4.4 L'épibiose bactérienne d'*Alvinella pompejana*

L'annélide polychète tubicole *Alvinella pompejana* colonise les parois des cheminées hydrothermales actives (Desbruyères *et al.*, 1998). Ces annélides sont inféodés à toutes les principales zones actives du Pacifique où ils ont été retrouvés le long de la dorsale orientale de l'océan Pacifique. Au stade adulte, *A. pompejana* présente systématiquement une association avec des bactéries épibiontes (Haddad *et al.*, 1995 ; Cary *et al.*, 1997 ; Desbruyères *et al.*, 1998). Cette association est spécifique et obligatoire. Les épibiontes qui dominent cette communauté microbienne sont des bactéries filamenteuses localisées au niveau des expansions des téguments dorsaux (Cary *et al.*, 1997 ; Desbruyères *et al.*, 1998). Le rôle de ces bactéries filamenteuses n'est pas encore déterminé.

Plusieurs tentatives ont été réalisées pour isoler ces bactéries filamenteuses de la communauté épisymbiotique d'*A. pompejana*. Différents types de bactéries ont été isolées, mais aucune d'entre elles ne leur correspondaient (Prieur *et al.*, 1990 ; Jeanthon, 2000 ; Miroschnichenko *et al.*, 2002 ; Campbell *et al.*, 2001). Des études moléculaires ont permis d'identifier la population symbiotique d'*A. pompejana*, montrant que cette population était dominée par un seul clade monophylétique groupant au sein de la sous-division des ϵ -*Proteobacteria*. (Haddad *et al.*, 1995 ; Cary *et al.*, 1997 ; Campbell & Cary, 2001). Par la suite, des résultats d'hybridation in situ ont indiqué que ces ϵ -*Proteobacteria* correspondaient à la bactérie filamenteuse associée aux expansions dorsales (Cary *et al.*, 1997).

Le rôle de l'association particulière entre les alvinelles et leurs épibiontes bactériens reste à être élucidé. Les symbiotes pourraient à la fois fournir une source de nourriture au polychète, mais également intervenir dans la détoxification de l'environnement du ver (Alayse-Danet *et al.*, 1987 ; Prieur & Jeanthon, 1987). Plusieurs études plaident pour que cette communauté symbiotique soit chimiolithoautotrophe (Desbruyères *et al.*, 1983 ; Van Dover & Fry, 1989 ; Desbruyères *et al.*, 1998). Cependant ce n'est pas par la voie classique de Calvin-Bassham-Benson (cycle de Calvin) que ces bactéries semblent fixer le carbone (Desbruyères *et al.*, 1983, 1998). Elles fixeraient plutôt le carbone par la voie du cycle inverse de l'acide tricarboxylique (rTCA) (Fig. 5).

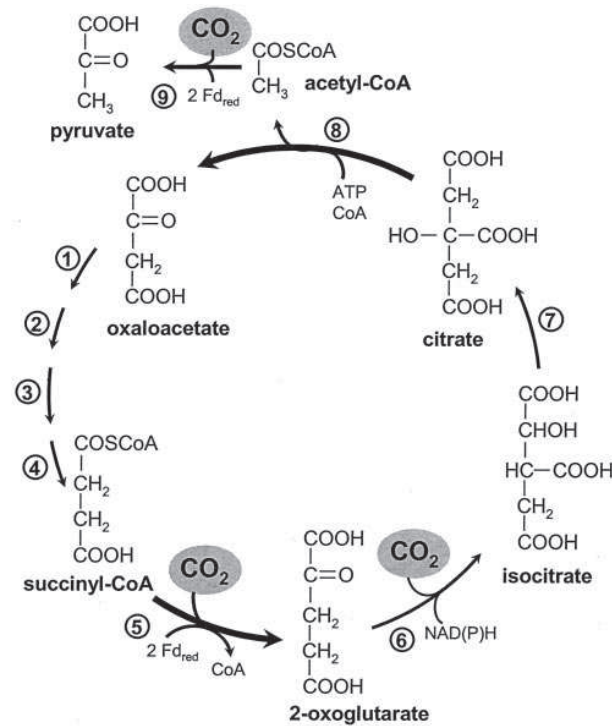


Fig. 5 – Fixation autotrophe du CO_2 via le cycle inverse de l'acide tricarboxylique (rTCA). Les réactions enzymatiques clés sont indiquées par des flèches marquées en gras. Activités enzymatiques : malate déhydrogénase (1), fumarate hydratase (2), fumarate réductase (3), succinyl-CoA synthase (4), 5,2-oxoglutarate : ferredoxin oxidoreductase (5), isocitrate déhydrogénase (6), aconitate hydratase (aconitase) (7), ATP citrate lyase (8) et pyruvate : ferredoxin oxidoreductase (9). Fd_{red} : reduced ferredoxin. (D'après Hugler *et al.*, 2005)

C'est en construisant des banques fosmidiques avec de l'ADN extrait de la biomasse épisymbiotique du polychaete, que Campbell et ses collaborateurs (2003) ont prouvé l'existence de cette voie de fixation du carbone chez ces épibiontes. Deux de leurs fosmides affiliés au phylotype des bactéries filamenteuses contenaient le gène codant l'ATP citrate lyase, une enzyme clé impliquée dans le cycle rTCA. Ce résultat est surprenant, étant donné que la majorité des chimiolithoautotrophes associés à des invertébrés marins (incluant les invertébrés des sources hydrothermales) utilisent le cycle de Calvin pour la fixation du CO_2 . Cette voie métabolique (rTCA) n'a été identifiée que chez quelques micro-organismes, de plus l'utilisation de cette voie et son importance dans les écosystèmes microbiens sont peu connues (Campbell & Cary, 2004). Une sonde a été construite à partir de la séquence nucléotidique du gène codant l'ATP citrate lyase afin, par reverse transcriptase PCR, afin de montrer la diversité et l'expression de ce gène dans les communautés symbiotiques de diverses origines. Les résultats ont indiqué que ce gène était exprimé non seulement chez les bactéries filamenteuses, mais également chez de nombreuses ϵ -Proteobacteria existantes sous forme libre dans l'environnement hydrothermal marin. La présente étude a permis de montrer que le cycle du rTCA jouait un rôle majeur dans la fixation du CO_2 pour l'écosystème hydrothermal marin.

5 Exemples d'analyses métagénomiques à l'échelle des communautés microbiennes : reconstruction de génomes et/ou de voies métaboliques

Les premières tentatives de reconstruction de génomes à partir de banques métagénomiques ont été initiées sur des communautés virales provenant d'océans et de fèces humains (Breitbart *et al.*, 2002, 2003). Depuis, d'autres études de métagénomique ont permis de reconstruire les génomes de micro-organismes incultivés. C'est le cas de l'impressionnante étude menée sur le métagénome de la mer des Sargasses (Venter *et al.*, 2004), et de l'analyse métagénomique de la communauté microbienne d'un biofilm collecté au niveau d'un drainage acide d'une mine (Tyson *et al.*, 2004).

5.1 Etude du métagénome de la mer des Sargasses

Le plus grand projet de métagénomique réalisé jusqu'à présent est celui mené par l'équipe de Venter sur des communautés microbiennes marines provenant d'échantillons de la mer des Sargasses (Venter *et al.*, 2004). Ce projet consistait à faire un inventaire le plus exhaustif possible des gènes et des génomes microbiens que renferme l'océan. La mer des Sargasses, située au large des Bermudes, est une région de l'Atlantique qui a été très bien caractérisée ces 50 dernières années aussi bien en océanographie physique qu'en géochimie (Conte, 1998 ; Steinberg *et al.*, 2001). Elle offre donc la possibilité d'interpréter des données de génomique dans un contexte océanographique bien connu (Venter *et al.*, 2004).

5.1.1 Une production de données considérables

Dans le cadre de ce projet de métagénomique, Venter et son équipe ont pompé quelques centaines de litres d'eau au travers de filtres ayant une porosité de 0.1 à $3\mu\text{m}$. Ils ont ensuite extrait l'ADN génomique de la microflore retenue dans ces filtres, puis ont construit plusieurs banques métagénomiques (avec des inserts d'ADN de petite taille, 2-6 kpb). Plus de deux millions de séquences d'ADN (soit environ 1.6 Mpb d'ADN) issues de ces banques ont été obtenues. Ces séquences ont été annotées, permettant d'identifier plus de 1.2 millions d'ORFs provenant d'environ 1800 espèces génomiques différentes. En avril 2004, 5% des séquences accessibles sur GenBank provenaient de la collection de la mer des Sargasses (Riesenfeld *et al.*, 2004).

La diversité et la distribution microbiennes des échantillons de la mer des Sargasses (Fig. 6) étaient plutôt conformes aux distributions et diversités microbiennes déjà connues pour les océans. Ainsi comme attendu, les groupes microbiens les plus abondants dans les eaux de surface marines, étaient tous représentés. C'est par exemple le cas du clade SAR11 affilié aux α -*Proteobacteria*, du clade SAR86 affilié aux γ -*Proteobacteria*, des cyanobactéries et des espèces

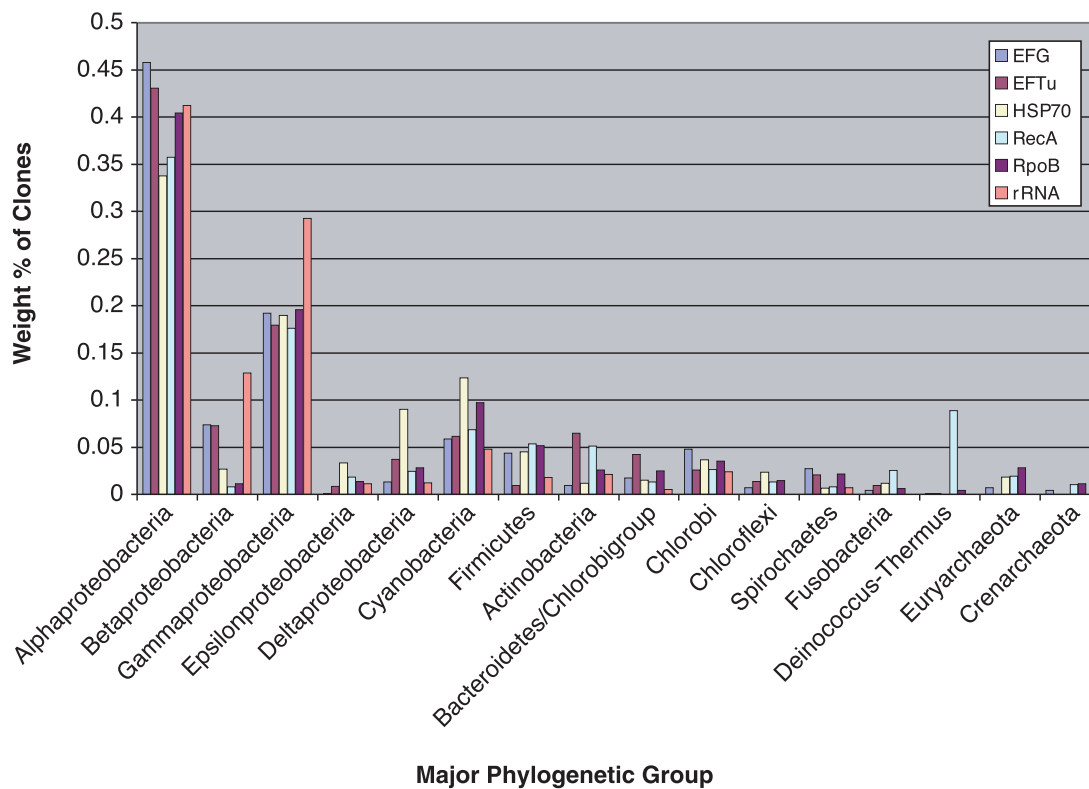


Fig. 6 – Affiliation phylogénétique des séquences provenant de l’analyse métagénomique de la mer des Sargasses. Cette analyse est basée sur l’utilisation de plusieurs marqueurs phylogénétiques : l’ARNr 16S, RecA, EF-TU, EF-G, HSP70 et l’ARN polymérase B (RpoB) (d’après Venter *et al.*, 2004).

appartenant au genre *Cytophaga*. D’autres bactéries marines planctoniques comme les espèces de *Roseobacter*, *Alteromonas*, le clade SAR116 et des crenarchaeotes marines ont également été détectées.

5.1.2 Photobiologie dans les océans

La présence de rhodopsines chez des bactéries avait été démontrée la première fois par Béjà (Béjà *et al.*, 2000a) (Cf § 4.1). Des études ultérieures ont par la suite montré que de nombreuses bactéries affiliées aux *Proteobacteria* avaient elles aussi des “proteorhodopsines”. En effet, 67 nouveaux homologues de protéorhodopsines ont pu être révélés (Sabehi *et al.*, 2003). Ces proteorhodopsines sont optimisées pour capter les différentes longueurs d’onde de la lumière en fonction de la profondeur de l’océan (Béjà *et al.*, 2001 ; de la Torre *et al.*, 2003 ; Sabehi *et al.*, 2003 ; Man *et al.*, 2003 ; Béjà *et al.*, 2002b). Parmi les 1.2 millions d’ORFs identifiés dans l’étude du métagénome de la mer des Sargasses, 782 d’entre eux présentaient une forte similarité avec des protéines photo-réceptrices de type rhodopsine (Venter *et al.*, 2004). Ces 782 rhodopsines identifiées dans le métagénome de la mer des Sargasses élargissent considérablement le spectre d’espèces ayant ce type de pompes à protons. Ceci suggère que la phototrophie par la rhodopsine est un processus marin microbien significatif et largement distribué, pouvant jouer un rôle considérable sur le flux

du carbone et de l'énergie dans les océans.

5.1.3 Biogéochimie des océans

L'analyse du métagénome de la mer des Sargasses a également révélé la présence de nombreux gènes impliqués dans le cycle du phosphore. La mer des Sargasses est caractérisée par un apport et une disponibilité en éléments nutritifs très faibles (Venter *et al.*, 2004). La présence de nombreux gènes impliqués dans l'utilisation du phosphore sous toutes ses formes (polyphosphates, phosphates, et autres phosphores inorganiques) dans cet environnement océanique limité en phosphore, est certainement un bon endroit pour identifier des nouveaux mécanismes de fixation du phosphore (Riesenfeld *et al.*, 2004). Enfin, contrairement au dogme qui affirmait que la nitrification dans les océans n'était accomplie que par des bactéries, les auteurs de cette étude ont identifié un scaffold affilié à une crenarchaeote marine qui contenait une ammonium monooxygénase (Venter *et al.*, 2004). Ceci suggère que les crenarchaeotes marines sont également capables d'oxyder l'ammonium. (Fig. 7)

5.1.4 Difficultés liées à l'analyse du métagénome de la mer des Sargasses

Si cette gigantesque étude a démontré la puissance de la génomique environnementale pour explorer le contenu en gènes et la diversité d'un écosystème donné, elle a également révélé des erreurs qui peuvent être rencontrées lors de l'échantillonnage, de l'assemblage des séquences pour la reconstruction de génomes et de l'interprétation des données (DeLong, 2005).

Tout d'abord, il a été particulièrement difficile de reconstruire des génomes complets d'organismes à partir des séquences générées par cette étude, ceci malgré la quantité énorme de données produites. La grande diversité microbienne présente dans la mer des Sargasses⁴ en est la principale cause (Riesenfeld *et al.*, 2004). Il semblerait de plus que les deux génomes reconstruits par Venter et ses collaborateurs soient certainement des contaminants présents dans l'échantillon d'eau de mer (DeLong, 2005). Des erreurs flagrantes d'assemblages de séquences ont également été repérées. Un exemple évident concerne des contigs contenant des gènes d'origine bactérienne codants l'ARNr 5S et l'ARNr 23S juxtaposés à un gène d'origine archéenne codant l'ARNr 16S (DeLong, 2005). Enfin, l'interprétation de cette quantité énorme de données demeure difficile, dans la mesure où aucune fonction connue n'a pu être inférée à 69% des ORFs identifiés dans cette étude (Venter *et al.*, 2004). Cela suggère qu'il faudra que de nombreux génomes soient entièrement séquencés afin que l'on puisse interpréter au mieux les données issues de cette étude.

⁴Les communautés marines sont caractérisées par une diversité microbienne très riche, de l'ordre de 100 à 200 espèces par millilitre d'eau (Curtis *et al.*, 2002).

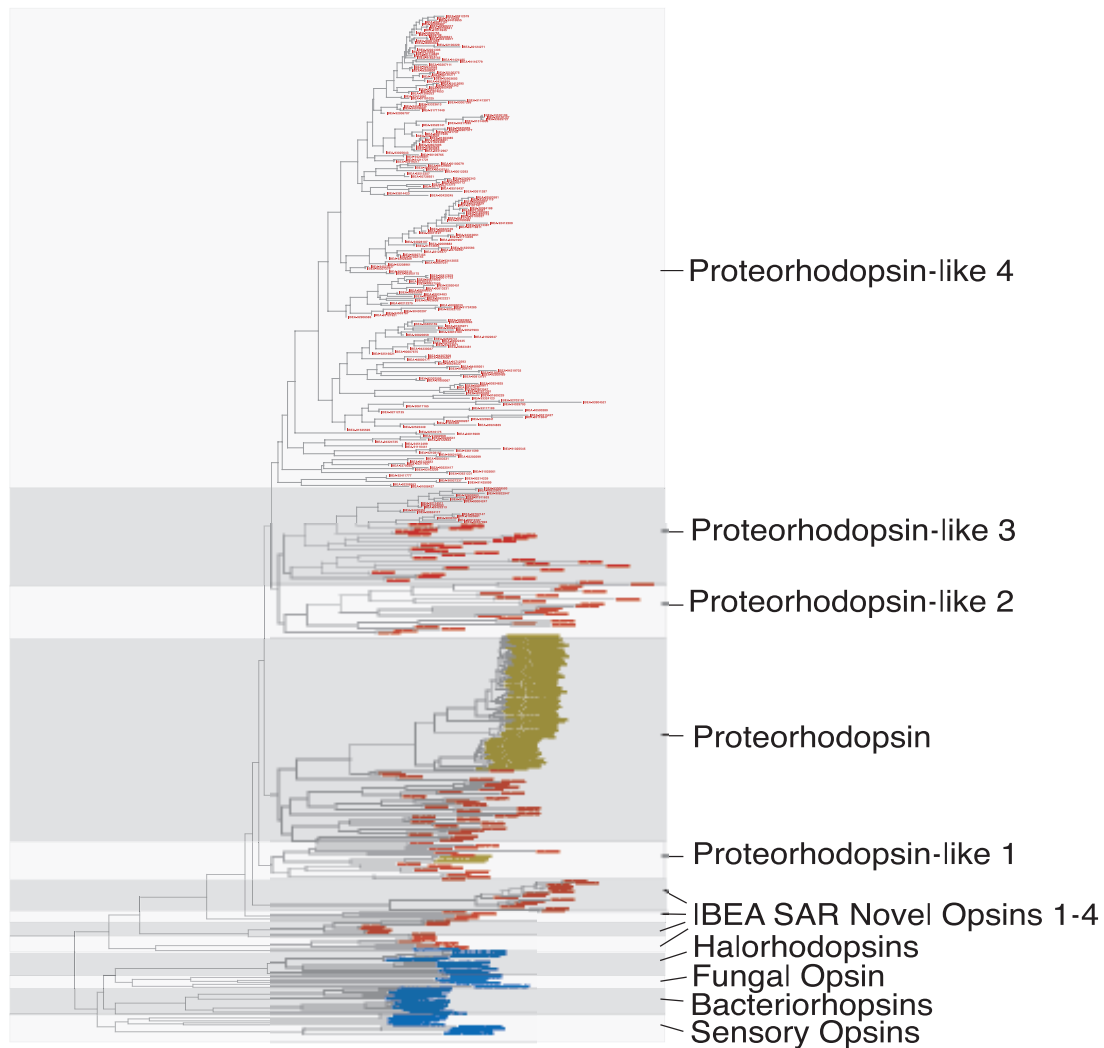


Fig. 7 – Arbre phylogénétique illustrant les différentes rhodopsines identifiées dans la mer des Sargasses ainsi que leurs homologues de la base de donnée GenBank (d'après Venter *et al.*, 2004). Le nom des différentes sous-familles de séquences est indiqué sur la droite de la figure. Les séquences ayant plus de 75 aa ont été alignées avec ClustalW. L'arbre a été construit grâce à la méthode du Neighbor-joining. Le code couleur est le suivant : en bleu, il s'agit de séquences provenant d'espèces cultivées ; en jaune, les séquences issues de micro-organismes incultivés provenant de divers échantillons environnementaux ; en rouge, les séquences d'espèces incultivées provenant de la mer des Sargasses.

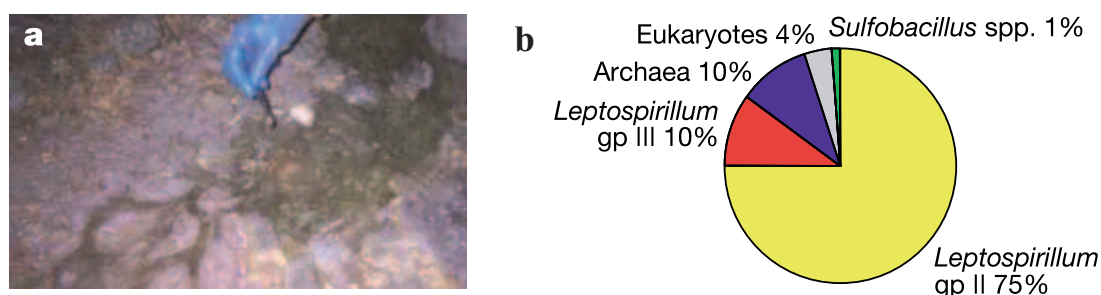


Fig. 8 – (a) Photographie du biofilm de la mine de Richmond; (b) abondance relative des micro-organismes du biofilm évaluée par FISH.(D'après Tyson *et al.*, 2004)

5.2 Etude métagénomique d'un biofilm microbien échantillonné dans le drainage acide d'une mine

5.2.1 L'étude de cycles biogéochimiques à l'aide de la métagénomique

L'un des aspects les plus prometteurs de la métagénomique est de fournir les clés du fonctionnement (voies métaboliques et cycles biogéochimiques) d'une communauté microbienne. L'analyse des fonctions spécifiques de chaque membre d'une communauté peut permettre de générer les modèles d'interactions entre micro-organismes qui maintiennent l'équilibre nutritionnel et énergétique d'une communauté (Handelsman, 2004). Le plus bel exemple est celui de l'analyse menée sur un biofilm flottant à la surface d'une eau de drainage acide provenant de la mine Richmond située dans la montagne "Iron Mountain", en Californie (Tyson *et al.*, 2004) (Fig. 8.a). Cette étude, la première du genre, a permis de reconstruire les génomes d'organismes incultivés d'une communauté microbienne, ainsi que de reconstituer les principales voies de fixation de carbone et d'azote ayant lieu au sein de cette communauté.

5.2.2 Caractéristiques de l'eau de drainage de la mine Richmond

L'effluent de drainage étudié représente l'un des environnements les plus extrêmes connu sur Terre. La solution est caractérisée par un pH très acide de 0.8, une forte concentration en métaux lourds, une température de 42°C, et des conditions de microaérophilie. Mis à part le carbone et l'azote présents dans l'air atmosphérique, il n'y a pas d'autres apports en source de carbone et d'azote dans cette solution. La mine est riche en sulfures parmi lesquels les pyrites (FeS_2) se trouvent sous forme dissoute suite à leur oxydation par des micro-organismes ferroxydants (Bond *et al.*, 2000 ; Baker & Banfield, 2003). La survie des micro-organismes dans cet environnement acide riche en métaux lourds nécessite que les membres de cette communauté microbienne aient des métabolismes chimolithoautotrophes. Des études préliminaires ont montré que la diversité microbienne de ce biofilm était faible (Tyson *et al.*, 2004).

5.2.3 Assemblage des séquences de la communauté microbienne du biofilm

A partir de l'ADN extrait de ce biofilm, Tyson et ses collaborateurs ont construit une banque métagénomique en clonant des inserts de petite taille (~ 3 kpb) dans le vecteur pUC18. La faible diversité microbienne du biofilm (Fig. 8.b) leur a permis de cloner tout le métagénome de cette communauté et de le séquencer avec un fort taux de recouvrement. Via le séquençage aléatoire de 76.2 Mpb de séquences, ils ont reconstitué les génomes presque complets des deux micro-organismes dominant le biofilm : celui de la bactérie *Leptospirillum* group II (avec une couverture de 10X, représentant ainsi le premier génome du phylum des *Nitrospirae*) et celui de l'archaea *Ferroplasma* type II (couverture de 10X)⁵. De plus, ils ont reconstruit partiellement le génome d'une autre espèce de *Leptospirillum* (*Leptospirillum* group III) et deux génomes apparentés aux *Ferroplasma* (*F. acidarmanus* type I et "G-plasma").

La reconstruction de ces génomes a été relativement facile étant donné que le biofilm avait une faible diversité microbienne et que le contenu en G+C des génomes des taxons dominants étaient significativement distincts les uns des autres (Tyson *et al.*, 2004).

5.2.4 Analyse métabolique

La reconstruction partielle ou presque complète des génomes du biofilm a révélé les voies de fixation de carbone et d'azote et de production d'énergie pour chaque organisme. Par exemple, les analyses ont montré que certaines enzymes comme la CO deshydrogénase et la formate hydrogénylase étaient apparemment présentes chez tous ces organismes, indiquant qu'ils étaient capables de fixer le carbone. D'autre part, seul l'organisme affilié au group III de *Leptospirillum* semblait pouvoir fixer l'azote, laissant sous-entendre que la survie du biofilm reposait sur cette bactérie (Galperin, 2004). *Ferroplasma* et *Leptospirillum* semblent tous deux tirer leur énergie de l'oxydation du fer (Fig. 9). Les génomes des deux groupes contiennent d'ailleurs leur propre système de chaînes de transports d'électrons. De plus, tous les génomes contiennent des gènes dont la fonction est associée à l'expulsion d'éléments toxiques, de gènes responsables de l'efflux des protons (permettant de maintenir un pH intracellulaire proche de la neutralité) et des gènes de résistance aux métaux.

L'analyse de la communauté de ce biofilm est un bon modèle pour l'analyse future de nouvelles communautés environnementales. Cependant, il sera certainement plus difficile de déterminer l'origine des fragments d'ADN et d'assigner des fonctions à des communautés qui sont phylogénétiquement ou physiologiquement plus complexes (Galperin, 2004).

⁵Ce génome de 1.82 Mpb contient un ARNr 16S ayant 99% de similarité avec celui de *F. acidarmanus* (isolé plus tôt sur ce même site (Dopson *et al.*, 2004)

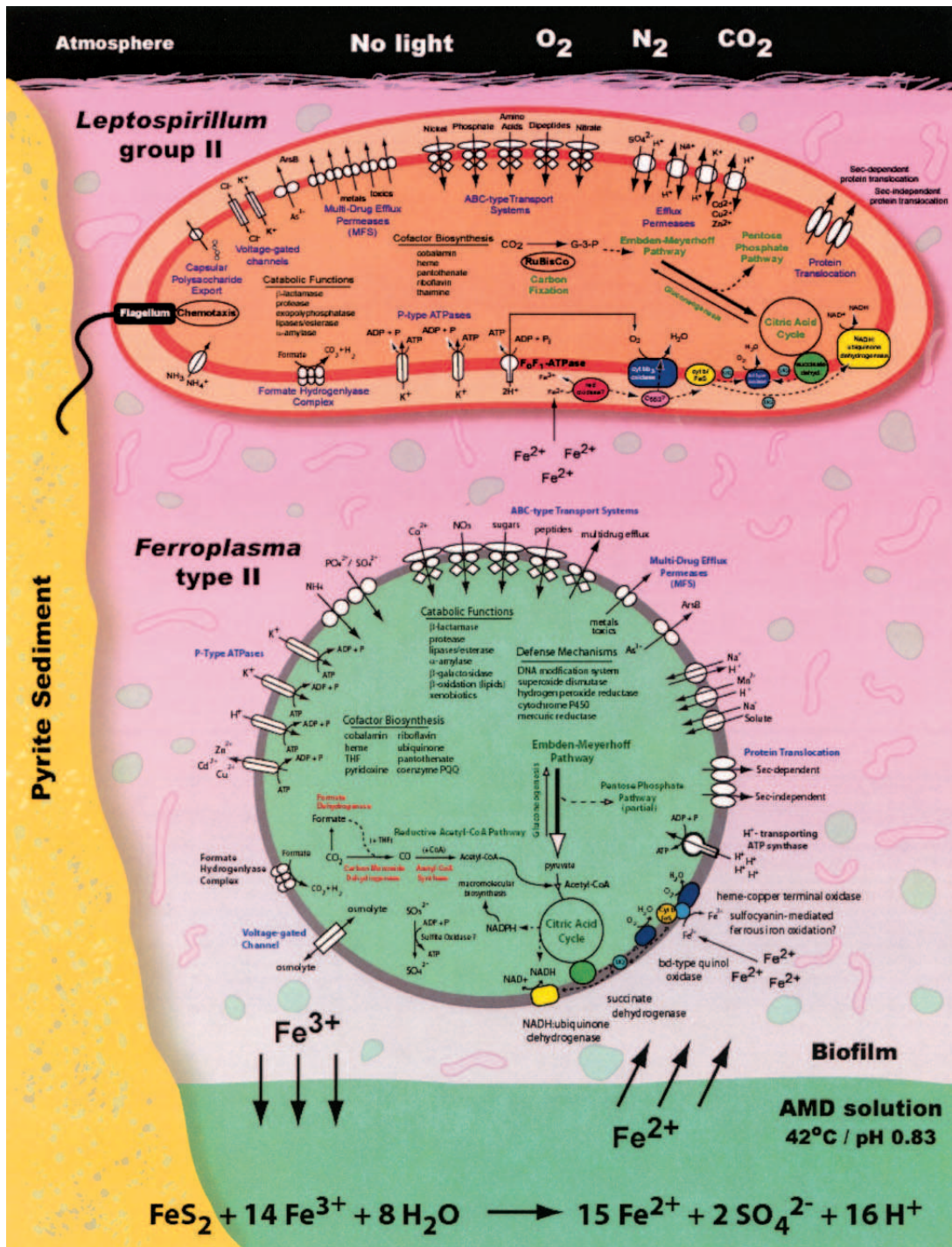


Fig. 9 – Schéma représentant les cycles biogéochimiques déduits de l'analyse métagénomique du biofilm microbien prélevé dans le drainage acide de la mine de Richmond. Ce schéma a été construit suite à l'analyse des 2,180 ORFs identifiés dans le génome de la bactérie *Leptospirillum* du groupe II ainsi que les 1 931 ORFs identifiés dans celui de la bactérie *Ferroplasma* de type II. AMD : pour Acid Mine Drainage. (D'après Tyson *et al.*, 2004)

5.3 Oxydation anaérobie du méthane dans les sédiments marins profonds

Le méthane piégé sous forme d'hydrates de méthane, est présent en grande quantité dans les sédiments marins (Kvenvolden, 1999). On trouve des gisements particulièrement vastes de ces hydrates de méthane au niveau des talus continentaux des océans, entre 500 et 2000 m de profondeur (Kvenvolden, 1988). Une grande partie de ce méthane enfermé dans les sédiments est oxydée par des micro-organismes qui occupent les régions anoxiques des océans (Judd *et al.*, 2002 ; Boetius & Suess, 2004). Etant donné que le méthane est un gaz à effet de serre encore plus puissant que le dioxyde de carbone, l'oxydation anaérobie du méthane est un processus microbien de toute première importance, car il permet de limiter le réchauffement du climat. Ce processus microbien a été attribué à des consortia d'archaea méthanotrophes et de bactéries sulfato-réductrices (Zehnder & Brock, 1979). Les archaea méthanotrophes ont été affiliées sur la base de leur ARNr 16S à de nouvelles lignées archaéennes appelées ANME-1, ANME-2, ANME-3 (Hinrichs *et al.*, 1999 ; Boetius *et al.*, 2000 ; Knittel *et al.*, 2005).

Comme toutes les tentatives d'isolement en culture pure de ces méthanotrophes ont échouées, les mécanismes de l'oxydation anaérobie du méthane (AOM) demeuraient jusqu'à peu à peine connus. Seule l'hypothèse suivante avait été formulée : l'oxydation anaérobie du méthane serait réalisée par des micro-organismes capables de conduire une voie "inversée" de la méthanogénèse (Hoehler *et al.*, 1994). Afin de comprendre les mécanismes de l'AOM, plusieurs études de métagénomique ont été menées ces dernières années (Hallam *et al.*, 2003 ; Krüger *et al.*, 2003 ; Hallam *et al.*, 2004) (Meyerdierks *et al.*, 2005).

Les premières analyses métagénomiques ont montré que des fragments génomiques assignés aux ANME-1 et ANME-2, contenaient une methyl-coenzyme-M réductase (MCR) modifiée (Hallam *et al.*, 2003 ; Krüger *et al.*, 2003). La MCR⁶ est une enzyme clé de la méthanogénèse ; elle en assure la dernière étape (Fig. 10.a). Une analyse biochimique de cette nouvelle MCR révéla qu'elle contenait un cofacteur F430 modifié (Krüger *et al.*, 2003). Plus récemment, presque tous les gènes codant les enzymes typiquement associées à la voie de la méthanogénèse ont été identifiés dans des fragments génomiques assignés aux ANME-1 et ANME-2 (Hallam *et al.*, 2004) (Fig. 10.b). L'ensemble de ces données confortent la théorie que les archaea ANME conduisent une méthanogénèse inverse, "reverse methanogenesis". La plus récente étude métagénomique traitant de l'oxydation anaérobie du méthane a, quant à elle, permis d'aller plus loin dans les caractéristiques génomiques des différents groupes d'ANME (Meyerdierks *et al.*, 2005). Les gènes et les opérons potentiellement impliqués dans cette voie y sont analysés en détail.

⁶Dans la voie de la méthanogénèse, la MCR catalyse la réduction de l'intermédiaire réactionnel méthyl-coenzyme-M, conduisant à la formation de méthane. La MCR est composée de trois sous-unités distinctes ($\alpha_2\beta_2\gamma_2$) qui sont fortement liées à deux molécules d'un composé nickel, appelé cofacteur F430 (Ermler *et al.*, 1997).

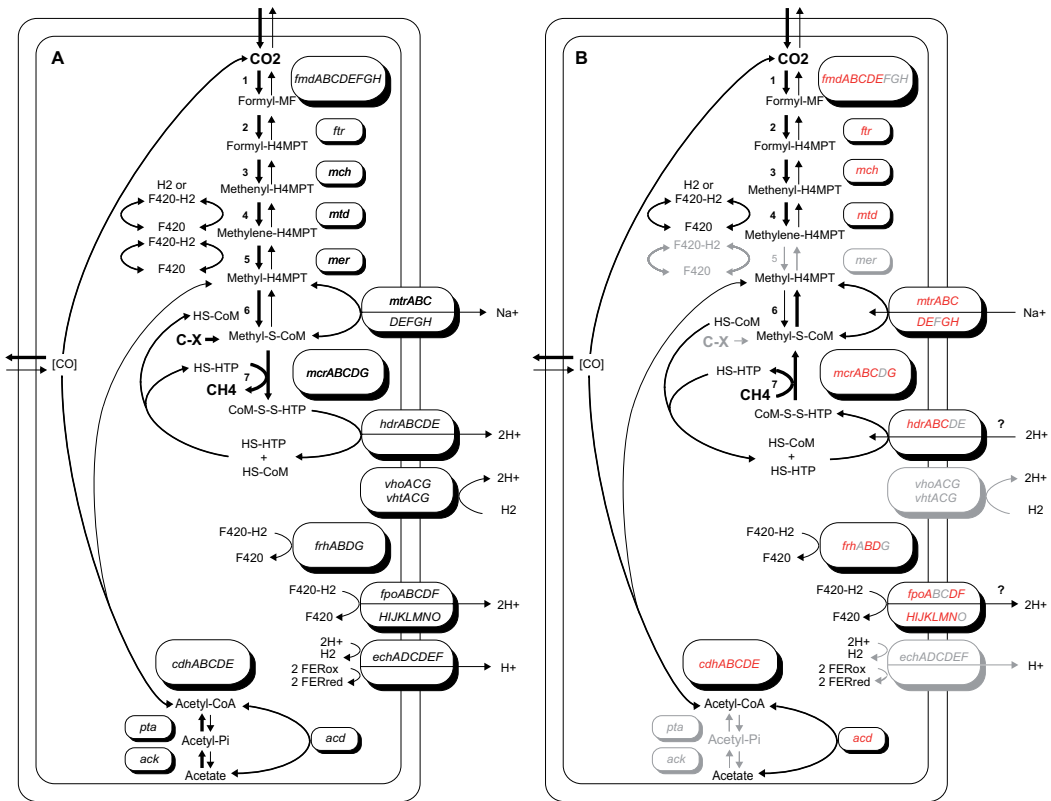
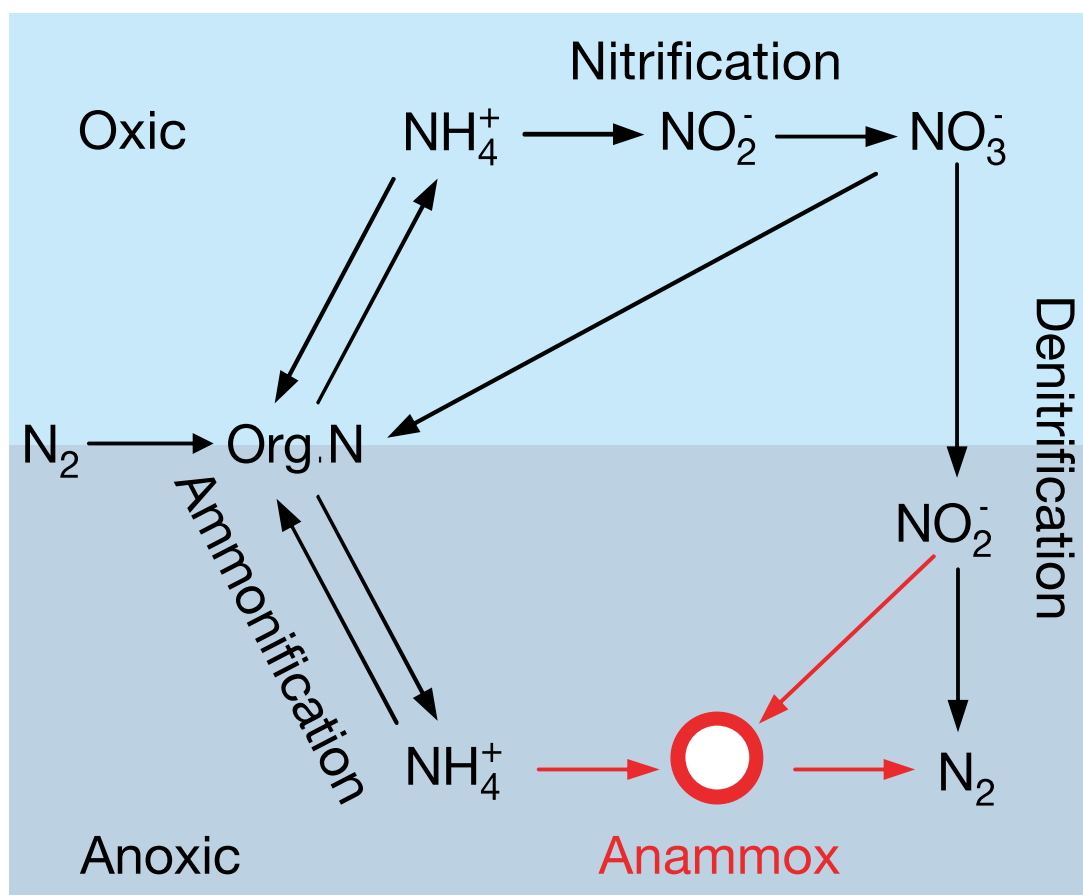


Fig. 10 – Modèle hypothétique de la méthanogénèse inverse chez les archaea de la lignée ANME-1. (A) Voie de la méthanogénèse. Les gènes identifiés sont indiqués en noir. (B) Voie de la méthanogénèse inverse chez les ANME-1. Cette voie est déduite de la prédiction des gènes identifiés dans les fosmides affiliés aux ANME-1. Les gènes identifiés dans les fosmides sont indiqués en rouge, ceux n'ayant pas été identifiés sont en gris (d'après Hallam *et al.*, 2004).

Fig. 11 – Cycle de l'azote (d'après Kuypers *et al.*, 2003).

5.4 Reconstitution du génome de la bactérie *Kuenenia stuttgartiensis*, un acteur majeur du cycle de l'azote

Le cycle de l'azote est étudié depuis le 19^{ème} siècle, mais il n'a pas encore livré tous ses secrets. Très récemment, on a observé que l'ammonium pouvait être oxydé par des bactéries anaérobies dites bactéries anammox (pour anaerobic ammonium oxidation) (Fig. 11) (Schmid *et al.*, 2000 ; Strous & Jetten, 2004). La réaction métabolique anammox permet, sans l'apport d'oxygène, de transformer l'ammonium et les nitrites en azote gazeux. Cette transformation produit comme intermédiaire l'hydrazine qui est une substance hautement toxique. Les cellules se protègent de l'hydrazine en la confinant dans une vésicule appelée "anammoxosome". Cette vésicule identifiée comme étant le siège de la réaction, est constituée de lipides particuliers appelés ladderanes, dont la biosynthèse était jusqu'à présent inconnue.

Le processus anammox est un élément majeur du cycle de l'azote. On estime qu'il contribue à plus de 50% de la transformation de l'azote dans les océans (Arrigo, 2005). Les quelques bactéries anammox identifiées jusqu'à présent, comme par exemple *Kuenenia stuttgartiensis* (Schmid

et al., 2000), appartiennent toutes à l'ordre des *Planctomycetales*. Ces bactéries n'ont pu être isolées en culture pure ; elles nécessitent donc d'être étudiées par des approches de génomique. Dans le cadre d'une étude métagénomique, le génome de la bactérie anammox *K. stuttgartiensis* a pu être dernièrement reconstitué permettant d'apporter plusieurs éléments de réponses sur le mécanisme du processus anammox (Strous *et al.*, 2006).

K. stuttgartiensis est une bactérie qui se divise très lentement (une division toutes les 2 à 3 semaines). Cette bactérie n'a pu être isolée en culture pure. Cependant, pendant plus d'un an, elle a pu être maintenue en coculture, dans un bioréacteur. Cette culture a été réalisée en condition d'anaérobiose, dans un bioréacteur alimenté par des eaux usées qui avaient la particularité d'être riches en ammonium et en nitrites (Strous *et al.*, 2006). L'ADN génomique contenu dans le bioréacteur a été extrait, puis cloné afin de construire des banques de banques de BAC et fosmidiques. Le génome de *K. stuttgartiensis* a pu être reconstitué à plus de 98%. L'analyse du génome de *K. stuttgartiensis* a permis :

- d'identifier les gènes responsables de la synthèse des ladderanes,
- d'identifier les gènes de la voie biochimique de l'oxydation anaérobie de l'ammonium, permettant ainsi de comprendre comment l'hydrazine est métabolisé,
- de mettre en évidence un très grand répertoire de voies métaboliques (plus de 200 gènes directement impliqués dans le catabolisme et la respiration ont été identifiés dans le génome),
- de confirmer que la seule voie de fixation du carbone chez cette bactérie anammox est celle de l'acétyl CoA (aucune autre voie connue et complète n'a pu être identifiée dans le génome),
- de mettre fin à une controverse sur l'origine évolutive du groupe des Planctomycètes, en montrant que les Planctomycètes ont plus de gènes en commun avec les *Chlamydiées* qu'avec n'importe quel autre groupe, et que de ce fait, elles descendent donc du même ancêtre commun que ces dernières.

L'ÉCOSYSTÈME HYDROTHERMAL MARIN

1	Découverte de l'écosystème hydrothermal marin	71
2	Localisation des sources hydrothermales	71
2.1	Tectonique des plaques	71
2.2	Dorsales océaniques	72
3	Origine des sources hydrothermales	74
4	Diversité des édifices hydrothermaux marins	76
4.1	Les fumeurs noirs	76
4.2	Les diffuseurs	76
4.3	Les autres structures	77
5	La biocénose associée à l'écosystème hydrothermal marin profond	77
5.1	La faune hydrothermale	78
5.1.1	<i>Alvinella pompejana</i>	78
5.1.2	<i>Riftia pachyptila</i>	80
5.1.3	<i>Rimicaris exoculata</i>	80
5.2	La microflore hydrothermale	80
5.2.1	Types d'habitats microbiens	81
5.2.2	Diversité métabolique des micro-organismes hydrothermaux	81
5.2.3	Contribution des méthodes moléculaires à la connaissance de la diversité des communautés microbiennes des sources hydrothermales marines profondes	87

1 Découverte de l'écosystème hydrothermal marin

L'océan profond constitue l'environnement le plus étendu de notre planète, et pourtant c'est l'un des moins étudiés. L'écosystème abyssal couvre 307 millions de km², soit les 2/3 de la surface du globe terrestre. A cause du froid et de l'absence de lumière qui y règnent, cet écosystème a longtemps été considéré comme l'un des plus pauvres de la planète. Cependant, cette vision d'un monde abyssal quasiment stérile fut ébranlée à la fin des années 70 lors d'une exploration sous marine menée sur la dorsale océanique des Iles Galápagos. C'est à 2500 m de profondeur, que des géologues embarqués à bord du sous-marin américain Alvin découvrirent avec stupéfaction l'existence d'une faune exubérante (Corliss & Ballard, 1977 ; Lonsdale, 1977 ; Corliss *et al.*, 1979). Ils observèrent des communautés luxuriantes de bivalves, de grands vers et autres organismes réparties tout autour de sources d'eau chaude. Cette découverte de ce qu'on appellera plus tard les sources hydrothermales marines profondes a bouleversé nos connaissances en océanographie biologique ainsi que notre vision de l'océan profond.

2 Localisation des sources hydrothermales

2.1 Tectonique des plaques

Depuis cette découverte, de nombreux sites hydrothermaux ont été localisés. On les retrouve généralement dans des zones à forte activité tectonique. La tectonique des plaques suppose que l'enveloppe externe de la Terre (lithosphère) est formée de plaques rigides et que toute l'activité, sismique et volcanique, est concentrée aux frontières entre plaques. Le mouvement des plaques est entretenu par la création et la destruction du plancher océanique qui participe à la convection dans le manteau terrestre. Il existe trois mouvements de plaques (Fig. 1) :

- Celui se produisant au niveau des zones d'expansion ou d'accrétion du plancher océanique, zones où deux plaques lithosphériques divergent. Ce mouvement de divergence résulte de forces de tension qui émanent des cellules de convection de l'asthénosphère. Il est à l'origine de la formation des dorsales océaniques. Les dorsales océaniques qui constituent une chaîne de montagnes sous-marine, s'étendent sur près de 60 000 km et sont situées à des profondeurs comprises entre 500 et 4 000 m avec une moyenne de 2 500 m.
- Celui se produisant au niveau des zones de subduction. Les zones de subduction résultent de la convergence de deux plaques lithosphériques. Il peut s'agir d'une plaque océanique qui glisse sous une plaque continentale (à l'origine des arcs continentaux volcaniques), ou de deux plaques océaniques qui convergent (à l'origine des arcs insulaires volcaniques). Dans ce dernier cas, c'est la plaque océanique la plus dense (généralement la plus vieille) qui s'enfonce sous l'autre. De bons exemples d'arcs insulaires volcaniques se retrouvent dans le Pacifique Ouest, avec les grandes fosses des Mariannes, de Tonga, des Kouriles et des Aléoutiennes, ainsi que dans l'océan Atlantique au niveau de la fosse de Puerto Rico ayant

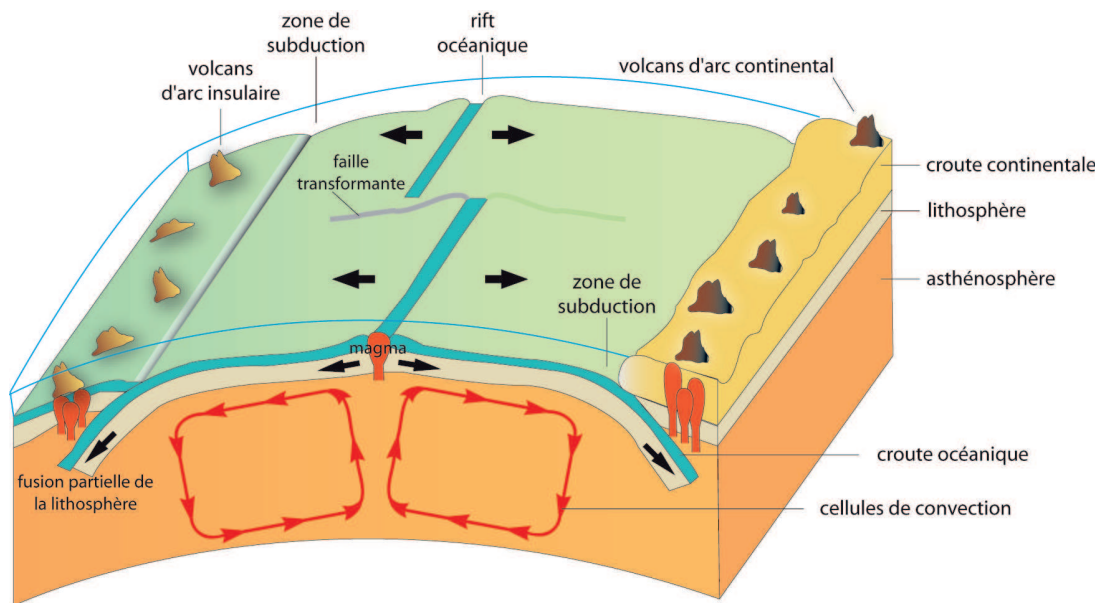


Fig. 1 – Principe général de la tectonique des plaques, source <http://ggl.ulaval.ca/personnel/bourque/sl/>.

donné naissance à l'arc des Antilles. La convergence de deux plaques lithosphériques peut parfois ouvrir un espace océanique en arrière et parallèlement à une chaîne volcanique, créant ce que l'on appelle un bassin arrière arc.

- Enfin, celui localisé au niveau des zones transformantes, le long desquelles coulissent des plaques ou des fragments de plaques, sans création ni résorption de croûte. C'est le cas par exemple de la faille de San Andreas en Californie.

2.2 Dorsales océaniques

Les dorsales océaniques parcourent les fonds océaniques sur près de 60 000 km avec une profondeur moyenne de 2 500 m. Elles forment le système volcanique le plus étendu sur Terre. Leur vitesse d'expansion varie d'un type de dorsale à un autre (Edmonds *et al.*, 2003) (Fig. 2). Ainsi, la dorsale médio-Atlantique (ou Mid-Atlantic Ridge, MAR), la dorsale Sud-Ouest de l'océan Indien et la dorsale Gakkel de l'océan Arctique s'ouvrent lentement (Fig. 3). Leur expansion est inférieure à 10-20 mm par an. D'autres dorsales ont des vitesses d'expansions intermédiaires de l'ordre de 50 à 80 mm par an, il s'agit par exemple du nord de la dorsale orientale de l'océan Pacifique (ou East Pacific Rise, EPR), ou du rift des Galápagos. Enfin, un dernier type de dorsales concerne celles qui ont des vitesses d'ouverture beaucoup plus rapides, c'est le cas des dorsales 9°N-10°N et 13°N EPR ainsi que la dorsale 17°S de l'EPR dont l'expansion est estimée entre 110 et plus de 150 mm par an.

A ce jour, seules quelques centaines de kilomètres de dorsales ont été explorées, et plus particulièrement celles ayant des expansions rapides. Au niveau de ces dorsales règne une activité

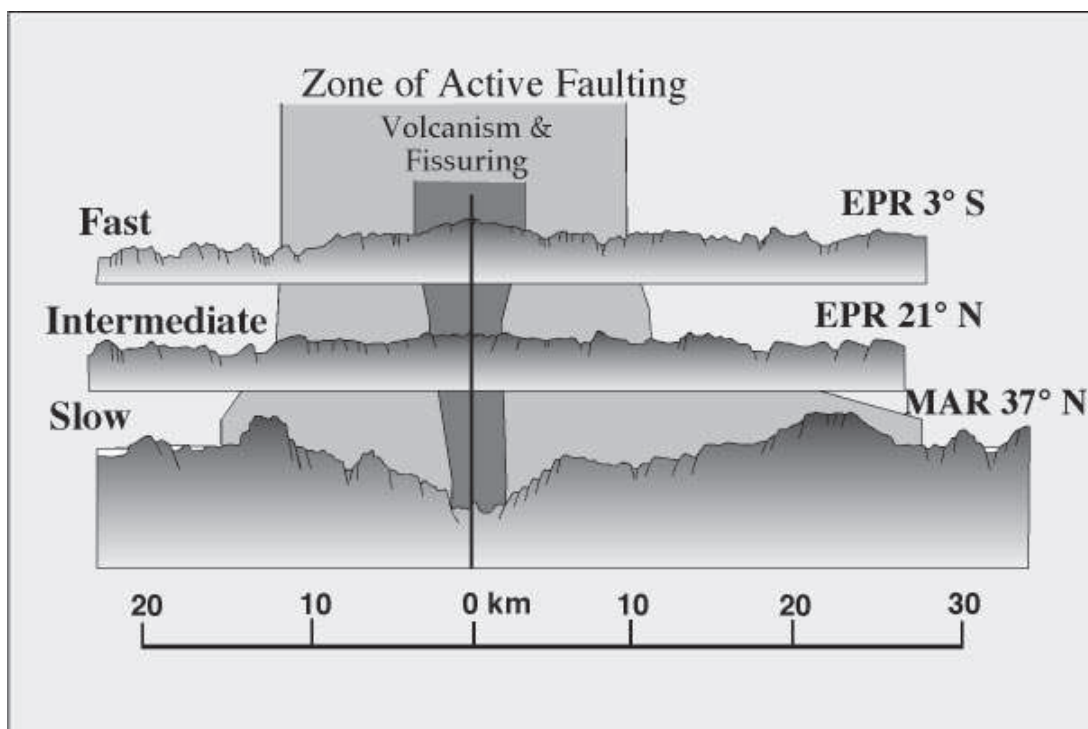


Fig. 2 – Les trois types de dorsales océaniques : celles à expansion rapide (en haut du schéma), à expansion intermédiaire (au milieu) et à expansion lente (en bas). La topographie, la profondeur de la chambre magmatique, les fréquences des activités volcaniques, les tailles des édifices hydrothermaux, la diversité et la composition de la faune hydrothermale dépendent tous de la vitesse d'expansion des océans. EPR, East Pacific Rise; MAR, Mid-Atlantic Ridge (d'après Kelley *et al.*, 2002).

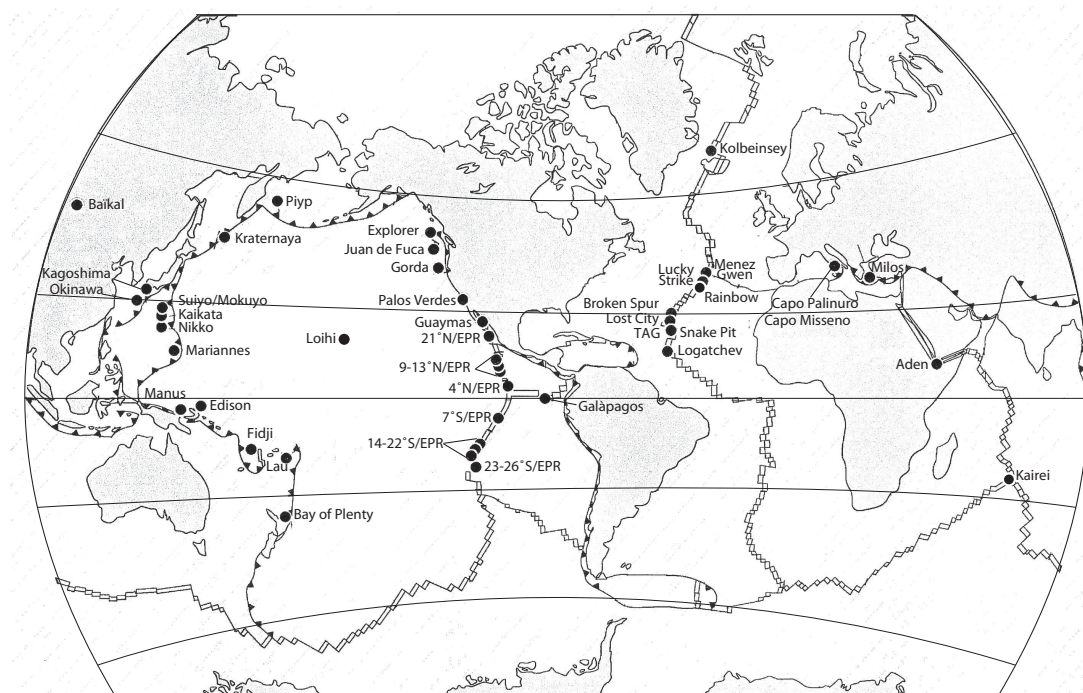


Fig. 3 – Localisation des dorsales océaniques et des sites hydrothermaux marins étudiés par les biologistes. Les double-traités représentent les zones d'expansion, les traits simples indiquent les failles transformantes, les points localisent les sites hydrothermaux et les triangles situent les zones de subduction (d'après Von Damm, 2001).

hydrothermale intense. La carte suivante présente une liste non exhaustive des sites hydrothermaux étudiés jusqu'à présent (Fig. 3). A l'image des sites hydrothermaux récemment découvertes dans l'océan Indien (Van Dover *et al.*, 2001) et l'océan Atlantique (Kelley *et al.*, 2001 ; Charlou *et al.*, 2002), de nombreux sites restent probablement à découvrir.

3 Origine des sources hydrothermales

Les sources hydrothermales marines profondes résultent d'une infiltration de l'eau de mer au niveau de fissures qui apparaissent à des endroits fragilisés par l'activité tectonique sous-jacente (Jannasch, 1995) (Fig. 4). L'eau de mer initialement froide ($\sim 2^{\circ}\text{C}$) et dense s'infiltré jusqu'à une profondeur de plusieurs kilomètres. A l'approche de la chambre magmatique et au contact des roches en surfusion, elle s'échauffe et voit sa densité diminuer au-delà de son point critique. On parle alors de fluide supercritique. Sous cet état, sa capacité calorifique, son pouvoir de solubilisation et sa fluidité augmentent. Sous l'effet des pressions hydrostatiques et lithosphériques, le fluide surchauffé va remonter à la surface en lessivant les roches basaltiques. Il se charge alors en ions métalliques (Cu^{2+} , Ca^{2+} , Mn^{2+} , Fe^{2+} , Zn^{2+} , Si^{+}) et en gaz dissous (H_2S , CH_4 , CO , CO_2 , H_2) et s'appauvrit en Mg^{2+} , SO_4^{2-} , NO_3^- et PO_4^{3-} (Annexe B). Finalement, le fluide qui peut atteindre des températures voisines de 400°C , rejaille au niveau du plancher océanique en des points focalisés qui constituent les sources hydrothermales. Les propriétés physiques et chi-

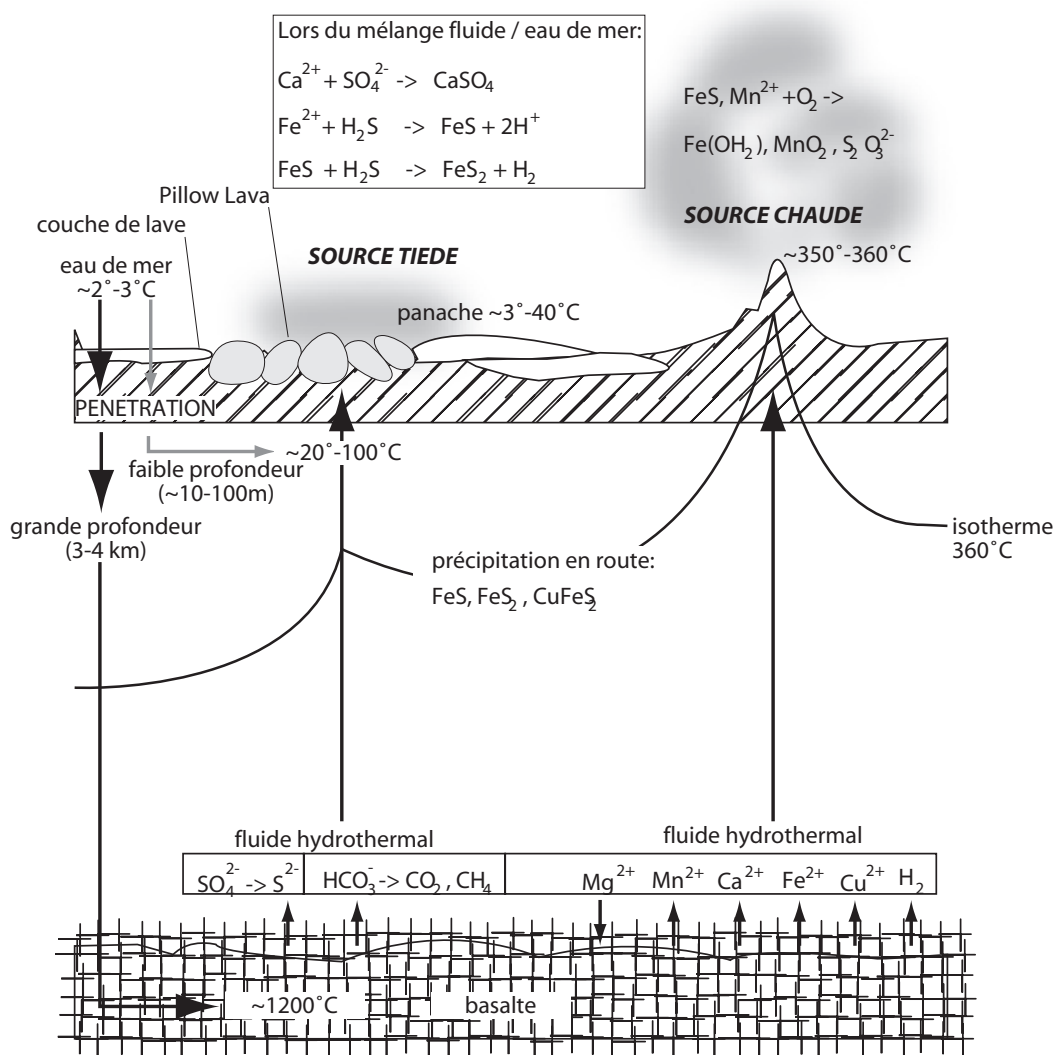


Fig. 4 – Principe de la circulation de l'eau de mer dans la croûte océanique, au niveau des zones d'expansion (d'après Jannasch, 1995).

miques de ce fluide hydrothermal anoxique et acide sont sans équivalent dans les milieux côtiers et contrastent très fortement avec celle de l'eau de mer.

Selon le degré de mélange du fluide avec l'eau de mer avant son émission, la composition, le débit et la température du fluide en sortie peuvent varier d'un site à l'autre conditionnant ainsi la forme sous laquelle le fluide va être émis et la distribution de la biocénose hydrothermale (Desbruyeres *et al.*, 2001).

4 Diversité des édifices hydrothermaux marins

4.1 Les fumeurs noirs

L'édification d'une cheminée hydrothermale se déroule en plusieurs étapes (Fouquet *et al.*, 1988). Lorsque le fluide ne subit pas de dilution, il est émis en surface à de très fortes températures (350°C-400°C) et avec un débit très important (0.4-3 m.s⁻¹). Ce fluide généralement acide, contient de l'hydrogène sulfuré (H₂S) et il est chargé en métaux tels que le fer, le zinc et le cuivre. La première étape de l'édification d'une cheminée hydrothermale, a lieu lorsque le fluide hydrothermal rentre au contact de l'eau de mer riche en ions sulfates. A ce moment, une paroi poreuse constituée de barytine (sulfates baryum, BaSO₄) et d'anhydrite (sulfate de calcium, CaSO₄) se forme par précipitation chimique. Ensuite, l'édifice croît verticalement suite à des dépôts de sulfures métalliques. Ces sulfures métalliques proviennent de la réaction chimique entre les sulfates de baryum, le calcium et les ions métalliques du fluide hydrothermal. Progressivement, les sulfates sont remplacés par des sulfures de fer, de zinc et de cuivre (chalcopyrite, pyrrhotite, pyrite et sphalérite) qui sont plus stables. La présence d'inclusions d'anhydrite persistant dans les sulfures métalliques témoigne de ce processus de remplacement. Le dépôt de ces sulfures métalliques réduit la porosité de l'édifice. La température augmente alors au centre de la cheminée, et de nouveaux sulfures se déposent à la place de l'anhydrite.

Ces édifices minéraux qui canalisent la sortie du fluide, peuvent atteindre de 1 à 20 m de hauteur. Leur morphologie et leur composition minéralogique varient en fonction du débit d'émission du fluide, des courants et de l'âge de la cheminée. Les cheminées sont généralement représentées comme des structures cylindro-coniques, mais dans la réalité elles forment des structures irrégulières présentant parfois des expansions latérales ou flanges.

4.2 Les diffuseurs

Lorsque le fluide est dilué par l'eau de mer avant son émission (lors de son passage à travers la lithosphère océanique), sa température est alors moins élevée ne permettant pas la précipitation des sulfures métalliques. Ce fluide diffuse à faible vitesse (0.1-0.5 m.s⁻¹) au travers de petits canaux. Si la dilution du fluide est supérieure à 70%, le fluide qui en résulte n'est plus assez chaud (<10°C à 50°C) pour permettre le transport d'importantes quantités de sulfures et de métaux dissous. La faible minéralisation du fluide et son faible débit entraîne la précipitation de minéraux (silice amorphe, oxydes de fer) qui engendre l'édification de structures appelées diffuseurs. Cependant, si la dilution du fluide ne dépasse pas les 50%, sa température est de l'ordre de 100°C à 250°C. A ces températures, la précipitation de particules blanches de silice, d'anhydrite et de barite participe à la formation d'édifices hydrothermaux appelés diffuseurs.

4.3 Les autres structures

Un nouveau type d'édifice a été récemment découvert au niveau du site hydrothermal de Lost City (Kelley *et al.*, 2001). Ce site est localisé non pas au niveau de l'axe d'une dorsale océanique, mais à 15 km d'un axe, à savoir celui de la dorsale médio-Atlantique. La chimie du fluide hydrothermal de ce site est différente de celle observée en contexte basaltique : le fluide est particulièrement basique (pH 9 à pH 11), relativement froid ($<40^{\circ}\text{C}$ à 90°C) et les édifices minéraux formés à cet endroit sont composés de carbonates (CaCO_3) et d'hydroxyde de magnésium ($\text{Mg}(\text{OH})_2$) (Kelley *et al.*, 2005). Ces édifices forment parfois de grandes cheminées blanches de carbonates pouvant atteindre 30 à 60 m de haut.

Enfin, un dernier exemple de système hydrothermal marin est celui que l'on observe par exemple dans le bassin de Guaymas. Dans le cas présent, le fluide percole au travers de couches sédimentaires pouvant atteindre 400 m d'épaisseur (sédiments d'origine planctonique très riches en matières organiques, en hydrocarbures et sulfates). Lorsque le fluide atteint la surface des sédiments, il diffuse lentement et sa température n'est que de quelques dizaines de degrés.

5 La biocénose associée à l'écosystème hydrothermal marin profond

Malgré des conditions environnementales très contraignantes et dites "extrêmes", la biomasse observée dans l'écosystème hydrothermal marin profond est 500 à 1000 fois plus élevée que celle estimée en milieu abyssal. La flore microbienne et la faune de l'écosystème hydrothermal sont souvent distribuées en auréoles concentriques autour des émissions hydrothermales, ceci en fonction de leur capacité de résistance aux conditions environnementales et de leurs besoins trophiques. Les substances chimiques dissoutes dans les fluides hydrothermaux fournissent l'énergie vitale pour ces communautés. Si la composition du fluide hydrothermal est relativement stable d'après des mesures obtenues sur plusieurs sites hydrothermaux de l'EPR à des intervalles de temps de 1 à 4 ans (Damm *et al.*, 1995), le mélange du fluide avec l'eau de mer environnante est cependant très variable et peut fluctuer dans le temps. La faune hydrothermale ne peut survivre sans la présence du fluide. On ne connaît pas la durée de vie des sources hydrothermales, mais on sait qu'elles sont éphémères. Si une source hydrothermale s'éteint à la suite d'un tremblement de terre, d'une éruption volcanique ou de l'accumulation de sulfures qui peuvent obstruer le réseau de conduits qui alimentent la source, toute la communauté animale qui vivait à proximité de la source, disparaît, montrant la très forte dépendance des organismes hydrothermaux vis-à-vis de l'émission des fluides.

5.1 La faune hydrothermale

A proximité des sources hydrothermales, on trouve une faune très abondante formant un véritable “oasis de vie”. Parmi les organismes métazoaires les plus étudiés, on peut citer : les annélides polychètes *Alvinellidae* dont les espèces les plus connues sont *Alvinella pompejana*, *A. caudata* et *Paralvinella palmiformis* ; les vestimentifères *Riftiidae* avec l'unique espèce *Riftia pachyptila* ; les mollusques bivalves de la famille des *Mytilidae* avec les espèces *B. thermophilus* et *B. japonicus* ainsi que de la famille des *Vesicomomyidae* avec l'espèce *C. magnifica* ; les crustacés de la famille des *Alvinocarididae* avec l'espèce *Rimicaris exoculata* ; et les gastéropodes *Alviniconcha hessleri* et *Ifremeria nautilei* (Fig. 5). Ces derniers ont été préférentiellement présentés pour leur association symbiotique bien caractérisée avec des micro-organismes.

5.1.1 *Alvinella pompejana*

L'annélide polychète *A. pompejana* est, comme les autres *Alvinellidae*, strictement endémique des sources hydrothermales du Pacifique (de 21°N à 23°S) (Desbruyères *et al.*, 1998). C'est parce qu'il vit sous une pluie constante de cendres que les scientifiques l'ont baptisé “ver de Pompéi”.

Ce polychète est pionnier dans la colonisation des cheminées hydrothermales de la dorsale EPR (Desbruyères *et al.*, 1985). Les adultes vivent en colonies dans les tubes organiques qu'ils secrètent. Ces tubes sont progressivement incrustés dans la structure minérale des cheminées. *A. pompejana* colonise le pôle le plus chaud des sources, à même les cheminées, près de la sortie des fluides hydrothermaux. Des données précises relatives à l'environnement immédiat de ces vers ont été mesurées à plusieurs reprises. Cet environnement est caractérisé par de très fortes concentrations en sulfures (pouvant être de l'ordre du millimolaire) (Le Bris *et al.*, 2003), un pH acide (pH~4) avec des températures allant de 30°C à 120°C (Desbruyères *et al.*, 1985 ; Chevaldonné *et al.*, 1992 ; Le Bris *et al.*, 2005) et de fortes concentrations en métaux toxiques (Desbruyères *et al.*, 1998). Les conditions physico-chimiques à l'intérieur du tube sont cependant encore mal connues (Le Bris *et al.*, 2005). La minéralogie des dépôts associés aux tubes suggèrent que les conditions internes sont très différents de celles qui prévalent à l'extérieur (Zbinden *et al.*, 2003). Bien que la température du fluide circulant à l'intérieur des tubes soit difficile à mesurer (températures sans doute comprises entre 40°C et 60°C) (Di Meo-Savoie *et al.*, 2004 ; Le Bris *et al.*, 2005), *A. pompejana* supporte des températures très élevées. Ce polychète est pour cela réputé comme étant l'un des métazoaires les plus thermotolérant connu sur Terre (Chevaldonné *et al.*, 1992 ; Cary *et al.*, 1998).

A. pompejana est remarquable pour son association obligatoire avec des bactéries filamenteuses épibiotiques qui recouvrent totalement la surface de leur corps. Cette communauté microbiennes épibiotique a fait l'objet d'études métagénomiques (chapitre 2, §4.4).

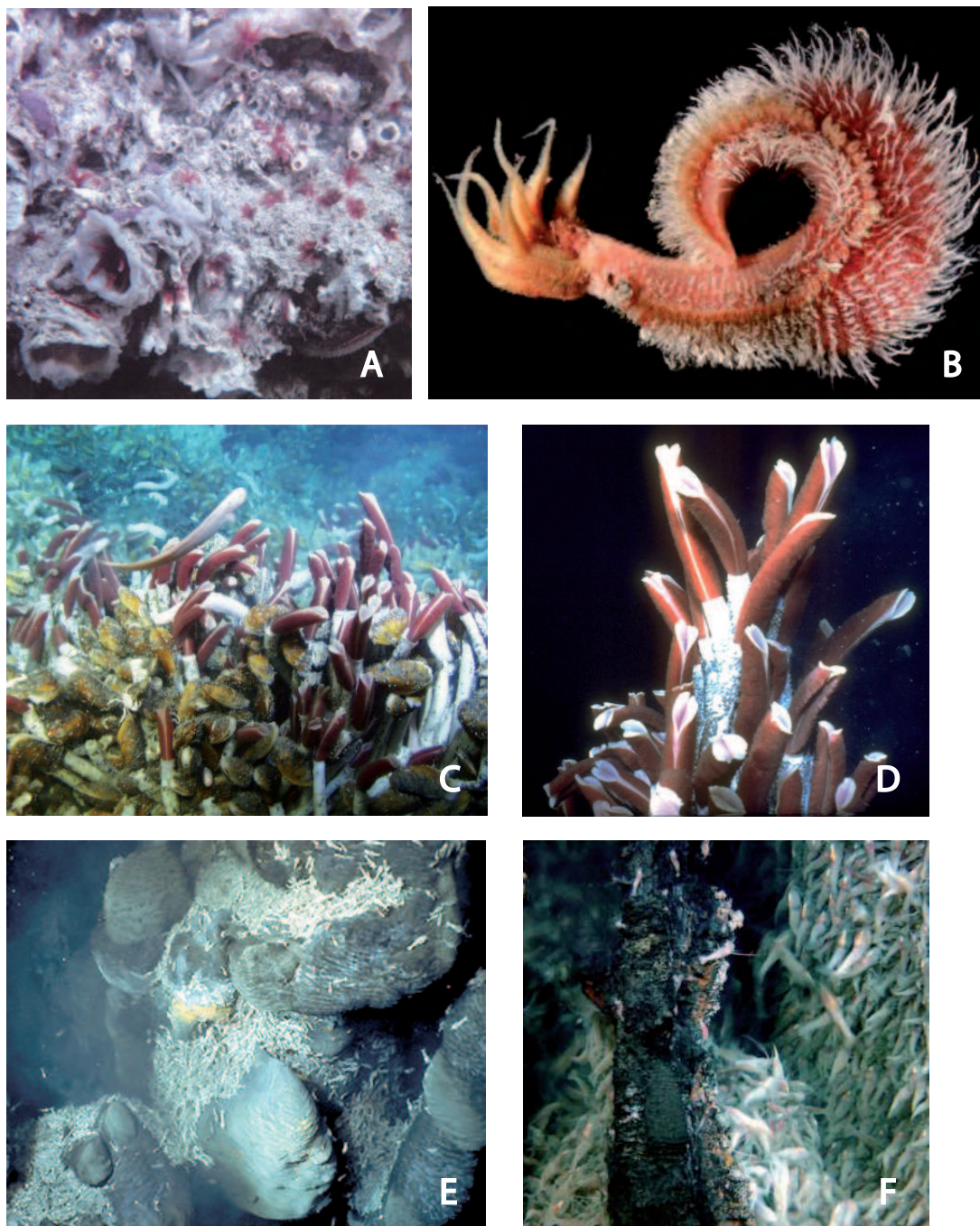


Fig. 5 – Quelques représentants de la faune hydrothermale marine profonde : A et B, *Alvinella pompejana* ; C et D, *Riftia pachytila* ; E et F, *Rimicaris exoculata*. Sources Ifremer

5.1.2 *Riftia pachyptila*

Le vestimentifère *R. pachyptila* est un long ver pouvant atteindre 2 m de long pour un diamètre de 4 à 5 cm. *R. pachyptila* se développe en véritables buissons pouvant atteindre 100 à 200 individus au m². Des colonies de *R. pachyptila* ont été observées le long de la dorsale Pacifique orientale, de celles des Galápagos et dans le bassin de Guaymas.

R. pachyptila vit dans des zones de températures variant entre 5 et 25°C. Son tube clos à la base, est ancré au sol. Ce ver est dépourvu de bouche et d'organe digestif. Pour assurer leur nutrition, *R. pachyptila* héberge dans un organe spécifique appelé trophosome, des bactéries endosymbiotiques sulfo-oxydantes (Fisher *et al.*, 1988) qui ont fait l'objet d'études métagénomiques (chapitre 2 §4.3). Ces bactéries transforment les molécules inorganiques H₂S, O₂ en sucres assimilables pour le vestimentifère. Ces molécules inorganiques semblent être véhiculées par les hémoglobines des *Riftia* (Arp & Childress, 1983 ; Fisher *et al.*, 1988 ; Flores *et al.*, 2005).

5.1.3 *Rimicaris exoculata*

R. exoculata est une crevette de 50 mm. Elle n'a pas de rostre et sa carapace est dépourvue d'épine. Elle est fréquemment rencontrée sur les sites hydrothermaux actifs de la ride médio-Atlantique, dominant la biomasse de la plupart de ces sites hydrothermaux (Desbruyeres *et al.*, 2001). Cette crevette a été récemment observée sur la ride indienne (Van Dover *et al.*, 2001). A la différence des espèces hydrothermales précédentes, cette espèce qui forme des essaims denses et très mobiles sur la surface des cheminées, représente un mode de colonisation "non fixée" (Segonzac, 1992 ; Gebruck *et al.*, 2000). Des densités considérables de cette crevette peuvent représenter jusqu'à 2500 individus au m². Cette crevette présente également une association symbiotique obligatoire avec des bactéries, notamment au niveau de la chambre branchiale. Cette épibiose implique des bactéries affiliées aux ϵ -*Proteobacteria* (Polz & Cavanaugh, 1995). Dans le tube digestif, on trouve une microflore distincte de la précédente et composée d' ϵ -*Proteobacteria*, *Defferibacteres* et *Entomoplasmales* (Zbinden & Cambon-Bonavita, 2003).

5.2 La microflore hydrothermale

En absence totale de lumière (la lumière ne pénètre approximativement que dans les 300 premiers mètres de l'océan), le réseau trophique de l'écosystème hydrothermal repose presque exclusivement sur la chimiosynthèse microbienne. La découverte de photorécepteurs chez une crevette hydrothermale (Van Dover *et al.*, 1989), ou des pigments photosynthétiques chez une bactérie isolée du panache d'un fumeur noir (Yurkov *et al.*, 1999) laissent cependant présager l'existence d'organismes pouvant convertir les radiations électromagnétiques en source d'énergie.

Par définition, la chimiosynthèse désigne la fixation de CO₂ à l'obscurité, par analogie avec le

terme photosynthèse (Jannasch & Wirsen, 1979 ; Jannasch, 1985). Ainsi, au sein de l'écosystème hydrothermal marin profond, les éléments apportés par le fluide hydrothermal (CO_2 , CO , CH_4 , H_2 , H_2S , N_2 ...) et par l'eau de mer (O_2 , SO_4) pourraient subvenir à la croissance d'organismes chimiolithoautotrophes, qui obtiendraient leur énergie de l'oxydation des composés réduits présents dans le fluide hydrothermal (Jannasch & Wirsen, 1979 ; Lonsdale, 1977 ; Jannasch, 1995 ; Karl, 1995).

5.2.1 Types d'habitats microbiens

Caractérisé par une variation spatiale et temporelle très importante, l'écosystème hydrothermal constitue un biotope particulièrement dynamique. Sur une échelle de quelques centimètres à quelques mètres autour d'une émission hydrothermale, la température peut passer de plus de 400°C à seulement 2°C . Ce gradient thermique abrupt et les forts gradients chimiques (potentiels d'oxydo-réduction, pH) qui en découlent définissent différentes niches écologiques (Fig. 6). Ces niches répertoriées dans la revue de Karl (1995) sont les suivantes :

- les communautés microbiennes en association ecto-, epi- ou endosymbiotique avec les métazoaires hydrothermaux tels que les vestimentifères, les mollusques, les crustacés ou les polychètes,
- les micro-organismes se développant sous forme de tapis ou mattes microbiennes sur les roches basaltiques, les cheminées et sédiments exposés aux fluides hydrothermaux,
- les micro-organismes inféodés au panache hydrothermal chaud. Cette microflore pourrait être issue d'une biosphère souterraine sous-jacente aux cheminées hydrothermales, ce concept de biosphère souterraine étant corroboré par de récentes études (Takai *et al.*, 2004a),
- les micro-organismes présents sous forme libre ou entraînés par le fluide hydrothermal, et provenant probablement d'une biosphère profonde.

Cette classification théorique permet de décrire l'organisation spatiale de communautés microbiennes et d'orienter les stratégies d'échantillonnage. Grâce au développement de techniques d'échantillonnage, les communautés microbiennes associées aux différents habitats proposés par Karl ont été étudiées en conjuguant les approches culturales et les approches moléculaires de la biodiversité microbienne.

5.2.2 Diversité métabolique des micro-organismes hydrothermaux

L'environnement hydrothermal, sous l'influence du mélange turbulent du fluide avec l'eau de mer, offre aux micro-organismes un vaste panel de conditions physico-chimiques. D'après nos connaissances actuelles sur les métabolismes microbiens et sur l'environnement hydrothermal, à savoir notamment les espèces chimiques présentes dans l'eau de mer ou dans le fluide hydrothermal (Annexe B), il est possible de lister les voies métaboliques susceptibles d'être utilisées par

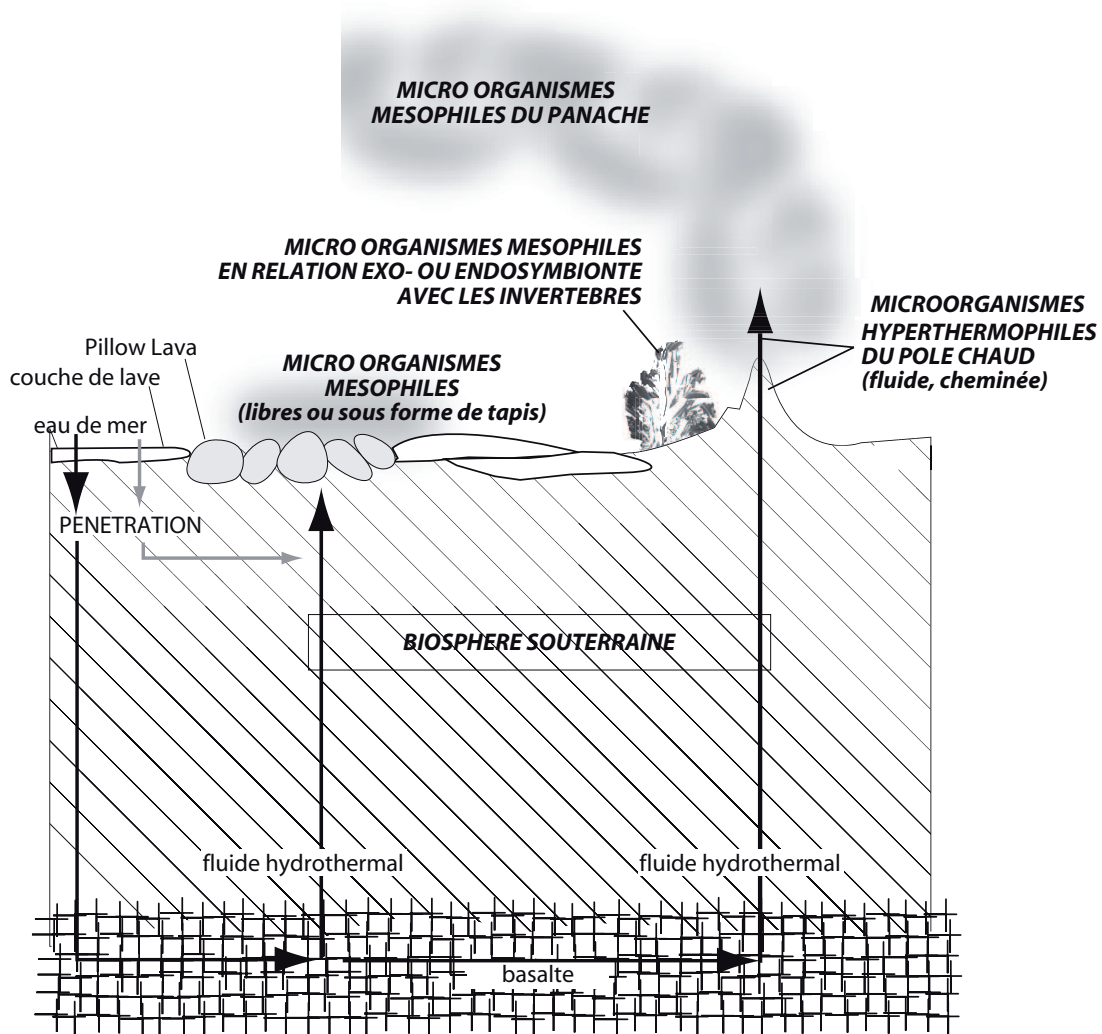


Fig. 6 – Localisation des principaux habitats microbiens des sources hydrothermales marines (d'après Karl, 1995).

les micro-organismes des sources hydrothermales (liste non-exhaustive) (Tableau 1).

La diversité métabolique et physiologique des espèces bactériennes et archaéennes qui ont été isolées d'environnements hydrothermaux profonds, est particulièrement grande. Plus d'une centaine de souches microbiennes ont été isolées et caractérisées. Les microbiologistes ont d'abord axé leurs travaux sur la recherche de micro-organismes thermophiles (température optimale de croissance $\geq 60^{\circ}\text{C}$) et hyperthermophiles (température optimale de croissance $\geq 80^{\circ}\text{C}$). Outre l'intérêt fondamental que revêt l'étude de ces micro-organismes thermophiles et hyperthermophiles, leur capacité à produire des enzymes thermostables suscite un vif intérêt de la part des industriels (Huber & Stetter, 1998 ; Fujiwara, 2002). La plupart des micro-organismes thermophiles et hyperthermophiles décrits à ce jour sont des méthanogènes appartenant à l'ordre des *Methanococcales* (Whitman *et al.*, 2001) et des hétérotrophes anaérobies réduisant le soufre. Ces derniers incluent des représentants de l'ordre bactérien des *Thermotogales* (Huber & Stetter, 1992) et de l'ordre archaéen des *Thermococcales* (Zillig & Reysenbach, 2001). Outre ces micro-organismes, de nombreuses espèces mésophiles / thermophiles / hyperthermophiles, aérobies / anaérobies / microaérophiles, hétérotrophes ou autotrophes ont également été cultivées. Ci-suit, une liste non exhaustive des principaux groupes de micro-organismes isolés des écosystèmes hydrothermaux marins profonds.

La microflore aérobie. Divers groupes de procaryotes aérobies ont été isolés. Ces derniers peuvent oxyder divers composés tels que les composés soufrés réduits, le dihydrogène, l'ammonium, le méthane et les composés en C_1 réduits (par ex. : le méthanol, les méthylamines, les halométhanes), le fer (Fe^{2+}) et le manganèse (Mn^{2+}) (Kelley *et al.*, 2002).

L'oxydation aérobie des composés soufrés a été particulièrement étudiée car elle pourrait être à la base de toute la chaîne trophique hydrothermale. La production primaire de l'écosystème hydrothermal semble en effet être essentiellement assurée par des bactéries chimio-autotrophes sulfo-oxydantes. Ces bactéries sulfo-oxydantes sont des γ - et ε -*Proteobacteria*. Ce sont des mésophiles affiliées aux genres *Thiomicrospira*, *Thiobacillus*, *Sulfurimonas* et *Sulfurovum*. La plupart de ces isolats ont été obtenus dans des échantillons de fluides ou de surfaces exposées aux fluides. Ces derniers constituent une zone de mélange entre le fluide hydrothermal (chaud, anoxique, riche en composés soufrés réduits et en CO_2) et l'eau de mer (froide et oxygénée). Dans cette zone de mélange, les micro-organismes disposent des oxydants O_2 et NO_3 ainsi que de nombreux composés soufrés comme le sulfure d'hydrogène, le soufre élémentaire, le thiosulfate, les sulfures minéraux (par ex. la pyrite) et les polythionates. Des bactéries sulfo-oxydantes, hétérotrophes et mixotrophes ont également été isolées de milieux hydrothermaux. Ce sont des γ -*Proteobacteria* mésophiles associées aux genre *Pseudomonas*, *Acinetobacter* et *Vibrio*.

Conditions	Donneur d'électrons	Accepteur d'électrons	Source de carbone	Processus métabolique
Aérobiose	H ₂	O ₂	CO ₂	Oxydation de l'hydrogène
	HS ⁻ , S ⁰ , S ₂ O ₃ ²⁻ , S ₄ O ₆ ²⁻	O ₂	CO ₂	Oxydation des composés soufrés
	Fe ²⁺	O ₂	CO ₂	Oxydation du fer
	Mn ²⁺	O ₂	[CH ₂ O]	Oxydation du manganèse
	NH ₃ , NO ₂ ⁻	O ₂	CO ₂	Nitrification
	CH ₄ , CO (et autres composés en C1) [CH ₂ O]	O ₂	CO ₂ [CH ₂ O]	Méthanotrophie et méthylotrophie Hétérotrophie aérobie
Anaérobiose	H ₂	NO ₃ ⁻	CO ₂	Dénitrification
	H ₂	S ⁰	CO ₂	Sulfo-réduction
	H ₂	SO ₄ ²⁻	CO ₂	Sulfato-réduction
	H ₂	CO ₂	CO ₂	Méthanogénèse
	CH ₄	SO ₄ ²⁻	CO ₂ ou CH ₄	Oxydation anaérobie du méthane ⁽¹⁾
	NH ₄ ⁺	NO ₂ ⁻ , (NO ₃ ⁻ , Fe ³⁺)	?	Oxydation anaérobie de l'ammoniac ⁽¹⁾
	S ²⁻ , S ⁰ , S ₂ O ₃ ²⁻	NO ₃ ⁻	CO ₂	Dénitrification, sulfo-oxydation
	[CH ₂ O]	NO ₃ ⁻	[CH ₂ O]	Dénitrification
	[CH ₂ O]	S ⁰	[CH ₂ O]	Hétérotrophie, réduction du soufre
	[CH ₂ O]	SO ₄ ²⁻	[CH ₂ O]	Hétérotrophie, sulfato-réduction
[CH ₂ O]	[CH ₂ O]	[CH ₂ O]	Fermentation	

(1) A ce jour, il n'y a pas d'isolats décrits capables d'oxydation aérobie ou anaérobie de l'ammoniac, et d'oxydation anaérobie du méthane.

Tableau 1 – Types de métabolismes microbiens pouvant être observés au niveau des sources hydrothermales océaniques profondes (d'après Karl, 1995 ; Alain, 2003).

L'oxydation aérobie de l'hydrogène n'a été que récemment mise en évidence dans le milieu hydrothermal marin profond. Cette voie métabolique est la réaction énergétique clé des *Aquificales* appartenant au genre *Persephonella*, mais certaines ϵ -*Proteobacteria* procèdent également à l'oxydation aérobie de l'hydrogène. Ces dernières sont associées aux genres *Hydrogenimonas*, *Caminibacter*, *Nitratiruptor* et *Nitratifractor*. Ces bactéries autotrophes sont des mésophiles ou thermophiles. Elles sont jusqu'à présent endémiques aux sources hydrothermales marines. Outre l'oxygène, elles peuvent également réduire les composés soufrés ou le nitrate. Au sein des *Archaea*, ce métabolisme a été retrouvé uniquement chez *Pyrolobus fumarii* qui est une espèce chimiolithoautotrophe stricte et aérobie facultative, dont l'optimum de température est de 106°C.

De nombreux micro-organismes chimioorganohétérotrophes aérobie ont également été isolés d'environnements hydrothermaux. Il s'agit de bactéries mésophiles et thermophiles. Les méso-

philes appartiennent entre autres aux genres *Halomonas* et *Alteromonas* ; et les thermophiles appartiennent aux genres *Thermus*, *Bacillus*, *Marinithermus*, *Oceanithermus* et *Vulcanithermus*. Ces bactéries utilisent une grande variété de composés organiques comme donneurs d'électrons. *Aeropyrum camini* est à ce jour la seule *Archaea* hyperthermophile aérobie stricte isolée de l'écosystème hydrothermal marin.

La microflore anaérobie. Les micro-organismes anaérobies isolés de sources hydrothermales marines incluent à la fois des bactéries et des archaea, dont certaines peuvent pousser à des températures supérieures à 90°C. Les groupes métaboliques les plus représentatifs sont les réducteurs de sulfates et soufre, les méthanogènes, et les hétérotrophes fermentaires.

La réduction des composés soufrés est probablement le mode privilégié de respiration anaérobie dans le pôle chaud de l'écosystème hydrothermal. La sulfato-réduction est conduite par des bactéries mésophiles et thermophiles et par des archaea hyperthermophiles. Certaines d'entre elles peuvent également réduire le thiosulfate, les sulfites et le soufre élémentaire. Ce type de métabolisme peut être conduit en condition d'autotrophie ou d'hétérotrophie. Les sulfato-réducteurs autotrophes incluent les archaea hyperthermophiles affiliées au genre *Archaeoglobus* et des bactéries thermophiles comme *Thermodesulfobacterium hydrogenophilum* et *Thermodesulfatator indicus* (Annexe C). Les sulfato-réducteurs hétérotrophes incluent, quant à eux, divers organismes mésophiles comme par exemple *Desulfovibrio hydrothermalis* et des organismes thermophiles comme *Desulfonauticus submarinus*. Les bactéries sulfato-réductrices hétérotrophes utilisent couramment l'acétate comme source de carbone, mais peuvent également utiliser d'autres acides organiques.

La respiration lithotrophique du soufre a été mise en évidence chez une crenarchaeote hyperthermophile du genre *Ignicoccus* et chez les bactéries thermophiles appartenant aux genres *Desulfurobacterium* et *Balnearium*. Jusqu'à ce jour, ces bactéries ont été trouvées uniquement dans les environnements hydrothermaux marins profonds. Il semble que ces bactéries jouent un rôle important en tant que producteurs primaires de matière organique dans les zones anaérobies du pôle hydrothermal chaud. Des ϵ -*Proteobacteria* respirant le soufre ont également été mises en évidence, il s'agit des espèces *Nautilia litotrophica*, *Caminibacter hydrogenophilus*, *Lebetimonas acidiphila* et *Thioreductor micantisoli*.

La culture d' ϵ -*Proteobacteria* issues d'échantillons hydrothermaux est très récente. Leur présence dans l'écosystème hydrothermal semblait évidente depuis de nombreuses années. Des inventaires moléculaires réalisés sur des échantillons hydrothermaux avaient en effet montré que de nombreux phylotypes affiliés aux ϵ -*Proteobacteria* étaient fréquemment et abondamment identifiés dans cet écosystème (cf §5.2.3). Cependant, ces phylotypes n'avaient jusqu'à ces dernières

années pas de représentants cultivés. Très récemment, plusieurs *ε-Proteobacteria* ont pu finalement être isolées de collecteurs in situ, de fragments de cheminées et de spécimens d'*A. pompejana* (Takai *et al.*, 2003 ; Miroshnichenko *et al.*, 2004 ; Alain *et al.*, 2002b ; Miroshnichenko *et al.*, 2002 ; Takai *et al.*, 2005b, 2004b ; Nakagawa *et al.*, 2005d ; Inagaki *et al.*, 2004, 2003 ; Campbell *et al.*, 2001 ; Nakagawa *et al.*, 2005a ; Voordeckers *et al.*, 2005). Ces dernières sont toutes des mésophiles ou thermophiles modérées chimiolithoautotrophes, anaérobies ou microaérophiles ayant la capacité d'utiliser le nitrate et/ ou des composés soufrés comme accepteurs d'électrons. Les *ε-Proteobacteria* semblent particulièrement adaptées aux conditions fluctuantes des zones de mélange entre le fluide hydrothermal et l'eau de mer. Leur abondance, leur diversité génétique et physiologique en font des producteurs primaires potentiels de premier ordre dans les zones anaérobies ou dans les zones de mélange entre le fluide hydrothermal et l'eau de mer. Ces bactéries participent au cycle du soufre en réduisant ou en oxydant le soufre élémentaire ou les composés soufrés.

Comme la sulfato-réduction, la méthanogénèse est assurée par des micro-organismes mésophiles à hyperthermophiles. Toutes les méthanogènes connues sont des archaea anaérobies strictes qui réduisent le CO₂ en méthane. Cependant, à la place du CO₂, elles peuvent éventuellement utiliser l'acétate ou le formate. La majorité des archaea méthanogènes isolées des sources hydrothermales marines appartient à l'ordre des *Methanococcales* qui comprend deux familles : les *Methanocaldococaceae* et les *Methanococaceae*. Les *Methanococcales* isolées d'environnements hydrothermaux appartiennent aux genres *Methanotorris*, *Methanocaldococcus* et *Methanothermococcus*. Ces méthanogènes sont toutes des thermophiles ou hyperthermophiles. Outre les *Methanococcales*, l'espèce *Methanopyrus kandleri* qui est l'unique représentant de l'ordre des *Methanopyrales* peut également assurer la méthanogénèse dans l'écosystème hydrothermal profond.

Les micro-organismes chimioorganohétérotrophes anaérobies isolés d'environnements hydrothermaux marins, sont pour la plupart des fermentaires réduisant le soufre. Il s'agit des espèces appartenant aux genres *Thermococcus*, *Pyrococcus* et *Palaeococcus* (ordre des *Thermococcales*, domaine des *Euryarchaeota*). Parmi les hyperthermophiles, c'est l'ordre dans lequel le plus d'espèces ont été à ce jour caractérisées. Ces micro-organismes sont considérés comme étant les principaux décomposeurs de matière organique de l'écosystème hydrothermal. Ils sont capables d'utiliser divers composés organiques complexes pour leur croissance : des substrats protéiques tels que l'extrait de levure et la peptone ainsi que des polysaccharides. Ces composés sont ensuite métabolisés par fermentation. La croissance des *Thermococcales* est stimulée par la présence de soufre élémentaire utilisé comme accepteur terminal d'électrons (S⁰ réduit en H₂S). Les archaea du genre *Staphylothermus*, *Pyrodictium* et *Desulfurococcus* (domaine des *Crenarchaeota*) sont également des hétérotrophes fermentaires soufre-réducteurs fréquemment rencontrés dans cet écosystème.

Les bactéries organohétérotrophes les plus thermophiles des sources hydrothermales appartiennent aux genres *Thermosipho*, *Thermotoga* et *Marinitoga* (ordre des *Thermotogales*).

La réduction du fer ferrique, qui est un composé abondant dans l'écosystème hydrothermal profond, a été montrée en présence d'hydrogène chez l'archaea hyperthermophile *Geoglobus ahangarii* et la bactérie thermophile *Deferribacter abyssi*.

Récemment, une archaea coccoïde de très petite taille (400 nm de diamètre) a été cultivée en coculture avec une archaea hyperthermophile chimiolithoautotrophe du genre *Ignicoccus*. Cette archaea "*Nanoarchaeum equitans*" possède le plus petit génome archaea connu (0.5 Mpb). Une étude récente a d'ailleurs permis de détecter ce type d'organisme dans des sources hydrothermales marines profondes (Hohn *et al.*, 2002).

5.2.3 Contribution des méthodes moléculaires à la connaissance de la diversité des communautés microbiennes des sources hydrothermales marines profondes

L'utilisation de méthodes indépendantes des techniques de culture pour l'étude des microorganismes de l'écosystème hydrothermal profond date du milieu des années 90. Ainsi, les communautés microbiennes ayant colonisé des collecteurs *in situ* (Reysenbach *et al.*, 2000 ; Corre *et al.*, 2001 ; Nercessian *et al.*, 2003 ; Alain *et al.*, 2004), la microflore des tapis microbiens (Moyer *et al.*, 1994, 1995, 1998 ; Longnecker & Reysenbach, 2001), des cheminées hydrothermales (Takai & Horikoshi, 1999 ; Takai *et al.*, 2001a ; Schrenk *et al.*, 2003), des sédiments hydrothermaux (Teske *et al.*, 2002), la microflore associée aux métazoaires hydrothermaux (Haddad *et al.*, 1995 ; Polz & Cavanaugh, 1995 ; Alain *et al.*, 2002a ; Lopez-Garcia *et al.*, 2002) ou celle associée au fluide hydrothermal (Huber *et al.*, 2002, 2003) ont été étudiées par le biais d'inventaires des gènes codant pour les ARNr 16S (Jeanthon, 2000 ; pour revue). Ces études ont pour la plupart révélé une extraordinaire diversité d'organismes et ont permis la découverte de nouvelles lignées phylogénétiques pour lesquelles aucun représentant cultivé n'est connu à ce jour.

Les études moléculaires portant sur des échantillons du pôle tiède de l'écosystème hydrothermal ont révélé que les ε -*Proteobacteria* représentaient un groupe extrêmement divers et abondant. Des séquences d'ARNr 16S affiliées aux ε -*Proteobacteria* ont été majoritairement retrouvées dans des banques de clones réalisées à partir d'échantillons de tapis microbiens (Moyer *et al.*, 1995 ; Longnecker & Reysenbach, 2001), dans des colonisateurs *in situ* (Reysenbach *et al.*, 2000 ; Corre *et al.*, 2001 ; Lopez-Garcia *et al.*, 2003a), dans la microflore épibiotique des métazoaires hydrothermaux (Haddad *et al.*, 1995 ; Campbell & Cary, 2001 ; Alain *et al.*, 2002a ; Lopez-Garcia *et al.*, 2002) et dans les émissions de fluide hydrothermal diffus (Huber *et al.*, 2003).

Outre les séquences affiliées aux ε -*Proteobacteria*, des séquences affiliées aux γ - α - δ - *Proteobacteria*, aux *Verrumicrobia*, aux *Aquificales*, au groupe des *Cytophaga-Flavobacterium-Bacteroides* (CFB), au genre *Desulfurobacterium* et aux divisions candidates OP8, OP11, WS6 et ABY ont été recensées dans les inventaires moléculaires du pôle tiède de l'écosystème hydrothermal. Un inventaire moléculaire portant sur un échantillon de tapis microbien se développant à proximité de sources hydrothermales marines du volcan sous-marin Loihi a révélé la présence d'archaea planctoniques affiliées au marine *Crenarchaeota* groupe I et marine *Euryarchaeota* groupe II (Moyer *et al.*, 1998).

Le premier inventaire moléculaire portant sur le pôle chaud de l'écosystème hydrothermal date de 1999 (Takai & Horikoshi, 1999). Cet inventaire réalisé à partir d'échantillons de fluides hydrothermaux, de fragments de fumeurs noirs, de diffuseurs et de sédiments provenant de sites actifs, a révélé une diversité phylogénétique très importante. De nombreuses séquences détectées correspondaient à de nouvelles lignées phylogénétiques ne possédant aucun représentant cultivé comme les DHVE (pour Deep-sea Hydrothermal Vent Euryarchaeota) (Fig. 7). Dans l'ensemble, les analyses phylogénétiques portant sur les communautés archéennes présentes au niveau des cheminées hydrothermales ont révélé la présence d'une très importante diversité d'*Archaea* appartenant aux *Thermococcales*, aux *Methanococcales*, aux *Methanopyrales*, aux *Archaeaoglobales*, aux *Desulfurococcales* et aux *Igniococcales* (Takai & Horikoshi, 1999 ; Reysenbach *et al.*, 2000 ; Takai *et al.*, 2001a ; Huber *et al.*, 2002 ; Nercessian *et al.*, 2003, 2004 ; Schrenk *et al.*, 2003). L'étude de Schrenk et ses collaborateurs a de plus montré que la proportion de bactéries était moins importante à l'intérieur des cheminées qu'à l'extérieur (Schrenk *et al.*, 2003).

Les inventaires moléculaires portants sur des échantillons hydrothermaux profonds ont donc mis en évidence une très grande diversité au sein des *Bacteria* et des *Archaea*, répartie dans quasiment tous les phyla identifiés à ce jour. Beaucoup de séquences détectées correspondent à de nouvelles lignées ne possédant aucun représentant cultivé, certaines d'entre elles semblant être inféodées à cet écosystème profond.

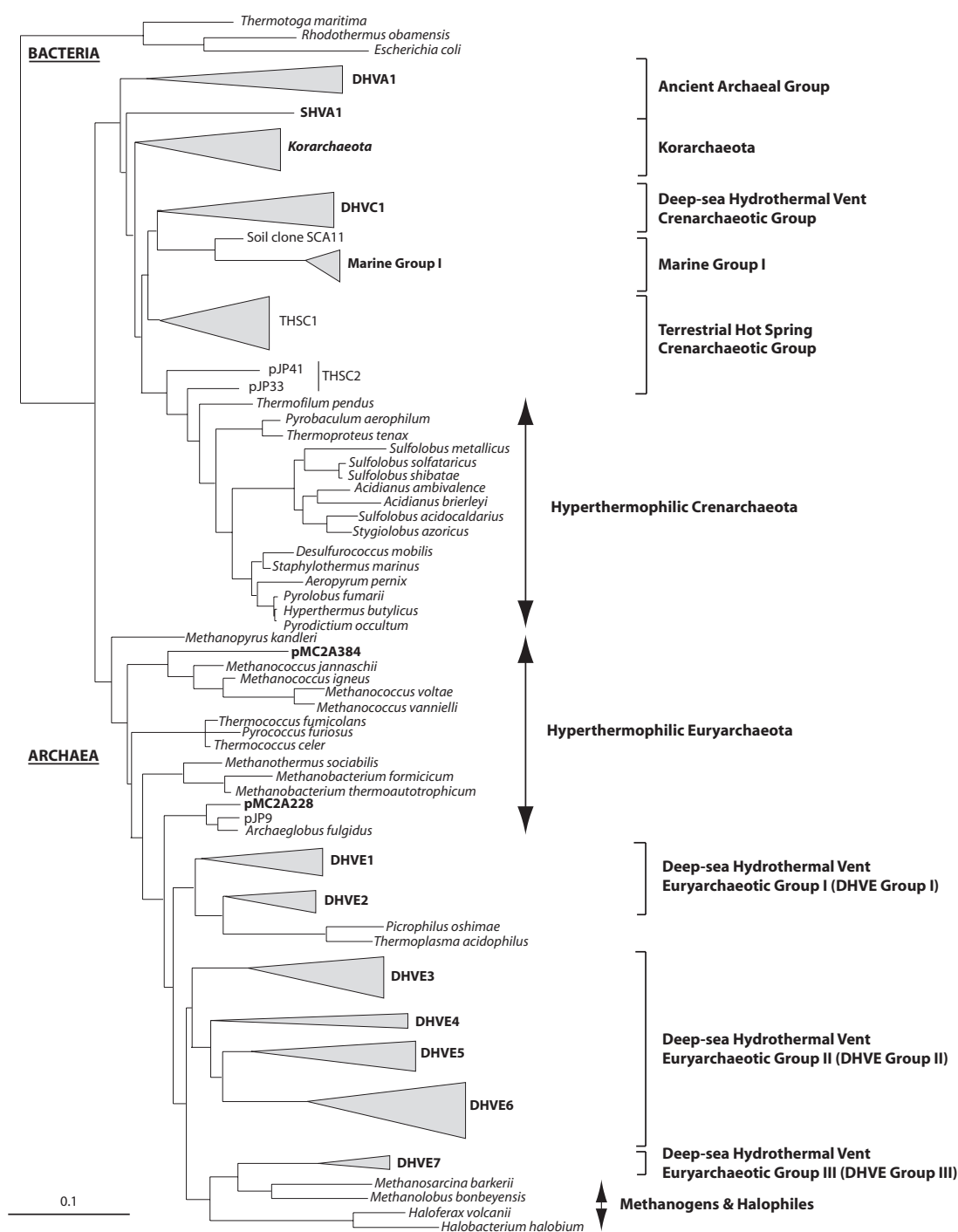


Fig. 7 – Arbre phylogénétique illustrant les nouvelles lignées d'*Archaea* (en gras) mises en évidence dans les sources hydrothermales marines profondes par Takai et Horikoshi (1999) (voir cette référence pour le détail des méthodes de phylogénétique utilisées). Les triangles suivis d'un nom en caractères gras indiquent les nouvelles lignées identifiées par les auteurs. DHVE, Deep-sea Hydrothermal *Euryarchaeota*. L'échelle indique 0.1 substitution par position (d'après Takai & Horikoshi, 1999).

CHAPITRE 4

OBJECTIFS DE LA THÈSE

L'une des activités du laboratoire dans lequel j'ai effectué ma thèse est centrée sur l'étude de la diversité microbienne de l'écosystème hydrothermal marin profond. La première phase de cette activité a tout d'abord consisté à faire appel à des techniques culturales pour étudier la diversité microbienne de cet écosystème. Cette approche a permis d'isoler des centaines de souches appartenant au domaine des *Archaea* et des *Bacteria*. Les premières souches isolées au laboratoire furent essentiellement des archaea hyperthermophiles anaérobies stricts appartenant à l'ordre des *Thermococcales* (Prieur *et al.*, 1995). Des bactéries aérobies appartenant aux genres *Bacillus* (Marteinsson *et al.*, 1996) et *Thermus* (Marteinsson *et al.*, 1995) furent également isolées. Par la suite, les travaux ont été orientés vers la culture des micro-organismes autotrophes. Ces travaux ont abouti à l'isolement d'archaea méthanogènes hyperthermophiles appartenant à l'ordre des *Methanococcales* (Jeanthon *et al.*, 1998, 1999), ainsi qu'à l'isolement de bactéries sulfo- et sulfato-réductrices telles que les souches *Desulfurobacterium thermolithotrophum* (L'Haridon *et al.*, 1998) et *Thermodesulfobacterium hydrogenophilum* (Jeanthon *et al.*, 2002).

Afin de compléter les informations issues de l'approche culturale, des techniques moléculaires d'écologie microbienne ont été dans un second temps développées au laboratoire. L'approche moléculaire a tout d'abord consisté à identifier phylogénétiquement les micro-organismes de l'écosystème hydrothermal. Pour cela, des inventaires d'ARNr 16S ont été générés par PCR et/ou RT-PCR, puis ont été analysés. La mise au point de sondes oligonucléotidiques ciblant l'ARNr 16S des micro-organismes fréquemment rencontrés dans les systèmes hydrothermaux a ensuite permis de les quantifier par Fluorescence In Situ Hybridization (FISH). Des échantillons hydrothermaux provenant des différentes niches de l'écosystème hydrothermal ont été ainsi étudiés au sein du laboratoire, faisant l'objet de plusieurs articles et thèses (Harmsen *et al.*, 1997 ; Corre, 2000 ; Corre *et al.*, 2001 ; Nercessian, 2003 ; Nercessian *et al.*, 2003, 2004, 2005b). Les résultats issus de ces études ont notamment contribué à révéler l'étendue de la diversité des micro-organismes incultivés. Des séquences d'ARNr 16S de bactéries et archaea dont la position phylogénétique est éloignée de celles d'organismes isolés ont été identifiées. Ce fut par exemple le cas, au sein des *Bacteria*, de la mise en évidence de nombreux phylotypes associés à des ϵ -*Proteobacteria* incultivées (Corre, 2000 ; Corre *et al.*, 2001). Au sein des *Archaea*, des séquences affiliées aux DHVE, aux *Crenarchaeota* marines du groupe I et aux *Korarchaeota* ont pu également être détectées (Nercessian, 2003 ; Nercessian *et al.*, 2003, 2004, 2005b).

La combinaison de l'approche culturale et moléculaire pour l'étude de la diversité microbienne de l'écosystème hydrothermal, a permis d'une part, de révéler l'étendue de la diversité physiologique, métabolique et génétique des micro-organismes hydrothermaux et, d'autre part, de mettre en évidence de nouvelles lignées d'organismes incultivés dont nous ignorions le rôle et l'importance écologique dans l'écosystème hydrothermal. Des marqueurs fonctionnels impliqués dans les processus de méthanogénèse, de méthanotrophie et de sulfato-réduction ont été par la

suite utilisés, afin d'étudier la diversité fonctionnelle des micro-organismes hydrothermaux (Nercessian *et al.*, 2005a). Plusieurs séquences obtenues lors de cette étude n'ont pu être affiliées à des séquences de gènes appartenant à des organismes cultivés, montrant l'existence de nouvelles lignées au sein des trois groupes fonctionnels étudiés.

Ma thèse s'est inscrite dans la continuité de ces travaux. Au commencement de ma thèse, de nombreux micro-organismes identifiés dans les inventaires moléculaires, n'avaient toujours pas de représentants cultivés. Nous ignorions donc leur fonction et leur importance écologique dans l'écosystème hydrothermal. La suite logique était de rechercher des informations supplémentaires sur ces micro-organismes incultivés de l'écosystème hydrothermal, et ce par des méthodes s'affranchissant de leur cultivabilité.

La microflore incultivée présente dans la zone de mélange entre le fluide hydrothermal et l'eau de mer a été plus particulièrement ciblée pour ce travail de thèse. Pour cela, deux types d'échantillons prélevés dans cette zone de mélange ont été étudiés. Le premier échantillon consistait en un biofilm microbien collecté lors d'une expérience de colonisation, à la surface d'un colonisateur *in situ*. Le deuxième échantillon consistait quant à lui, en des tubes d'*Alvinellidae* prélevés sur la paroi d'une cheminée hydrothermale active. Les deux échantillonnages ont été effectués lors de la campagne océanographique PHARE qui s'est déroulée au printemps 2002, sur un site hydrothermal de la dorsale du Pacifique oriental (13°N EPR).

J'ai commencé ma thèse par l'étude de la communauté microbienne du biofilm. Pour cela, j'ai utilisé les outils moléculaires déjà mis en place avant mon arrivée au laboratoire. Des banques d'ADNr 16S générées par PCR ont été construites afin d'inventorier les micro-organismes du biofilm, et leur abondance a été déterminée par FISH. Pour compléter les informations issues de cette approche moléculaire, les micro-organismes ont été observés par microscopie électronique à balayage (MEB) et la composition chimique de l'échantillon été déterminée par spectroscopie dispersive en énergie (EDS) et par diffraction aux rayons X.

Par la suite, j'ai choisi d'utiliser une seconde approche afin d'étudier la microflore incultivée associée aux tubes d'*Alvinellidae*. Pour cela, j'ai mis au point au laboratoire l'utilisation d'une nouvelle technique d'écologie microbienne, la métagénomique. Cette approche indépendante des méthodes de culture, permet non seulement de s'affranchir des biais inhérents à la PCR, mais également d'accéder aux génomes des micro-organismes incultivés. L'accès à ces génomes laisse présager l'identification de gènes fonctionnels pouvant être placés sur le génome à proximité d'un marqueur phylogénétique, permettant ainsi de faire un lien entre l'identité moléculaire, la fonction et l'activité métabolique d'un organisme incultivé.

Au moyen de ces méthodes d'études de communautés microbiennes, les objectifs de ma thèse étaient d'apporter des éléments de réponse aux questions suivantes :

- Quelle est la composition de communautés microbiennes de divers substratums exposés aux fluides hydrothermaux ?
- Quels sont en termes d'abondance, les principaux organismes incultivés de l'écosystème hydrothermal ?
- Dans ces communautés microbiennes, quelles sont les potentialités métaboliques, les caractéristiques physiologiques et génétiques des micro-organismes incultivés ?
- Quels rôles jouent ces micro-organismes incultivés dans les cycles biogéochimiques de l'écosystème hydrothermal ?

Ces différents points sont abordés dans mes travaux de thèse. Après avoir détaillé le matériel et les méthodes utilisés pour l'approche méthodologique de la métagénomique (deuxième partie du manuscrit de thèse), ces différents points sont présentés et discutés dans la troisième partie de ce manuscrit, sous forme d'articles parus ou sous presse. Enfin les résultats les plus importants issus de ces travaux de thèse, sont commentés et mis en perspective dans une dernière partie intitulée "Discussion et Perspectives".

Matériels et Méthodes

Construction et criblage de la banque métagénomique

1	Quantification et estimation de la taille de l'ADN génomique .	102
2	Réparation de l'extrémité des inserts d'ADN	103
3	Sélection de la taille des inserts d'ADN réparés	104
4	Elution des fragments d'ADN génomique de 40 kpb contenus dans l'agarose	105
5	Réaction de ligation	106
6	Réaction d'empaquetage des clones dans des particules de phage λ	106
7	Calcul du titre des particules phagiques	107
8	Étalement et sélection de la banque métagénomique	108
9	Conservation de la banque	108
10	Criblage de la banque	108
11	Induction du nombre de copies de fosmides dans les clones . . .	109

Dans le cadre de ma thèse, j'ai construit une banque métagénomique à partir d'ADN extrait de la microflore associée à des tubes d'*A. pompejana*. Cette banque, qui est présentée dans les études 2 et 3 de ce manuscrit, a été réalisée en utilisant le kit "CopyControlTM Fosmid Library Production Kit" (Epicentre). Avant celle-ci, la construction d'une banque métagénomique n'avait jamais été tentée dans le laboratoire où j'ai effectué ma thèse.

En plus des techniques liées à la construction et à l'analyse de cette banque métagénomique, d'autres techniques de biologie moléculaire ont été utilisées, telles que celles permettant d'amplifier les gènes codant les ARNr 16S, de construire des banques de clones d'ADNr 16S, d'extraire de l'ADN plasmidique et génomique, d'analyser des clones par RFLP (Restriction Fragment Length Polymorphism), de séquencer des fragments d'ADN et de faire de l'hybridation in situ. Ces techniques sont couramment utilisées au laboratoire. Elles sont d'autre part décrites, bien que sommairement, dans les articles inclus dans ce manuscrit de thèse. Pour ces raisons, j'ai choisi de ne parler dans ce chapitre que du matériel et des méthodes relatifs à la construction et à l'analyse de la banque métagénomique. Le détail fourni dans ce chapitre sera probablement une ressource intéressante pour de futurs étudiants.

Préparation de l'ADN source

Lors de la mission océanographique PHARE (2002, dorsale du pacifique oriental, 13°N EPR), des spécimens d'*Alvinella pompejana* associés à leur tube ont été récoltés à une profondeur de 2624 m. Ces polychètes ont été prélevés sur les parois d'une cheminée hydrothermale active du site Elsa (12°48.180 N, 103°56.315 W) à l'aide du robot téléopéré Victor. Ils ont été ensuite transférés dans une boîte étanche, afin de minimiser leur contamination lors de la remontée à la surface.

A bord, les polychètes ont été extraits de leurs tubes. Ces tubes ont été déposés dans un flacon contenant de l'eau de mer stérile. Cette préparation a été agitée vigoureusement au vortex afin de détacher les cellules procaryotes se trouvant à la surface des tubes. Ces derniers ont été retirés de la préparation et les cellules procaryotes ont été culotées par centrifugation pendant 15 min à 12000 g. Pour extraire l'ADN génomique de ces cellules, on a alors utilisé le kit UltraClean DNA (MoBio), selon les indications du manufacturier.

Afin de purifier au maximum cet ADN génomique, une extraction supplémentaire de l'ADN a ensuite été réalisée en utilisant des solvants (phénol, chloroforme et acide isoamylique), comme décrit ci-dessous :

- ajouter un volume de PCI (phénol /chloroforme / acide isoamylique : 25 /24 /1),
- agiter par retournement pendant ~45 sec (une émulsion jaune laiteuse doit se former),
- centrifuger à 20000 g pendant 15 min à 4°C,

- récupérer la phase aqueuse supérieure contenant les acides nucléiques, avec un cône bleu (dont on aura coupé le bout au ciseau de manière à en élargir l'ouverture pour ne pas casser l'ADN).

Cette opération a été effectuée une deuxième fois, puis fut suivie d'un traitement au chloroforme. Ce traitement permet d'éliminer les traces de phénol pouvant inhiber les futures réactions enzymatiques. Il a consisté à :

- ajouter un volume du mélange chloroforme /alcool isoamylique (24 /1),
- agiter vigoureusement à la main,
- centrifuger à 20000 g pendant 10 min à 4°C,
- récupérer la phase aqueuse supérieure qui contient les acides nucléiques, comme décrit précédemment.

L'ADN contenu dans la phase aqueuse a été ensuite précipité à température ambiante avec de l'isopropanol, en procédant de la manière suivante :

- ajouter 0.1 volume d'acétate de sodium (3 M, pH 5.2) et 0.7 volume d'isopropanol 100%, puis agiter doucement par retournement (des filaments translucides d'ADN apparaissent dans la solution),
- laisser reposer environ 5 min à température ambiante,
- pêcher l'ADN en enroulant délicatement les filaments sur une pipette Pasteur coudée.

L'ADN a alors été lavé en trempant l'extrémité de la pipette Pasteur dans un microtube contenant 500 μL d'éthanol 75%. Après avoir laissé sécher à l'air l'ADN enroulé sur la pipette pendant environ 5 min, l'extrémité de la pipette Pasteur portant l'ADN a été cassée puis placée dans un autre microtube contenant 150 μL de tampon TE 1X (10 mM Tris/HCl, 2 mM EDTA, pH 8). L'ADN se réhydrate ainsi dans le tampon pendant toute une nuit à 4°C.

L'extraction de l'ADN génomique est une étape cruciale pour la réussite de la construction d'une banque métagénomique. La difficulté de cette étape tient dans le fait que l'on doit obtenir de l'ADN de haut poids moléculaire (>40 kpb), en quantité suffisante (un minimum de 20 μg à une concentration de 0.5 $\mu\text{g}\cdot\mu\text{L}^{-1}$ est requis) et avec un excellent degré de pureté. Afin d'éviter de la dégrader, l'ADN doit être manipulé (congélation/ décongélation, pipetage) un minimum de fois et ne doit surtout pas être exposé aux rayons UV. En effet, même une brève exposition aux rayons UV peut endommager l'ADN et réduire l'efficacité de clonage de 100 à 1000 fois.

1 Quantification et estimation de la taille de l'ADN génomique

La quantification de l'ADN génomique extrait a été effectuée par spectrophotométrie ultraviolette. Sachant qu'une unité de densité optique mesurée à 260 nm correspond à une solution

d'ADN double brin à $50 \mu\text{g.mL}^{-1}$

Pour construire la banque métagénomique présentée dans ce travail de thèse, nous avons choisi d'utiliser un vecteur de clonage de type fosmide. Ce type de vecteur permet de cloner des fragments d'ADN dont la taille est comprise entre 35 et 45kpb (cf. introduction, chapitre 2, §3.3.1). Nous nous sommes donc assurés que l'ADN génomique que nous avons extrait avait un poids moléculaire ≥ 40 kpb.

L'intégrité et l'estimation de la taille de l'ADN génomique ont été vérifiées par électrophorèse en gel d'agarose. Un gel d'agarose LMP (Low Melting Point, Eurobio) à 1% (poids / vol.) dans du tampon TAE 1X (0.04 M Tris-acétate, 0.001 M EDTA) a été coulé dans un moule rectangulaire de 20 cm de long. Un échantillon de l'ADN génomique et le marqueur T7 (marqueur de taille d'ADN migrant à 40 kpb, commercialisé par Epicentre) ont été chargés en parallèle dans ce gel toute la nuit, sous une tension de 30 V. Une fois la migration terminée, le gel a été coloré pendant 15 min dans un bain de bromure d'éthidium (BET, agent intercalant de l'ADN) à 0.5 mg.L^{-1} , puis rincé pendant 20 min dans de l'eau distillée. L'échantillon d'ADN génomique a été visualisé en exposant le gel aux rayons UV.

L'ADN génomique ayant migré à la même vitesse que le marqueur de taille T7, nous avons procédé directement à l'étape suivante qui consiste à réparer les extrémités de l'ADN génomique. Si l'ADN génomique avait migré plus lentement que le marqueur de taille T7 (dans ce cas l'ADN génomique aurait eu un poids moléculaire > 40 kpb), une étape supplémentaire de fragmentation mécanique de l'ADN aurait été nécessaire. Cette fragmentation mécanique aurait consisté à pipeter plusieurs fois l'ADN au travers d'un cône de $200 \mu\text{L}$. Les étapes qui suivent ont été réalisées en utilisant le matériel du kit "CopyControlTM Fosmid library production" (Epicentre) en suivant les indications du fabricant.

2 Réparation de l'extrémité des inserts d'ADN

Cette étape a permis de générer des extrémités d'ADN à bout franc, phosphorylées en 5 min. Cette réaction a été réalisée à température ambiante, en incubant pendant 45 min les réactifs suivants :

- tampon de réparation 10X (Epicentre)⁷ : $8 \mu\text{L}$
- mélange de dNTPs (dATP, dCTP, dGTP et dTTP, 2.5mM pour chaque) : $8 \mu\text{L}$
- ATP 10 mM : $8 \mu\text{L}$
- $20 \mu\text{g}$ d'ADN génomique extrait

⁷Le tampon de réparation 10 X contient : 330 mM de Tris-acétate [pH 7.8], 660 mM d'acétate de potassium, 100 mM d'acétate de magnésium, 5 mM de DTT.

- mélange d’enzymes de réparation des extrémités incluant les enzymes T4 ADN polymérase et T4 polynucléotide kinase (Epicentre) : 4 μL
- eau milliQ : quantité pour compléter le volume à 80 μL

Ajouter du tampon de charge puis incubé pendant 10 min à 70°C pour arrêter la réaction de réparation.

3 Sélection de la taille des inserts d’ADN réparés

Les fragments d’ADN génomique dont les extrémités ont été réparées, ont ensuite été séparés par électrophorèse en champ pulsé (Pulsed Field Gel Electrophoresis, PFGE à l’aide d’un appareil CHEF DRIII (Bio-Rad, Richmond, California, USA). Pour cela, un gel d’agarose LMP (Eurobio) à 1% préparé dans du tampon TAE 1 X a été confectionné. Plusieurs dents du peigne servant à former les puits dans le gel ont été scotchés afin de ne faire qu’un seul puits large dans lequel sera chargé l’échantillon d’ADN (le puits doit permettre de charger un volume de 100 μL). Une fois solidifié, le gel a été positionné dans le cadre de la cuve pour électrophorèse et cette dernière a été délicatement remplie de tampon TAE 1 X. L’ADN génomique réparé a été chargé dans le gel. De part et d’autre du puits réservé à cet échantillon d’ADN, 100 ng de marqueur de taille T7 ainsi que le marqueur de poids moléculaire Low Range PFG Marker 0.1-200 kb (New England, Biolabs) ont été déposés dans les puits voisins. La migration a ensuite été lancée avec les paramètres suivants : voltage de 6 V/cm, angle de 120°, et temps de pulse évoluant de 5 à 15 secondes pendant 13 heures puis de 15 à 25 secondes pendant 3 heures. Après s’être assuré que l’échantillon d’ADN était bien rentré dans le gel (après environ 10 min de migration), la pompe permettant la circulation du tampon d’électrophorèse (positionnée sur 70) puis le système de refroidissement (réglé sur 12°C) ont été mis en route.

Une fois la migration électrophorétique terminée, le protocole suivant a été réalisé afin d’éviter d’exposer l’ADN génomique aux rayons UV et de sélectionner malgré tout, les fragments d’ADN génomique ayant une taille de 40 kpb. Les parties du gel contenant les marqueurs T7 et Low Range PFG Marker (bandes se trouvant de part et d’autre de celle contenant l’ADN génomique) ont été découpées avec des lames stériles. La bande contenant l’ADN génomique a été mise de côté ; les bandes contenant les marqueurs ont été quant à elles colorées dans un bain de BET, puis visualisées sous lumière UV. Un repère indiquant la position du marqueur T7 a été pointé sur les bandes de gel à l’aide d’une pipette Pasteur. La lumière UV a été éteinte, puis toutes les bandes ont été réassemblées afin de reconstituer le gel. Une bande de gel de 2 à 4 mm d’épaisseur qui contient l’ADN génomique ayant migré à hauteur des repères du marqueur T7, a été découpée à l’aide d’une lame stérile puis placée dans un tube Falcon 15 mL qui aura été préalablement taré. A cette étape, la bande de gel contenant les fragments d’ADN génomique qui ont migré à 40 kpb peut être stockée à 4°C ou à -20°C, pendant un an.

4 Elution des fragments d'ADN génomique de 40 kpb contenus dans l'agarose

Afin d'éluer les fragments d'ADN génomique de 40 kpb contenus dans la bande de gel d'agarose LMP, effectuer le protocole suivant :

- peser le tube Falcon 15mL qui a été préalablement taré, pour déterminer le poids de la bande d'agarose contenant les fragments d'ADN génomique de 40 kpb (considérer que 1 mg d'agarose solide donnera 1 μL d'agarose fondu),
- préchauffer le tampon GELase 50X (2mM Tris-Bis [pH 6], 2 M NaCl; Epicentre) à 45°C,
- faire fondre l'agarose dans un bain-marie préchauffé à 70°C pendant 10 à 15 min, puis transférer rapidement le tube dans un bain-marie préchauffé à 45°C,
- ajouter la quantité appropriée de tampon GELase 50X préchauffé de manière à obtenir une concentration finale de 1X,
- ajouter 1 unité (1 μL) d'enzyme GELase (Epicentre) pour 100 μL d'agarose fondu, maintenir le tube à 45°C et l'agiter doucement,
- incuber la solution à 45°C pendant au moins 1 heure (une incubation pendant toute la nuit est possible),
- transférer la réaction à 70°C pendant 10 min afin d'inactiver l'enzyme GELase,
- écarter 15 μL de cette solution pour une quantification ultérieure de l'ADN sur gel d'agarose et répartir dans des microtubes stériles de 1.5 mL, le restant de la solution en fractions de 500 μL
- refroidir le(s) tube(s) dans un bain de glace pendant 5 min,
- centrifuger pendant 20 min à 20000 g afin de culoter les oligosaccharides insolubles,
- prélever 90 à 95% du surnageant (l'ADN se trouve dans le surnageant) puis le transférer dans un microtube de 1.5 mL. Prendre garde à ne pas prélever le culot gélatineux.

Afin de précipiter l'ADN qui se trouve dans le surnageant, procéder de la manière suivante :

- ajouter 0.1 volume d'acétate de sodium (3 M, pH 6) et agiter doucement,
- ajouter 2.5 volumes d'éthanol 100% et agiter doucement par retournement du tube,
- laisser reposer 10 min à température ambiante,
- centrifuger à la vitesse maximum (~ 20000 g) pendant 20 min, puis aspirer le surnageant,
- laver le culot d'ADN deux fois en ajoutant 500 μL d'éthanol 70%, en vortexant à faible vitesse puis en centrifugeant à 20000 g pendant 10 min,
- inverser délicatement le tube et laisser sécher le culot d'ADN à l'air, pendant 5 à 10 min,
- laisser l'ADN se réhydrater et se dissoudre dans du tampon TE (ajouter le volume nécessaire pour obtenir une concentration finale d'ADN d'environ 0.1 $\mu\text{g} \cdot \mu\text{L}^{-1}$), pendant au moins 4 heures à 4°C.

La concentration de l'ADN avant précipitation a été évaluée sur un gel d'agarose, ceci afin de connaître la quantité de tampon TE qu'il faut ajouter pour obtenir, après la précipitation de l'ADN une concentration finale de $0.1 \mu\text{g} \cdot \mu\text{L}^{-1}$. Les $15 \mu\text{L}$ de solution qui ont été écartés lors de l'éluion de l'ADN, ont été déposés dans un gel d'agarose à 0.9% préparé dans du tampon TAE 1X. La migration électrophorétique a été effectuée à un voltage de 75 V pendant 1h30. Des concentrations connues de marqueur d'ADN T7 ont été mises à migrer en parallèle avec l'ADN afin d'estimer par comparaison sa concentration.

5 Réaction de ligation

La réaction de ligation permet de cloner les fragments d'ADN génomiques de 40 kpb dans des vecteurs de clonage de type fosmide. Cette réaction a été effectuée en mélangeant les réactifs dans l'ordre indiqué ci-dessous :

- eau milliQ stérile : quantité suffisante pour compléter le volume final à $10 \mu\text{L}$
- tampon de ligation Fast-Link 10X (Epicentre) : $1 \mu\text{L}$
- ATP 10 mM : $1 \mu\text{L}$
- vecteur CopyControl pCC1FOS1 $0.5 \mu\text{g} \cdot \mu\text{L}^{-1}$: $1 \mu\text{L}$
- ADN génomique : $0.25 \mu\text{g}$ d'ADN génomique de 40 kpb
- ADN ligase Fast-Link ($2 \text{ U} \cdot \mu\text{L}^{-1}$, Epicentre) : $1 \mu\text{L}$

Cette réaction a été réalisée dans un volume final de $10 \mu\text{L}$, à température ambiante, pendant 2 heures. Le ratio de 10 :1 de vecteur CopyControl pCC1FOS1 pour de l'ADN génomique est optimal pour la réaction de ligation ($0.5 \mu\text{g}$ de vecteur CopyControl pCC1FOS équivaut à 0.09 pmole et $0.25 \mu\text{g}$ d'ADN génomique de 40 kpb équivaut à 0.009 pmole). Sous ces conditions, une réaction permet d'obtenir entre 10^3 et 10^6 clones, en fonction de la qualité de l'insert d'ADN.

Afin d'inactiver la ligase, le mélange a été ensuite incubé pendant 10 min à 70°C . A ce moment, on a le choix de conserver l'échantillon à -20°C ou bien de procéder à l'étape suivante.

6 Réaction d'empaquetage des clones dans des particules de phage λ

Le jour de la réaction d'empaquetage, inoculer 50 mL de milieu LB dans lequel a été ajouté 10 mM de MgSO_4 , avec 5 mL d'une culture ayant poussé toute la nuit d'*E. coli* EPI300 ([F-mcrA Δ (mrr-hsdRMS-mcrBC) Φ 80dlacZ Δ M15 Δ lacX74 recA1 endA1 araD139 Δ (ara, leu)7697 galU galK λ - rpsL nupG trfA tonA dhfr], Epicentre). Cette culture est mise à incuber à 37°C sous agitation, jusqu'à l'obtention d'une $DO_{600} = 0.8^{-1}$, puis est stockée à 4°C .

La réaction d'empaquetage consiste à emballer les vecteurs de clonage ayant des inserts

de 40 kbp, dans des particules de phage λ . Cette réaction a été réalisée de la façon suivante :

- décongeler dans de la glace, un tube d'extraits de phage MaxPlax λ Packaging (Epicentre),
- une fois décongelé, transférer immédiatement la moitié du contenu du tube (soit 25 μL d'extraits de phage) dans un second tube de 1.5 mL qu'on placera dans de la glace,
- recongeler immédiatement les 25 μL restants d'extrait de phage à -70°C ,
- ajouter les 10 μL de la réaction de ligation dans les 25 μL d'extraits de phage décongelés,
- mélanger plusieurs fois la préparation par pipetage, en évitant de former des bulles, puis centrifuger brièvement le tube,
- incubé 90 min à 30°C ,
- ajouter les 25 μL d'extraits de phage qui ont été recongelés,
- incubé à nouveau les réactions à 30°C pendant 90 min,
- ajouter du tampon de dilution de phage (10 mM de Tris-HCl [pH 8.3], 100 mM de NaCl, 10 mM de MgCl_2), de manière à obtenir un volume final de 1 mL, et mélanger doucement,
- ajouter 25 μL de chloroforme (un précipité visqueux peut se former) et conserver à 4°C .
- déterminer le titre des particules de phage et procéder à l'étape suivante, ou bien conserver les particules de phage à 4°C (conservation possible pour quelques semaines).

7 Calcul du titre des particules phagiques

A cette étape de la construction de la banque métagénomique, il est recommandé de calculer le titre des particules de phage ayant empaqueté le vecteur de clonage. Ce titre correspond à une évaluation chiffrée de l'efficacité infectieuse de la préparation virale. Son calcul repose sur le fait que seules les cellules d'*E. coli* EPI300 qui ont été infectées par des particules phagiques ayant empaqueté le vecteur CopyControl pCC1FOS Fosmid (vecteur qui porte un gène de résistance au chloramphénicol), peuvent se développer sur un milieu solide contenant du LB et du chloramphénicol.

Le titre des particules de phage a été calculé de la façon suivante :

- réaliser une série de dilution (10^{-1} , 10^{-2} , 10^{-3} , 10^{-4} , 10^{-5}) des particules phagiques avec du tampon de dilution de phage (10 mM Tris-HCl [pH 8.3], 100 mM NaCl, 10 mM MgCl_2),
- pour chaque dilution, prendre 10 μL de particules phagiques diluées pour infecter 100 μL d'une culture d'*E. coli* EPI300 préparée précédemment ($\text{DO}_{600} = 0.8$ à 1, cf § VII), incubé à 37°C pendant 20 min
- étaler ces cultures sur des boîtes de Pétri contenant du LB et du chloramphénicol à $12.5 \mu\text{g.mL}^{-1}$ et incubé toute la nuit à 37°C
- dénombrer les colonies et calculer le titre des particules de phage selon la formule suivante :

$$\frac{(\text{nombre de colonies}) \times (\text{facteur de dilution}) \times (1000 \mu\text{L.mL}^{-1})}{\text{volume du phage étalé } [\mu\text{L}]}$$

8 Etalement et sélection de la banque métagénomique

Connaissant le titre des particules phagiques et le nombre de clones que l'on aimerait obtenir, il faut alors calculer le volume de particules phagiques dont on a besoin pour préparer la banque métagénomique. Sachant cela et de manière à obtenir une densité de clones adéquate sur boîte de Pétri, il convient de diluer convenablement les particules phagiques dans du tampon de dilution de phage. On infecte alors les cellules d'*E. coli* EPI300 préparées lors de l'étape VII avec ces particules phagiques diluées, en respectant le ratio suivant : 10 μL de particules phagiques diluées pour 100 μL de cellules à infecter. L'adsorption est menée pendant 20 min à 37°C. On étale ensuite les cellules sur des boîtes contenant du LB et du chloramphénicol à 12.5 $\mu\text{g.mL}^{-1}$ pour sélectionner les clones contenant le vecteur fosmidique. Les boîtes sont mises à incuber toute la nuit à 37°C.

9 Conservation de la banque

Les clones ont tous été individuellement repiqués dans des plaques de 96 puits contenant 200 μL de milieu de culture sélectif et mises à incuber pendant 17 heures à 37°C. Le milieu de culture contenait du LB, du chloramphénicol à 12.5 $\mu\text{g.mL}^{-1}$ et du milieu Hogness 1X. L'ajout de milieu Hogness au LB permet de congeler directement les clones après les avoir fait pousser dans ce milieu de culture. Les cultures ont d'abord été congelées à -20°C puis à -80°C. Le milieu Hogness a été préparé à une concentration de 10X, aliquoté, autoclavé puis conservé à température ambiante. A cette concentration, il contient : 36 mM de $\text{K}_2\text{HPO}_4 \cdot 3\text{H}_2\text{O}$, 13 mM de KH_2PO_4 , 20 mM de citrate- N_{a3} , 2H₂O 10mM $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ et 44% (v/v) de glycérol.

10 Criblage de la banque

Le criblage de la banque métagénomique a été réalisé par PCR. Les séquences cibles de ces PCR étaient le gène codant l'ARNr 16S des *Archaea* et le gène codant l'ARNr 18S des eucaryotes. Les clones de la banque ayant dans leur insert l'un de ces gènes ont ainsi pu être détectés.

Pour faciliter le criblage de la banque, les clones de la banque ont été mélangés de façon à former des pools ; un pool étant constitué de 24 clones différents (9 μL de culture par clone). Le but de ce "pooling" était de minimiser le nombre de PCR nécessaire pour cribler la banque. Les pools ont été centrifugés à 8000 g pendant 5 min, puis resuspendus dans 215 μL de tampon TE [pH 8]. Un petit volume (2.5 μL) de chaque pool a été lysé à 95°C pendant 10 min puis a servi comme matrice d'ADN pour la réaction de PCR. Le mélange réactionnel des PCR était le suivant :

- Tampon 10X : 1.25 μL (1X final),
- MgCl_2 50 mM : 0.375 μL (1.5 mM final),

Nom	Spécificité	Sens	Position	Séquence (5' - 3')
21F	<i>Archaea</i>	S	7-26	TTC CGG TTG ATC CTG CCG GA
1492R	Procaryotes	AS	1492-1509	GGT TAC CTT GTT ACG ACT T
18S-42F	Eucaryotes	S	~42-61	CTC AAR GAY TAA GCC ATG CA
18S-1498R	Eucaryotes	AS	~1498-1479	CAC CTA CGG AAA CCT TGT TA

Tableau 1 – Liste des amorces utilisées pour l'amplification par PCR de fragments d'ADNr 16S archaéens et de fragments d'ARNr 18S eucaryotes. S= amorce sens, AS=amorce anti-sens. La position des amorces correspond à *Thermococcus celer* pour les amorces d'*Archaea*, et à *Saccharomyces cerevisiae* pour les amorces d'eucaryotes.

- mélange de dNTPs 10mM chaque (Interchim) : 0.3125 μL (250 μM final),
- amorce sens 10 μM : 0.25 μL (0.2 μM final),
- amorce anti-sens 10 μM : 0.25 μL (0.2 μM final),
- UptiTherm DNA polymérase 5U. μL^{-1} (Interchim) : 0.5 U,
- eau milliQ : quantité suffisante pour 12.5 μL ,
- matrice ADN : 2.5 μL de pool de cellules lysées.

Le programme de cette réaction d'amplification des gènes par PCR était le suivant :

30 sec à 95°C (dénaturation)	}	40 cycles.
30 sec à 53°C (hybridation)		
1min 30 à 72°C (élongation)		
7 min à 72°C (élongation initiale)		

Les amorces suivantes (Tableau 1) ont été utilisées pour les différentes réactions d'amplification.

Les produits d'amplifications ont été ensuite visualisés sur gel d'agarose 0.8% préparé dans du tampon TAE 1X. Seuls les pools contenant un ou plusieurs clones dont les fosmidés ont l'un des gènes ciblés dans leur insert, doivent présenter une bande d'amplification. Un second criblage par PCR, a ensuite permis d'identifier le(s) clone(s) positif(s) dans les pools présentant des produits d'amplification.

11 Induction du nombre de copies de fosmidés dans les clones

Une fois que les clones d'intérêts ont été identifiés (clones dont les fosmidés portent le gène codant l'ARNr 16S d'archaea ou le gène codant l'ARNr 18S d'eucaryote), l'amplification du nombre de copies de leur fosmide a été induite pour faciliter leur extraction en vue d'un séquençage ultérieur. Cette opération s'est effectuée de la manière suivante :

- faire une pré-culture de chaque clone d'intérêt dans 5 mL de milieu LB contenant 12.5 $\mu\text{g.mL}^{-1}$ de chloramphénicol.
- incubé la ou les pré-culture(s) à 37°C, pendant toute une nuit et sous agitation,
- dispenser 4.5 mL de milieu frais LB + 12.5 $\mu\text{g.mL}^{-1}$ de chloramphénicol dans des tubes

Falcon de 15 mL, puis ajouter 0.5 mL de pré-culture et 5 μ L de la solution d'induction 1000X (Epicentre)

- incuber les cultures devant être induites à 37°C, sous forte agitation, pendant 5 heures
- centrifuger les cellules et extraire les fosmides au moyen d'une méthode standard d'extraction de l'ADN plasmidique. Nous avons pour notre part utilisé une méthode de lyse alcaline Sambrook *et al.* (1989).

Résultats

Étude 1

Novel uncultured *Epsilonproteobacteria* dominate a filamentous sulphur mat from the 13°N hydrothermal vent field, East Pacific Rise 113

Étude 2

Characterization of a 40 kbp uncultivated euryarchaeotal fosmid from a microbial community associated to the hydrothermal vent polychaete annelid, *Alvinella pompejana* 139

Étude 3

Thermophilic lifestyle for an uncultured archaeon from hydrothermal vents : evidence from environmental genomics 167

NOVEL UNCULTURED
Epsilonproteobacteria DOMINATE A
FILAMENTOUS SULPHUR MAT FROM THE
13°N HYDROTHERMAL VENT FIELD,
EAST PACIFIC RISE

Novel uncultured *Epsilonproteobacteria* dominate a filamentous sulphur mat from the 13°N hydrothermal vent field, East Pacific Rise.

Hélène Moussard, Erwan Corre, Marie-Anne Cambon-Bonavita, Yves Fouquet, Christian Jeanthon

Pendant la campagne océanographique PHARE qui se déroulait sur le site hydrothermal de la dorsale du Pacifique oriental (13°N EPR), nous avons eu l'occasion d'observer la formation très rapide d'un biofilm microbien. Ce biofilm d'aspect filamenteux, blanc, dense et épais, est apparu en seulement quatre jours sur le flotteur d'un module de colonisation qui avait été déposé sur la paroi d'une cheminée hydrothermale active. Si des biofilms sont fréquemment observés sur la plupart des sites hydrothermaux marins profonds (Longnecker & Reysenbach, 2001), les micro-organismes qui les composent demeurent peu connus.

La formation très rapide du biofilm sur notre colonisateur in situ, nous a fait penser aux dépôts très spectaculaires de floculats blancs répertoriés à plusieurs reprises sur des sites hydrothermaux profonds (Kelley *et al.*, 2002 ; pour revue) (Fig. 1.A). Ces floculats blancs sont entraînés et émis par le fluide hydrothermal peu de temps après des éruptions volcaniques. Leur émission est parfois tellement intense et massive, que pour imager le phénomène, on le compare à une tempête de neige, et les édifices hydrothermaux sont appelés "snowblowers" (Fig. 1.B).

Des analyses de ces floculats ont révélé qu'il s'agissait de filaments de soufre biogénique associés à des micro-organismes. Ces derniers n'ont été ni isolés en culture pure, ni phylogénétiquement caractérisés. Toutefois, une expérience de culture réalisée en laboratoire a permis d'enrichir ces micro-organismes (Taylor *et al.*, 1999). Pour cela, des échantillons de biofilms hydrothermaux (Fig. 2.A à D) ont été incubés en condition de microaérophilie, dans un réacteur. Ce réacteur qui est alimenté en continu par de l'eau de mer, permet d'établir un gradient d'H₂S allant d'une concentration de 400 μM à 1200 μM (Fig. 2.F). Des filaments de soufre identiques à ceux des sites hydrothermaux, ont été rapidement observés dans le réacteur (comparer Fig. 2.D et G). Les micro-organismes responsables de leur formation étaient des vibrions mobiles de type v (Fig. 2.E et H). Les micro-organismes de type c (bâtonnets assemblés en chaînes) qui étaient également associés aux filaments de soufre hydrothermaux, n'ont pas été enrichis dans ce réacteur (Fig. 2.E). La caractérisation des vibrions responsables de la formation des filaments de soufre,

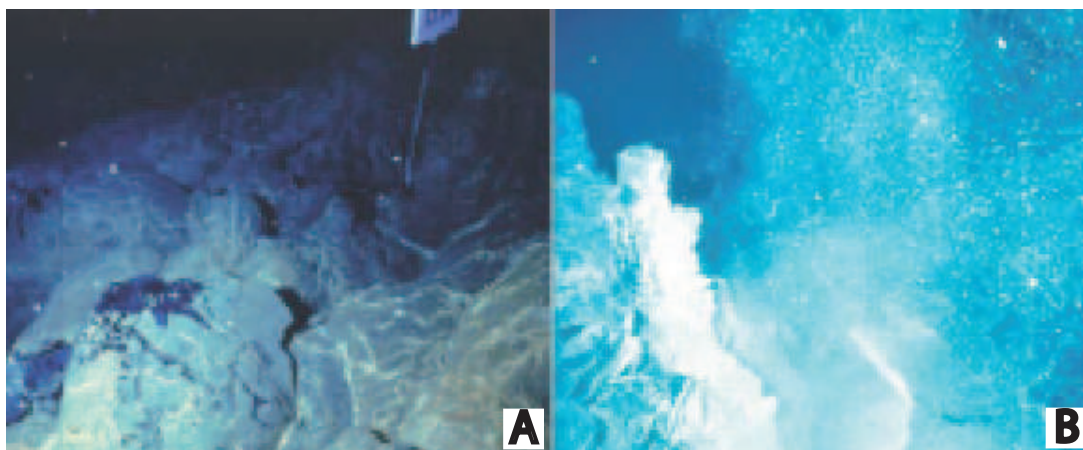


Fig. 1 – Quelques photographies de mattes microbiennes observées sur des sites hydrothermaux profonds. (A) Photographie de mattes bactériennes (matériel blanc) tapissant des roches basaltiques provenant d’une éruption volcanique sur la dorsale Juan de Fuca, en 1993. (B) Photographie montrant l’expulsion de flocculats blancs entraînés par le fluide hydrothermal, peu de temps après une éruption volcanique sur la dorsale du Pacifique oriental (9°N EPR) (d’après Haymon *et al.*, 1993). Ce phénomène donne l’impression d’assister à une tempête de neige (source Kelley *et al.*, 2002).

n’est pas allée plus loin. Cependant, leur morphologie est très comparable à celle d’un organisme isolé d’eau de mer côtière qui produit des filaments de soufre identiques à ceux observés sur les sites hydrothermaux (comparer Fig. 2.H et I) (Taylor & Wirsen, 1997 ; Taylor *et al.*, 1999). Cet organisme marin a été par la suite caractérisé (Wirsen *et al.*, 2002). Il s’agit d’une bactérie microaérophile, sulfo-oxydante et autotrophe, qui est affiliée aux *Epsilonproteobacteria*, et pour laquelle le nom de “*Candidatus Arcobacter Sulfidicus*” a été proposé.

Le biofilm qui s’est formé sur le flotteur de notre module de colonisation ressemblait à celui qui avait été caractérisé par Taylor et ses collaborateurs (1999). Cet article présente l’étude approfondie de ce biofilm. Ce dernier a été observé en microscopie électronique à balayage. Sa composition chimique a été caractérisée par diffraction aux rayons X et par observation microscopique électronique à balayage couplée à une analyse de spectroscopie à dispersion d’énergie (SEM-EDS). La composition microbienne du biofilm a été ensuite étudiée par des techniques phylogénétiques moléculaires (inventaires des gènes *aclB* et des gènes codant les ARNr 16S). Enfin, une sonde ciblant les micro-organismes présents dans ce biofilm a été mise au point, afin de les quantifier.

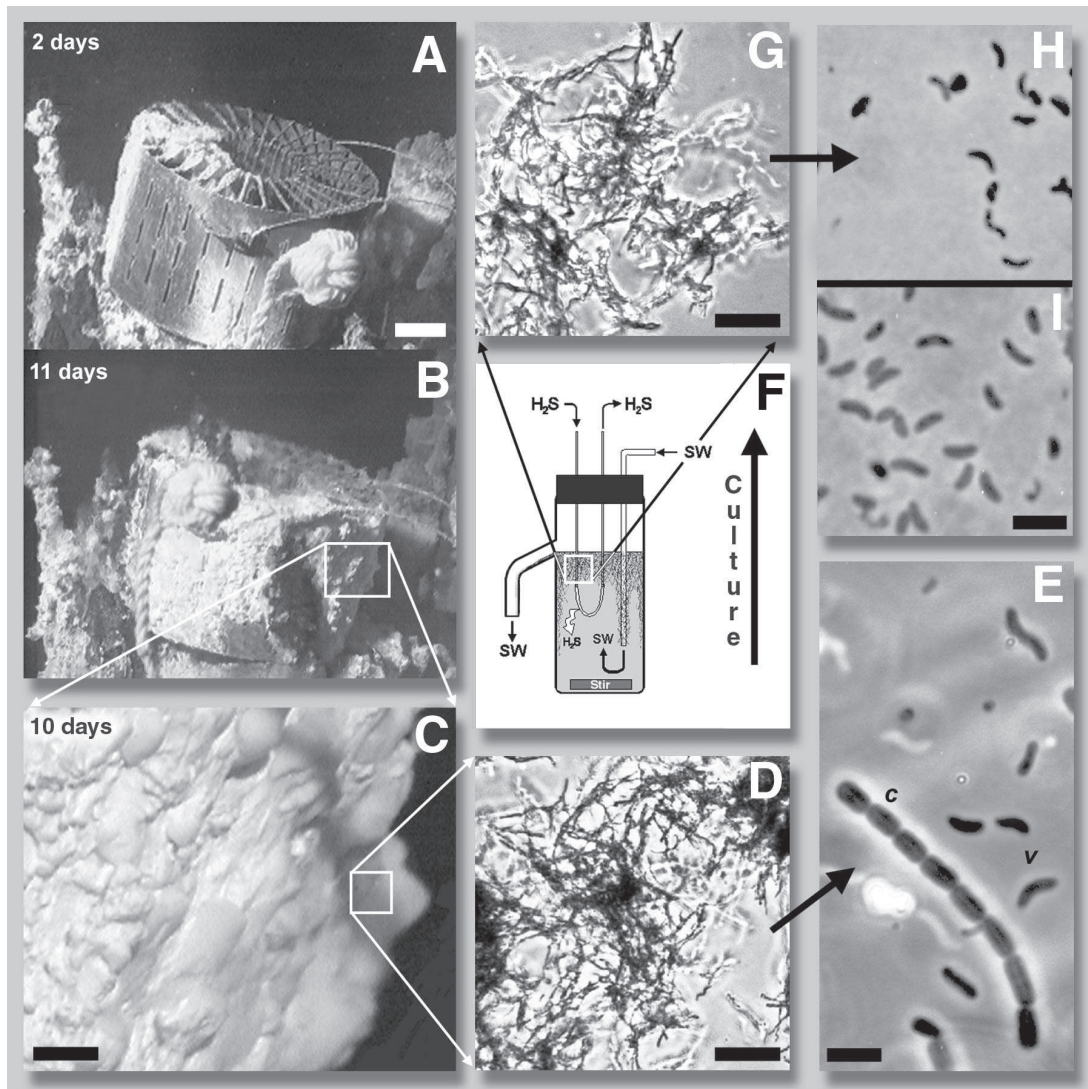


Fig. 2 – Formation de biofilms contenant des filaments de soufre d'origine microbienne. Ces biofilms ont été observés sur le site hydrothermal 9°N EPR et dans une culture de micro-organismes réalisée à bord d'un bateau. (A et B) : Observation du biofilm contenant des filaments de soufre, sur un module de colonisation déployé au niveau du site hydrothermal 9°N EPR ; barre : 5 cm. (C) : Agrandissement photographique du biofilm ; barre : 1 cm. (D) : Photographie à contraste de phase du biofilm ; barre : 10 μm . (E) : Observation microscopique des micro-organismes du biofilm ; barre : 2 μm . (F) : Schéma du réacteur ayant permis la culture en continue des micro-organismes du biofilm. L'inoculum provient d'un échantillon du biofilm prélevé sur le module de colonisation. SW : eau de mer oxygénée. (G) : Microphotographie à contraste de phase des filaments de soufre produits à bord du bateau, dans le réacteur ; barre : 10 μm . (H) : Micro-organismes enrichis dans le réacteur. (I) : Micro-organisme isolé d'eau de mer cotière peu profonde, produisant le même type de biofilm que celui observé sur le site hydrothermal 9°N EPR ; barre : 2 μm (d'après Taylor *et al.*, 1999).



Novel uncultured *Epsilonproteobacteria* dominate a filamentous sulphur mat from the 13°N hydrothermal vent field, East Pacific Rise

Hélène Moussard¹, Erwan Corre², Marie-Anne Cambon-Bonavita¹, Yves Fouquet³ & Christian Jeanthon¹

¹Laboratoire de Microbiologie des Environnements Extrêmes, Centre National de la Recherche Scientifique, IFREMER, Université de Bretagne Occidentale, Institut Universitaire Européen de la Mer, Plouzané, France; ²FR 2424, Station Biologique, Roscoff, France; and ³Département de Géosciences Marines, IFREMER Centre de Brest, Plouzané, France

Correspondence: Christian Jeanthon, UMR 7144, Equipe Phytoplancton Océanique, Station Biologique, Place Georges Teissier, 29682 Roscoff Cedex, France. Tel.: +33 298 292 379; fax: +33 298 292 324; e-mail: jeanthon@sb-roscoff.fr

Received 4 July 2005; revised 14 June 2006; accepted 16 June 2006.

DOI: 10.1111/j.1574-6941.2006.00192.x

Editor: Michael Wagner

Keywords

deep-sea hydrothermal vent; biofilm; 16S rRNA gene; *Epsilonproteobacteria*; *Arcobacter*; FISH

Abstract

Rapid growth of microbial sulphur mats have repeatedly been observed during oceanographic cruises to various deep-sea hydrothermal vent sites. The microorganisms involved in the mat formation have not been phylogenetically characterized, although the production of morphologically similar sulphur filaments by a *Arcobacter* strain coastal marine has been documented. An *in situ* collector deployed for 5 days at the 13°N deep-sea hydrothermal vent site on the East Pacific Rise (EPR) was rapidly colonized by a filamentous microbial mat. Microscopic and chemical analyses revealed that the mat consisted of a network of microorganisms embedded in a mucous sulphur-rich matrix. Molecular surveys based on 16S rRNA gene and *aclB* genes placed all the environmental clone sequences within the *Epsilonproteobacteria*. Although few 16S rRNA gene sequences were affiliated with that of cultured organisms, the majority was related to uncultured representatives of the *Arcobacter* group ($\leq 95\%$ sequence similarity). A probe designed to target all of the identified lineages hybridized with more than 95% of the mat community. Simultaneous hybridizations with the latter probe and a probe specific to *Arcobacter* spp. confirmed the numerical dominance of *Arcobacter*-like bacteria. This study provides the first example of the prevalence and ecological significance of free-living *Arcobacter* at deep-sea hydrothermal vents.

Introduction

At deep-sea hydrothermal vents, steep temperature and chemistry gradients created by the mixing of anoxic hydrothermal fluids with oxygenated seawater are known to offer a variety of habitats for the growth of diverse microorganisms. The microbial communities occupy both aerobic and anaerobic environments having *in situ* temperatures ranging from ambient seawater temperature to temperatures greater than 110 °C (Karl, 1995). Cultivation of metabolically and phylogenetically diverse thermophilic and hyperthermophilic bacteria and archaea has been extensively reported from the hottest ecological niches of the hydrothermal environment (Jones *et al.*, 1983; Harmsen *et al.*, 1997a; Takai *et al.*, 2001; Huber *et al.*, 2002). In part, this is because black smokers and related sulphide structures are the most extensively studied vent environments. Microorganisms of other thermal classes have however been isolated from these heterogeneous and fluctuating habitats although most of the

microbial diversity has only been characterized by molecular methods (Takai & Horikoshi, 1999; Jeanthon, 2000; Reysenbach *et al.*, 2000; Teske *et al.*, 2002; Nercessian *et al.*, 2003).

Microbial mats growing on rocks, chimneys, sediments or animal surfaces which are exposed to vent waters represent another hydrothermal vent habitat with associated microbial communities (Karl, 1995). Most of the reports on microbial mats involved microscopic observations (Jannasch & Wirsén, 1981; Nelson *et al.*, 1989; Taylor *et al.*, 1999), or fatty acid (Hedrick *et al.*, 1992) and phylogenetic analyses (Moyer *et al.*, 1995; Longnecker & Reysenbach, 2001; Alain *et al.*, 2004). In certain locations, the mat-forming organisms were identified as large filamentous sulphur-oxidizing bacteria of the genera *Beggiatoa* and *Thiothrix* (Jannasch & Wirsén, 1981; Jannasch *et al.*, 1989; Nelson *et al.*, 1989). These bacteria are chemoautotrophic; they fix CO₂ by oxidizing hydrogen sulphide present in hydrothermal fluids. Flocculent material entrained out of the seafloor in the venting fluids and rapid growth of white

filamentous bacteria have also often been observed after recent eruptive events or at sites of extensive emission discharges at Loihi hydrothermal vents, Guaymas Basin, on the East Pacific Rise (EPR) and Juan de Fuca ridge (Nelson *et al.*, 1991; Haymon *et al.*, 1993; Embley *et al.*, 1995, 2000; Moyer *et al.*, 1995; Shank *et al.*, 1998; Taylor *et al.*, 1999). Subsequent studies of the emitted particles showed that they were composed of inorganic, sulphur-rich filaments with a low organic carbon content (Nelson *et al.*, 1991). The associated microorganisms have never been isolated or phylogenetically characterized. However, the formation of filamentous sulphur of morphology nearly identical to the material discharged from diffuse-flow hydrothermal vents was documented for a sulphide-oxidizing bacterium isolated from shallow coastal marine waters (Taylor & Wirsen, 1997; Taylor *et al.*, 1999). An extensive characterization revealed that this motile vibrioid autotrophic organism clustered with members of the genus *Arcobacter* within the *Epsilonproteobacteria* (Wirsen *et al.*, 2002).

Molecular examination of mats attached to chimneys, animal surfaces and various sampling devices also revealed microbial communities dominated by a wide diversity of uncultured groups of *Epsilonproteobacteria* (Polz & Cavanaugh, 1995; Cary *et al.*, 1997; Corre *et al.*, 2001; Longnecker & Reysenbach, 2001; Huber *et al.*, 2003; Lopez-Garcia *et al.*, 2003; Alain *et al.*, 2004). Although there is generally no evidence available that they are also numerically abundant *in situ*, it is now recognized that the *Epsilonproteobacteria* are the predominant bacteria in the global deep-sea hydrothermal systems (Polz & Cavanaugh, 1995; Cary *et al.*, 1997; Corre *et al.*, 2001; Longnecker & Reysenbach, 2001; Huber *et al.*, 2003; Lopez-Garcia *et al.*, 2003; Alain *et al.*, 2004). Genes for key enzymes of the reductive tricarboxylic acid (rTCA) cycle, e.g. *aclB* – the β subunit of the ATP citrate lyase, were detected in cultured *Epsilonproteobacteria* and recovered in various hydrothermal vent habitats (Campbell *et al.*, 2003; Campbell & Cary, 2004; Hügler *et al.*, 2005; Takai *et al.*, 2005). These reports suggested that autotrophic CO₂ fixation via the rTCA cycle operate in diverse *Epsilonproteobacteria* and might sustain the predominant primary production *in situ*.

In this study we investigated the bacterial community from a white filamentous mat formed within a few days on a syntactic foam float associated with an *in situ* collector deployed at the 13°N hydrothermal site (EPR). The goal in this study was to determine the microbial compositions of the mat that was morphologically comparable to that observed and described by Taylor *et al.* (1999). On the basis of molecular phylogenetic surveys based on 16S rRNA and *aclB* genes and quantitative fluorescence *in situ* hybridization (FISH), we demonstrate the actual *in situ* dominance of yet uncultured *Epsilonproteobacteria* including the prevalence of *Arcobacter* relatives in the hydrothermal sulphur mat.

Material and methods

Hydrothermal site description

The hydrothermal sample was collected at the 13°N hydrothermal vent field on the EPR during the oceanographic cruise PHARE conducted with the research vessel 'L'Atalante' and the remotely operated vehicle (ROV) 'Victor 6000'. The portion of the EPR situated at 13°N is about 300 km south of the Orozco fracture zone and 100 km north of a small transform fault located at 11°49'N. At 13°N, the dome-shaped ridge has a central rift valley varying between 200 m and more than 600 m in width and with a mean depth of about 2600 m (Hekinian & Fouquet, 1985). The volcanic activities at 13°N give rise to a wide range of hydrothermal venting structures from diffuse vents (up to 100 °C) to black smokers with temperatures above 350 °C. Our study focuses on the PP12 chimney of the Genesis hydrothermal site (12°48'66" N, 103°56'44" W). The chimney was colonized by dense populations of the polychaetous annelid *Alvinella pompejana* that builds tubes directly in contact with the sulphides. Few specimens of the tube-worm *Riftia pachyptila* were also observed.

Sample collection

An *in situ* collector (Nercessian *et al.*, 2003), designed to concentrate the microorganisms discharged by hydrothermal emissions, was deployed on the PP12 chimney (Fig. 1). The *in situ* collector was brought to the deployment site into an insulated container and settled down by the arm of the ROV. Just before deployment of the *in situ* collector, the chimney's surface was scraped by the arm of the ROV during the sampling of *Alvinella pompejana* specimens. The mechanical removing of the overlying sulphide structure resulted in an increase of the vent emission. The collector was settled at the exact location where alvinellids were taken, a few centimeters above the vent emission (Fig. 1). Immediately after deployment and prior to removal, discrete temperature measurements were achieved at the bottom and at the top of the collector using the thermoprobe (Micrel, Hennebont, France) held by the teleoperated arm of the ROV. Video imagery of the *in situ* collector was acquired during successive dives. The collector was surveyed during the following Victor dives. At the fourth day of deployment, a thick and dense white biofilm was observed on the syntactic foam float attached to the collector (Fig. 1).

After 5 days of exposure, the *in situ* collector was transferred into enclosed containers to minimize contamination from surrounding seawater during transportation to surface. Once onboard the ship, the microbial mat covering the marker was aseptically collected by scraping and immediately subsampled. Material for scanning electron microscopy (SEM) was fixed with 2.5% glutaraldehyde in

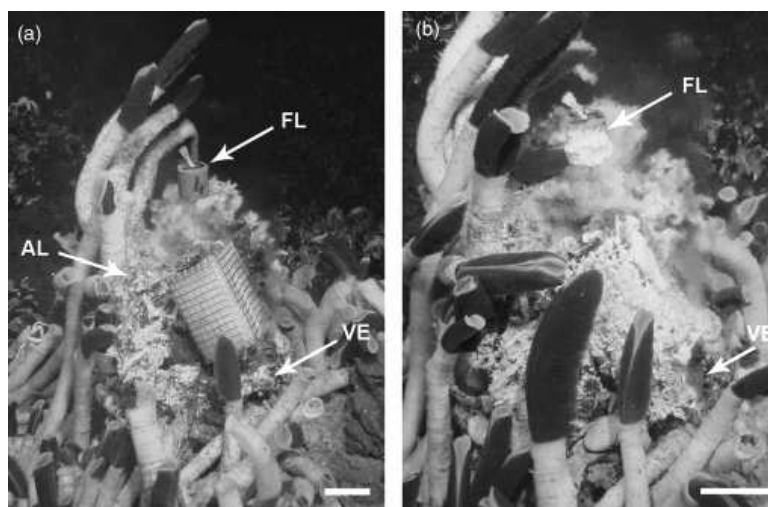


Fig. 1. Photographs of the *in situ* collector deployed for 5 days at Genesis vent site. (a) Deposition of the collector near a *Riftia pachyptila* colony; the syntactic foam float (FL) is attached to the collector; white tubes of *Alvinella pompejana* (AL) are visible around the vent orifice (VE); bar, 10 cm (copyright Ifremer/PHARE 2002). (b) Recovery of the *in situ* collector after 5 days of incubation; the dense white biofilm formed on the syntactic foam float was scrapped once on-board.

30 g L⁻¹ sea salts (Sigma) and stored at 4 °C. Aliquot for molecular and X-ray diffraction (XRD) analyses was stored in ethanol at -20 °C. For FISH experiments, biomass was fixed for 3 h at 4 °C with 4% paraformaldehyde in 3 × phosphate-buffered saline (PBS) (Sambrook *et al.*, 1989). The fixation buffer was then replaced by a mixture of equal volumes of 3 × PBS and 100% ethanol and the fixed sample was stored at -20 °C until further use.

Scanning electron microscopy coupled to energy-dispersive X-ray spectroscopy

Macroscopic mat filaments were deposited on a 10-µm-pore-size membrane filter (Millipore) and fixed overnight at 5 °C in 2.5% formaldehyde. After salt was eliminated, the preparation was dehydrated in increasing ethanol concentrations (50%, 70%, 90% and 100%). Samples were then critical point dried and gold coated. Observation was carried out with a Philips (XL 30-LaB6) scanning electron microscope operated at 25 kV.

The elemental composition of the sample was analyzed by energy-dispersive X-ray spectroscopy (EDS) coupled to the scanning electron microscope.

X-ray diffraction

XRD analysis was accomplished with a Bruker D-500 goniometer coupled with the automated DACO-MP system for data acquisition. Mineral phases were identified on the diffraction spectra using the EVA program. Powder XRD scans were performed on dried crushed samples under the following conditions: Cu K α radiation, 25 mA, 35 kV, and 5 s per step with a step size of 0.02° (2 θ). The scan range for all samples was 5–70°. Background and Kalph2 lines were removed from the diagrams. The mineralogical composition of the sample

was determined by comparing sample diffraction patterns to mineral standards provided by the JCPDS files.

Nucleic acid isolation and purification

The sample was centrifuged at 4000 g for 10 min. The pellet was suspended in 150 µL of homogenization buffer [50 mM Tris (pH 8), 25% sucrose]. Cells were lysed at room temperature with 1 mg mL⁻¹ lysozyme for 10 min, followed by 10 min of incubation after addition of 1% SDS. Then, proteinase K (1 mg mL⁻¹) were added to the lysis suspensions and the tubes were incubated at 50 °C for 60 min. Nucleic acid extraction was performed with prewarmed (55 °C) phenol-chloroform-isoamyl alcohol mixture as previously described (Teske *et al.*, 2002). Nucleic acids were precipitated with NaCl 5 M (0.1 vol) and 100% ethanol (2 vol). After overnight incubation at -20 °C, the precipitated DNA was centrifuged (10 000 g for 30 min) and washed with 75% ethanol. Nucleic acids was dried and suspended in 50 µL of sterile water.

PCR amplifications and cloning

Small-subunit rRNA genes were amplified by PCR using universal primer 1492R (5'-CGGTTACCTTGTTACGACTT-3'), *Bacteria*-specific primer 8F (5'-AGAGTTTGATYMTGGCTCAG-3'), and *Archaea*-specific primer 4F (5'-TCCGGTTGATCCTGCCRG-3'). The final 25 µL PCR mixture consisted in 1 × buffer (Promega), 1.5 mM MgCl₂, 0.25 mM of each dNTP, 0.2 µM of each primer, 1 U of *Taq* DNA Polymerase (Promega) and 2 µL of DNA. Contaminants in PCR reaction mixtures were avoided as previously described (Goldenberger & Altwegg, 1995). PCR cycles were performed on a Perkin Elmer 9700 thermal cycler. After a denaturation step at 95 °C for 5 min, the following conditions were repeated for 30 cycles: denaturation at 95 °C for 1 min, annealing at

53 °C for 1 min, and chain extension at 72 °C for 2 min. A final extension step of 8 min at 72 °C was performed. Under these conditions, no PCR products were visible on 0.8% (w/v) agarose gels stained with ethidium bromide. To obtain visible PCR products, primary amplifications using primer pairs 8F-1492R and 4F-1492R were stopped at 15 cycles. Aliquots (1 µL) of the primary amplifications were then used as templates for secondary amplifications. Nested PCRs were performed using forward primers 50F (5'-AA CACATGCAAGTCGAACG-3') and 63F (5'-CAGGCCTAAC ACATGCAAGTC-3') specific for *Bacteria* and 341F (5'-CCT AYGGGGYGCASCAGGCG-3') specific for *Archaea* with universal reverse primer 1407R (5'-GACGGGCGGTGWGT RCAA-3'). The PCR mixtures and conditions were as described before except that 30 cycles were applied.

PCR detection of the ATP citrate lyase beta subunit (*acLB*) was based on previously designed primers (Campbell *et al.*, 2003). The cycling parameters for amplification of *acLB* genes were that described by Campbell & Cary (2004). Negative controls were performed without DNA template or PCR products.

PCR products were cloned using the TOPO XL PCR Cloning Kit (Invitrogen) into in *Escherichia coli* TOP 10 cells. Plasmids containing 16S rRNA genes were extracted using the QIAprep Miniprep kit (Qiagen).

Restriction fragment length polymorphism analysis, clone sequencing and phylogenetic analysis

Screening of clones containing 16S rRNA gene sequences was performed by restriction fragment length polymorphism (RFLP) analysis. The PCR products of the clone inserts amplified with primers M13F (5'-GTAAAACGACGGC CAG-3') and M13R (5'-CAGGAAACAGCTATGAC-3') were cut with the restriction endonuclease *Bst*U1 (New England Biolabs) according to manufacturer instructions. The DNA fragments were separated by electrophoresis on 2% agarose gels run in 1 × Tris-acetate buffer (Sambrook *et al.*, 1989).

One to three clones per RFLP pattern were sequenced with a 4200 IR² Licor DNA sequencer, using the Thermo Sequenase Primer Cycle sequencing kit (Amersham Pharmacia Biotech) and the infrared-labeled primers M13F and M13R (MWG, Germany). DNA sequences were submitted to the CHECK_CHIMERA program of the Ribosomal Database Project (RDP) (<http://rdp.cme.msu.edu/html/>). Sequences were subjected to BLAST searches within the GenBank database (<http://www.ncbi.nlm.nih.gov/>) to determine 16S rRNA gene sequence similarities to cultured and not yet cultured organisms. Sequence alignments were performed using the CLUSTALW program (Thompson *et al.*, 1994) and refined manually using the SEAVIEW program (Galtier *et al.*, 1996).

Phylogenetic trees were constructed using the PHYLO-WIN program (Galtier *et al.*, 1996). Distance trees were constructed using the neighbor-joining algorithms (Saitou & Nei, 1987) with the Kimura two-parameters correction (Kimura, 1980) for nucleotides and PAM correction for proteins. Maximum-likelihood analyses on protein sequences were performed using the PHYML program (Kimura, 1980). The robustness of inferred topologies was evaluated using a bootstrap analysis (Felsenstein, 1985). The sequences obtained in this study have been deposited in the GenBank/EMBL/DDBJ database with the following accession numbers: DQ071274–DQ071290 (16S rRNA gene) and AM181241–AM181291 (*acLB*).

16S rRNA oligonucleotide probe design and optimization

The 16S rRNA oligonucleotide probe EPSI682 [S-*-Epsi-0682-a-A-19 according to (Alm *et al.*, 1996)] specific to all Epsilonproteobacterial lineages identified in this study was designed using the PROBE_DESIGN function of the ARB software package (<http://www.arb-home.de>). The probe EPSI682 (5'-CGGATTTTACCCCTACACM-3') was checked for specificity using the PROBE_MATCH function in the RDP (<http://rdp.c.e.msu.edu>).

Dot-blot hybridization was performed to estimate the specificity of the designed probe. The probe was labeled at its 5' end with T4 polynucleotide kinase (phosphatase-free; Boehringer Mannheim Biochemicals, Meylan, France) and [γ -³²P] ATP (Amersham, Les Ulis, France).

For determination of probe specificity, PCR-amplified 16S rRNA gene fragments from *Alvinella pompejana* epibiont clone APG13B (Haddad *et al.*, 1995), and environmental clones VC1.2 (Corre *et al.*, 2001) and VC2.2 (Corre, 2000) were used as targets. Total nucleic acids from *Sulfur-ospirillum barnesii* (DSM 10660^T), *Wolinella succinogenes* (DSM 1740^T), *Arcobacter nitrofragilis* (DSM 7299^T), *Campylobacter coli* (DSM 4689^T), *C. jejuni* (DSM 4688^T), *Thiomicrospira denitrificans* (DSM 1251^T), *Helicobacter pylori* (ATCC 43579), *Zobellia galactanivorans* (DSM 12802^T), *E. coli* (DSM 3950) and *Bacillus sp.* strain SG9 (Marteinsson *et al.*, 1996) were also used.

The probe specificity was determined as previously described (Harmsen *et al.*, 1997b) with slight modifications. Denatured PCR-amplified 16S rRNA gene fragments and total nucleic acids from cultured strains were applied on nylon membrane (Hybond-N⁺, Amersham Biosciences, Orsay, France) using a dot-blot unit (SRC 96D, MinifoldII, Schleicher and Schuell, Dassel, Germany). The dissociation temperature (T_d) was determined for the [γ -³²P]ATP-labeled probe and hybridization was carried out as described by (Franks *et al.*, 1998). In addition to the newly designed probe, the 16S rRNA-targeting oligonucleotide probe

EUB338 (S-D-Bact-0338-a-A-18) specific for bacteria (Amann *et al.*, 1990) was used as control in the dot-blot analysis. Hybridization responses of the radioactively labeled probe were detected with a Phosphor Imager (Molecular Dynamics, Sunnyvale) and printed with Adobe (Seattle, USA) Photoshop 3.0.

Campylobacter jejuni (DSM 4688^T), *Arcobacter skirrowii* (DSM 7302^T), *Nautilia lithotrophica* (DSM 13520^T) and *Desulfurobacterium thermolithotrophum* (DSM 11699^T) were used for specificity studies in whole-cell hybridization experiments with the newly designed oligonucleotide probe ARC94 which targets the genus *Arcobacter* (Snaird *et al.*, 1997) and the bacterial probe EUB338. Fluorescein-, Cy-3- and rhodamine-labeled and unlabeled oligonucleotides were synthesized and purified by Eurogentec (Seraing, Belgium). Cells were fixed for 3 h in 4% paraformaldehyde in 1 × PBS (3 × PBS for marine strains) at 4 °C, washed in 1 × (or 3 ×) PBS and stored at -20 °C in 50% (v/v) ethanol-1 × PBS. FISH analysis was performed as previously described (Harmsen *et al.*, 1997b). The optimal formamide concentration of probe EPSI682 was evaluated by varying the formamide concentration from 0% to 60%.

Fluorescence *in situ* hybridization on the hydrothermal mat sample

To quantify the amount of cells hybridizing with probes EPSI682, ARC94, EUB338 and ARCH945, 250 µL of mat sample fixed in 3 × PBS containing 4% paraformaldehyde were centrifuged 6 min at 3500 g. The pellet was resuspended in 50 µL of the hybridization buffer used above with 250 ng of each labeled probe used and incubated at 46 °C for 3 h. Cells were then washed for 30 min with 5 mL of prewarmed (48 °C) wash-buffer before being filtered onto 0.2-µm-pore-size Nucleopore polycarbonate filters (Costar, Dutscher, France). The cells were viewed and counted with an Olympus BX60 epifluorescence microscope equipped with a U-MWIB (for fluorescein) and U-MWG (for rhodamine) filter cubes (Olympus). A minimum of 25 cells per field, 15 fields per filter and a minimum of three filters per probe tested were counted.

Results

Habitat characterization of the mat-forming organisms

In order to better understand microbial diversity, dynamics and colonization processes at vents, an *in situ* collector was deployed on the chimney PP12 at the Genesis site of the 13°N vent field (EPR). The deployment site was selected based on the maximum temperature observed in the venting fluid (<100 °C). The collector was placed a few centimeters above the vent orifice within the zone of active mixing of

diffuse flow and seawater (Fig. 1a). The temperatures recorded in the shimmering vent fluid near the collector ranged between 11 and 45 °C, in the range of, or exceeding, temperatures recorded a few days earlier (Alain *et al.*, 2004). No temperature measurements were performed near the float, but given that the distance between the collector and the float was about 15 cm, we can reasonably assume that average temperatures at which the biofilm formed were below 20 °C. Chemical analyses had also previously shown that diffuse-flow fluids of this chimney contained high sulphide levels and negligible amounts of iron (Alain *et al.*, 2004). At the fourth day of deployment, a thick and dense white biofilm visible on video imagery was observed on the syntactic foam float attached to the collector and at the collector's base (data not shown). The collector was recovered after 5 days of incubation (Fig. 1b). The *in situ* collector had been colonized by several specimens of juvenile and adult *Alvinella* worms in their freshly secreted organic tubes. The white mat that remained on the marker was collected for further detailed analyses. SEM analysis revealed that the mat consisted in a dense network of cells (mainly rods) embedded within a mucous matrix (Supplementary Fig. 1a and b). Cells were interconnected by the matrix throughout the biofilm.

The elemental microanalysis of the mat was investigated by EDS and XRD. The SEM-EDS analysis indicated that the network of microorganisms was coated with elemental sulphur. XRD diagrams confirmed the dominance of elemental sulphur (with 2θ peaks at about 23.1°, 25.9°, 27.8°, 28.7°, 28.95° and 31.4°) (data not shown). No traces of sulphide minerals, such as pyrite, were detected.

Molecular characterization of the mat-associated microbial communities

The genetic diversity of the microbial assemblages associated with the mat sample was analyzed by constructing 16S rRNA gene. Regardless of the PCR conditions used, no archaeal 16S rRNA genes could be amplified from the DNA extractions. 16S rRNA gene amplification products were obtained using two different bacterial primer pairs. A total of 78 clones containing full-length inserts (59 clones from the library obtained using 50F-1407R and 19 clones from the library generated with primers 63F-1407R) were categorized by RFLP. The clones were then grouped into distinct clusters on the basis of their restriction pattern. Verification of the RFLP clone family groupings was obtained by comparing full 16S rRNA gene sequences of up to three representative clones from each family. Sequences that were ≥97% similar to each other were grouped as a sequence type (Table 1). None of the sequences were identified as chimeras from RDP analysis. Homologous coverage values (Singleton *et al.*, 2001) of the libraries generated using primers 50F-1407R

Table 1. Phylogenetic affiliations of bacterial type sequences detected in the hydrothermal mat sample

Sequence type*	Other representatives	Clone numbers in libraries		Closest cultivated strains (% identity)
		50F-1407R	63F-1407R	
L50-WB6	L50-WB9, L50-WB56, L63-WB2, L63-WB6	48	15	<i>Arcobacter</i> strain KT0913 (AF235110) (92%)
L50-WB53		1		<i>Arcobacter</i> strain KT0913 (AF235110) (93%)
L50-WB13	L63WB19	1	1	Epsilonproteobacterial strain 49MY (AB091293) (96%)
L50-WB17		1		Epsilonproteobacterial strain 49MY (AB091293) (94%)
L50-WB20	L50-WB49	5		Epsilonproteobacterial strain 49MY (AB091293) (95%)
L50-WB2		1		<i>Nitratifactor salsuginis</i> (AB175500) (93%)
L63-WB1	L63-WB8, L63-WB16		3	<i>Sulfurovum lithotrophicum</i> (AB091292) (96%)
L50-WB11	L50-WB7	2		<i>Sulfurimonas autotrophica</i> (AB088431) (98%)

*Sequence types were identified by RFLP analysis and phylogenetic analysis. All sequence types had generally different RFLP profiles. However, clones L50-WB20 and L63-WB1 displayed the same RFLP profile and were categorized in distinct sequence types on the basis of the phylogenetic analysis. Clones L50-WB6, L50-WB9 and L50-WB56 that had different RFLP types were categorized in one sequence type as their sequence similarity was $\geq 97\%$. For the same reason, L50-WB20 and L50-WB49 were assigned to a unique sequence type.

($C_{50F} = 93.2\%$) and 63F-1407R ($C_{60F} = 94.7\%$) indicated that a significant part of their 16S rRNA gene diversity was assessed.

All clone sequence types were affiliated to the *Epsilonproteobacteria* (Fig. 2; Table 1). A majority of the clones ($\sim 80\%$) had only 92% sequence identity with an uncharacterized *Arcobacter* strain isolated from marine surface water (Eilers *et al.*, 2000) and was more closely related (95% similarity) with an environmental clone sequence (AY225610) recovered from a Mid-Atlantic ridge hydrothermal system (Lopez-Garcia *et al.*, 2003). Five sequence types clustered within the group F defined by Corre *et al.* (2001) (Fig. 2). Three of them (L50-WB13, L50-WB17, and L50-WB20) had clones obtained from the hydrothermal environment as closest relatives (97–99% similarity) (Lopez-Garcia *et al.*, 2002) (K.A. Kormas, M.K. Tivey, K. von Damm, and A. Teske, unpublished data). Their closest cultivated relative was a yet uncharacterized mesophilic microaerophile (strain 49MY) isolated from a yellow mat sample from the Okinawa Trough (Takai *et al.*, 2003). The closest relatives of the two other clone sequences from group F (L50-WB2 and L63-WB1) were an environmental clone sequence obtained from tubes of *A. pompejana* (M.A. Cambon-Bonavita, V. Riou, K. Alain, V. Cuff, F. Lesongeur, G. Barbier, and J. Querellou, unpublished data) and *Sulfurovum lithotrophicum*, respectively. The latter organism, recently isolated from a deep-sea hydrothermal sediment core, is a mesophilic chemolithoautotroph capable of sulphur and thiosulphate oxidation and nitrate reduction (Inagaki *et al.*, 2004).

The remaining sequence type (L50-WB11) clustered into the family *Helicobacteraceae* (Corre *et al.*, 2001) and had *Sulfurimonas autotrophica* as closest relative (98% similarity); this strain was recently isolated from hydrothermal vent sediment and is a microaerophilic chemolithoautotroph

capable of growth on sulphur, thiosulphate and sulphide (Inagaki *et al.*, 2003).

To get more insight into the ecophysiological characteristics of microbial components of the mat sample community, the presence and the diversity of *aclB* genes were evaluated. A total of 51 *aclB* clones were obtained and they showed high identity level between them ($> 95\%$ on protein level). All were affiliated to the *Epsilonproteobacteria* (Fig. 3) and clustered with sequences of *Nautilia* strain Am-H, *Lebetimonas acidophila*, and environmental clones from alvinellid tubes (Inagaki *et al.*, 2003) regardless the reconstruction methods used. Although Takai *et al.* (2005) recognized that phylogenies for the 16S rRNA and *aclB* genes are incongruent, no significant differences could be made between tree topologies when the 16S rRNA gene sequence dataset matched that of the available *aclB* sequences (cultured *Epsilonproteobacteria*) (Nye *et al.*, 2006). This suggests that the *aclB* sequences retrieved from the mat sample were not that of the *Arcobacter*-like populations.

Design and specificity of probe EPSI682

The designed probe EPSI682 (5'-CGGATTTTACCCCTACACM-3') matched perfectly all the sequences identified in the study (Fig. 2). The specificity of the probe encompassed the 16S rRNA gene sequences of environmental clones and cultured strains of families *Campylobacteraceae*, *Helicobacteraceae* and '*Hydrogenimonaceae*'. The probe EPSI682 also matched all the sequences affiliated to the group F defined by Corre *et al.* (2001), except that of the epibionts of hydrothermal invertebrates (one M: A mismatch at position 682; *E. coli* numbering). Members of the new family *Nautiliaceae* (Miroshnichenko *et al.*, 2004) and of group G defined by Takai *et al.* (2003) contained a T:G mismatch at position 693.

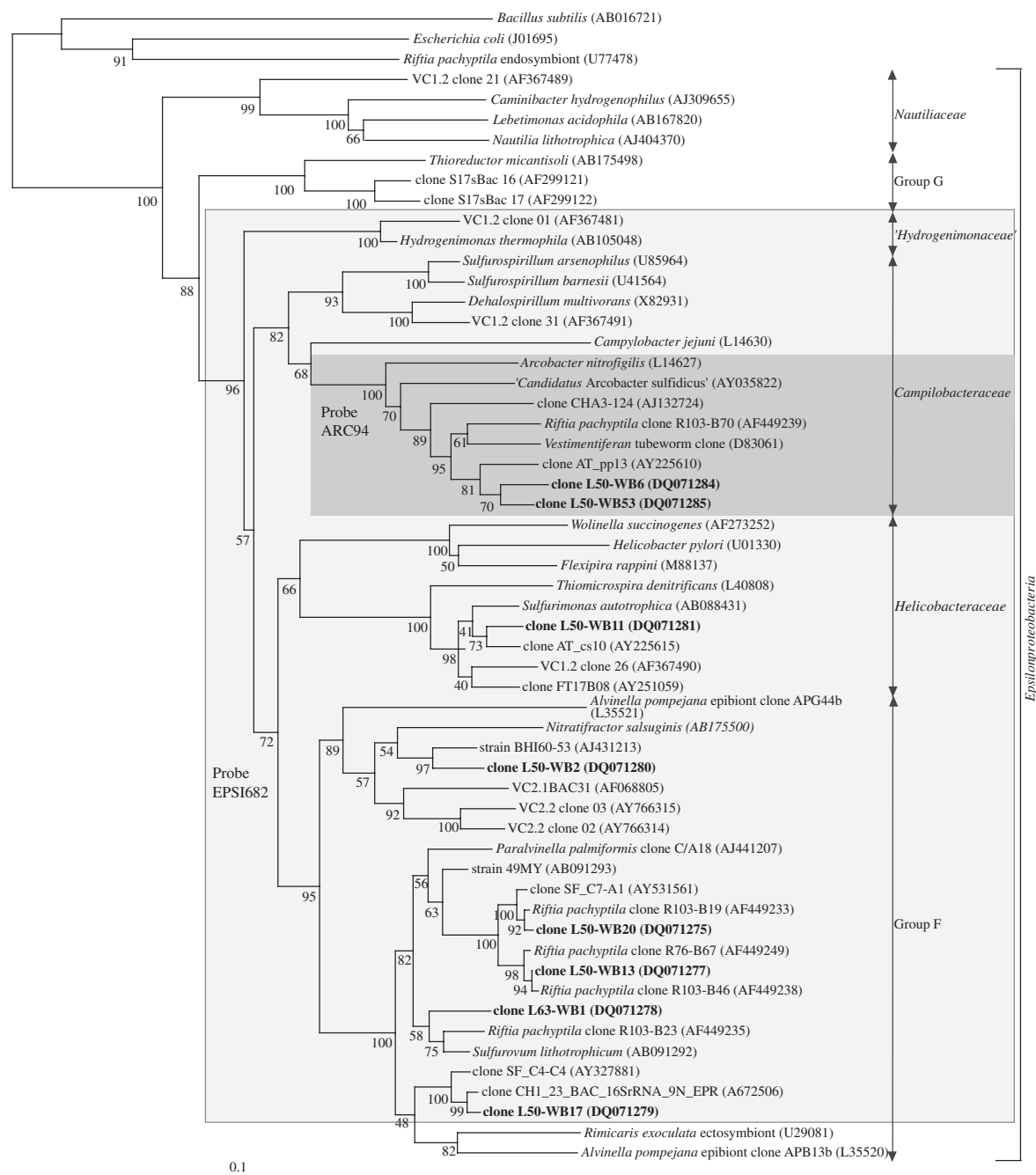


Fig. 2. Phylogenetic position of the epsilonproteobacterial sequences retrieved from the hydrothermal mat sample (in bold). The tree topology was obtained using the neighbour-joining method. Two 'Gammaproteobacteria', *Escherichia coli* and *Riftia pachyptila* endosymbiont, and *Bacillus subtilis* were used as outgroup. The reference 16S rRNA gene sequences were obtained from GenBank; all environmental clone sequences presented in the tree have been retrieved from hydrothermal vents. The numbers at the nodes represent the bootstrap values ($\geq 50\%$) obtained with 500 resamplings. Scale bar indicates the expected number of changes per sequence position. Environmental clone group F and group G are according to Corre *et al.* (2001) and Takai *et al.* (2003), respectively. The light grey and dark grey boxes encompass sequences targeted by probes EPSI682 and ARC94, respectively.

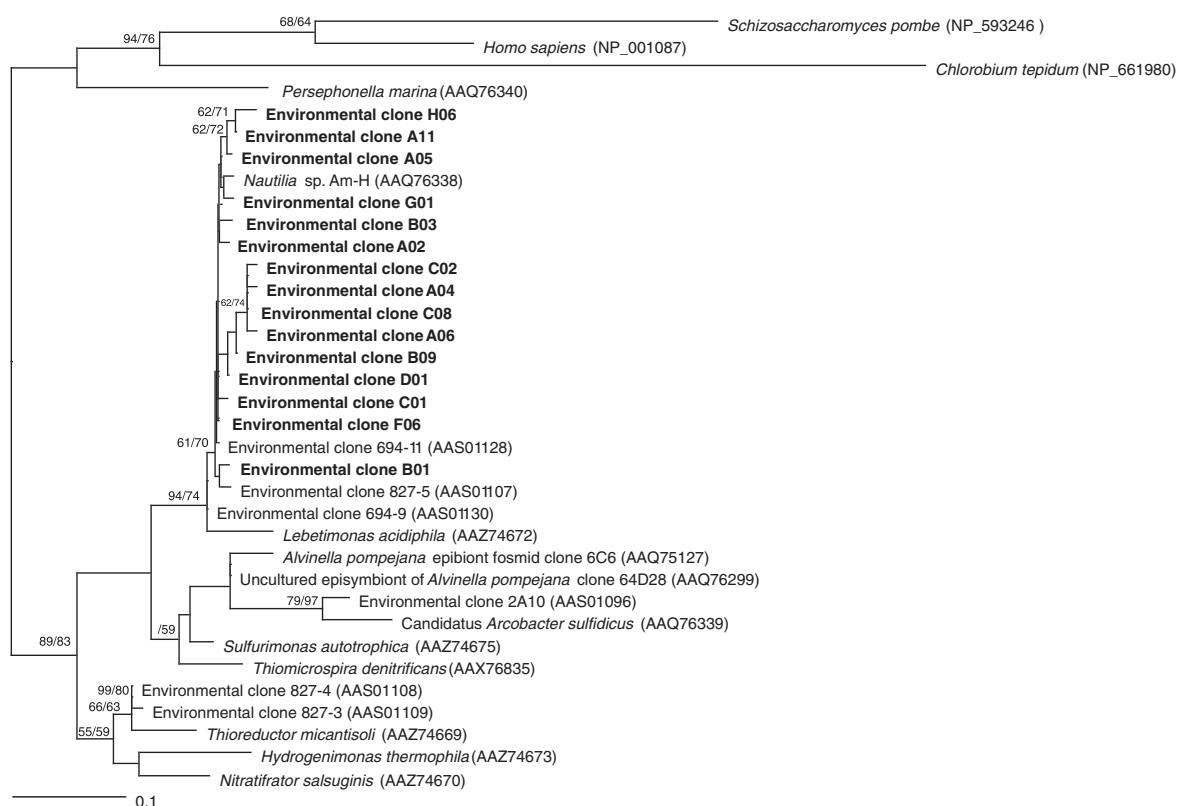


Fig. 3. Phylogenetic position of translated *alcB* clone sequences retrieved from the hydrothermal mat sample (in bold). Other *alcB* sequences were obtained from GenBank. The tree topology was obtained using the maximum likelihood method. The numbers at the nodes represent the bootstrap values ($\geq 50\%$) obtained respectively with 500 resamplings in neighbour-joining and 100 resampling in maximum likelihood method. Scale bar indicates the expected number of changes per amino acid position.

The specificity of radioactively labeled probe EPSI682 was also ensured by optimization of experimental hybridization conditions. The T_d value for probe EPSI682 experimentally measured at 46°C was used as the posthybridization washing temperature to test its specificity by dot-blot hybridization of nucleic acids from 27 target and nontarget species and environmental clones (Fig. 4). Confirming the *in silico* analysis, probe EPSI682 gave positive signals for environmental clones and strains belonging to families *Helicobacteraceae* and *Campylobacteraceae*. As expected, no hybridization signals were obtained for members of the family 'Nautiliaceae' (Fig. 4c, blots A4 and B4) and nontarget strains (Fig. 4c, line 9).

The specificity of the fluorescently labeled probe EPSI682 was tested by whole-cell hybridization of paraformaldehyde-fixed cells of reference strains. The formamide concentration was optimized using mixtures of target (*Arcobacter skirrowii* and *Campylobacter jejuni*) and nontarget strains (*Nautilia lithotrophica* and *Desulfurobacterium thermolithotrophum*). The hybridizations were performed simultaneously with either bacterial probe EUB338 or *Arcobacter*-specific probe

ARC94 (Snaird *et al.*, 1997). The probe ARC94 was also tested as a majority of the phylotypes retrieved in both libraries clustered within the *Arcobacter* group. The new designed probe showed the brightest fluorescence when hybridization buffers with 20% formamide were used. Only *C. jejuni* and *A. skirrowii* cells, specifically as target strains, had positive hybridization signals with probe EPSI682 (Fig. 5). The simultaneous hybridization with probes EPSI682 and ARC94 distinguished *Arcobacter* cells from that of *C. jejuni* (Fig. 5c).

Quantification of cells in the mat sample

The probes EPSI682 and ARC94 were used simultaneously with bacterial specific EUB338 or the archaeal specific probe ARCH915 to identify and to quantify the microorganisms in the mat sample. No signals were obtained when paraformaldehyde-fixed cells were hybridized with probe ARCH915. This result and the failure to amplify archaeal 16S rRNA genes from the mat DNA suggest strongly that *Archaea* were absent, or are present in amounts undetectable by the currently applied methods.

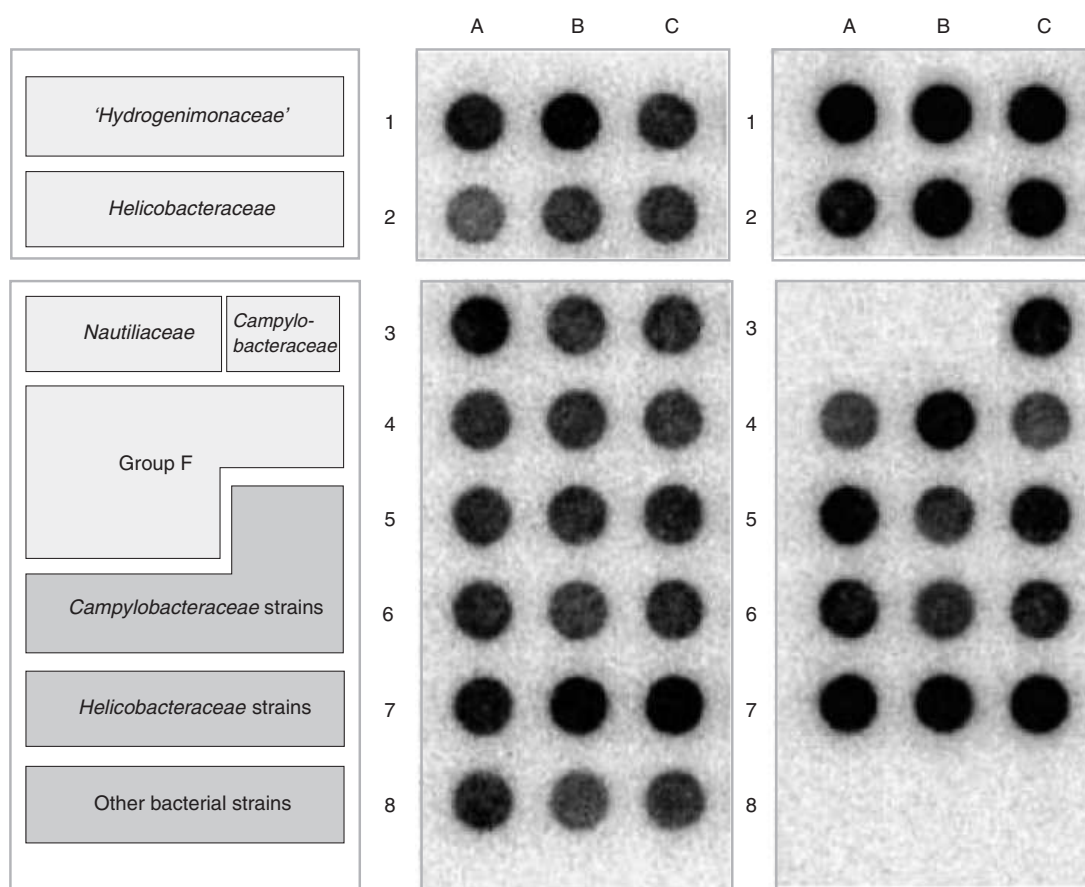


Fig. 4. Dot blot analysis of the probe specificities. Hybridization responses of radioactively labeled probes to nucleic acids from 27 target and nontarget DNAs (left) of environmental clones (light grey background) and strains (dark grey background). The blots were hybridized with the following probes: BACT338 (middle) and EPSI682 (right). The layout of nucleic acid samples on the blot is the following: 1A, VC1.2 clone 01; 1B, VC1.2 clone 10; 1C, VC1.2 clone 42; 2A, VC1.2 clone 26; 2B, VC1.2 clone 32; 2C, VC1.2 clone D; 3A, VC1.2 clone 21; 3B, VC1.2 clone 57; 3C, VC1.2 clone 31; 4A, VC1.2 clone 69; 4B, VC2.2 clone 01; 4C, VC2.2 clone 02; 5A, VC2.2 clone 03; 5B, *Alvinella* epibiont APB13B; 5C, *Campylobacter jejuni*; 6A, *Campylobacter coli*; 6B, *Geospirillum barnesii*; 6C, *Arcobacter nitrofigilis*; 7A, *Thiomicrospira denitrificans*; 7B, *Wolinella succinogenes*; 7C, *Helicobacter pylori*; 8A, *Escherichia coli*; 8B, *Zobellia galactanivorans*; 8C, *Bacillus* sp. strain SG9).

FISH experiments (Fig. 6a and b) revealed that ~95% of the bacterial cells reacted with probe EPSI682. Cells hybridizing with probe EPSI682 were morphologically diverse. Simultaneous hybridization with probes EPSI682 and ARC94 (Fig. 6c–e) showed that cells of the *Arcobacter* group represented around 60% of the cells detected by probe EPSI682. The cells that had positive hybridization signals with both probes had the same morphologies.

Discussion

One of the objectives of the oceanographic cruise PHARE was to identify the microbial communities involved in the first steps of colonization processes on virgin surfaces exposed to the deep-sea hydrothermal vent fluids at the 13°N (Le Bris *et al.*, 2002). Visible structures, macroscopi-

cally comparable to that observed at 9°N and Guaymas Basin hydrothermal sites (Taylor *et al.*, 1999) formed in a few days during our colonization experiments. Our goal in this study was to examine the chemical composition of this material and the diversity of the microbial community involved in its formation. The microscopic and chemical analyses of the sample revealed that it consisted in a network of microorganisms embedded in a mucous sulphur-rich matrix. Given that hydrothermal chemical precipitates from the 13°N hydrothermal site do not generally contain such amounts of elemental sulphur (Y. Fouquet, unpublished results), we hypothesized that the sulphur formation had a microbial origin.

Our goal in this study was to determine the diversity of the microbial community involved in the mat formation. We applied the full-cycle rRNA approach, including the

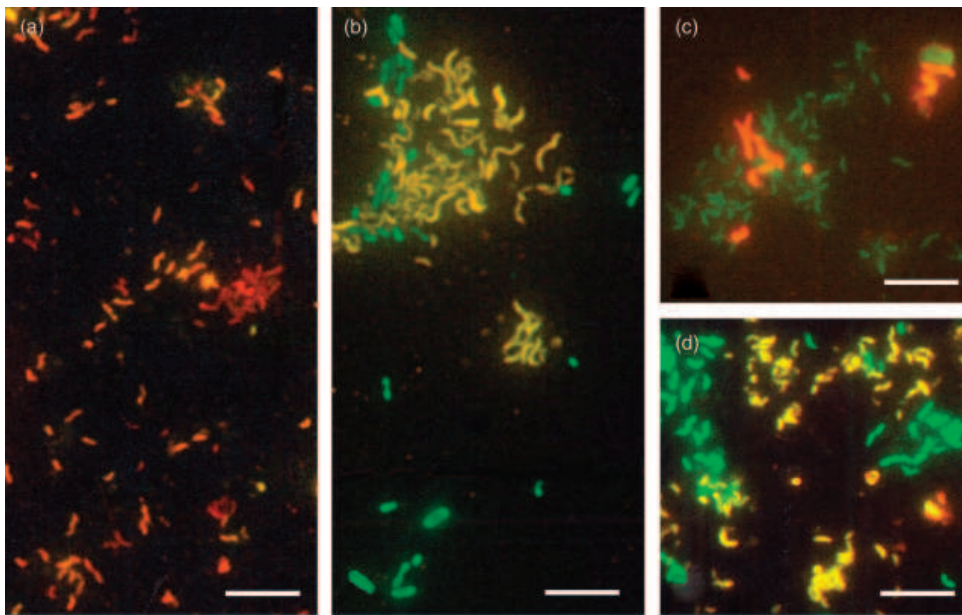


Fig. 5. Whole cell hybridizations of different mixtures of fixed cells hybridized simultaneously with fluorescein-labeled (green fluorescence) and a rhodamine- or Cy3-labeled probe (red fluorescence). Epifluorescence photomicrographs (double exposures) are shown at a magnification of $\times 1140$ (scale bar, 10 μm). Cells that hybridized with both types of probes appeared yellowish or orange on the double-exposed film. (a) Cells of *Campylobacter jejuni* and *Nautilia lithotrophica* simultaneously hybridized with probes EPSI682 (green fluorescence) and EUB338 (red fluorescence). (b) Cells of *Arcobacter skirrowii* and *Desulfurobacterium thermolithotrophum* simultaneously hybridized with probes EPSI682 (red fluorescence) and EUB338 (green fluorescence). (c) Cells of *A. skirrowii* and *C. jejuni* simultaneously hybridized with probes EPSI682 (green fluorescence) and ARC94 (red fluorescence). (d) Cells of *A. skirrowii* and *C. jejuni* simultaneously hybridized with probe EUB338 (green fluorescence) and ARC94 (red fluorescence).

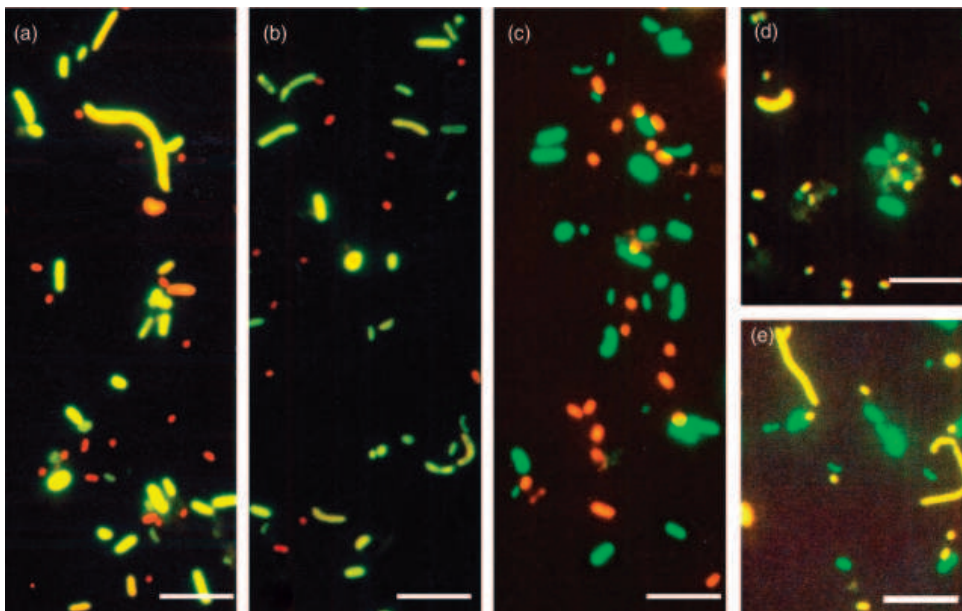


Fig. 6. Whole-cell hybridizations of cells from the mat collected on the 13°N EPR with probes EPSI682 (green fluorescence) and EUB338 (red fluorescence) (a–b), with probes EPSI682 (green fluorescence) and probe ARC94 (red fluorescence) (c) and with probes ARC94 (red fluorescence) and EUB338 (green fluorescence) (d and e). Epifluorescence photomicrographs (double exposures) are shown at a magnification of $\times 1,140$ (scale bar, 10 μm).

construction of 16S rRNA gene libraries and FISH experiments to identify the components of the mat community and to assess quantitatively the dominant microbial populations. The libraries were entirely composed of clones assigned to the *Epsilonproteobacteria*. These results were in line with many phylogenetic molecular surveys that showed that *Epsilonproteobacteria* can represent 40–98% of the clone libraries from hydrothermal environments sampled throughout the world's oceans (Haddad *et al.*, 1995; Moyer *et al.*, 1995; Polz & Cavanaugh, 1995; Cary *et al.*, 1997; Reysenbach *et al.*, 2000; Corre *et al.*, 2001; Longnecker & Reysenbach, 2001; Alain *et al.*, 2002, 2004; Lopez-Garcia *et al.*, 2002, 2003; Teske *et al.*, 2002; Takai *et al.*, 2003; Goffredi *et al.*, 2004). However, contrary to previous studies that showed that the *Epsilonproteobacteria* of groups F and B (family 'Hydrogenimonaceae') were the most frequently retrieved phylotypes (Corre *et al.*, 2001; Alain *et al.*, 2002, 2004; Lopez-Garcia *et al.*, 2002, 2003; Takai *et al.*, 2003; Goffredi *et al.*, 2004), the majority of our sequences grouped into a phylotype related to the genus *Arcobacter*. This finding was somewhat unusual as no *Arcobacter* isolates have been reported from deep-sea vents and *Arcobacter*-related phylotypes account generally for a minor fraction of the sequences obtained from this environment (Lopez-Garcia *et al.*, 2002, 2003; Huber *et al.*, 2003; Alain *et al.*, 2004).

The genus *Arcobacter* is unique among the *Epsilonproteobacteria* as it contains validly published species isolated from both animal and environmental sources, including marine habitats (McClung *et al.*, 1983; Vandamme *et al.*, 1992; Mansfield & Forsythe, 2000; Donachie *et al.*, 2005; Houf *et al.*, 2005). Furthermore, a number of potentially novel species have been described from environments such as oil fields, marine waters and sediments (Eilers *et al.*, 2000; Gevertz *et al.*, 2000; Kaeberlein *et al.*, 2002; Wirsen *et al.*, 2002). *Arcobacter* species are generally known as common human and animal pathogens or free-living organisms that respond poorly in conventional biochemical and growth tests (On, 1996). Typical features of characterized species are optimal growth under microaerophilic conditions at room temperature (Vandamme *et al.*, 1991). Given that the 16S rRNA gene sequence of the microorganisms dominant in the mat libraries shared ~92% similarity with that of validly published *Arcobacter* species, we can only speculate about their ecophysiology. However, because XRD analyses identified elemental sulphur as a main component of the mat, it is tempting to suggest that our dominant group might be sulphur-oxidizing bacteria. Consistent with this hypothesis, the ability to oxidize sulphur compounds and to use oxygen as electron acceptor is a common trait in various members of the *Epsilonproteobacteria* isolated from deep-sea hydrothermal vents (Inagaki *et al.*, 2003, 2004; Takai *et al.*, 2003, 2004; Nakagawa *et al.*, 2005). Although direct oxidation of hydrogen sulphide has been demonstrated for a restricted number

of *Epsilonproteobacteria* isolated from hydrothermal vents (Inagaki *et al.*, 2003; Takai *et al.*, 2003), it seems reasonable to suggest that the mat dominant organisms might also share this metabolic capacity as '*Candidatus Arcobacter sulfidicus*', which is the only described organism able to excrete sulphur filaments comparable to that of our mat, is a sulphide-oxidizer (Wirsen *et al.*, 2002).

Given that autotrophic CO₂ fixation through the rTCA cycle has been recently shown in *Epsilonproteobacteria*, the mat sample was analysed for the presence and diversity of the ATP citrate lyase gene. Although the recovered *aclB* environmental clones were also all affiliated to *Epsilonproteobacteria*, phylogenetic analyses showed that they did not correspond to that of *Arcobacter*-like populations. It could suggest that the main microbial components of the mat sample may use a different carbon fixation pathway. This hypothesis seems however unlikely as it has been demonstrated that the rTCA cycle operates in a significant number of representative strains from widespread phylogenetic subgroups and physiological types of *Epsilonproteobacteria* (Campbell *et al.*, 2003; Hügler *et al.*, 2005; Takai *et al.*, 2005). Since the primers used in this study were designed using a limited number of *aclB* sequences, it is also possible that a range of *aclB* sequence types have escaped molecular detection. These hypotheses need to be evaluated in the future to clarify the main microbial pathways used in the hydrothermal sulphur mat formations.

The ecological dominance of the *Epsilonproteobacteria* in the hydrothermal environment has been mostly hypothesized on the basis of their distribution in 16S rRNA gene libraries. However, for a variety of reasons (copy number, bias during DNA extraction, PCR, cloning), the frequency of 16S rRNA gene clone appearance does not reflect the actual microbial community structure (Farrelly *et al.*, 1995; von Wintzingerode *et al.*, 1997). Several studies have reported that biases in PCR amplification could just explain the overwhelming abundance of *Epsilonproteobacteria* (Corre, 2000; Watanabe *et al.*, 2002). Therefore, quantitative characterization was required to test if this dominance was also reflected in the makeup of the mat-associated populations *in situ*. Oligonucleotide probes specific to uncultured groups of *Epsilonproteobacteria* have been designed over recent years (Polz & Cavanaugh, 1995; Cary *et al.*, 1997; Snaird *et al.*, 1997; Longnecker & Reysenbach, 2001; Engel *et al.*, 2003). However, some 16S rRNA gene sequences identified in this study were not targeted by the existing probes, justifying the development of a probe with a wider coverage. Although probe EPSI682 demonstrated the actual omnipresence of *Epsilonproteobacteria*, it is likely that nontargeted members of this lineage (Fig. 2) may also have contributed to the bacterial cells detected with the probe EUB338.

It is interesting to notice that striking differences exist between our results and that of Alain *et al.* (2004) although

both colonization experiments took place on the same diffuse vent for the same time period. The very thin microbial veils that grew on the colonization device deployed by Alain *et al.* (2004) were macroscopically different from our thick sulphur mats. The absence of *Archaea*, or their very low representation, is one of the only common findings between both studies. This is probably due to the fact that biofilms had formed in environmental conditions (low temperatures, presence of oxygen) that are not compatible with the lifestyle of most known archaea (Alain *et al.*, 2004). Although *Epsilonproteobacteria* also prevailed in the clone libraries derived from the veils, the dominant phylotypes belonged to group F and no *Arcobacter*-related sequences were retrieved. Several indications suggest that the two biofilm communities had settled under distinct environmental parameters (such as temperature, oxygen and sulphide concentrations). Just before the deployment of our collector, the mechanical removing of the overlying sulphide structure resulted in an increase of the vent emission (Fig. 1). This explains why the fluid temperatures measured with the ROV thermocouple and probably the sulphide concentrations to which our collector was exposed were generally higher than that recorded earlier (14–860 μM) (Alain *et al.*, 2004). Consistent with this hypothesis, filamentous sulphur formation by '*Candidatus Arcobacter sulfidicus*' required high sulphide concentrations (> 400 μM) and its sulphide optimum was higher (1.5 mM) than those reported for other sulphur-oxidizers (Wirsen *et al.*, 2002). The capacity to grow optimally and selectively at high sulphide concentrations may also explain how these bacteria can locally out-compete other sulphur-oxidizers as we observed in our sample.

This report further expands the known diversity of the *Epsilonproteobacteria* from the initial reports of the prevalence of this group at deep-sea vents. Given that the ecological dominance of these microorganisms had been hypothesized on the only basis of their distribution in 16S rRNA gene libraries, we provide here a first demonstration of the numerical abundance of these microorganisms *in situ*. We showed that bacteria taxonomically different from '*Candidatus Arcobacter sulfidicus*' are probably involved in filamentous sulphur formation. This process can be ecologically significant in environments where opposing oxygen-sulphide gradients are available at the necessary concentrations and can contribute significantly to the overall organic matter production at hydrothermal vents. It would therefore be significant now to delineate the environmental parameters that govern this efficient sulphide oxidation and the extent of the diversity and distribution of the responsible organisms.

Acknowledgements

We thank Nadine Le Bris and Françoise Gaill, chief scientist and project leader of the Phare cruise for having invited us

on board. We thank the captain and crews of '*L'Atalante*' and the pilots and support crew of the ROV '*Victor*'. We are grateful to Philippe Crassous (Ifremer, Brest) and Monique Briand for their valuable contribution in SEM experiments and figure drawings, respectively. We also thank the two anonymous reviewers for their constructive criticisms of an earlier draft of the manuscript. We thank Stéphanie Gouriou (Faculté de Médecine, Brest), Tristan Barbeyron (Station Biologique, Roscoff), Cathy Ragimbeau (CNEVA, Ploufragan), Jan Kuever (Max Plank Institute, Bremen), Barbara Campbell (University of Delaware, USA) for having provided us in cells or DNA of reference strains. HM was supported by a grant from the Ministère de la Recherche.

References

- Alain K, Olagnon M, Desbruyères D, Pagé A, Barbier G, Juniper SK, Quérelou J & Cambon-Bonavita MA (2002) Phylogenetic characterization of the bacterial assemblage associated with mucous secretions of the hydrothermal vent polychaete *Paralvinella palmiformis*. *FEMS Microbiol Ecol* **42**: 463–476.
- Alain K, Zbinden M, Le Bris N, Lesongeur F, Querellou J, Gaill F & Cambon-Bonavita MA (2004) Early steps in microbial colonization processes at deep-sea hydrothermal vents. *Environ Microbiol* **6**: 227–241.
- Alm EW, Oerther DB, Larsen N, Stahl DA & Raskin L (1996) The oligonucleotide probe database. *Appl Environ Microbiol* **62**: 3557–3559.
- Amann RI, Binder BJ, Olson RJ, Chisholm SW, Devereux R & Stahl DA (1990) Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. *Appl Environ Microbiol* **56**: 1919–1925.
- Campbell BJ, Stein JL & Cary SC (2003) Evidence of chemolithoautotrophy in the bacterial community associated with *Alvinella pompejana*, a hydrothermal vent polychaete. *Appl Environ Microbiol* **69**: 5070–5078.
- Campbell BJ & Cary SC (2004) Abundance of reverse tricarboxylic acid cycle genes in free-living microorganisms at deep-sea hydrothermal vents. *Appl Environ Microbiol* **70**: 6282–6289.
- Cary SC, Cottrell MT, Stein JL, Camacho F & Desbruyères D (1997) Molecular identification and localization of filamentous symbiotic bacteria associated with the hydrothermal vent annelid *Alvinella pompejana*. *Appl Environ Microbiol* **63**: 1124–1130.
- Corre E (2000) Approches moléculaires de la diversité microbienne de deux environnements extrêmes: les sources hydrothermales profondes et les réservoirs pétroliers. PhD Thesis, Université de Bretagne Occidentale.
- Corre E, Reysenbach AL & Prieur D (2001) ϵ -Proteobacterial diversity from a deep-sea hydrothermal vent on the mid-Atlantic ridge. *FEMS Microbiol Lett* **205**: 329–335.

- Donachie SP, Bowman JP, On SL & Alam M (2005) *Arcobacter halophilus* sp. nov., the first obligate halophile in the genus *Arcobacter*. *Int J Syst Evol Microbiol* **55**: 1271–1277.
- Eilers H, Pernthaler J, Gloeckner FO & Amann R (2000) Culturability and in situ abundance of pelagic bacteria from the North Sea. *Appl Environ Microbiol* **66**: 3044–3051.
- Embley RW Jr, Chadwick WW, Jonasson IR, Butterfield DA & Baker ET (1995) Initial results of the rapid response to the 1993 CoAxial event: relationships between hydrothermal and volcanic processes. *Geophys Res Lett* **22**: 143–146.
- Embley RW, Chadwick WW Jr, Perfit MR, Smith MC & Delaney JR (2000) Recent eruptions on the CoAxial segment of the Juan de Fuca ridge: implications for mid-ocean ridge accretion processes. *J Geophys Res* **105**: 16501–16526.
- Engel AS, Lee N, Porter ML, Stern LA, Bennett PC & Wagner M (2003) Filamentous '*Epsilonproteobacteria*' dominate microbial mats from sulfidic cave springs. *Appl Environ Microbiol* **69**: 5503–5511.
- Farrelly V, Rainey FA & Stackebrandt E (1995) Effect of genome size and rrn gene copy number on PCR amplification of 16S rRNA genes from a mixture of bacterial species. *Appl Environ Microbiol* **61**: 2798–2801.
- Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **30**: 783–791.
- Franks AH, Harmsen HJM, Raangs GC, Jansen GJ, Schut F & Welling GW (1998) Variations of bacterial populations in human feces measured by fluorescent in situ hybridization with group-specific 16S rRNA-targeted oligonucleotide probes. *Appl Environ Microbiol* **64**: 3336–3345.
- Galtier N, Gouy M & Gautier C (1996) SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny. *Comput Appl Biosci* **12**: 543–548.
- Gevertz D, Telang AJ, Voordouw G & Jenneman GE (2000) Isolation and characterization of strains CVO and FWKO B, two novel nitrate-reducing, sulfide-oxidizing bacteria isolated from oil field brine. *Appl Environ Microbiol* **66**: 2491–2501.
- Goffredi SK, Waren A, Orphan VJ, Van Dover CL & Vrijenhoek RC (2004) Novel forms of structural integration between microbes and a hydrothermal vent gastropod from the Indian ocean. *Appl Environ Microbiol* **70**: 3082–3090.
- Goldenberger D & Altwegg M (1995) Eubacterial PCR: contaminating DNA in primer preparations and its elimination by UV light. *J Microbiol Methods* **21**: 27–32.
- Haddad A, Camacho F, Durand P & Cary SC (1995) Phylogenetic characterization of the epibiotic bacteria associated with the hydrothermal vent polychaete *Alvinella pompejana*. *Appl Environ Microbiol* **61**: 1679–1687.
- Harmsen H, Prieur D & Jeannot C (1997a) Distribution of microorganisms in deep-sea hydrothermal vent chimneys investigated by whole-cell hybridization and enrichment culture of thermophilic subpopulations. *Appl Environ Microbiol* **63**: 2876–2883.
- Harmsen HJM, Prieur D & Jeannot C (1997b) Group-specific 16S rRNA-targeted oligonucleotide probes to identify thermophilic bacteria in marine hydrothermal vents. *Appl Environ Microbiol* **63**: 4061–4068.
- Haymon RM, Fornari DJ, Von Damm KL *et al.* (1993) Volcanic eruption of the mid-ocean ridge along the East Pacific Rise crest at 9 degree 45-52'N: direct submersible observations of seafloor phenomena associated with an eruption event in April, 1991. *Earth Planet Sci Lett* **119**: 85–101.
- Hedrick DB, Pledger RD, White DC & Baross JA (1992) In situ microbial ecology of hydrothermal vent sediments. *FEMS Microbiol Lett* **101**: 1–10.
- Hekinian R & Fouquet Y (1985) Volcanism and metallogenesis of axial and off-axial structures on the East Pacific Rise near 13°N. *Econ Geol* **80**: 221–249.
- Houf K, On SL, Coenye T, Mast J, Van Hoof J & Vandamme P (2005) *Arcobacter cibarius* sp. nov., isolated from broiler carcasses. *Int J Syst Evol Microbiol* **55**: 713–717.
- Huber JA, Butterfield DA & Baross JA (2002) Temporal changes in archaeal diversity and chemistry in a Mid-Ocean Ridge seafloor habitat. *Appl Environ Microbiol* **68**: 1585–1594.
- Huber JA, Butterfield DA & Baross JA (2003) Bacterial diversity in a seafloor habitat following a deep-sea volcanic eruption. *FEMS Microbiol Ecol* **43**: 393–409.
- Hugler M, Wirsén CO, Fuchs G, Taylor CD & Sievert SM (2005) Evidence for autotrophic CO₂ fixation via the reductive tricarboxylic acid cycle by members of the ϵ subdivision of *proteobacteria*. *J Bacteriol* **187**: 3020–3027.
- Inagaki F, Takai K, Kobayashi H, Neelson KH & Horikoshi K (2003) *Sulfurimonas autotrophica* gen. nov., sp. nov., a novel sulfur-oxidizing ϵ -proteobacterium isolated from hydrothermal sediments in the Mid-Okinawa trough. *Int J Syst Evol Microbiol* **53**: 1801–1805.
- Inagaki F, Takai K, Neelson KH & Horikoshi K (2004) *Sulfurovum lithotrophicum* gen. nov., sp. nov., a novel sulfur-oxidizing chemolithoautotroph within the epsilon-Proteobacteria isolated from Okinawa Trough hydrothermal sediments. *Int J Syst Evol Microbiol* **54**: 1477–1482.
- Jannasch HW & Wirsén CO (1981) Morphological survey of microbial mats near deep-sea thermal vents. *Appl Environ Microbiol* **41**: 528–538.
- Jannasch HW, Nelson KE & Wirsén CO (1989) Massive natural occurrence of unusually large bacteria (*Beggiatoa* sp.) at a hydrothermal deep-sea vent. *Nature* **342**: 834–836.
- Jeannot C (2000) Molecular ecology of hydrothermal vent microbial communities. *Antonie Van Leeuwenhoek* **77**: 117–133.
- Jones WJ, Leigh JA, Mayer F, Woese CR & Wolfe RS (1983) *Methanococcus jannaschii*, sp. nov., an extremely thermophilic methanogen from a submarine hydrothermal vent. *Arch Microbiol* **136**: 254–261.
- Kaerberlein T, Lewis K & Epstein SS (2002) Isolating 'uncultivable' microorganisms in pure culture in a simulated natural environment. *Science* **296**: 1127–1129.
- Karl DM (1995) Ecology of free-living, hydrothermal vent microbial communities. *The Microbiology of Deep-Sea*

- Hydrothermal Vents* (Karl DM, ed), pp. 35–124. CRC Press Inc, Boca Raton.
- Kimura M (1980) A simple model for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* **16**: 111–120.
- Le Bris N, Gaill F, Alain K *et al.* (2002) Biogeochemical interactions in extreme environmental conditions: an integrated study (PHARE 2002/EPR 13°N). *InterRidge News* **11**: 20–23.
- Longnecker K & Reysenbach AL (2001) Expansion of the geographic distribution of a novel lineage of ϵ -*Proteobacteria* to a hydrothermal vent site on the Southern East Pacific Rise. *FEMS Microbiol Ecol* **35**: 287–293.
- Lopez-Garcia P, Gaill F & Moreira D (2002) Wide bacterial diversity associated with tubes of the vent worm *Riftia pachyptila*. *Environ Microbiol* **4**: 204–215.
- Lopez-Garcia P, Duperron S, Philippot P, Foriel J, Susini J & Moreira D (2003) Bacterial diversity in hydrothermal sediment and epsilonproteobacterial dominance in experimental microcolonizers at the Mid-Atlantic Ridge. *Environ Microbiol* **5**: 961–976.
- Mansfield LP & Forsythe SJ (2000) *Arcobacter butzleri*, *A. skirrowii* and *A. cryaerophilus* -potential emerging human pathogens. *Rev Med Microbiol* **11**: 161–170.
- Marteinsson VT, Birrien J-L, Jeanthon C & Prieur D (1996) Numerical taxonomic study of thermophilic *Bacillus* isolated from three geographically separated deep-sea hydrothermal. *FEMS Microbiol Ecol* **21**: 255–266.
- McClung CR, Patriquin DG & Davis RE (1983) *Campylobacter nitrofragilis* sp. nov., a nitrogen-fixing bacterium associated with roots of *Spartina alterniflora*. *Int J Syst Microbiol* **133**: 605–612.
- Miroshnichenko ML, L'Haridon S, Schumann P, Spring S, Bonch-Osmolovskaya EA, Jeanthon C & Stackebrandt E (2004) *Caminiobacter profundus* sp. nov., a novel thermophile of *Nautiliales* ord. nov. within the class '*Epsilonproteobacteria*', isolated from a deep-sea hydrothermal vent. *Int J Syst Evol Microbiol* **54**: 41–45.
- Moyer CL, Dobbs FC & Karl DM (1995) Phylogenetic diversity of the bacterial community from a microbial mat at an active, hydrothermal vent system, Loihi seamount, Hawaii. *Appl Environ Microbiol* **61**: 1555–1562.
- Nakagawa S, Takai K, Inagaki F, Horikoshi K & Sako Y (2005) *Nitratiruptor tergaricus* gen. nov., sp. nov. and *Nitratifactor salsuginis* gen. nov., sp. nov., nitrate-reducing chemolithoautotrophs of the ϵ -*Proteobacteria* isolated from a deep-sea hydrothermal system in the Mid-Okinawa Trough. *Int J Syst Evol Microbiol* **55**: 925–933.
- Nelson DC, Wirsén CO & Jannasch HW (1989) Characterization of large, autotrophic *Beggiatoa* spp. abundant at hydrothermal vents of the Guaymas basin. *Appl Environ Microbiol* **55**: 2909–2917.
- Nelson D, Haymon RM, Lilley M & Lutz R (1991) Rapid growth of unusual hydrothermal bacteria observed at new vents during ADVENTURE dive program to the EPR crest at 9°45'–52'N. *EOS Trans Am Geophys Union* **72**: 481.
- Nercessian O, Reysenbach AL, Prieur D & Jeanthon C (2003) Archaeal diversity associated with in situ samplers deployed on hydrothermal vents on the East Pacific Rise (13 degrees N). *Environ Microbiol* **5**: 492–502.
- Nye TMW, Lio P & Gilks WR (2006) A novel algorithm and web-based tool for comparing two alternative phylogenetic trees. *Bioinformatics* **22**: 117–119.
- On SLW (1996) Identification methods for campylobacters, helicobacters, and related organisms. *Clin Microbiol Rev* **9**: 405–422.
- Polz MF & Cavanaugh CM (1995) Dominance of one bacterial phylotype at a Mid-Atlantic Ridge hydrothermal vent site. *Proc Natl Acad Sci USA* **92**: 7232–7236.
- Reysenbach AL, Longnecker K & Kirshtein J (2000) Novel bacterial and archaeal lineages from an in situ growth chamber deployed at a mid-Atlantic ridge hydrothermal vent. *Appl Environ Microbiol* **66**: 3798–3806.
- Saitou N & Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**: 406–425.
- Sambrook J, Fritsch EF & Maniatis T (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, New York.
- Shank TM, Fornari DJ, Von Damm KL, Lilley MD, Haymon RM & Lutz RA (1998) Temporal and spatial patterns of biological community development at nascent deep-sea hydrothermal vents (9°50'N, East Pacific Rise). *Deep-Sea Res II* **45**: 465–515.
- Singleton DR, Furlong MA, Rathbun SL & Whitman WB (2001) Quantitative comparisons of 16S rRNA gene sequence libraries from environmental samples. *Appl Environ Microbiol* **67**: 4374–4376.
- Snaidr J, Amann R, Huber I, Ludwig W & Schleifer KH (1997) Phylogenetic analysis and in situ identification of bacteria in activated sludge. *Appl Environ Microbiol* **63**: 2884–2896.
- Takai K & Horikoshi K (1999) Genetic diversity of archaea in deep-sea hydrothermal vent environments. *Genetics* **152**: 1285–1297.
- Takai K, Komatsu T, Inagaki F & Horikoshi K (2001) Distribution of *Archaea* in a black smoker chimney structure. *Appl Environ Microbiol* **67**: 3618–3629.
- Takai K, Inagaki F, Nakagawa S, Hirayama H, Nunoura T, Sako Y, Neelson KH & Horikoshi K (2003) Isolation and phylogenetic diversity of members of previously uncultivated ϵ -*Proteobacteria* in deep-sea hydrothermal fields. *FEMS Microbiol Lett* **218**: 167–174.
- Takai K, Neelson KH & Horikoshi K (2004) *Hydrogenimonas thermophila* gen. nov., sp. nov., a novel thermophilic, hydrogen-oxidizing chemolithoautotroph within the ϵ -*Proteobacteria*, isolated from a black smoker in a Central Indian Ridge hydrothermal field. *Int J Syst Evol Microbiol* **54**: 25–32.
- Takai K, Campbell BJ, Cary SC *et al.* (2005) Enzymatic and genetic characterization of carbon and energy metabolisms by

- deep-sea hydrothermal chemolithoautotrophic isolates of *ε-proteobacteria*. *Appl Environ Microbiol* **71**: 7310–7320.
- Taylor CD & Wirsén CO (1997) Microbiology and ecology of filamentous sulfur formation. *Science* **277**: 1483–1485.
- Taylor CD, Wirsén CO & Gail F (1999) Rapid microbial production of filamentous sulfur mats at hydrothermal vents. *Appl Environ Microbiol* **65**: 2253–2255.
- Teske A, Hinrichs KU, Edgcomb V, de Vera Gomez A, Kysela D, Sylva SP, Sogin ML & Jannasch HW (2002) Microbial diversity of hydrothermal sediments in the Guaymas basin: evidence for anaerobic methanotrophic communities. *Appl Environ Microbiol* **68**: 1994–2007.
- Thompson JD, Higgins DG & Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**: 4673–4680.
- Vandamme P, Falsen E, Rossau R, Hoste B, Segers P, Tytgat R & De Ley J (1991) Revision of *Campylobacter*, *Helicobacter*, and *Wolinella* taxonomy: Emendation of generic descriptions and proposal of *Arcobacter* gen. nov. *Int J Syst Bacteriol* **41**: 88–103.
- Vandamme P, Vancanneyt M, Pot B *et al.* (1992) Polyphasic taxonomic study of the emended genus *Arcobacter* with *Arcobacter butzleri* comb. nov. and *Arcobacter skirrowii* sp. nov., an aerotolerant bacterium isolated from veterinary specimens. *Int J Syst Bacteriol* **42**: 344–356.
- von Wintzingerode F, Gobel UB & Stackebrandt E (1997) Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS Microbiol Rev* **21**: 213–229.
- Watanabe K, Kodama Y & Kaku N (2002) Diversity and abundance of bacteria in an underground oil-storage cavity. *BMC Microbiol* **2**: 23.
- Wirsén CO, Sievert SM, Cavanaugh CM, Molyneux SJ, Ahmad A, Taylor LT, DeLong EF & Taylor CD (2002) Characterization of an autotrophic sulfide-oxidizing marine *Arcobacter* sp. that produces filamentous sulfur. *Appl Environ Microbiol* **68**: 316–325.

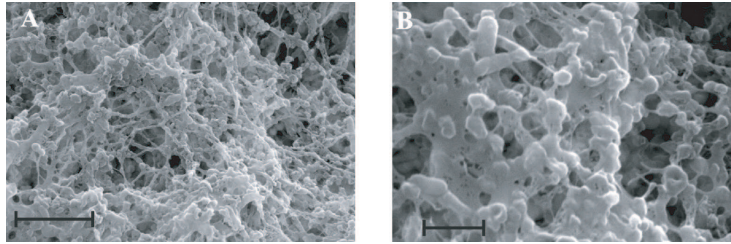
Supplementary material

The following supplementary material is available for this article:

Figure S1. Scanning electron microphotographs of the bacterial mat. (A) Micrograph showing the dense network of cells; bar: 5 μm; (B) Enlargement showing mostly rod cells embedded in a polymeric matrix, bar: 2 μm.

This material is available as part of the online article from: <http://www.blackwell-synergy.com/doi/abs/10.1111/j.1574-6941.2006.00192.x> (This link will take you to the article abstract).

Please note: Blackwell Publishing are not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.



Supplementary Fig. 1. Scanning electron microphotographs of the material mat.
(A) Micrograph showing the dense network of cells; bar: 5 μm ;
(B) Enlargement showing mostly rod cells embedded in a polymeric matrix, bar: 2 μm .

Quelle part de la communauté microbienne du biofilm utilise le cycle rTCA pour fixer le carbone ?

Avant que de récentes études ne soient réalisées sur des ε -*Proteobacteria* chimiolithoautotrophes, tout portait à croire que dans l'écosystème hydrothermal marin profond, la principale voie de fixation du carbone était celle du cycle de Calvin. En effet, des souches hydrothermales sulfoxydantes, autotrophes obligatoires ou facultatives, principalement affiliées aux γ -*Proteobacteria* avaient été isolées et caractérisées (*Thiomicrospira* spp., *Beggiatoa* spp., par ex.); et des γ -*Proteobacteria* sulfo-oxydantes avaient été mises en évidence dans des invertébrés hydrothermaux (Cavanaugh *et al.*, 1981 ; Felbeck *et al.*, 1981 ; Cavanaugh, 1983 ; Karl *et al.*, 1988 ; Nelson *et al.*, 1989 ; Ruby *et al.*, 1981). Dans tous les cas, l'utilisation du cycle de Calvin chez ces organismes, avait été expérimentalement validée ou bien supposée. Des mesures d'activité réalisées sur l'enzyme clé de cette voie, la RubisCO, avaient permis de montrer l'utilisation de cette voie chez quelques uns de ces isolats ainsi que sur des échantillons environnementaux hydrothermaux (Wirsén *et al.*, 1993).

Alors, que tout laissait présager que la chaîne trophique reposait sur le cycle de Calvin, dans le milieu hydrothermal marin profond, il a été récemment montré que les ε -*Proteobacteria* chimiolithoautotrophes utilisaient la voie du cycle inverse de l'acide tricarboxylique (rTCA) pour fixer leur carbone (Campbell *et al.*, 2003 ; Campbell & Cary, 2004 ; Hugler *et al.*, 2005 ; Takai *et al.*, 2005a). Avant ces études, le cycle rTCA n'avait été démontré que chez quelques micro-organismes appartenant à des espèces du genre *Chlorobium*, certains membres des δ -*Proteobacteria* (par ex., *Desulfobacter hydrogenophilus*), des *Aquificales* et des archaea de la famille des *Thermoproteaceae* (Evans *et al.*, 1966 ; Fuchs *et al.*, 1980 ; Hugler *et al.*, 2003). Des analyses d'expressions par RT-PCR des gènes clés du cycle rTCA, ainsi que l'abondance et la diversité phylogénétique de ces mêmes gènes ont permis de suggérer que les micro-organismes vivants sous forme libre utilisaient préférentiellement cette voie pour fixer le carbone dans le milieu hydrothermal, et que ces organismes étaient des ε -*Proteobacteria* (Campbell & Cary, 2004 ; Campbell *et al.*, 2006). Nous avons donc utilisé des marqueurs fonctionnels du cycle rTCA, pour savoir si les ε -*Proteobacteria* que nous avons détectées dans le biofilm, utilisaient elles aussi la voie rTCA pour fixer leur carbone.

Deux marqueurs du cycle rTCA ont été utilisés pour ce travail de thèse : le gène *aclB* codant pour l'ATP citrate lyase et le gène *oorA* codant pour la 2-oxoglutarate : ferredoxin oxidoreductase (également connue sous le nom de alpha-ketoglutarate synthase). Les résultats liés à l'utilisation du marqueur *aclB* ont été détaillés dans l'étude 1, ceux concernant le marqueur *oorA* sont détaillés dans cette partie "Résultats complémentaires".

Organisme ou source de l'ADN amplifié	Gène ciblé pour amplification par PCR		
	<i>aclB</i>	<i>oorA</i>	ADNr 16S
<i>Nautilia lithotrophica</i>	+	+	+
<i>Arcobacter skirowii</i>	-	-	+
<i>Campylobacter jejuni</i>	-	+	+
<i>Sulfurospirillum halorespirans</i>	+	+	+
<i>Sulfurimonas autotrophica</i>	+	+	+
Biofilm prélevé sur le colonisateur	+	-	+

Tableau 1 – Récapitulatif des résultats obtenus après amplification par PCR des gènes codant pour la sous-unité beta de l'ATP citrate lyase (gène *aclB*), la sous-unité alpha de la 2-oxoglutarate ferredoxine oxidoreductase (gène *oorA* et le gène codant pour les ARNr 16S, à partir de l'ADN extrait de cinq souches d' ϵ -*Proteobacteria* et de l'ADN extrait du biofilm. Le signe + désigne qu'une bande spécifique de la taille attendue a été obtenue. Le signe - désigne qu'aucune bande spécifique de la taille attendue n'a été obtenue.

L'ATP citrate lyase et la 2-oxoglutarate : ferredoxin oxidoreductase sont deux enzymes clés du cycle rTCA. A l'aide d'amorces spécifiques mises au point par Campbell et ses collaborateurs (2003), nous avons tenté d'amplifier les gènes de ces deux enzymes, à partir de l'ADN extrait du biofilm.

Des produits d'amplification ayant la taille attendue, ont été obtenus avec le couple d'amorces spécifiques du gène *aclB* (voir étude 1). L'analyse phylogénétique des séquences des gènes *aclB* amplifiés, a révélé qu'elles étaient affiliées à des ϵ -*Proteobacteria*. Cependant, aucune de ces séquences groupait avec celles des *Arcobacter*, groupe pourtant dominant dans le biofilm si l'on considère l'inventaire des séquences codant pour les ARNr 16S. Lorsque nous avons voulu amplifier les gènes *oorA* à partir de l'ADN extrait du biofilm, nous n'avons pas réussi à obtenir des produits d'amplification.

Nous avons donc voulu vérifier l'efficacité des amorces utilisées sur des ADN d' ϵ -*Proteobacteria* phylogénétiquement diverses. Les souches suivantes ont été sélectionnées : *Sulfurospirillum halorespirans*, *Arcobacter skirowii*, *Campylobacter jejuni*, *Nautilia lithotrophica* et *Sulfurimonas autotrophica*. Parallèlement à cela, nous avons également tenté d'amplifier les gènes codant pour les ARNr 16S de ces mêmes souches à l'aide d'amorces spécifiques des *Bacteria* (voir Tableau 1).

Comme attendu, les gènes codant pour ARNr 16S des différentes ϵ -*Proteobacteria* que nous avons sélectionnées, ont tous été amplifiés. Des produits d'amplification des gènes *aclB* et *oorA* ont été obtenus pour les souches *N. lithotrophica*, *S. halorespirans* et *S. autotrophica*. Cependant, l'amplification des gènes *aclB* et/ou *oorA* a posé des problèmes pour les souches *C. jejuni* et *A. skirowii*. En effet seul, seul le gène *oorA* a pu être amplifié pour la souche *C. jejuni*. Nous n'avons pas non plus réussi à amplifier les gènes *aclB* et *oorA* à partir de l'ADN extrait d'*A. skirowii*.

Les gènes *aclB* et *oorA* ont pourtant été mis en évidence et amplifiés chez la souche “*Candidatus Arcobacter sulfidicus*” (Hugler *et al.*, 2005).

Cette expérience a donc révélé que les amorces préalablement dessinées pour cibler les gènes *aclB* et *oorA* ne couvraient pas l’ensemble des ϵ -*Proteobacteria*. L’absence de gènes *aclB* pouvant être affiliés à des *Arcobacter* dans l’ADN extrait du biofilm ainsi que notre incapacité à amplifier des gènes *oorA* à partir de ce même ADN, n’indiquent donc pas forcément leur absence de l’échantillon. On peut penser que cela est dû aux amorces utilisées qui n’ont pas une spécificité suffisamment étendue pour couvrir l’ensemble des ϵ -*Proteobacteria*. Ces amorces ont d’ailleurs été dessinées à partir du faible nombre de séquences disponibles à l’époque (celles provenant de deux fragments génomiques affiliés à des ϵ -*Proteobacteria*, ainsi que celles de *Chlorobium limicola* et de quelques organismes eucaryotes). Il serait donc intéressant de refaire cette expérience en redessinant des amorces spécifiques des gènes *aclB* et *oorA* à l’aide des nouvelles séquences environnementales qui ont été récemment inventoriées (Takai *et al.*, 2005a).

Nous pourrions également essayer d'amplifier le gène codant pour l'enzyme clé du cycle de Calvin, c'est-à-dire la RubisCO, à partir de l'ADN extrait du biofilm. L'absence d'un produit d'amplification spécifique de ce gène permettrait de soutenir l'hypothèse selon laquelle les membres de la communauté microbienne du biofilm ont bien fixé leur carbone via le cycle rTCA et non pas via le cycle de Calvin.

CHARACTERIZATION OF A 40 KBP
UNCULTIVATED EURYARCHAEOTAL
FOSMID FROM A MICROBIAL COMMUNITY
ASSOCIATED TO THE HYDROTHERMAL
VENT POLYCHAETE ANNELID, *Alvinella*
pompejana

Characterization of a 40 kbp uncultivated euryarchaeotal fosmid from a microbial community associated to the hydrothermal vent polychaete annelid, *Alvinella pompejana*

Le développement des techniques moléculaires a permis de mettre en évidence une très grande diversité microbienne au sein des écosystèmes hydrothermaux marins profonds (Cary *et al.*, 1997 ; Harmsen *et al.*, 1997 ; Takai & Horikoshi, 1999 ; Corre *et al.*, 2001 ; Longnecker & Reysenbach, 2001 ; Slobodkin *et al.*, 2001 ; Takai *et al.*, 2001a ; Nercessian *et al.*, 2003 ; Schrenk *et al.*, 2003). Ces études basées sur l'amplification puis l'analyse phylogénétique de séquences d'ARNr 16S, ont également démontré l'existence, parfois prépondérante, de nouvelles lignées au sein des *Archaea* (Takai & Horikoshi, 1999 ; Takai *et al.*, 2001a ; Nercessian *et al.*, 2003). Certaines de ces lignées semblent être inféodées à l'écosystème hydrothermal profond et représentent des groupes d'organismes incultivés jusqu'alors totalement inconnus. La métagénomique représente une nouvelle approche prometteuse pour mieux caractériser ces micro-organismes incultivés.

Cette étude avait pour objet d'analyser la diversité spécifique d'une communauté microbienne associée à des tubes d'*A. pompejana*. Ces tubes ancrés aux parois de cheminées hydrothermales actives sont exposés aux fluides hydrothermaux. Ils forment un substrat idéal pour la colonisation de populations microbiennes thermophiles. La diversité de la communauté microbienne a été analysée dans un premier temps par l'amplification et le séquençage des ADNr 16S. Cette analyse a révélé que la population archéenne était dominée par des *Euryarchaeota* incultivées. Dans un deuxième temps, nous avons choisi d'analyser cette communauté microbienne par une approche métagénomique, dans l'espoir d'obtenir des fragments génomiques de ces *Euryarchaeota* incultivées.

A ce jour, les seules banques métagénomiques construites à partir d'échantillons hydrothermaux, visaient à étudier les associations symbiotiques obligatoires entre des bactéries et des métazoaires (Hughes *et al.*, 1997 ; Robinson *et al.*, 1998 ; Millikan *et al.*, 1999 ; Campbell *et al.*, 2003). Cette étude métagénomique est la première du genre à explorer un groupe d'*Archaea* incultivées de l'environnement hydrothermal marin profond.



Uncultured *Archaea* in a hydrothermal microbial assemblage: phylogenetic diversity and characterization of a genome fragment from a euryarchaeote

Hélène Moussard¹, David Moreira², Marie-Anne Cambon-Bonavita¹, Purificación López-García² & Christian Jeanthon¹

¹Laboratoire de Microbiologie des Environnements Extrêmes, UMR 6197, Centre National de la Recherche Scientifique, IFREMER, Université de Bretagne Occidentale, Institut Universitaire Européen de la Mer, Plouzané, France and ²Equipe Diversité et Evolution Microbiennes, Unité d'Ecologie, Systématique & Evolution, UMR 8079, Université Paris-Sud, Orsay, France

Correspondence: Christian Jeanthon, UMR 7144, Equipe Phytoplancton Océanique, Station Biologique, Place Georges Teissier, 29680 Roscoff Cedex, France. Tel.: +33 298 292 379; fax +33 298 292 324; e-mail: jeanthon@sb-roscoff.fr

Received 6 December 2005; revised 30 January 2006; accepted 31 January 2006.

doi:10.1111/j.1574-6941.2006.00128.x

Editor: Julian Marchesi

Keywords

16S rRNA gene diversity; metagenome; *Archaea*; deep-sea hydrothermal vent; horizontal gene transfer.

Abstract

The polychaete *Alvinella pompejana* lives in organic tubes on the walls of active hydrothermal chimneys along the East Pacific Rise. To examine the diversity of the archaeal community associated with the polychaete tubes, we constructed libraries by direct PCR amplification and cloning of 16S rRNA genes. Almost half of the sequences of the 16S rRNA gene libraries clustered with uncultured archaeal groups. In an effort to access genomic information from uncultured archaeal members we further constructed a fosmid library from the same DNA source. One of the clones, Alv-FOS5, was sequenced completely. Its sequence analysis revealed an incomplete rRNA operon and 32 predicted ORFs. Seventeen of these ORFs have been assigned putative functions, including transcription and translation, cellular processes and signalling, transport systems and metabolic pathways. Phylogenetic analyses of the 16S rRNA gene suggested that Alv-FOS5 formed a new lineage related to members of Deep-Sea Hydrothermal Vent *Euryarchaeota* group II. Phylogenetic analyses of predicted proteins revealed the existence of likely cases of horizontal gene transfer, both between *Crenarchaeota* and *Euryarchaeota* and between *Archaea* and *Bacteria*. This study is the first step in using genomics to reveal the physiology of an as yet uncultured group of archaea from deep-sea hydrothermal vents.

Introduction

At deep-sea hydrothermal vents, the most extensively studied thermal environments are the sulfide edifices that serve as conduits for the hot, anaerobic, metal-rich venting fluids. The microbial diversity present at the surfaces and within the walls of vent chimneys and associated structures has been investigated using several complementary methods. Numerous thermophilic and hyperthermophilic microorganisms have been cultured from fluids and fragments of chimneys. Most of the microorganisms known to thrive in the hottest parts of the ecosystem belong to metabolically and phylogenetically diverse groups of anaerobic archaea and are predominantly affiliated to the kingdom *Euryarchaeota*. The heterotrophic sulfur-metabolizers of the order *Thermococcales* are the predominant cultured group (Zillig & Reysenbach, 2001). The other archaea frequently isolated from this environment include hyperthermophilic metha-

nogens and sulfate-reducers (Huber & Stetter, 2001; Whitman *et al.*, 2001). In addition to the direct isolation of strains from sulfide structures, the abundance, composition and distribution of archaeal populations have been assessed by whole-cell hybridization studies (Harmsen *et al.*, 1997; Schrenk *et al.*, 2003). Meanwhile, culture-independent phylogenetic studies have considerably expanded our view of the archaeal diversity at deep-sea vents and revealed the existence of unique and previously unrecognized lineages (Takai & Horikoshi, 1999; Takai *et al.*, 2001; Huber *et al.*, 2002a, b; Nercessian *et al.*, 2003a, b). These molecular phylogenetic assessments have provided guides to the development of 16S rRNA gene-targeted oligonucleotide probes for different groups of archaea and their application has revealed the widespread distribution of certain uncultured lineages (Nercessian *et al.*, 2004). In order to characterize some of these key metabolic processes, PCR-based

approaches have been extended to the characterization of functional genes (Nercessian *et al.*, 2005).

Single gene phylogenetic surveys do not, however, provide information about the functional role of the different microorganisms within the community and the genetic information they contain. The use of small- and large-insert genomic libraries has been recently initiated as a powerful approach for isolating and identifying genes from uncultured microorganisms and complex microbial assemblages (Stein *et al.*, 1996; Beja *et al.*, 2000a, b; Rondon *et al.*, 2000; Venter *et al.*, 2004). Such metagenomic libraries have been successfully applied to identify novel genes of uncultured archaea from soils and diverse marine habitats (Stein *et al.*, 1996; Schleper *et al.*, 1997, 1998; Beja *et al.*, 2000a, b; Quaiser *et al.*, 2002; Hallam *et al.*, 2003, 2004; Kruger *et al.*, 2003; Lopez-Garcia *et al.*, 2004; Moreira *et al.*, 2004; Treusch *et al.*, 2004; Tyson *et al.*, 2004; Venter *et al.*, 2004; Erkel *et al.*, 2005). To date, two fosmid libraries obtained from hydrothermal samples have contributed to our current understanding of obligate associations between invertebrates and bacteria (Hughes *et al.*, 1997; Campbell *et al.*, 2003).

In addition to their thermophilic microorganisms, the surfaces of deep-sea hydrothermal chimneys are also colonized by polychaete worms. One of these, *Alvinella pompejana*, inhabits organic tubes secreted along the walls of active black smokers on the East Pacific Rise (Desbruyères *et al.*, 1998). The microenvironment of the worm colonies is characterized by spatial and temporal variability with extreme temperature gradients and high concentrations of sulphide and heavy metals. Although thermal conditions of the fluids inside the tubes are difficult to characterize accurately (Di Meo-Savoie *et al.*, 2004; Le Bris *et al.*, 2005), *A. pompejana* is considered to be one of the most thermotolerant metazoans known (Chevaldonné *et al.*, 1992; Cary *et al.*, 1998). One of the most spectacular features of *A. pompejana* is the unique epibiotic community associated with its dorsal integument. Although these microbial assemblages are composed of morphologically and metabolically diverse microorganisms (Gaill *et al.*, 1987; Jeanthon & Prieur, 1990; Campbell & Cary, 2001; Campbell *et al.*, 2001), their predominant members are filamentous *Epsilonproteobacteria* that have eluded all attempts at culturing (Haddad *et al.*, 1995; Cary *et al.*, 1997). Isolation of *Epsilonproteobacteria* has, however, been reported from diverse hydrothermal samples, including from *A. pompejana* tubes and epibiotic microbiota (Alain *et al.*, 2002; Miroshnichenko *et al.*, 2002; Takai *et al.*, 2003). All these isolates are mesophilic to thermophilic chemolithotrophs using hydrogen or sulfur compounds as electron donor and oxygen, nitrate and/or elemental sulfur as electron acceptors.

Microscopic observations have shown that inner and outer surfaces of alvinellid tubes are densely covered with morphologically diverse microorganisms (Desbruyères *et al.*, 1985), but their molecular identification has not been yet assessed. Because the tubes of *A. pompejana* are directly exposed to the hydrothermal fluid emissions, we hypothesized that their surfaces might provide suitable microhabitats for uncultured microorganisms adapted to high temperatures. In the process of extending our studies of hydrothermal archaea, we evaluate here the archaeal diversity associated with the alvinellid tubes through a phylogenetic analysis of a 16S rRNA gene library. A large-insert fosmid library constructed with the DNA of the same microbial assemblage was screened for archaeal 16S rRNA genes to identify clones that can be used to explore the genetic and metabolic potential of specific phylogenetic groups. We report the identification of six archaeal 16S rRNA gene-containing genomic fragments, five of them belonging to uncultured lineages. Finally, we also present the first genomic sequence isolated from an uncultured hydrothermal euryarchaeon.

Materials and methods

Sample collection and nucleic acid extraction

Alvinella pompejana specimens in their associated tubes were collected at the 13°N hydrothermal vent field on the East Pacific Rise during the oceanographic cruise Phare (Le Bris *et al.*, 2002). The samples were collected by the manipulator arm of the remotely operated vehicle (ROV) *Victor* from the wall of an active sulfide chimney of the Elsa hydrothermal site (PP-Hot3 marker; 12°48.180'N, 103°56.315'W; 2624 m depth). They were placed into enclosed containers to minimize contamination from surrounding seawater during transportation to the surface. Once aboard ship, the worms were removed from their tubes and discarded. The alvinellid tubes were rinsed with sterile seawater and vortexed vigorously to separate the prokaryotic cells from their surfaces. The alvinellid tubes were discarded from the homogenate and the remaining content was centrifuged (12 000 g for 15 min). DNA was immediately extracted on board from this preparation using an UltraClean DNA kit (MoBio Laboratories, Carlsbad, CA), according to the manufacturer's instructions. This step was followed by two supplementary extractions with a phenol–chloroform–isoamyl alcohol and once with chloroform–isoamyl alcohol. Nucleic acids were precipitated at room temperature with 3 M acetate sodium (0.1 vol) and 100% ethanol (2 vol). The precipitate was washed with cold ethanol (70%) and resuspended in milliQ water.

Construction of the archaeal 16S rRNA gene libraries

Small-subunit RNA genes were amplified by PCR using the archaeal-specific forward primers 4F (5'-TCCGGTTGATCC TGCCRG-3') and W36 (5'-TCCAGGCGCTACGGGG-3') with the prokaryotic reverse primer 1492R (5'-CGGTTCCTTGTTACGACTT-3'). The final 25- μ L PCR mixtures consisted of 1 \times DNA polymerase buffer (Interchim, Wörgel, Austria), 1.5 mM MgCl₂, 0.25 mM of each dNTP, 0.2 μ M of each primer, 1 U of *Uptitherm* DNA Polymerase (Interchim) and 1 μ L of DNA. PCR cycles were performed on a Perkin Elmer 9700 thermal cycler. After a denaturation step at 95 °C for 5 min, the following conditions were repeated for 40 cycles: denaturation at 95 °C for 30 s, annealing at 53 °C for 30 s and chain extension at 72 °C for 1.5 min. A final extension step of 8 min at 72 °C was performed. Under these conditions, no PCR products were visible on 0.8% (weight in volume) agarose gels stained with ethidium bromide.

To obtain visible PCR products, primary amplifications using primer pairs 4F-1492R and W36-1492R were stopped at 15 cycles. Aliquots (1 μ L) of the primary amplifications were then used as templates for secondary amplifications. Nested PCRs were performed using forward primers W36 and 1407R (5'-GACGGGGGTGWGTRCAA-3'). The PCR mixtures and the thermal cycles were the same as described above, except that 25 cycles were applied. All amplifications were performed in triplicate, and the resulting products were pooled. Negative controls were performed without DNA template or PCR products. The archaeal 16S rRNA gene libraries were constructed using the TOPO TA Cloning kit (Invitrogen) according to the manufacturer's instructions. These libraries were designed as '4F' (constructed using the primer pair 4F-1492R in the primary PCR) and 'W36' (using the primer pair W36-1492R in the primary PCR). Clones were arrayed in 96-well plates containing 175 μ L of kanamycin (50 μ g mL⁻¹) – Luria-Bertani (LB) medium. Cultures were kept frozen in 12% glycerol at -80 °C.

Fosmid library construction

The fosmid genomic library was constructed using the CopyControlTM Fosmid Library Production Kit (Epicentre, Madison, WI) according to the manufacturer's instructions. Briefly, 20 μ g of the extracted DNA was end-repaired (End-Repair Enzyme Mix, Epicentre). The totality of the end-repaired DNA was loaded in a 1% low-melting-point agarose gel. Pulsed-field gel electrophoresis was performed at 12 °C in 0.5 \times Tris borate EDTA (TBE) buffer in a CHEF-DR III (Bio-Rad, Hercules, CA), with an included angle of 120°, 6 V cm⁻¹, for 13 h with 5–15 s pulses, followed by 3 h with 15–25 s pulses. Agarose slices containing DNA in the size range 35–45 kb were cut off the gel and digested with

GELase (Epicentre). The DNA (250 ng) excised from the gel was cloned in 0.5 μ g of the fosmid vector CopyControl pCC1FOS (Epicentre). The ligated fosmid (2 μ g) was packaged in one tube of phage particles (MaxPlax Lambda Packaging Extracts, Epicentre) and used to transfect *Escherichia coli* EPI300-T1 cells. Clones were arrayed in 96-well plates containing 200 μ L of chloramphenicol (12.5 μ g mL⁻¹) – LB medium with 1 \times Hogness buffer (Werner *et al.*, 1997). The 96-well plates were kept frozen at -80 °C after re-growth. A total of 5017 fosmid clones were obtained, which corresponds to 175–225 Mbp environmental DNA assuming an average insert size of 35–45 kb.

Screening of the fosmid library and sequence of fosmid clone Alv-FOS5

The fosmid library was pooled into groups of 24 clones that served for PCR screening. A pool consisted of 9 μ L of each clone that was centrifuged at 8000 g for 5 min and resuspended in 215 μ L of TE buffer (pH 8). A small volume (2.5 μ L) of each pool was lysed by heat (95 °C for 10 min) and served as DNA template for the PCR. The fosmid library was screened by PCR (under the conditions described above except that the final volume was 12.5 μ L and the number of cycles was 40) using the 16S rRNA gene *Archaea*-specific forward primer 21F (5'-TTCCGGTTGATCCTGCCGGA-3') and the prokaryotic reverse primer 1492R (5'-GGTTTCCTTGTTACGACTT-3'). *Eukarya*-specific primers 42F (5'-CTCAARGAYTAAGCCATGCA-3') and 1498R (5'-CACCTACGGAAACCTTGTTA-3') were also used to screen the fosmid library. Amplicons of the expected sizes were obtained for five pools using the archaeal primers and for six pools using the eukaryal primers. Further PCR screening of individual clones integrating these pools led to the identification of the 11 clones containing either 16S or 18S rRNA genes. One of these clones, designed Alv-FOS5, was entirely sequenced (Genome Express) as described previously (Lopez-Garcia *et al.*, 2004). The average coverage of the sequence was $\times 10.7$.

Analysis of the fosmid sequence Alv-FOS5: tRNA gene, DNA repeat and ORF search, protein identification

The sequence of Alv-FOS5 was annotated using the GenoAnnot module of the Genostar platform (<http://www.genostar.org>). tRNAs were searched using the FAStrRNA program, which is integrated into the GenoAnnot module, and the tRNA-scan-SE 1.21 program (<http://www.genetics.wustl.edu/eddy/tRNAscan-SE>). The ORF identification was carried out using the CDS_prediction task of the GenoAnnot module with the following Markov model parameters: 50 amino acids for the minimum ORF length, 80 as score threshold and the genome of *Thermoplasma acidophilum*

(NC_002578) for the codon usage matrix. Predicted proteins were classified according to the clusters of orthologous groups of proteins (COGs) database (Tatusov *et al.*, 2001). The programs TBLASTN and BLASTP were used to look for sequence similarity with published protein sequences (Altschul *et al.*, 1997). If putative ORFs were detected in different reading frames, only those that had homologues in databases (e value $< e-05$) were selected. In the case of several ORFs that showed no homologues, the longer was conserved. Signal peptides were predicted using the SignalP program (<http://www.cbs.dtu.dk/services/SignalP/>) and transmembrane segments were predicted using the TMHMM program (<http://www.cbs.dtu.dk/services/TMHMM>). Conserved domains of the predicted protein sequences were detected using the program InterProScan (<http://www.ebi.ac.uk/InterProScan>). Comparative analyses of the predicted protein sequences were performed by RPS-BLAST, using the CDART tool (<http://www.ncbi.nlm.nih.gov/Structure/lexington/lexington.cgi?cmd=rps>). Comparative genome analyses were done with GENOMAPPER (<http://www-archbac.u-psud.fr/Genomap/GenomapBrowser.html>).

Phylogenetic analysis of 16S and 23S rRNA gene sequences

Archaeal clones were sequenced by cycle sequencing using the DYEnamic ET terminator cycle sequencing kit (Amersham Biosciences) and M13F (5'-GTAAAACGACGGCCAG-3') and M13R (5'-CAGGAAACAGCTATGAC-3') primers. The sequences were determined using an ABI Prism[®] 3730xl DNA analyser (Applied Biosystems).

The closest relatives to the 16S rRNA gene sequences were identified in the GenBank database by BLASTN (<http://www.ncbi.nlm.nih.gov/BLAST/>). Sequences were aligned using CLUSTALW (Thompson *et al.*, 1994) and distance matrices were constructed using DNADIST from the PHYLIP package with the Jukes–Cantor correction for multiple substitutions (<http://evolution.genetics.washington.edu/phylip.html>). Sequences diverging more than 3% in similarity were defined as having unique phylotype. At least one sequence per phylotype was chosen for complete clone sequencing. The 16S rRNA gene sequences were submitted to the program CHECK-CHIMERA of the Ribosomal Database Project (RDP) (<http://rdp8.cme.msu.edu/cgis/chimera.cgi?-su=SSU>). The relative richness of the two archaeal 16S rRNA gene libraries was estimated by calculating their homologous coverage (C) (Singleton *et al.*, 2001).

To construct the phylogenetic trees, sequences were aligned using the program ED of the MUST package (Philippe, 1993). After removal of gaps and unambiguously aligned positions, a total of 795 positions were used for all the phylogenetic analyses. Bayesian Markov Chain Monte Carlo analysis was carried out with MrBayes v3.0 (Huelsen-

beck & Ronquist, 2001), under a general time reversible (GTR) model assuming a gamma distribution with six rate categories and a proportion of invariable sites. Bayesian posterior probabilities were computed by running four chains for 1 000 000 generations. Trees were sampled every 100 generations and 2000 trees were discarded as 'burn-in'. The maximum-likelihood (ML) analysis was carried out using the program TREEFINDER (Jobb *et al.*, 2004) using the same model described above. ML bootstrap proportions were estimated based on 200 replicates.

The 23S rRNA gene sequences of Alv-FOS5 and selected *Archaea* were also aligned using the program ED. Then, they were manually concatenated to the aligned 16S rRNA gene sequences. A total of 3502 positions were used for the phylogenetic 16S plus 23S rRNA gene analyses, after removal of gaps and unambiguously aligned positions. Phylogenetic analyses of the concatenated sequences were performed using Bayesian and ML methods as described above.

Phylogenetic analyses of the Alv-FOS5 predicted proteins

In the case of proteins, the sequences homologous to Alv-FOS5 identified by BLAST were retrieved from GenBank and aligned using CLUSTAL W (Thompson *et al.*, 1994). Amino acid alignments were inspected manually using the ED program of the MUST package (Philippe, 1993). All ambiguously aligned regions were excluded from phylogenetic analysis. For all alignments, neighbour-joining phylogenetic trees were constructed using MUST. In several cases, additional Bayesian analyses were carried out with MrBayes v3.0, using the JTT model with a gamma distribution and a proportion of invariant sites. The number of generations and trees sampled were the same as for the rRNA gene phylogenetic Bayesian analysis. The numbers of amino acid positions considered in the Bayesian analyses corresponded to 259 (transcriptional factor B), 225 (Sua5-related protein) and 105 (predicted permease).

Nucleotide sequence accession numbers

The 16S rRNA gene sequences reported in this study are deposited in the GenBank database under accession numbers DQ082929 to DQ082984. The complete sequence of the fosmid Alv-FOS5 is available at GenBank under accession number DQ178753.

Results and discussion

Archaeal 16S rRNA gene diversity associated with *A. pompejana* tubes

To identify the archaeal communities associated with the tubes of alvinellid polychaetes, a cultivation-independent approach was used. Archaeal clone libraries 4F and W36

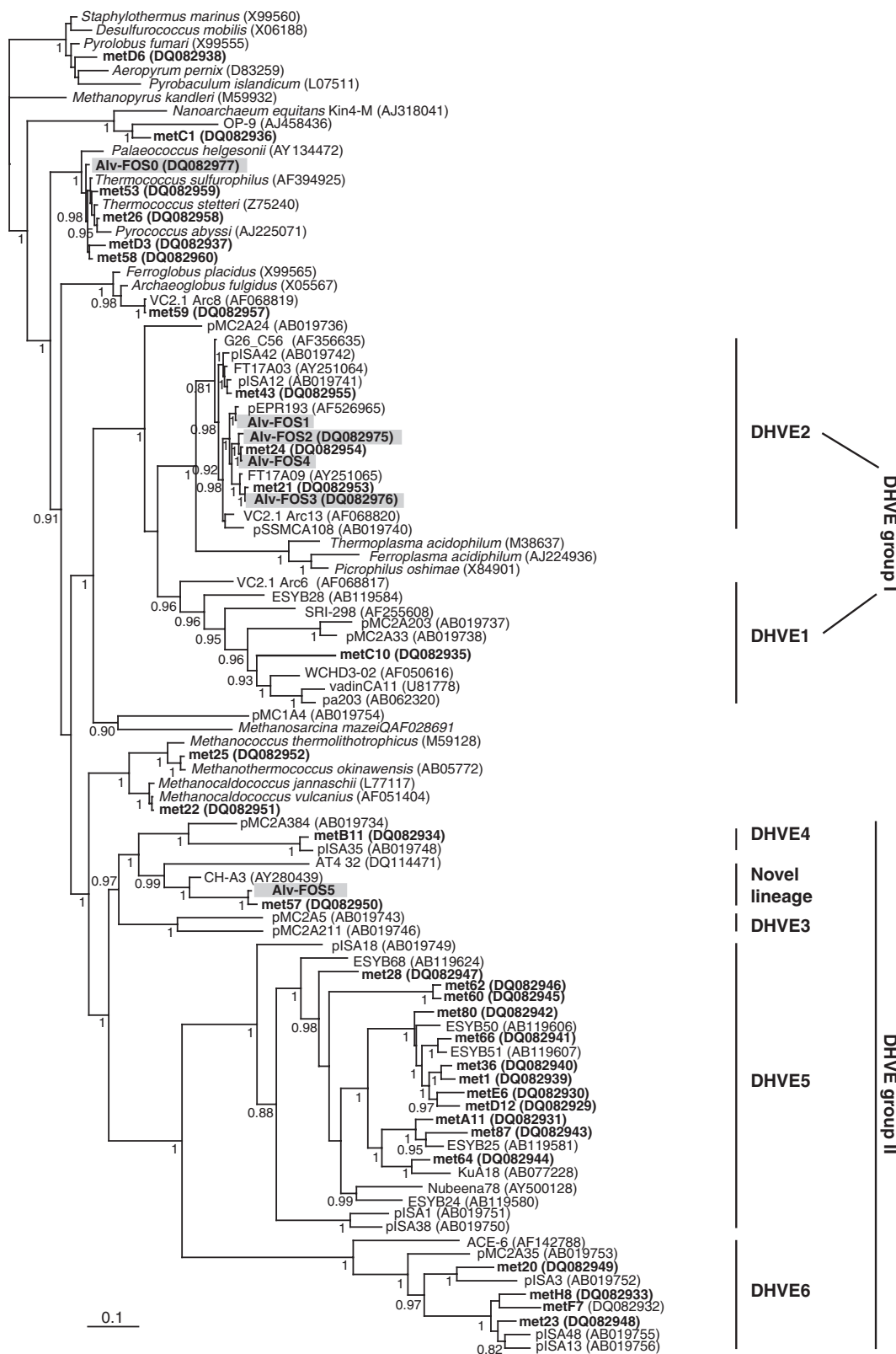
were constructed and a total of 115 clones were sequenced. Two sequences were found to be chimeric and were omitted from subsequent analysis. The libraries were composed of 26 phylotypes (e.g. sequences showing less than 97% similarity) (Table 1). Homologous coverage values ($C_{W36} = 0.83$ and $C_{4F} = 0.82$) indicated that a significant part of the 16S rRNA archaeal gene diversity of the libraries was detected in both libraries. Phylogenetic analysis of the 16S rRNA gene sequences showed that all phylotypes except two were affiliated to the *Euryarchaeota*.

Of the 24 euryarchaeal phylotypes (Table 1), four representing over 50% of the sequences retrieved from both libraries were closely related to cultured species. These phylotypes belonged to the orders *Thermococcales*, *Methanococcales* and *Archaeoglobales* (Fig. 1), members of which are known as common inhabitants of deep-sea hydrothermal vents (Huber *et al.*, 2002a, b; Nercessian *et al.*, 2003a, b, 2004). The 20 remaining euryarchaeal phylotypes were not related (<90%) to any known cultured organism. Phylogenetic analyses placed them among the uncultured Deep-sea

Table 1. Distribution of the representative archaeal 16S rRNA gene sequences identified from tubes of *Alvinella pompejana* collected at 13°N on the East Pacific Rise

Phylogenetic group	Sequence type	G+C content (mol%)	Other representatives	Closest relatives	Sequence similarity (%)	Libraries	
						W36* (n = 55)	4F* (n = 58)
Crenarchaeota							
<i>Desulfurococcales</i>	metD6	66.67	–	<i>Staphylothermus marinus</i>	94	1.8 (1)	–
Euryarchaeota							
<i>Thermococcales</i>							
<i>Thermococcus</i>	met58	65.8	metF5, metA3, metA5, met71, met26, met53, met19, metD3	<i>Thermococcus sulfurophilus</i>	98	36.4 (20)	24.1 (14)
<i>Archaeoglobales</i>							
<i>Archaeoglobus</i>	met59	64.25	met18, met30	Clone pEPR796	99	3.6 (2)	1.7 (1)
Methanococcales							
<i>Methanocaldococcus</i>	met22	64.09	metH4, metD1, metD11	<i>Methanocaldococcus vulcanius</i>	99	5.5 (3)	1.7 (1)
<i>Methanothermococcus</i>	met25	59.65	met65, met81	Clone G26_C45	99	7.3 (4)	24.1 (14)
DHVE group I							
DHVE1	metC10	57.29	met9, met56	Clone WCHD3-02	88	9.1 (5)	8.6 (5)
DHVE2	met24	61.44	met43, met50, metF2, metF8	Clone VC2.1 Arc13	97	9.1 (5)	19 (11)
	met21	61.44	–	Clone FT17A09	97	5.5 (3)	–
DHVE group II							
DHVE4	metB11	50.14	–	Clone pISA35	97	1.8 (1)	–
DHVE5	met62	51.11	met60	Clone ESYB24	85	–	3.4 (2)
	met28	53.97	–	Clone ESYB68	89	1.8 (1)	–
	metA11	51.50	metD4	Clone ESYB25	93	3.6 (2)	–
	met87	55.01	met13, metF11	Clone ESYB25	93	1.8 (1)	3.4 (2)
	met64	53.57	–	Clone KuA18	91	–	1.7 (1)
	metE6	52.50	–	Clone ESYB51	91	–	1.7 (1)
	metD12	53.78	met35	Clone ESYB51	93	3.6 (2)	–
	met80	53.28	–	Clone ESYB50	94	–	1.7 (1)
	met66	52.03	–	Clone ESYB51	96	–	1.7 (1)
	met36	53.38	–	Clone ESYB50	93	1.8 (1)	–
	met1	54.05	metG3	Clone ESYB50	93	1.8 (1)	1.7 (1)
DHVE6	met20	49.12	–	Clone pISA3	90	1.8 (1)	–
	met23	48.97	–	Clone pISA48	93	1.8 (1)	–
	metF7	48.87	–	Clone pISA48	92	–	1.7 (1)
	meth8	50.39	–	Clone pISA13	94	–	1.7 (1)
Novel lineage	met57	59.85	–	Clone CH-A3	88	–	1.7 (1)
Nanoarchaeota	metC1	67.44	–	Clone CU-1	87	1.8 (1)	–

*Numbers correspond to the percentages of clones that appeared in the library.



Hydrothermal Vent *Euryarchaeota* (DHVE) groups I and II (Takai & Horikoshi, 1999) (Table 1 and Fig. 1). Within DHVE group I, most of our sequences were closely related (97% similarity) to members of the DHVE2 lineage (14.5% and 19% of libraries W36 and 4F, respectively). The distribution of this lineage is to date restricted to the deep-sea hydrothermal vent environment and its members have been detected only in the hottest parts of this ecosystem (Reysenbach *et al.*, 2000; Takai *et al.*, 2001; Huber *et al.*, 2002a,b; Hoek *et al.*, 2003; Nercessian *et al.*, 2003a,b, 2004). One phylotype (metC10) grouped within the DHVE1, the second cluster of the DHVE group I (Table 1). This phylotype shared distant relationship (88% similarity) with its closest relative, a sequence (clone WCHD3-02) recovered from a contaminated aquifer (Dojka *et al.*, 1998). Other DHVE1 sequences available in databases have been retrieved from diverse environments, including hydrothermal vent systems (Takai & Horikoshi, 1999), coastal marine sediments (N. Kaku & K. Watanabe, unpublished results) and terrestrial hot springs (Skirnisdottir *et al.*, 2000).

The sequence types that grouped within the DHVE group II were primarily related to the DHVE5. They consisted of 11 phylotypes (Table 1) and most of them had environmental clones retrieved from coastal marine sediments (N. Kaku & K. Watanabe, unpublished results) as closest relatives. Five minor phylotypes found in either library 4F or library W36 clustered within the DHVE4 and DHVE6 lineages and had sequences retrieved from hydrothermal sediments as closest relatives (Takai & Horikoshi, 1999). Lastly, a unique sequence (phylotype met57) that could not be assigned a previously described lineage formed a cluster with clone sequences recovered from deep-sea hydrothermal vents (Pagé *et al.*, 2004; P. López-García and D. Moreira, unpublished results).

The two remaining phylotypes that did not group within the *Euryarchaeota* were only present in the library W36. The phylotype metD6 was related (94% similarity) to the heterotrophic sulfur-reducing crenarchaeote *Staphylothermus marinus*, a common inhabitant of deep-sea hydrothermal vents (Huber & Stetter, 2001). Finally, the phylotype metC1 clustered with the archaeal symbiont *Nanoarchaeum equitans* (Huber *et al.*, 2002a, 2002b) and the uncultured nanoarchaeote OP-9 (Hohn *et al.*, 2002) (87% similarity).

To gain insight into the genomic features of some of the uncultured organisms associated with the *A. pompejana* tubes, a fosmid library was generated from the DNA

previously used for the construction of the 16S rRNA gene libraries.

Identification of fosmid clones containing archaeal and eukaryotic small-subunit rRNA genes

A total of 5017 fosmid clones were obtained. Assuming that the size of the inserts ranged from 35 to 45 kb, an estimated total of 200 Mb of environmental DNA has been cloned in our library. We screened the library by using a primer set specific for 16S rRNA genes of *Archaea*. Five 16S rRNA gene-containing euryarchaeal genome fragments (Alv-FOS0, 1, 2, 3 and 4) were recovered in the library. Sequences analyses of the archaeal 16S rRNA genes showed that Alv-FOS0 belonged to the order *Thermococcales* (Fig. 1). The four other archaeal 16S rRNA genes were closely related (>97% sequence similarity) and formed a cluster within the DHVE2 lineage. Interestingly, these 16S rRNA gene sequences were identical or almost identical (99–100% similarity) to 16S rRNA gene sequences recovered in the PCR-derived libraries.

To look for the presence of microbial eukaryotes, the library was further screened by using 18S rRNA gene eukaryal-specific primers. Six clones exhibited PCR products of the expected size. Four of them contained 18S rRNA gene sequences highly related to that of hydrothermal annelids (97–99% similarity), and one had that of a *Candida fermentati* strain as closest relative (98% similarity) (Suh & Blackwell, 2004). Surprisingly, the last clone recovered in the library by this screening contained an archaeal 16S rRNA gene sequence. This clone, designated Alv-FOS5, had the phylotype met57 obtained in this study as closest relative (>97% similarity). These clones formed an independent cluster distantly related (88–90%) to the phylotypes CH-A3 and AT4-32 (Fig. 1) recovered from deep-sea hydrothermal vents. Interestingly, clone AT4-32 had also been retrieved using 18S rRNA gene-specific primers during a molecular survey of eukaryotic diversity in hydrothermal sediment from the Mid-Atlantic ridge (Lopez-Garcia *et al.*, 2003; P. López-García and D. Moreira, unpublished results). The failure to identify the clone Alv-FOS5 with the archaeal-specific primers was probably due to the presence of five mismatches between primer 21F used in the library screening and its targeted sequence in the Alv-FOS5 genomic fragment. The reasons why we identified the clone Alv-FOS5 with the eukaryal-specific primer pair are still unclear as

Fig. 1. Bayesian phylogenetic tree of archaeal 16S rRNA gene sequences retrieved from *Alvinella pompejana* tubes. The 16S rRNA gene sequences derived from the PCR-based libraries W36 and 4F are shown in bold type and those identified in the fosmid library are in bold type and shaded in grey. Accession numbers of sequences retrieved from GenBank are given in parentheses. In total, 795 unambiguously aligned positions were used in the analysis. Posterior probabilities > 0.8 are indicated at nodes; the scale bar represents the estimated number of nucleotide substitutions per site.

only 12 of the 20 bases of the 18S-42F primer perfectly matched with the Alv-FOS5 16S rRNA gene.

Given that only six archaeal 16S rRNA gene-containing genomic fragments have been identified, the comparison between the composition of the metagenomic library and PCR libraries is impossible. However, the fact that four of the six clones from the metagenomic library belonged to the DHVE2 lineage (Alv-FOS1 to Alv-FOS4) was surprising as DHVE2-related sequences were not dominant in our two PCR-derived libraries and in previous surveys (Takai & Horikoshi, 1999; Reysenbach *et al.*, 2000; Takai *et al.*, 2001; Nercessian *et al.*, 2003a,b). This could suggest that these organisms may be more abundant in the hydrothermal vent environment than previously thought. Their real significance could be addressed via fluorescence *in situ* hybridization or DNA–DNA hybridization experiments (Nercessian *et al.*, 2004).

Fosmids Alv-FOS1 and Alv-FOS4 were selected for full sequencing of their inserts and the analysis of the sequences flanking their 16S rRNA gene sequences will be presented elsewhere. The fosmid Alv-FOS5 was also chosen for full sequencing as representative of a novel lineage of the DHVE II group.

Sequence analysis of the euryarchaeal fosmid clone Alv-FOS5

The entire sequence of the genome insert of clone Alv-FOS5 was determined. The genome fragment had a size of 39843 bp with an average G+C content of 43.95 mol%. This value is in the range of the G+C content of *Thermococcales* (40–51 mol%), some of whose members are the closest cultured relatives to Alv-FOS5 based on 16S rRNA gene sequence similarity (<90%). The rRNA operon of clone Alv-FOS5 consisted only of 16S and 23S rRNA genes, without linked 5S rRNA genes (Fig. 2). This 16S + 23S rRNA gene arrangement is common in crenarchaeotal genomes and also exists in *Euryarchaeota*. Among the completely euryarchaeal sequenced genomes, *Archaeoglobus fulgidus*, *Pyrococcus* species, *Thermococcus kodakarensis* and *Methanocaldococcus jannaschii* (for one of its two rRNA

operons) displayed a similar arrangement to that found in Alv-FOS5. However, this feature had not been observed in genomic fragments from uncultured *Euryarchaeota* of group II, which have the 16S and 23S rRNA genes separated in the genome, nor their closest cultured relatives the *Thermoplasma* (Beja *et al.*, 2000a,b; Moreira *et al.*, 2004). The 16S–23S intergenic transcribed spacer (ITS) of Alv-FOS5 (228 bp in length) contained an Ala-tRNA (Table 2 and Fig. 2). The occurrence of an Ala-tRNA in the 16S–23S ITS has been described for all complete euryarchaeal sequenced genomes that displayed at least the 16S + 23S rRNA gene arrangement. The genome fragment contained two additional tRNAs, a Thr- and a Leu-tRNA gene, the latter being located directly downstream of the 3'-end of the 23S rRNA gene.

The G+C content was quite homogeneous along the Alv-FOS5 sequence, except for the region that contained the 23S and the 16S rRNA genes and linked tRNA genes (Table 2). The low sequence similarity between the Alv-FOS5 16S rRNA gene sequence and its closest cultured relatives makes any metabolic prediction unreasonable. By contrast, the high G+C content of its 16S rRNA gene sequence (60.2 mol%) may be indicative of a thermophilic way of life (Dalgaard & Garrett, 1993; Galtier *et al.*, 1999). Interestingly, the Alv-FOS5 16S rRNA gene sequence formed a cluster with other sequences (clones met57 and CH-A3) that also have high G+C contents (>59 mol% on the basis of 1000 bases).

A total of 32 ORFs longer than 50 amino acids were predicted in Alv-FOS5 (Table 2 and Fig. 2). ORFs were densely packed (92% of the complete Alv-FOS5 was predicted to be coding sequences), as has been previously reported for archaeal genomes. Among the 32 predicted ORFs, 17 showed significant sequence similarity with products of genes of known functions, six were homologous to uncharacterized proteins, and nine were predicted proteins of unknown function without homologues in databases. Among the proteins of unknown function, ORF27 was found to have one transmembrane segment and to harbour a signal peptide while ORF28 exhibited a weak domain signature of the nucleotidyltransferase superfamily.

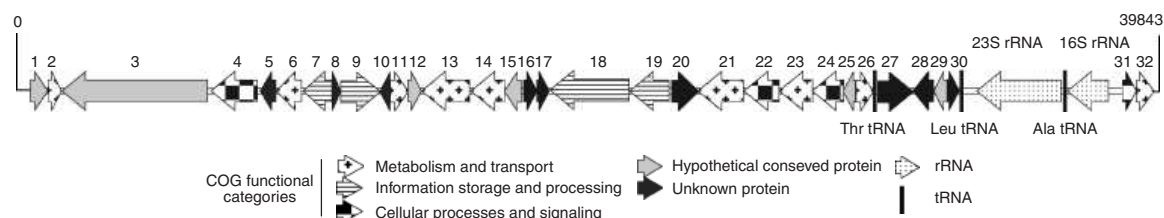


Fig. 2. Schematic representation of the euryarchaeal clone Alv-FOS5. ORFs categorized in proteins with assigned functions (COGs), conserved hypothetical proteins, unknown proteins and rRNA genes are labelled as indicated in the inset. The tRNA genes are indicated by black bars. See Table 1 for ORF numbers of the respective fragments.

Table 2. Predicted RNA and protein coding genes in the euryarchaeotal fosmid AlV-FOS5. ORFs are numbered 1 to 32 following the sense 5' to 3' in Fig. 2

ORF	Nucleotide range	Protein size (amino acids)	Putative function (COG accession if available)	Most similar homologue (BLAST e-value < e-05)	Comments	G+C content (mol%)
1	< 1–622	+ 206	Conserved hypothetical protein	<i>Methanosarcina mazei</i> NP_632407 (4e-47)	ORF truncated. 6 predicted transmembrane helices. Only present in <i>Methanosarcinaceae</i> and in few low-G+C gram-positive <i>Bacteria</i> . Assigned as predicted integral membrane protein (COG5658)	38.81
2	624–983	+ 119	Predicted permease (COG0730)	<i>Nostoc</i> sp. NP_478267 (2e-19)	4 predicted transmembrane helices. Only present in <i>Archaea</i> and <i>Bacteria</i> . Likely transferred from <i>Bacteria</i>	40.83
3	1061–6271	1736	Fibronectin type III domain protein	<i>Geobacter sulfurreducens</i> (2e-22)*	2 predicted transmembrane helices and 1 predicted signal peptide. Very poorly resolved phylogeny. Present mostly in <i>Bacteria</i> and in few <i>Euryarchaeota</i>	38.51
4	6411–8057	– 548	Archaeal thermosome of the HSP60 family (COG0459)	<i>Thermococcus litoralis</i> AAP37564 (0.0)	Widely distributed in <i>Archaea</i> and <i>Eukarya</i>	46.81
5	8171–8722	– 183	Unknown	No homologues		41.12
6	8746–9651	– 301	Sugar kinase of the ribokinase family (COG0524)	<i>Methanopyrus kandleri</i> NP_614100 (9e-22)	Poorly resolved phylogeny. Only present in <i>Archaea</i> (most of the cases in two copies) and <i>Bacteria</i>	47.46
7	9706–10734	– 342	Putative translation factor, SUA5-related protein (COG0009)	<i>Pyrococcus horikoshii</i> NP_142416 (e-103)	Distributed in the three domains	47.42
8	10829–11047	+ 72	Unknown	No homologues		41.55
9	11052–12476	+ 474	Archaeal transcriptional initiation factor TFB (COG1405)	<i>Methanosarcina mazei</i> NP_633796 (1e-92)	Present only in <i>Archaea</i> and <i>Eukarya</i> . FOS5 protein longer than other homologues (~45 amino acids)	43.16
10	12481–12786	– 101	Unknown	No homologues		41.5
11	12832–13470	+ 212	Thymidilate kinase (COG0125)	<i>Pyrococcus furiosus</i> NP_579459 (1e-23)	Widely distributed in the three domains	41.94
12	13458–13946	+ 162	Conserved hypothetical protein	<i>Pyrococcus abyssi</i> NP_127369 (4e-46)	Only present in <i>Archaea</i> and few <i>Bacteria</i> . ORF assigned as uncharacterized conserved protein (COG1839)	42.94
13	13948–15672	– 574	Kef-type K ⁺ transport system (COG0475)	<i>Pseudomonas syringae</i> AAO58938 (2e-32)	15 predicted transmembrane helices. Present in various <i>Bacteria</i> and <i>Euryarchaeota</i> . In <i>Grenarchaeota</i> , only present in <i>Pyrobaculum aerophilum</i> and in <i>Sulfolobales</i>	42.2
14	15714–16922	– 402	Kef-type K ⁺ transport system, membrane component (COG0475)	<i>Lactobacillus acidophilus</i> YP_194758 (1e-12)	12 predicted transmembrane helices. Only present in <i>Bacteria</i> and <i>Archaea</i>	44.67

Table 2. Continued.

ORF	Nucleotide range	Protein size (amino acids)	Putative function (COG accession if available)	Most similar homologue (BLAST e-value < e-05)	Comments	G+C content (mol%)
15	16919–17527	– 202	Conserved hypothetical protein	<i>Thermotoga maritima</i> NP_228743 (4e-11)	Only present in few <i>Bacteria</i> and in <i>Thermococcales</i> (in multiple copies), methanogens, <i>Pyrobaculum aerophilum</i> and <i>Aeropyrum pernix</i>	44.33
16	17601–18083	+ 160	Unknown	No homologues		42.03
17	18174–18419	+ 81	Unknown	No homologues		39.02
18	18530–21367	– 945	Isoleucyl-tRNA synthetase (COG0060)	<i>Methanopyrus kandleri</i> NP_614182 (0.0)	Widely distributed in the three domains	43.9
19	21402–22820	– 472	Prolyl-tRNA synthetase (COG0442)	<i>Methanopyrus kandleri</i> NP_614512 (e-143)	Widely distributed in the three domains	43.76
20	22939–23886	+ 315	Unknown	No homologues	35 overlapping nucleotides with ORF22	41.08
21	23851–25485	– 544	Glycogen debranching enzyme (COG3408)	<i>Methanosarcina mazeri</i> NP_634042 (2e-53)	Present in <i>Euryarchaeota</i> , in <i>Sulfolobales</i> and in few <i>Bacteria</i>	44.95
22	25503–26762	– 419	Glycyl transferase group I (COG0438)	<i>Moorella thermoacetica</i> ZP_00330795 (2e-43)	Only present in <i>Archaea</i> (in multiple copies) and in <i>Bacteria</i>	43.81
23	26762–27943	– 393	Alpha-amylase/alpha-mannosidase (COG1449)	<i>Methanocaldococcus jannaschii</i> NP_248621 (2e-95)	Only present in <i>Archaea</i> and in <i>Crocospaera watsonii</i> and in four <i>Bacteroidetes</i> members	42.07
24	27945–29075	– 376	Glycyl transferase group I (COG0438)	<i>Picrophilus torridus</i> YP_024017 (2e-64)	Only present in <i>Archaea</i> (in multiples copies) and in <i>Bacteria</i>	42.44
25	29072–29293	– 73	Conserved hypothetical protein	<i>Methylococcus capsulatus</i> AAU90863 (5e-07)	1 predicted signal peptide. Only present in <i>Methylococcus capsulatus</i> and <i>Bacteroides fragilis</i>	43.69
26	29518–30144	+ 208	SEC59, dolichol kinase (COG0170)	<i>Methanosarcina acetivorans</i> NP_616933 (5e-12)	6 predicted transmembrane helices. Present in few <i>Bacteria</i> and in most of the <i>Euryarchaeota</i> . Phylogenetic analysis suggested a very divergent protein	43.38
tRNA	30157–30230	+	Threonine tRNA	No homologues	anticodon CGT	71.62
27	30273–31544	+ 423	Unknown	No homologues	1 predicted transmembrane helix and 1 predicted signal peptide	34.43
28	31562–32299	– 245	Unknown	No homologues	Domain superfamily nucleotidyltransferase (SSF81301)	41.03
29	32304–32774	– 156	Conserved hypothetical protein	<i>Methanothermobacter thermoautotrophicus</i> NP_275786 (3e-19)	Distributed in the three domains. Protein containing two cystathionine beta synthase (CBS) domains (COG0517)	41.19

30	32863-33162	+ 99	Unknown	No homologues		41.33
tRNA	33254-33341	+	Leucine tRNA		Anticodon CAA	64.77
rRNA	33845-36869	-	23S rRNA		Anticodon TGC	58.41
rRNA	36964-37040	-	Alanine tRNA			64.94
rRNA	37097-38567	-	16S rRNA			60.23
31	39062-39565	+ 167	Molecular chaperone, small heat shock protein (COG0071)	<i>Sulfolobus solfataricus</i> NP_343781 (1-e21)	Only present in Archaea	36.11
32	39570 > 39043	+ 91	Predicted HD superfamily hydrolase (COG1418)	<i>Methanopyrus kandleri</i> NP_614753 (2e-11)	Truncated ORF. Only present in Euryarchaeota and few Bacteria	41.61

D, direction of transcription; COG, cluster of orthologous groups of proteins.

* A conserved hypothetical protein of *T. kodakarensis* was the closest Blast hit (1e-23) on a short segment.

The predicted genes were classified into the functional categories defined by the COG database (Fig. 2) (Tatusov *et al.*, 2001).

Four predicted proteins were homologous to components involved in the transcription and translation processes. They included a transcription factor IIB (ORF9), a putative translation factor Sua5 (ORF7) and two aminoacyl-tRNA synthetases (ORF18 and 19). ORF18 and 19 showed a tandem gene arrangement that suggests cotranscription. This situation has not been detected so far in archaeal genomes available to date.

The predicted proteins involved in cellular processes and signalling were a chaperonin of the hsp60 family (ORF4), a small heat shock protein (ORF31) and two glycosyl transferases (ORF22 and 24). ORF4 is a putative homologue of class II molecular chaperonins, typically found in archaeal genomes (i.e. forming the thermosome) and in the eukaryotic cytosol (Trent *et al.*, 1991; Archibald *et al.*, 2000). Sequence comparison of the predicted chaperonin revealed high sequence similarity to heat-shock-induced thermosomes of *Thermococcales* (68–69% identity). The small heat shock protein (HSP) predicted in Alv-FOS5 was identified on the basis of its similarity with other small archaeal HSPs of the HSP20 family (37% identity with that of the thermophilic crenarchaeote *Sulfolobus solfataricus*). A compact cluster was formed by a glycogen debranching enzyme (ORF21), a glycosyl transferase (ORF22) and an α -amylase (ORF23). These predicted proteins did not contain any potential signal sequence, suggesting a cytoplasmic location in the cell. Whereas the glycosyl transferase encoded by ORF22 exhibited significant similarity with a bacterial homologue (*Moorella thermoacetica*), the other components of the cluster were most closely related with euryarchaeal homologues (Table 2). Interestingly, the gene arrangement in this cluster is conserved in genomes of the *Thermoplasmatales* and of several members of the phylum *Bacteroidetes* (*Cytophaga hutchinsonii*, *Porphyromonas gingivalis*, *Bacteroides fragilis* and *Bacteroides thetaiotaomicron*). In contrast to the genome of *Picrophilus torridus* genome, which exhibited a 6-phosphofructokinase gene upstream of this conserved gene cluster, a glucoamylase was found in *Thermoplasma* spp. and *Ferroplasma acidarmanus* (Angelov, 2004). No homologues of this gene were found in Alv-FOS5. The potential operon of the three functionally related genes is thought to be involved in glycogen turnover. Interestingly, an additional related gene (a glycosyl transferase, ORF24) is located downstream of this cluster in Alv-FOS5. This unique arrangement could be suggestive of a specific or optimized metabolic pathway.

Nine predicted proteins were presumably involved in transport systems and metabolic pathways. Besides the glycogen debranching enzyme and the alpha-amylase described above (ORF21 and 23), they included a sugar kinase

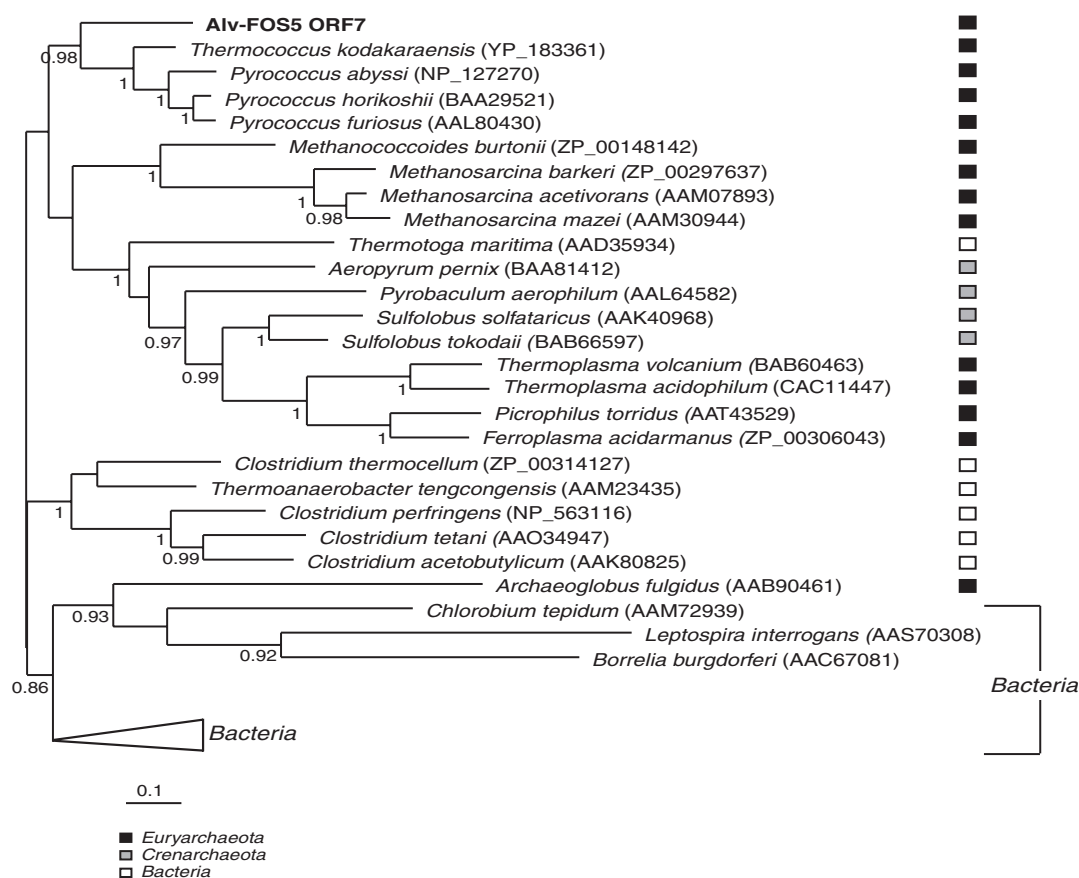


Fig. 4. Phylogenetic tree of the Sua5-related proteins showing a possible case of horizontal gene transfer between Alv-FOS5 and the *Thermococcales*. This tree was determined by a Bayesian phylogenetic analysis of 259 amino acids. Posterior probabilities >0.8 are indicated at nodes; the scale bar represents the estimated number of amino acid substitutions per site.

proteins. The KefB transporters are transmembrane channels involved in potassium efflux by a K^+/H^+ antiporter transport system. Based on their global alignment, the two predicted KefB proteins were 31.3% similar. These two ORFs were contiguous in the genomic fragment, separated by an intergenic region of 41 nt, suggesting a duplication. They showed a cluster organization with a cystathionine beta synthase (CBS) domain-containing protein (ORF15) for which no function was predicted. CBS domains are defined as sequence motifs that occur in CBS and in several other proteins in all kingdoms of life. Their functions are still unknown but point mutations within them cause several hereditary diseases in humans (Scott *et al.*, 2004). The three latter ORFs were most closely related to bacterial proteins. Lastly, the predicted protein involved in phospholipid biosynthesis (ORF26) shared significant similarities with only few homologues but all were euryarchaeal protein sequences.

Phylogenetic analysis of the 16S and 23S rRNA genes

The Bayesian phylogenetic analysis of the 16S rRNA gene sequences showed that the euryarchaeotal fosmid Alv-FOS5 strongly clustered within a group of environmental sequences retrieved from deep-sea hydrothermal vents (Fig. 1). This group included sequences met57 obtained in this study, CH-A3 (Pagé *et al.*, 2004), AT4-32 (P. Lopez-Garcia and D. Moreira, unpublished results), and pMC2A and pISA belonging to the DHVE3 and DHVE4 groups defined by Takai & Horikoshi (1999). These hydrothermal vent sequences formed a branch sister to diverse environmental sequences clustering within the DHVE5 and DHVE6 groups (Takai & Horikoshi, 1999). According to the Bayesian tree, the closest cultured microorganisms related to Alv-FOS5 appeared to be members of the hyperthermophilic *Methanococcales*. This relationship was strongly supported by a

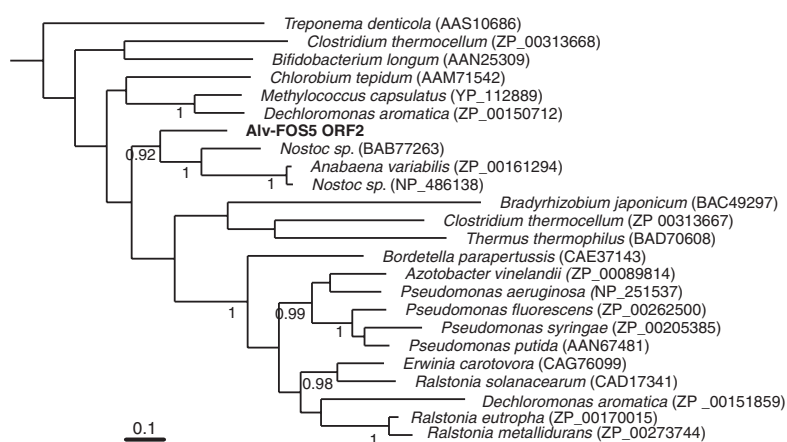


Fig. 5. Phylogenetic tree of the permeases showing a possible case of horizontal gene transfer between *Archaea* and *Bacteria*. This tree was determined by a Bayesian phylogenetic analysis of 105 amino acids. Posterior probabilities >0.8 are indicated at nodes; the scale bar represents the estimated number of amino acid substitutions per site.

posterior probability (PP) of 1 (Fig. 1). However, the phylogenetic tree resulting from an ML analysis did not support this affiliation (see supplementary Fig. S1). To check this disagreement and to increase the phylogenetic resolution, we carried out phylogenetic analyses with concatenated 16S and 23S rRNA gene sequences. The Bayesian tree using these combined sequences showed that Alv-FOS5 was related to the *Thermoplasmatales*, whereas the ML tree placed Alv-FOS5 as a sister branch to diverse euryarchaeotal species including the halophiles, the mesophilic *Methanosarcinales* and the *Thermoplasmatales* (see supplementary Figs S2 and S3). Thus, despite the use of a large amount of data (3502 positions in the concatenated set), the phylogenetic position of Alv-FOS5 within the *Euryarchaeota* remained unclear, probably because it defines a single-species, long branch without close relatives in the tree, which is a difficult situation to resolve using phylogenetic analysis (Hendy & Penny, 1989). Therefore, additional 23S rRNA gene sequences from related species will be necessary to resolve the unstable phylogenetic position of Alv-FOS5.

Phylogenetic analysis of protein-coding genes and detection of horizontal gene transfers

To clarify further the branching position of Alv-FOS5 and to identify possible horizontal gene transfer (HGT) events, phylogenetic analyses were carried out on the 23 predicted protein-coding genes of Alv-FOS5 having homologues in databases (all trees and alignments are available upon request). Unfortunately, the majority of these proteins failed to exhibit a clear phylogenetic profile, for several reasons. Sequence conservation for a few proteins (ORF3 and the dolichol kinase, ORF26) was so poor that the phylogenetic analyses were unreliable. In other cases, these genes have multiple copies in each species (this was particularly noticeable for those species with available complete genome

sequences), pointing to the existence of many paralogues. In general, these paralogues appear to have followed complex evolutionary histories with independent duplications and losses and probable HGT events. As a consequence, we obtained very complex trees, in which the monophyly of many groups (including even the three domains of life) was not retrieved. This problem was particularly severe for the phylogenetic analyses of the ribokinase-family protein (ORF6), the two KefB-related carriers (ORF13 and ORF14), the two glycosyl transferases (ORF22 and ORF24) and the two conserved hypothetical proteins encoded by ORF15 and 29 (data not shown).

Nevertheless, several ORFs encoded proteins with a good representation in databases and without apparent problems of hidden paralogues or parallel gene duplications and/or losses, such that they were useful for phylogenetic reconstruction. A good example was the archaeal transcription factor TFB (ORF9), given that its phylogenetic tree did not support any clear case of HGT and was fully congruent with the classical 16S rRNA gene archaeal trees (Woese *et al.*, 1990), with the exception of several independent duplications of this gene in different archaeal groups (e.g. *Thermoplasmatales*, *Haloarchaea*), which have already been reported in previous analyses (Archibald & Roger, 2002). In our Bayesian analysis (Fig. 3), the TFB sequence of Alv-FOS5 emerged without clear close relatives, further supporting the idea that it may belong to an independent group of euryarchaeota and in agreement with the phylogeny based on the 16S rRNA gene sequences (Fig. 1).

The phylogenetic analysis of the predicted Sua5-related protein (ORF7) showed that all archaeal sequences clustered together, except that of *A. fulgidus*, which branched within a group of bacterial sequences (Fig. 4). Within the archaeal sequences, two groups could be defined. The first included the Alv-FOS5 sequence and its homologues from the *Thermococcales* (with a strong PP of 0.98), while the second

one grouped other euryarchaeotal (*Thermoplasmatales* and some methanogens) with the crenarchaeotal homologues. Interestingly, the *Thermotoga maritima* homologue branched robustly within the archaeal sequences, which suggested that *T. maritima* has gained this gene from *Archaea*, in one of the numerous HGT events that have occurred between the thermophilic bacteria and the *Archaea* (Nelson *et al.*, 1999). Conversely, it seems that *A. fulgidus* has acquired its Sua5 homologue from *Bacteria*, suggesting that HGT of this gene is not infrequent. The position of Alv-FOS5 in our tree was unexpected, as in other trees based on conserved markers (see the 16S and 23S rRNA genes and the TFB factor above) it did not emerge so close to the *Thermococcales*. This result might suggest that Alv-FOS5 has acquired this gene from a member of the order *Thermococcales*. Interestingly, the region of this fosmid upstream to this gene contained two ORFs with a G+C content higher than the average for the rest of ORFs (41.65 mol%). These were ORF4 (46.81 mol%) and ORF6 (47.46 mol%). As for the Sua5-related protein, the phylogenetic analysis of ORF4, encoding a chaperonin of the Hsp60 family, also supported a close relationship between Alv-FOS5 and the *Thermococcales* (data not shown), while ORF6 (encoding a sugar kinase) yielded a complex phylogeny that was difficult to interpret (see above). The phylogenetic signal shown by the ORF4 and ORF6 together with the concurrent observation of an anomalous G+C content suggested that this region of Alv-FOS5 might have been acquired by HGT from a member of the *Thermococcales*.

Potential HGTs involving *Bacteria* were also observed. The best example concerned the predicted permease (ORF2). Its BLAST hit was of a protein encoded by a plasmid of *Nostoc* sp. (Table 2) and the phylogenetic analysis confirmed this relationship (Fig. 5). Although the G+C content of this ORF (40.83 mol%) did not deviate significantly from the average, the possibility of an HGT event was further supported by inspection of the alignment of ORF2 and its homologues, as only the Alv-FOS5 protein and few bacterial homologues (including that of *Nostoc* sp.) possessed several distinctive deletions, as well as several shared amino-acid sequence signatures (data not shown). The fact that this protein was encoded by plasmids in certain bacteria was also in agreement with the possibility of its dissemination by HGT.

Conclusion

We examined the diversity of the archaeal community associated with tubes of the hydrothermal polychaete *Alvinella pompejana* by constructing 16S rRNA gene libraries. Almost half of the identified sequences clustered with uncultured groups. To explore the genetic and metabolic potential of these groups, we constructed a large-insert

fosmid library with the DNA of the same microbial assemblage. We analysed the sequence of Alv-FOS5, one of the archaeal genome fragments identified in the library.

Phylogenetic analyses showed that the 16S rRNA gene sequences of Alv-FOS5 belong to a new cluster within the DHVE group II, the placement of which was strongly influenced by the sequences used in the datasets. The accurate determination of the position of the large hydrothermal group containing Alv-FOS5 will most likely be possible only when conserved sequences (small- and large-subunit rRNA and protein coding genes) from other species of this group become available. Regardless, it is clear that Alv-FOS5 and its relatives define a large, independent phylogenetic group within the *Euryarchaeota*. The G+C content of the 16S rRNA gene sequence of Alv-FOS5 was relatively high (>60 mol%), a feature shared by hyperthermophilic bacteria and archaea. The presence of a gene encoding a putative homologue of class II chaperonins highly similar to that of hyperthermophilic archaea also supports the hypothesis of a possible thermophilic lifestyle for Alv-FOS5.

Our analysis revealed a number of other genes in the immediate surroundings of the 16S and 23S rRNA genes. They include genes essential in the transport and energy metabolism of the archaeal organism. We identified a potential operon probably involved in the breakdown of intracellular glycogen and a sugar kinase. Although the presence of glycogen in archaea has been demonstrated (König *et al.*, 1982), little is known about glycogen metabolism in this group. Because sugar catabolic pathways are absent in nonheterotrophic archaea, except for glycogen-debranching methanogens (Verhees *et al.*, 2003), the finding of these genes may be suggestive of a heterotrophic lifestyle. To validate these predictions, it would be interesting to express these genes and to perform a detailed biochemical analyses to identify their precise functions. Additional genomic sequence data could also help to gain greater insight into the physiology of this organism. Better knowledge of the environmental distribution of Alv-FOS5 and its close relatives may suggest strategies for capturing new genomic sequences of this particular phylogenetic group and/or direct selective isolation attempts. Our study has demonstrated that mismatches in a widely used archaeal-specific 16S rRNA gene primer (21F) prevented their molecular detection. The genes identified in Alv-FOS5 might therefore serve as templates to design specific probes useful to identify the ecological niches inhabited by these organisms.

Acknowledgements

We thank Nadine Le Bris and Françoise Gaill, chief scientist and project leader of the Phare cruise, for inviting us on

board. We thank the captain and crew of *L'Atalante* and the pilots and support crew of the ROV *Victor*. We greatly acknowledge Emmanuelle Morin and Patrick Durand (Irisa-Inria) for their help in the use of the Genostar platform. This work was financed by the CNRS programme 'Geomex' and by the CNRS-Ministère de la Recherche programme 'Séquençage à Grande Echelle'. H.M. was supported by a grant from the Ministère de la Recherche.

References

- Alain K, Querellou J, Lesongeur F, Pignet P, Crassous P, Raguene G, Cuff V & Cambon-Bonavita MA (2002) *Caminibacter hydrogeniphilus* gen. nov., sp. nov., a novel thermophilic, hydrogen-oxidizing bacterium isolated from an East Pacific Rise hydrothermal vent. *Int J Syst Evol Microbiol* **52**: 1317–1323.
- Altschul SE, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W & Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389–3402.
- Angelov A (2004) Genome sequence analysis and characterization of recombinant enzymes from the thermoacidophilic archaeon *Picrophilus torridus*. PhD thesis, Georg-August-Universität Göttingen (<http://webdoc.sub.gwdg.de/diss/2004/angelov/angelov.pdf>).
- Archibald JM & Roger AJ (2002) Gene duplication and gene conversion shape the evolution of archaeal chaperonins. *J Mol Biol* **316**: 1041–1050.
- Archibald JM, Logsdon JM Jr & Doolittle WF (2000) Origin and evolution of eukaryotic chaperonins: phylogenetic evidence for ancient duplications in CCT genes. *Mol Biol Evol* **17**: 1456–1466.
- Beja O, Aravind L, Koonin EV, Suzuki MT, Hadd A & Nguyen LP (2000a) Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. *Science* **289**: 1902–1906.
- Beja O, Marcelino TS, Eugene VK, *et al.* (2000b) Construction and analysis of bacterial artificial chromosome libraries from a marine microbial assemblage. *Environ Microbiol* **2**: 516–529.
- Le Bris N, Gaill F, Alain K, *et al.* (2002) Biogeochemical interactions in extreme environmental conditions: an integrated study (PHARE 2002/EPR 13°N). *InterRidge News* **11**: 20–23.
- Le Bris N, Zbinden M & Gaill F (2005) Processes controlling the physico-chemical micro-environments associated with Pompeii worms. *Deep Sea Res I* **52**: 1071–1083.
- Campbell BJ & Cary SC (2001) Characterization of a novel spirochete associated with the hydrothermal vent polychaete annelid, *Alvinella pompejana*. *Appl Environ Microbiol* **67**: 110–117.
- Campbell BJ, Jeanthon C, Kostka JE, Luther GW & Cary SC (2001) Growth and phylogenetic properties of novel bacteria belonging to the epsilon subdivision of the Proteobacteria enriched from *Alvinella pompejana* and deep-sea hydrothermal vents. *Appl Environ Microbiol* **67**: 4566–4572.
- Campbell BJ, Stein JL & Cary SC (2003) Evidence of chemolithoautotrophy in the bacterial community associated with *Alvinella pompejana*, a hydrothermal vent polychaete. *Appl Environ Microbiol* **69**: 5070–5078.
- Cary SC, Cottrell MT, Stein JL, Camacho F & Desbruyères D (1997) Molecular identification and localization of filamentous symbiotic bacteria associated with the hydrothermal vent annelid *Alvinella pompejana*. *Appl Environ Microbiol* **63**: 1124–1130.
- Cary SC, Shank T & Stein J (1998) Worms bask in extreme temperatures. *Nature* **391**: 545–546.
- Chevaldonné P, Desbruyères D & Childress JJ (1992) Some like it hot . . . and some even hotter. *Nature* **359**: 593–594.
- Dalgaard J & Garrett R (1993) Archaeal hyperthermophilic genes. *The biochemistry of Archaea (Archaeobacteria)*, Vol. 26 (Kates M, Kushner DJ & Matheson AT, eds), pp. 535–563. Elsevier, Amsterdam, The Netherlands.
- Desbruyères D, Gaill F, Laubier L & Fouquet Y (1985) Polychaetous annelids from hydrothermal vent ecosystems: an ecological overview. *Biol Soc Wash Bull* **6**: 103–106.
- Desbruyères D, Chevaldonné P, Alayse AM, *et al.* (1998) Biology and ecology of the 'Pompei worm' (*Alvinella pompejana* Desbruyères and Laubier), a normal dweller of an extreme deep-sea environment: a synthesis of current knowledge and recent developments. *Deep-Sea Res II* **45**: 383–422.
- Dojka MA, Hugenholtz P, Haack SK & Pace NR (1998) Microbial diversity in a hydrocarbon- and chlorinated-solvent-contaminated aquifer undergoing intrinsic bioremediation. *Appl Environ Microbiol* **64**: 3869–3877.
- Erkel C, Kemnitz D, Kube M, Ricke P, Chin K-J, Dedysh S, Reinhardt R, Conrad R & Liesack W (2005) Retrieval of first genome data for rice cluster I methanogens by a combination of cultivation and molecular techniques. *FEMS Microbiol Ecol* **53**: 187–204.
- Gaill F, Desbruyères D & Prieur D (1987) Bacterial communities associated with 'Pompei worms' from the East Pacific Rise hydrothermal vents: SEM, TEM observations. *Microb Ecol* **13**: 129–139.
- Galtier N, Tourasse N & Gouy M (1999) A nonhyperthermophilic common ancestor to extant life forms. *Science* **283**: 220–221.
- Haddad A, Camacho F, Durand P & Cary SC (1995) Phylogenetic characterization of the epibiotic bacteria associated with the hydrothermal vent polychaete *Alvinella pompejana*. *Appl Environ Microbiol* **61**: 1679–1687.
- Hallam SJ, Girguis PR, Preston CM, Richardson PM & DeLong EF (2003) Identification of methyl coenzyme M reductase A (mcrA) genes associated with methane-oxidizing Archaea. *Appl Environ Microbiol* **69**: 5483–5491.
- Hallam SJ, Putnam N, Preston CM, Detter JC, Rokhsar D, Richardson PM & DeLong EF (2004) Reverse methanogenesis: testing the hypothesis with environmental genomics. *Science* **305**: 1457–1462.
- Harmsen H, Prieur D & Jeanthon C (1997) Distribution of microorganisms in deep-sea hydrothermal vent chimneys investigated by whole-cell hybridization and enrichment

- culture of thermophilic subpopulations. *Appl Environ Microbiol* **63**: 2876–2883.
- Hendy M & Penny D (1989) A framework for the quantitative study of evolutionary trees. *Syst Zool* **38**: 297–309.
- Hoek J, Banta A, Hubler F & Reysenbach A-L (2003) Microbial diversity of a sulphide spire located in the Edmond deep-sea hydrothermal vent field on the Central Indian Ridge. *Geobiology* **1**: 119–127.
- Hohn MJ, Hedlund BP & Huber H (2002) Detection of 16S rDNA sequences representing the novel phylum 'Nanoarchaeota': indication for a wide distribution in high temperature biotopes. *Syst Appl Microbiol* **25**: 551–554.
- Huber H & Stetter KO (2001) *Desulfurococcales*. *Bergey's Manual of Systematic Bacteriology, Vol. 1* (Boone DR, Castenholz RW & Garrity G, eds), pp. 179–197. Springer-Verlag, New York.
- Huber JA, Butterfield DA & Baross JA (2002a) Temporal changes in archaeal diversity and chemistry in a Mid-Ocean ridge seafloor habitat. *Appl Environ Microbiol* **68**: 1585–1594.
- Huber H, Hohn MJ, Rachel R, Fuchs T, Wimmer VC & Stetter KO (2002b) A new phylum of *Archaea* represented by a nanosized hyperthermophilic symbiont. *Nature* **417**: 63–67.
- Huelsenbeck JP & Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**: 754–755.
- Hughes DS, Felbeck H & Stein JL (1997) A histidine protein kinase homolog from the endosymbiont of the hydrothermal vent tubeworm *Riftia pachyptila*. *Appl Environ Microbiol* **63**: 3494–3498.
- Jeanthon C & Prieur D (1990) Susceptibility to heavy metals and characterization of heterotrophic bacteria isolated from two hydrothermal vent polychaete, *Alvinella pompejana* and *Alvinella caudata*. *Appl Environ Microbiol* **56**: 3308–3314.
- Jobb G, von Haeseler A & Strimmer K (2004) TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. *BMC Evol Biol* **4**: 18.
- König H, Skorko RW, Zillig W & Reiter WD (1982) Glycogen in thermoacidophilic archaeobacteria of the genera *Sulfolobus*, *Thermoproteus*, *Desulfurococcus* and *Thermococcus*. *Arch Microbiol* **132**: 297–303.
- Kruger M, Meyerdierks A, Glockner FO, *et al.* (2003) A conspicuous nickel protein in microbial mats that oxidize methane anaerobically. *Nature* **426**: 878–881.
- Lopez-Garcia P, Philippe H, Gail F & Moreira D (2003) Autochthonous eukaryotic diversity in hydrothermal sediment and experimental microcolonizers at the Mid-Atlantic Ridge. *Proc Natl Acad Sci USA* **100**: 697–702.
- Lopez-Garcia P, Brochier C, Moreira D & Rodriguez-Valera F (2004) Comparative analysis of a genome fragment of an uncultivated mesopelagic crenarchaeote reveals multiple horizontal gene transfers. *Environ Microbiol* **6**: 19–34.
- Di Meo-Savoie CA, Luther GWR & Cary SC (2004) Physicochemical characterization of the microhabitat of *Alvinella pompejana*, a hydrothermal vent annelid. *Geochim Cosmochim Acta* **68**: 2055–2066.
- Miroshnichenko ML, Kostrikina NA, L'Haridon S, Jeanthon C, Hippe H, Stackebrandt E & Bonch-Osmolovskaya EA (2002) *Nautilia lithotrophica* gen. nov., sp. nov., a thermophilic sulfur-reducing epsilon proteobacterium isolated from a deep-sea hydrothermal vent. *Int J Syst Evol Microbiol* **52**: 1299–1304.
- Moreira D, Rodriguez-Valera F & Lopez-Garcia P (2004) Analysis of a genome fragment of a deep-sea uncultivated Group II euryarchaeote containing 16S rDNA, a spectinomycin-like operon and several energy metabolism genes. *Environ Microbiol* **6**: 959–969.
- Nelson KE, Clayton RA, Gill SR, *et al.* (1999) Evidence for lateral gene transfer between *Archaea* and bacteria from genome sequence of *Thermotoga maritima*. *Nature* **399**: 323–329.
- Nercissian O, Reysenbach AL, Prieur D & Jeanthon C (2003a) Archaeal diversity associated with *in situ* samplers deployed on hydrothermal vents on the East Pacific Rise (13 degrees N). *Environ Microbiol* **5**: 492–502.
- Nercissian O, Reysenbach AL, Prieur D & Jeanthon C (2003b) Archaeal diversity associated with *in situ* samplers deployed on hydrothermal vents on the East Pacific Rise (13 degrees N). *Environ Microbiol* **5**: 492–502.
- Nercissian O, Prokofeva M, Lebedinski A, L'Haridon S, Cary C, Prieur D & Jeanthon C (2004) Design of 16S rRNA-targeted oligonucleotide probes for detecting cultured and uncultured archaeal lineages in high-temperature environments. *Environ Microbiol* **6**: 170–182.
- Nercissian O, Bienvenu N, Moreira D, Prieur D & Jeanthon C (2005) Diversity of functional genes of methanogens, methanotrophs and sulfate reducers in deep-sea hydrothermal environments. *Environ Microbiol* **7**: 118–132.
- Pagé A, Juniper SK, Olagnon M, Alain K, Desrosiers G, Querellou J & Cambon-Bonavita MA (2004) Microbial diversity associated with a *Paralvinella sulfincola* tube and the adjacent substratum on an active deep-sea vent chimney. *Geobiology* **2**: 225–238.
- Philippe H (1993) MUST, a computer package of management utilities for sequences and trees. *Nucleic Acids Res* **21**: 5264–5272.
- Quaiser A, Ochsenreiter T, Klenk HP, Kletzin A, Treusch AH, Meurer G, Eck J, Sensen CW & Schleper C (2002) First insight into the genome of an uncultivated crenarchaeote from soil. *Environ Microbiol* **4**: 603–611.
- Reysenbach AL, Longnecker K & Kirshtein J (2000) Novel bacterial and archaeal lineages from an *in situ* growth chamber deployed at a Mid-Atlantic Ridge hydrothermal vent. *Appl Environ Microbiol* **66**: 3798–3806.
- Rondon MR, August PR, Bettermann AD, *et al.* (2000) Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Appl Environ Microbiol* **66**: 2541–2547.
- Schleper C, Swanson RV, Mathur EJ & DeLong EF (1997) Characterization of a DNA polymerase from the uncultivated psychrophilic archaeon *Cenarchaeum symbiosum*. *J Bacteriol* **179**: 7803–7811.
- Schleper C, DeLong EF, Preston CM, Feldman RA, Wu K-Y & Swanson RV (1998) Genomic analysis reveals chromosomal variation in natural populations of the uncultured

- psychrophilic archaeon *Cenarchaeum symbiosum*. *J Bacteriol* **180**: 5003–5009.
- Schrenk MO, Kelley DS, Delaney JR & Baross JA (2003) Incidence and diversity of microorganisms within the walls of an active deep-sea sulfide chimney. *Appl Environ Microbiol* **69**: 3580–3592.
- Scott JW, Hawley SA, Green KA, Anis M, Stewart G, Scullion GA, Norman DG & Hardie DG (2004) CBS domains form energy-sensing modules whose binding of adenosine ligands is disrupted by disease mutations. *J Clin Invest* **113**: 274–284.
- Singleton DR, Furlong MA, Rathbun SL & Whitman WB (2001) Quantitative comparisons of 16S rRNA gene sequence libraries from environmental samples. *Appl Environ Microbiol* **67**: 4374–4376.
- Skirmisdottir S, Hreggvidsson GO, Hjorleifsdottir S, Marteinsonn VT, Petursdottir SK, Holst O & Kristjansson JK (2000) Influence of sulfide and temperature on species composition and community structure of hot spring microbial mats. *Appl Environ Microbiol* **66**: 2835–2841.
- Stein JL, Marsh TL, Wu KY, Shizuya H & DeLong EF (1996) Characterization of uncultivated prokaryotes: isolation and analysis of a 40-kilobase-pair genome fragment from a planktonic marine archaeon. *J Bacteriol* **178**: 591–599.
- Suh S-O & Blackwell M (2004) Three new beetle-associated yeast species in the *Pichia guilliermondii* clade. *FEMS Yeast Res* **5**: 87–95.
- Takai K & Horikoshi K (1999) Genetic diversity of Archaea in deep-sea hydrothermal vent environments. *Genetics* **152**: 1285–1297.
- Takai K, Komatsu T, Inagaki F & Horikoshi K (2001) Distribution of Archaea in a black smoker chimney structure. *Appl Environ Microbiol* **67**: 3618–3629.
- Takai K, Inagaki F, Nakagawa S, Hirayama H, Nunoura T, Sako Y, Nealson KH & Horikoshi K (2003) Isolation and phylogenetic diversity of members of previously uncultivated ϵ -*Proteobacteria* in deep-sea hydrothermal fields. *FEMS Microbiol Lett* **218**: 167–174.
- Tatusov RL, Natale DA, Garkavtsev IV, Tatusova TA, Shankavaram UT, Rao BS, Kiryutin B, Galperin MY, Fedorova ND & Koonin EV (2001) The COG database: new developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res* **29**: 22–28.
- Thompson JD, Higgins DG & Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**: 4673–4680.
- Trent JD, Nimmegern E, Wall JS, Hartl FU & Horwich AL (1991) A molecular chaperone from a thermophilic archaeobacterium is related to the eukaryotic protein t-complex polypeptide-1. *Nature* **354**: 490–493.
- Treusch AH, Kletzin A, Raddatz G, Ochsenreiter T, Quaiser A, Meurer G, Schuster SC & Schleper C (2004) Characterization of large-insert DNA libraries from soil for environmental genomic studies of Archaea. *Environ Microbiol* **6**: 970–980.
- Tyson GW, Chapman J, Hugenholtz P, Allen EE, Ram RJ, Richardson PM, Solovyev VV, Rubin EM, Rokhsar DS & Banfield JF (2004) Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **428**: 37–43.
- Venter JC, Remington K, Heidelberg JE, *et al.* (2004) Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**: 66–36.
- Verhees CH, Kengen SW, Tuininga JE, Schut GJ, Adams MW, De Vos WM & Van Der Oost J (2003) The unique features of glycolytic pathways in Archaea. *Biochem J* **375**: 231–246.
- Werner E, Holder AA & Hoheisel JD (1997) Growth and storage of YAC clones in Hogness freezing medium. *Nucleic Acids Res* **25**: 1467–1468.
- Whitman WB, Boone DR & Koga Y (2001) *Methanocaldococcaeae*. *Bergey's Manual of Systematic Bacteriology, Vol. 1* (Boone DR, Castenholz RW & Garrity G, eds), pp. 242–246. Springer-Verlag, New York.
- Woese CR, Kandler O & Wheelis ML (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci USA* **87**: 4576–4579.
- Zillig W & Reysenbach AL (2001) Thermococcales. *Bergey's Manual of Systematic Bacteriology, Vol. 1* (Boone DR, Castenholz RW & Garrity G, eds), pp. 341–348. Springer-Verlag, New York.

Supplementary material

The following supplementary material is available online:

Fig. S1. Maximum-likelihood (ML) phylogenetic tree of 16S rRNA archaeal genes.

Fig. S2. Maximum-likelihood (ML) phylogenetic tree of concatenated 16S and 23S rRNA archaeal genes.

Fig. S3. Bayesian phylogenetic tree of concatenated 16S and 23S rRNA archaeal genes.

This material is available as part of the online article from <http://www.blackwell-synergy.com>

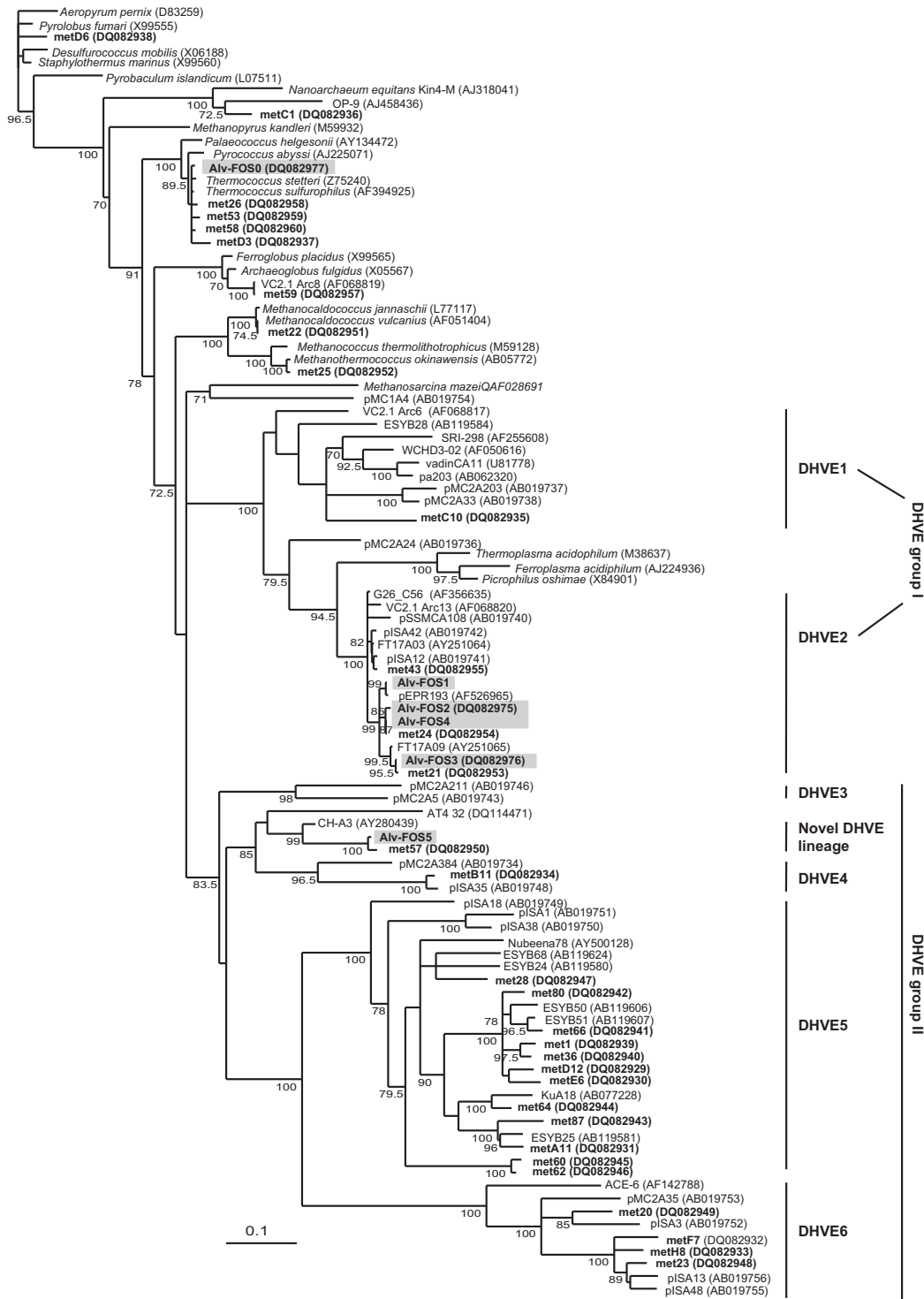


Fig. S1. Maximum-likelihood (ML) phylogenetic tree of 16S rRNA archaeal genes. A total of 795 positions were used in the phylogenetic analysis. Numbers at the nodes are ML bootstrap proportions estimated upon 200 samplings (only those > 70% are indicated); the scale bar represents the estimated number of nucleotide substitutions per site.

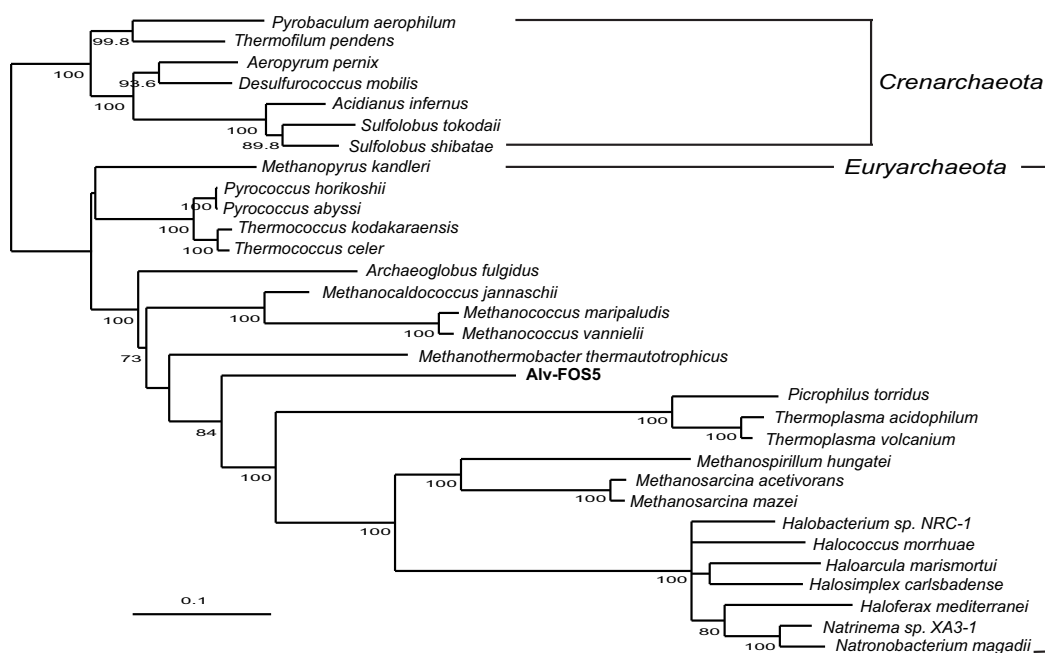


Fig. S2. Maximum-likelihood (ML) phylogenetic tree of concatenated 16S and 23S rRNA archaeal genes. A total of 3502 positions were used in the phylogenetic analysis. Numbers at the nodes are ML bootstrap proportions estimated upon 200 samplings (only those > 70% are indicated); the scale bar represents the estimated number of nucleotide substitutions per site.

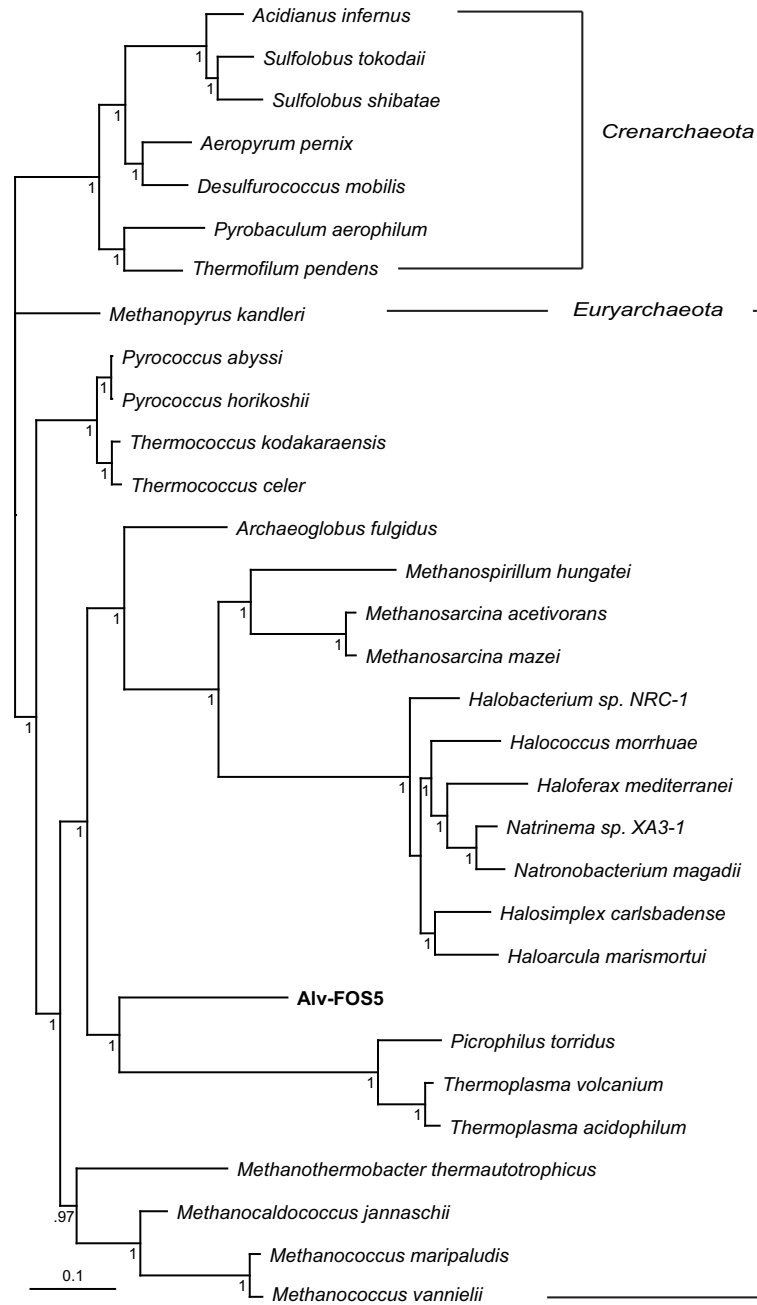


Fig. S3. Bayesian phylogenetic tree of concatenated 16S and 23S rRNA archaeal genes. A total of 3502 positions were used in the phylogenetic analysis. Posterior probabilities > 0.8 are indicated at nodes; the scale bar represents the estimated number of nucleotide substitutions per site.

Diversité bactérienne associée aux tubes d' *A. pompejana*

Dans l'étude 2, la diversité de la communauté archéenne associée à des tubes d'*A. pompejana* a été analysée à l'aide de techniques moléculaires. Pour cela, des tubes de ce polychète ont été prélevés lors de la campagne PHARE, sur la paroi d'une cheminée hydrothermale active du site 13°N EPR. Après extraction de l'ADN génomique issu de la communauté microbienne des tubes d'*A. pompejana*, les gènes d'ARNr 16S ont été amplifiés par PCR. Deux couples d'amorces spécifiques du domaine *Archaea* ont été utilisés pour amplifier les gènes d'ARNr 16S : le couple W36/1492R et le couple 4F/1492R. Le clonage de ces gènes d'ARNr 16S ainsi amplifiés a respectivement permis de construire la banque W36 contenant 83 clones et la banque 4F contenant 81 clones.

Tous les clones de ces deux banques ont été partiellement séquencés (séquences d'ADNr 16S d'environ 750 nucléotides). Quelques séquences chimériques détectées par le programme CHECK-CHIMERA (<http://rdp8.cme.msu.edu/cgis/chimera.cgi?su=SSU>), ont été éliminées de l'analyse (10 chimères identifiées dans la banque W36 et 9 dans la banque 4F). Au total, 73 séquences de la banque W36 et 72 séquences de la banque 4F ont été retenues pour une analyse phylogénétique. Considérant que les séquences d'ADNr 16S qui ont plus de 97% d'identité, appartiennent à la même espèce ; les séquences partielles qui partageaient au moins 97% d'identité ont été regroupées au sein d'un même phylotype représenté par une séquence type. Les organismes cultivés et incultivés les plus proches de chaque phylotype ont ensuite été identifiés par BLAST, dans la banque de données GenBank (<http://www.ncbi.nih.gov/BLAST/>). Comme attendu, la majorité des séquences analysées avait des organismes affiliés aux *Archaea* comme plus proches représentants (Etude 2). Cependant, une proportion non négligeable de séquences, à savoir 21.68% pour la banque W36 (soit 8 séquences) et 17.28% pour la banque 4F (soit 14 séquences), avait pour plus proches représentants des organismes affiliés au domaine des *Bacteria* (Tableau 1).

1 Diversité phylogénétique des clones bactériens

Parmi les séquences d'ADNr 16S bactériens, deux d'entre elles n'ont pu être affiliées à un groupe phylogénétique particulier. Il s'agit des séquences metB6 et met90, respectivement détectées dans les banques W36 et 4F. La séquence metB6 est phylogénétiquement très proche (98% d'identité) de la séquence environnementale IBC2-7 qui provient du site hydrothermal Mid-Okinawa (Nakagawa *et al.*, 2005b). Cependant metB6 ne présente que 90% d'identité avec *Caldithrix abyssi*, le micro-organisme cultivé le plus proche. *C. abyssi*, est une bactérie thermo-

ETUDE 2. RÉSULTATS COMPLÉMENTAIRES

Affiliation phylogénétique	Séquence type	Autres représentants ^A	Représentant le plus proche (% identité)	Organisme cultivé le plus proche (% identité)	Banques ^B	
					W36 ^a n=18	4F ^a n=14
Proteobacteria						
<i>ε-Proteobacteria</i> (<i>Helicobacteraceae</i>)	met6	metE9 metA2 metA8 metH12 metC5 metF1 metB12 metB7 metB8 met37 met15 met75 met77 met5 met76 met92 metE1 met82	Clone L50-WB7 (DQ071286) (98%)	<i>Sulfurimonas paralvinella</i> (AB252048) (98%)	55,55 (10)	64,28% (9)
<i>ε-Proteobacteria</i> (Groupe F)	met10	-	Clone a2b004 (AF420345) (96%)	<i>Proteobacterium</i> BHI80-20 (AJ431223) (96%)	5,55 (1)	-
<i>δ-Proteobacteria</i>	metB4	metC4 metG7 metF3 met42	Clone R4B9 (AF482443) (87%)	<i>Desulfacinum hydrothermale</i> (AF170417) (87%)	16,66 (3)	14,28 (2)
	metG6	-	Clone CS_B022 (AF420354) (97%)	<i>Geobacter bremsensis</i> (U96917) (89%)	-	7,14 (1)
	metC12	-	Clone 112 (AY835388) (98%)	<i>Desulfuromonas thiophila</i> (Y11560) (97%)	5,55 (1)	-
CFB	met33	met11	Clone P. palm C/A 100 (AJ441216) (98%)	<i>Flavobacterium</i> sp. 3034 (AM110988) (91%)	5,55 (1)	7,14 (1)
	met52	-	Clone TANB53 (AY667260) (95%)	<i>Cytophaga</i> sp. souche BHI60-95B (AJ431254) (91%)	5,55 (1)	-
Affiliation incertaine	met90	-	Clone FW17 (AF524026) (85%)	<i>Pediococcus dextrinicus</i> (D87679) (77%)	-	7,14 (1)
	metB6	-	Clone IBC2-7 (AB175571) (98%)	<i>Caldothrix abyssi</i> (AJ430587) (90%)	5,55 (1)	-

^A > 97% d'identité entre les séquences d'un même groupe, sur 750 nucléotides approximativement

^B Les nombres correspondent aux pourcentages de clones bactériens détectés dans la banque.

CFB: *Cytophaga-Flavobacteria-Bacteroides*

Tableau 1 – Affiliations phylogénétiques des séquences d'ADNr 16S bactériennes de la communauté microbienne associée aux tubes d'*A. pompejana*.

phile modérée, anaérobie et mixotrophe, qui a été isolée du site hydrothermal Logatchev de la dorsale médio-atlantique (Miroshnichenko *et al.*, 2003).

La deuxième séquence qui n'a pu être affiliée à un groupe phylogénétique, met90, a comme plus proche représentant la séquence environnementale FW17 (85% d'identité). Cette dernière provient d'un échantillon de sol d'une forêt de pin (Brofft *et al.*, 2002). L'organisme cultivé le plus proche de met90 est *Pediococcus dextrinicus*. Il n'y a seulement que 77% d'identité entre met90 et *P. dextrinicus*.

Mis à part les deux séquences précédemment évoquées, l'analyse des séquences bactériennes d'ADNr 16S provenant des banques W36 et 4F a révélé qu'elles étaient principalement apparentées aux ε -*Proteobacteria* et aux δ -*Proteobacteria*. Quelques séquences apparentées au groupe des *Cytophaga-Flavobacterium-Bacteroides* (groupe CFB) ont également été détectées dans les banques 4F et W36.

1.1 Clones affiliés aux ε -*Proteobacteria*

Deux phylotypes ont été identifiés parmi les ε -*Proteobacteria* de la banque 4F et W36. Les séquences types de ces phylotypes, met6 et met10, sont respectivement apparentés à *Sulfurimonas paralvinella* (98% d'identité) et à la souche BHI80-20 (96% d'identité). Ces ε -*Proteobacteria* ont toutes deux été isolées d'environnements hydrothermaux profonds.

Avec 55% des clones de la banque W36 et 64% des clones de la banque 4F, le phylotype met6 représente à lui seul, plus de la moitié des séquences des clones bactériens obtenues dans les deux banques. Cette étude confirme de précédentes études qui ont montré que les ε -*Proteobacteria* étaient largement distribuées dans les environnements hydrothermaux marins profonds (cf. étude 1).

1.2 Diversité des δ -*Proteobacteria*

Au sein des δ -*Proteobacteria*, trois phylotypes ont été identifiés. L'un d'entre eux, n'est représenté que par une seule séquence, à savoir metC12. Cette séquence a comme organisme cultivé le plus proche l'espèce *Desulfuromonas thiophila* (97% d'identité). *D. thiophila* est une bactérie mésophile sulfo-réductrice, isolée de sédiments anoxiques d'eau douce (Finster *et al.*, 1997).

Les deux autres séquences types apparentées aux δ -*Proteobacteria*, metB4 et metG6, n'ont pas de représentant cultivé proche. MetB4 qui regroupe 16,7% de clones bactériens de la banque W36 et 14,3% des clones bactériens de la banque 4F, ne présente que 87% d'identité avec *Desulfacinum hydrothermale*, le micro-organisme cultivé le plus proche. *D. hydrothermale* est une

bactérie thermophile sulfato-réductrice qui a été isolée de sédiments hydrothermaux près de l'île Milos (Grèce) (Sievert & Kuever, 2000). La séquence metG6, uniquement détectée dans la banque 4F, ne présente quant à elle que 89% d'identité avec *Geobacter bremensis*, l'organisme cultivé le plus proche. *G. bremensis* est une bactérie anaérobie isolée de sédiments d'eau douce, qui réduit le fer (Straub & Buchholz-Cleven, 2001). Notons cependant que la séquence metG6 est très proche de séquences environnementales d'origine hydrothermale océanique profonde, comme la séquence CS_B022 (97% d'identité) identifiée dans la bassin de Guaymas (Teske *et al.*, 2002).

1.3 Diversité des clones bactériens apparentés au groupe CFB

Deux séquences types, met52 et met33, ont été apparentées au groupe CFB. Les organismes cultivés les plus proches de ces séquences types sont encore une fois éloignés de ces séquences types, et ne présentent que 91% d'identité avec elles. Les organismes cultivés en question, sont des souches isolées d'échantillons hydrothermaux marins profonds, à savoir la souche BHI-95B affiliée au genre *Cytophaga* (Cambon-Bonavita, M.A., Riou, V., Alain, K., Cueff, V., Lesongeur, F., Barbier, G. et Querellou, J., résultats non publiés), et la souche 3034 affiliée au genre *Flabovacterium* (Zeng, R., résultats non publiés). Si ces séquences types sont éloignées des organismes cultivés les plus proches, la séquence type met33 est cependant très proche de la séquence environnementale P. palm C/A 100 détectée lors d'un inventaire moléculaire de la diversité microbienne associée au mucus du polychète *Paralvinella palmiformis*, avec 98% d'identité (Alain *et al.*, 2002a).

2 Conclusion

Pour conclure, le compartiment bactérien des tubes d'*A. pompejana* semble également particulièrement intéressant à analyser d'un point de vue métagénomique. En effet, sur les quelques séquences d'ADNr 16S bactériennes qui ont été analysées, toutes avaient pour plus proche représentant des organismes incultivés. Certaines d'entre elles, étaient parfois même très éloignées des organismes cultivés les plus proches. Grâce à la métagénomique, il serait donc possible d'accéder à une partie des informations contenues dans les génomes de ces micro-organismes incultivés.

**THERMOPHILIC LIFESTYLE FOR AN
UNCULTURED ARCHAEON FROM
HYDROTHERMAL VENTS : EVIDENCE
FROM ENVIRONMENTAL GENOMICS**

Thermophilic lifestyle for an uncultured archaeon from hydrothermal vents : evidence from environmental genomics

Les inventaires moléculaires conduits depuis quelques années au niveau de l'écosystème hydrothermal, ont révélé la présence de plusieurs groupes phylogénétiques d'incultivés semblant être inféodés à l'écosystème hydrothermal profond. C'est par exemple le cas du groupe DHVE2 (pour Deep-Sea Hydrothermal Vent Euryarchaeotic) qui a été découvert sur des champs hydrothermaux de l'arc insulaire japonais (Takai & Horikoshi, 1999). Ce groupe d'*Euryarchaeota* a été depuis mis en évidence dans la plupart des sites hydrothermaux échantillonnés sur les dorsales du Pacifique oriental, du Pacifique Nord et de l'Atlantique (Reysenbach *et al.*, 2000 ; Takai *et al.*, 2001a ; Huber *et al.*, 2002 ; Nercessian *et al.*, 2003, 2004). Ces études ont montré l'ubiquité des DHVE2 dans le pôle chaud de l'écosystème hydrothermal, mais elles ont apporté peu d'informations sur la physiologie, le mode de vie ou le rôle de ces micro-organismes dans leur écosystème. Seule l'hypothèse selon laquelle les DHVE2 seraient des organismes thermophiles a pu être formulée en raison du fort pourcentage en guanine et cytosine de leur ARNr 16S (Nercessian *et al.*, 2003). En effet, la plupart des organismes thermophiles et hyperthermophiles présentent des ARNr 16S dont le contenu en G et C est supérieur à 60% (Dalgaard & Garrett, 1993 ; Galtier *et al.*, 1999).

Dans l'article précédent de ce travail de thèse (article 2), nous avons construit une banque métagénomique à partir de l'ADN extrait d'une communauté microbienne associée à des tubes d'*A. pompejana*. Cette banque avait été criblée par PCR pour rechercher les clones portant des gènes d'ADNr 16S archaéens d'intérêt phylogénétique. Un séquençage préliminaire des ADNr 16S identifiés avait révélé que 4 clones positifs de la banque avaient pour plus proches voisins des représentants du groupe incultivé DHVE2. Deux de ces clones ont été entièrement séquencés. Cet article présente l'organisation génétique et l'annotation de ces fragments génomiques, qui ont permis d'assigner des fonctions à la majorité des gènes prédits.

L'objet de cette étude était également de fournir des hypothèses testables expérimentalement pour nous éclairer sur la physiologie et/ou le métabolisme des DHVE2. Des expériences de clonage et d'expression d'une ADN polymérase de la famille B (ADN polB) identifiée sur l'un des deux clones, ont été ainsi réalisées. Ces expériences ont permis de confirmer l'hypothèse selon laquelle les DHVE2 étaient des micro-organismes thermophiles.

Cette étude présente pour la première fois des fragments génomiques d'*Archaea* incultivées fréquemment détectées lors d'inventaires moléculaires dans les habitats du pôle chaud de l'écosystème hydrothermal. L'exploitation des données génomiques issues de cette étude ont permis de confirmer le mode de vie thermophile de ces *Archaea* incultivées.

A Thermophilic Lifestyle for an Uncultured Archaeon of the DHVE2 Lineage : Evidence from Environmental Genomics

HÉLÈNE MOUSSARD¹, GHISLAINE HENNEKE¹, DAVID MOREIRA²,
VINCENT JOUFFE¹, PURIFICACIÓN LÓPEZ-GARCÍA², CHRISTIAN JEANTHON¹ ‡

¹ Laboratoire de Microbiologie des Environnements Extrêmes, UMR 6197, Centre National de la Recherche Scientifique, Ifremer & Université de Bretagne Occidentale, Institut Universitaire Européen de la Mer, 29280 Plouzané, France

² Equipe Diversité et Evolution Microbiennes, Unité d'Ecologie, Systématique & Evolution, UMR 8079, Centre National de la Recherche Scientifique, Université de Paris-Sud, 91405 Orsay, France

‡ Corresponding author. Present address : Christian Jeanthon, UMR 7144, Equipe Phytoplancton Océanique, Station Biologique, Place Georges-Teissier, 29680 Roscoff Cedex. Tel : 33 298 292 323 ; Fax : 33 298 292 324 ; E-mail : Christian.Jeanthon@sb-roscoff.fr

Abstract

Molecular phylogenetic surveys of 16S rRNA genes have revealed the existence of several lineages of uncultured archaea in deep-sea hydrothermal vents. Here, we present a comparative analysis of two genome fragments (Alv-FOS1 and Alv-FOS4) from uncultured archaea retrieved from tubes of the hydrothermal polychaete *Alvinella pompejana*. Both clone inserts (~40 kb) contained a 16S rRNA gene separated from the 23S rRNA gene in the genome, a trait shared by *Methanopyrus kandleri*, planktonic group II *Euryarchaeota* and *Thermoplasmatales*. The clones exhibited 98% similarity in their 16S rRNA gene sequences and contained a co-linear region over 16.6 kb. A significant fraction of genes encoded proteins involved in housekeeping functions. Most phylogenetic reconstructions with orthologs predicted on a single or both genomic fragments supported the affiliation of the DHVE2 lineage as a sister group of the *Thermoplasmatales* as deduced from 16S rRNA gene sequences. Other phylogenetic analyses revealed the existence of several likely cases of horizontal gene transfer from *Thermococcales*. The deduced amino acid sequence of a DNA polymerase gene identified in Alv-FOS1 shared high similarity with those of the archaeal family B DNA polymerases. When expressed in *Escherichia coli*, this protein exhibited an optimal activity and a thermostability comparable to that of the thermophilic archaeon *Thermoplasma acidophilum*, its most closely related homolog. The partial biochemical characterization of this protein therefore substantiated the prediction that organisms belonging to the DHVE2 lineage are adapted to high temperatures. These results provide the first genomic insight into the mode of life of an uncharacterized group of archaea that may have a significant ecological role in hydrothermal high-temperature habitats.

1 Introduction

Molecular phylogenetic surveys of microbial communities in deep-sea hydrothermal vents have recently revealed the existence of archaeal lineages that were unanticipated from classical microbiological methods. These organisms form several monophyletic lineages and cluster, for most of them, within the Deep-sea Hydrothermal Vent *Euryarchaeota* (DHVE) groups I and II defined by Takai and Horikoshi (56). The molecular methods based on 16S rRNA or functional gene analyses have provided extensive information about the presence of archaeal taxa and species in diverse habitats of this extreme environment (29, 39, 56). However, these data (basically phylogenetic) generally provide limited information about the physiology, the mode of life, and the role of the organisms within the communities.

Several molecular techniques have been developed in order to overcome these limitations. Instead of cataloguing single genes, the capture of large environmental DNA fragments from naturally occurring microbial assemblages has recently been developed to explore the physiological potential of uncultured microorganisms and is now becoming a common method to characterize microbial communities. The first discoveries in diverse organisms and habitats and the potential applications of microbial community genomics have been nicely covered by recent reviews (1, 12, 14, 19, 26, 34, 52). Sequence analysis can be guided by the identification of phylogenetic relevant marker genes in order to link the genomic fragment to specific taxa. As often nothing but the 16S rRNA gene sequences of novel lineages are known, they have been generally used as phylogenetic anchors to identify genome fragments from complex environmental libraries.

The analyses of large genomic fragments of uncultured microorganisms have provided the basis for functional studies, including the monitoring of protein activities in the environment (4, 5) and the biochemical characterization of proteins after expression in *Escherichia coli* (4, 53). They also have provided new insights into population ecology, evolutionary processes such as lateral gene transfer (LGT) events, and microbial interactions (4, 6-7, 33, 36, 51, 61).

In an effort to get a more comprehensive view of uncultured archaea from deep-sea hydrothermal vents, we recently constructed a fosmid library from microbial assemblages associated to the tubes of the alvinellid polychaete *Alvinella pompejana* (37). *A. pompejana* colonizes the walls of black smokers on the East Pacific Rise and is considered to be one of the most thermotolerant metazoans known (16). Six archaeal 16S rRNA-containing fosmid clones were detected within the fosmid library. We have described the genomic fragment of a new lineage related to members of DHVE group II (37). Based on the analysis of their 16S rRNA gene sequences, four clones shared more than 97% sequence identity and were ascribed to the DHVE2 lineage within the DHVE group I. Phylogenetic surveys indicated that members of this lineage are widespread in submarine hydrothermal systems since they have been detected in samples from the East Pacific Rise, the Juan de Fuca Ridge, the western Pacific, the Indian Ocean, and

the Mid-Atlantic ridge (28-29, 40-41, 46, 56-57). The distribution of this lineage is restricted to deep-sea vents and limited to the hottest parts of these habitats. No successful attempts at cultivation and characterization of these microorganisms (41) have been reported until now.

In this study, we analyzed the sequences of two genomic fragments from representatives of the DHVE2 lineage and compared their gene organization. These clones contained a co-linear region over 16.6 kb and a significant fraction of genes encoding proteins involved in housekeeping functions. Phylogenetic analyses of individual genes suggest the existence of horizontal gene transfers between archaeal kingdoms and between both prokaryotic domains. A DNA polymerase gene identified in one genome fragment was cloned and expressed in *Escherichia coli*. The optimal activity temperature of the DNA polymerase and its thermostability were comparable to that of the thermophile *Thermoplasma acidophilum*, a close phylogenetic relative.

2 MATERIALS AND METHODS

Construction of the fosmid library from environmental DNA.

Fosmid library preparation has been described previously (37). Briefly, tubes of *A. pompejana* collected at the 13°N hydrothermal vent field (East Pacific Rise) were vortexed vigorously to separate the prokaryotic cells from their surfaces. The alvinellid tubes were discarded from the homogenate and nucleic acids were immediately extracted from the remaining mixture using the UltraClean DNA kit (MoBio Laboratories). The DNA was used to construct the environmental genomic library using the CopyControlTM Fosmid Library Production Kit (Epicentre) following the manufacturer's instructions. End-repaired DNA (End-Repair Enzyme Mix, Epicentre) whose size ranged from 35 to 45 kb was ligated into fosmid vector CopyControl pCC1FOS (Epicentre). The ligated fosmid (2 µg) was packaged in one tube of phage particles (MaxPlax Lambda Packaging Extracts, Epicentre) and used to transfect *E. coli* EPI300-T1 cells. The environmental genomic library contained 5,017 fosmid clones, which corresponded to about 200 Mb of environmental DNA assuming an average insert size of 35-45 kb.

Identification of fosmids Alv-FOS1 and Alv-FOS4.

The library was pooled in groups of 24 clone pools and pools that contained archaeal 16S rRNA were identified by PCR using archaeal-specific forward primer 21F (5'-TTCCGGTTGATCCTGCCGGA-3') and reverse primer 1492R (5'-GGTTACCTTGTTACGACTT-3') as previously described (37). Further PCR screening of individual clones integrating positive pools led to the identification of the 5 clones containing archaeal 16S rRNA genes. Two of them, designed Alv-FOS1 and Alv-FOS4, were entirely sequenced (Genome Express) as described previously (33). The average coverage of the se-

quence was 11.6X for Alv-FOS1 and 11X for Alv-FOS4, respectively.

Analysis of the Alv-FOS1 and Alv-FOS4 sequences : tRNA gene, open reading frame search and protein identification.

The sequences of Alv-FOS1 and Alv-FOS4 were annotated using the GenoAnnot module of the Genostar platform (<http://www.genostar.org>). tRNAs were searched using the FAStrNA program which is integrated into the GenoAnnot module and the tRNA-scan-SE 1.21 program (<http://www.genetics.wustl.edu/eddy/tRNAscan-SE>). The ORF identification was carried out using the CDS_prediction task of the GenoAnnot module with the following Markov model parameters : 50 amino acids for the minimum ORF length, 80 as score threshold and the genome of *T. acidophilum* (NC_002578) for the codon usage matrix. Predicted proteins were classified according to the clusters of orthologous groups of proteins (COGs) database (59). The programs TBLASTN and BLASTP were used to look for sequence similarity with published protein sequences (2). If putative ORFs were detected in different reading frames, only those that had homologs in databases (e value < e-05) were selected. In the case of several ORFs showed no homologs, the longest was conserved. Signal peptides were predicted using the SignalP program (<http://www.cbs.dtu.dk/services/SignalP/>) and trans-membrane segments were predicted using the TMHMM program (<http://www.cbs.dtu.dk/services/TMHMM>). Conserved domains of the predicted protein sequences were detected using the program InterProScan (<http://www.ebi.ac.uk/InterProScan>). Comparative analyses of the predicted protein sequences were performed by RPS-BLAST, using the CDART tool (<http://www.ncbi.nlm.nih.gov/Structure/lexington/lexington.cgi?cmd=rps>). Comparative genome analyses were done with GENOMAPPER (<http://www.archbac.u-psud.fr/Genomap/GenomapBrowser.html>).

Phylogenetic analysis of the 16S rDNA sequences.

Closest relatives to our sequences were identified in the GenBank database by BLAST (<http://www.ncbi.nlm.nih.gov/BLAST/>). To construct the phylogenetic tree, sequences were aligned using the program ED of the MUST package (43). After removal of gaps and unambiguously aligned positions, a total of 947 positions were used for the phylogenetic analyses. Bayesian Markov Chain Monte Carlo analysis was carried out with MrBayes v3.0 (30), under a general time reversible (GTR) model assuming a gamma distribution with six rate categories and a proportion of invariable sites. Bayesian posterior probabilities were computed by running 4 chains for 1,000,000 generations. Trees were sampled every 100 generations. Two thousand trees were discarded as "burn-in".

Phylogenetic analyses of the Alv-FOS1 and Alv-FOS4 predicted proteins.

In the case of proteins, the sequences homologous to Alv-FOS1 and Alv-FOS4 identified by BLAST were retrieved from GenBank and aligned using CLUSTALW (60). Amino acid alignments were inspected manually using the ED program of the MUST package. All ambiguously aligned regions were excluded from phylogenetic analysis. For all alignment, phylogenetic trees were constructed using MUST. If necessary, Bayesian analyses were carried out with MrBayes v3.0 with a mixed model of amino acid substitution and a gamma distribution with six rate categories plus a proportion of invariable sites. The number of generations and trees sampled were the same than as for the 16S rDNA phylogenetic Bayesian analysis. The number of amino acid positions considered in the Bayesian analyses corresponded to 330 (DNA polymerase B, Alv-FOS1 ORF16), 399 (permease of the major facilitator superfamily, Alv-FOS4 ORF10), 273 (subunit A of topoisomerase VI, Alv-FOS4 ORF34), and 260 (aspartate/tyrosine/aromatic aminotransferase Alv-FOS1 ORF35, Alv-FOS4 ORF22).

Cloning of Alv-FOS1 DNA polymerase.

Oligonucleotide primers were designed to PCR-amplify the gene of the predicted DNA polymerase encoded by ORF16. The two primers polB-S1 (Forward, 5'-CACAGAACGCAATTATCATATGCTATCCTGTGA-3') and polB-AS1 (Reverse, 5'-TCTAGGATCCTCACTCTTCGAAATCAAAGAGTG-3') were designed to contain the *NdeI* and *BamHI* restriction sites (underlined), respectively. The *NdeI* site contained the initiation codon ATG and the *BamHI* was immediately after the stop codon. The PCR reaction was performed using the high fidelity *Pfu* DNA polymerase (Promega). The PCR product was digested with *NdeI* and *BamHI* (New England Biolabs) and ligated using the T4 DNA ligase (New England Biolabs) into the *E. coli* pARHS expression vector (15) and transformed into *E. coli* DH5 α (Clontech) according to standard procedures (49). Positive clones were identified by restriction analysis and the identity of the gene was confirmed by nucleotide sequencing. The recombinant plasmid was designated pAlv-FOS1DNApol.

Expression and purification of Alv-FOS1 DNA polymerase.

(i) Buffers.

The following buffers were used in the course of the Alv-FOS1 DNA polymerase purification :

- Buffer A : 50 mM Tris-HCl [pH 8], 2 mM β -mercaptoethanol and a complete protease inhibitor cocktail tablet (Roche).
- Buffer B : 50 mM Tris-HCl [pH 8.8], 1 mM DTT.
- Buffer C : 50 mM Tris-HCl [pH 8.8], 1 mM DTT, 1M NaCl.
- Buffer D : 50 mM Tris-HCl [pH 8.8], 1 mM DTT, 0.1M NaCl.
- Buffer E : 50 mM Tris-HCl [pH 7.7], 1 mM DTT.

– Buffer F : 50 mM Tris-HCl [pH 7.7], 1 mM DTT, 1M NaCl

(ii) Production and purification.

pAlv-FOS1DNApol was used for transformation of *E. coli* HMS 174 (DE3) (Stratagene). Since cell lysis was observed in 500-ml cultures, transformed cells were grown at 37°C in several 100-ml flasks containing ampicillin (100 µg/ml)-LB medium. The cells were grown to an A_{600nm} of 0.6 and the overexpression of the recombinant protein was induced by the addition of 1 mM IPTG (final concentration). The induced culture was incubated at 37°C for an additional period of 4 hours with gentle agitation (150 rpm). The cells were harvested by centrifugation and suspended in 30 ml of buffer A. Cells were disrupted by French press (1000 p.s.i.) and the insoluble cell debris were removed by centrifugation. The supernatant was heated for 30 min at 70°C and the precipitated *E. coli* proteins were removed by further centrifugation (15,000 X g, 45 min, 4°C). The resulting supernatant was dialyzed overnight against buffer B. The dialysate (25 ml) was applied to an anion-exchange column [HiPrep 16/10 Q FF column, 20 ml (Amersham Biosciences)] pre-equilibrated with buffer B. Bound proteins were eluted with a linear gradient of buffer B to buffer C containing 1 M NaCl. The active fractions were pooled and dialysed against buffer B. The dialysate (12.5 ml) was applied to an affinity column [HiTrap Heparin, 5 ml (Amersham Biosciences)] pre-equilibrated with buffer B. Elution of proteins was carried out with a linear gradient from buffer D to buffer C. The active fractions were pooled, dialysed against buffer E and the fraction (2 ml) was applied to a cation exchanger column [MonoS, 1 ml (Amersham Biosciences)] pre-equilibrated with buffer E. The Alv-FOS1 DNA polymerase was eluted with a linear gradient from buffer E to buffer F, containing 1M NaCl. The active fractions were concentrated, dialysed against 50 mM Tris-HCl [pH 7.5], 1 mM DTT, 50% (v/v) glycerol, and stored at -20°C.

(iii) Protein determination.

The purity of the protein samples was evaluated on 10% SDS-polyacrylamide gel using the method of Laemmli (32) and proteins bands were revealed by Coomassie staining (polyacrylamide gel electrophoresis reagents were molecular biology grade reagents from Sigma-Aldrich ; molecular weight markers were obtained from Amersham Biosciences). Polymerase concentration was determined by the method of Bradford (8) using bovine serum albumin as standard. The theoretical molecular mass and the isoelectric point of the polymerase were calculated using the ProtParam tool on the ExPASy server (<http://www.expasy.org/tools/protparam.html>).

Polymerase activity assays.

DNA polymerase activity was performed in a final volume of 20 µl containing 20 mM Tris-HCl [pH 9], 25 mM KCl, 1.5 mM MgSO₄, 10 mM (NH₄)₂SO₄, 0.1 mg/ml BSA, 0.1% Tween 20 [this mixture corresponded to Isis DNA polymerase incubation buffer (Q-Biogene)], 0.35

μM [methyl, 1',2'- ^3H]-Thymidine 5'-triphosphate, ammonium salt (^3H]-dTTP, 119 Ci/mmol, Amersham Biosciences), 200 μM of each four deoxynucleotide triphosphate (dNTP, Q-Biogene) and 4 μg of activated calf thymus DNA (Sigma-Aldrich). The *Pyrococcus abyssi* polymerase D incubation buffer containing 20 mM Bis-Tris-HCl [pH 6.5], 10 mM MgCl_2 , 1 mM DTT, 0.4 mg/ml BSA was also tested as alternative incubation buffer as described by Henneke *et al.* (27).

The assay mixture (20 μl) was incubated at 50°C for 20 min with 1.4 pmol of Alv-FOS1 polymerase and the reaction was stopped on ice. DNA products were precipitated with 10% (w/v) TCA and insoluble radioactive material was determined by scintillation counting. Controls were made in the absence of enzyme.

To determine its optimal temperature and thermostability, the Alv-FOS1 polymerase was diluted ten-fold into the Isis DNA polymerase incubation buffer and the reactions were performed in the mixture described above. The optimal temperature was determined by carrying out the polymerase reactions for 20 min from 40°C to 80°C. For thermostability, 1.4 pmol of Alv-FOS1 polymerase was pre-incubated alone for 30 min at temperatures ranging from 4°C to 90°C. The pre-incubated polymerase was added to the assay buffer (20 μl) containing the appropriated concentrations of substrates as outlined above and the reaction was carried out for 20 min at 65°C. The thermostability of Alv-FOS1 DNA polymerase was compared to that of the Klenow fragment of *E. coli* DNA polymerase I and the Taq and Isis DNA polymerases that were assayed in their respective buffers (Q-Biogene). One unit of DNA polymerase was defined as the amount of enzyme catalysing the incorporation of 1 nmol of dNTP into acid-insoluble product.

Nucleotide sequence accession numbers.

The sequences of fosmid Alv-FOS1 and Alv-FOS4 have been deposited at GenBank under accession DQ118403 and DQ118404, respectively.

3 RESULTS AND DISCUSSION

We have previously reported the construction of an environmental library from microbial assemblages associated to the tubes of the hydrothermal polychaete *A. pompejana* (37). Four fosmid clones containing an archaeal 16S rRNA gene (Alv-FOS1, Alv-FOS2, Alv-FOS3, and Alv-FOS4) clustered within the DHVE2 lineage (> 97% sequence similarity between them) (Fig. 1). Interestingly, their 16 rRNA gene sequences were identical or almost identical (99 to 100% similarity) to sequences from PCR-derived libraries constructed using the same DNA (37). Bayesian phylogenetic analyses strongly confirmed the placement of the DHVE2 lineage as a sister group to the *Thermoplasmatales*, their closest cultured relatives (41, 56). To date,

cultured members of the *Thermoplasmatales* include the genera *Thermoplasma*, *Picrophilus*, and *Ferroplasma* (23, 45). Most of them are heterotrophic aerobes, although *Thermoplasma* and *Ferroplasma* are able to grow under anaerobic conditions. These acidophilic and mesophilic to thermophilic organisms have only been isolated from terrestrial environments. To get more insight into the physiology and genetics of members of the DHVE2 lineage, the ~40 kb-inserts of the two fosmids that appeared the most distant in the 16S rRNA tree, Alv-FOS1 and Alv-FOS4, were sequenced completely.

Sequence analysis and genomic organization of fosmids FOS1 and FOS4.

The inserts of clones Alv-FOS1 and Alv-FOS4 comprised 40,482 bp and 41,331 bp, respectively. In both genome fragments, neither 23S nor 5S rRNA genes but a Met-tRNA gene were present in the immediate surroundings of the 16S rRNA gene. In archaea, rRNA genes are generally organized in operons or at least exhibit a 16S + 23S gene arrangement. The situation found in the DHVE2 genome fragments is not common but similar to that observed in the genomes of *Thermoplasmatales* and *Methanopyrus kandleri*, a methanogenic hyperthermophile, and in two genomic fragments of uncultured planktonic euryarchaeota whose closest cultured neighbors are the *Thermoplasmatales* (4, 20, 36, 44, 55). *M. kandleri* and *Thermoplasmatales* contained only a single copy of each rRNA gene and these are separated in the genome. It is not possible to predict if the Alv-FOS1 and Alv-FOS4 contained only one copy of their rRNA genes. However, as the DHVE2 lineage shares a common origin with the *Thermoplasmatales* (based on their 16S rRNA sequence analysis), their ancestor might contain separated genes, and possibly one single copy.

A total of 41 and 44 predicted open reading frames (ORFs) longer than 50 amino acids were identified in Alv-FOS1 and Alv-FOS4, respectively (Table 1, Table 2, Fig. 2). The ORFs were generally efficiently packed with 88 and 91.42% of predicted coding sequences in Alv-FOS1 and Alv-FOS4, respectively. The total G+C content of Alv-FOS1 (49.2%) and Alv-FOS4 (47.5%) were slightly higher than that of *Thermoplasmatales*, ranging from 36 to 46% (23, 45). The G+C content was not homogenous along both sequences. Generally, the G+C values were < 40% for intergenic regions, > 60% for the 16S rRNAs and Met-tRNAs, while coding sequences displayed values close to the average. Since 16S rRNA molecules with a G+C content > 60% are typical of thermophilic microorganisms (11, 22), a thermophilic lifestyle has been suggested for members of the DHVE2 lineage (41).

Most of the predicted ORFs of Alv-FOS1 (24) and Alv-FOS4 (25) showed sequence homologies to the products of known genes (Table 1, Table 2). Others were homologous to conserved hypothetical proteins (5 and 6 for Alv-FOS1 and Alv-FOS4, respectively) or were predicted proteins of unknown function without homologs in databases (12 and 13, for Alv-FOS1 and Alv-FOS4, respectively).

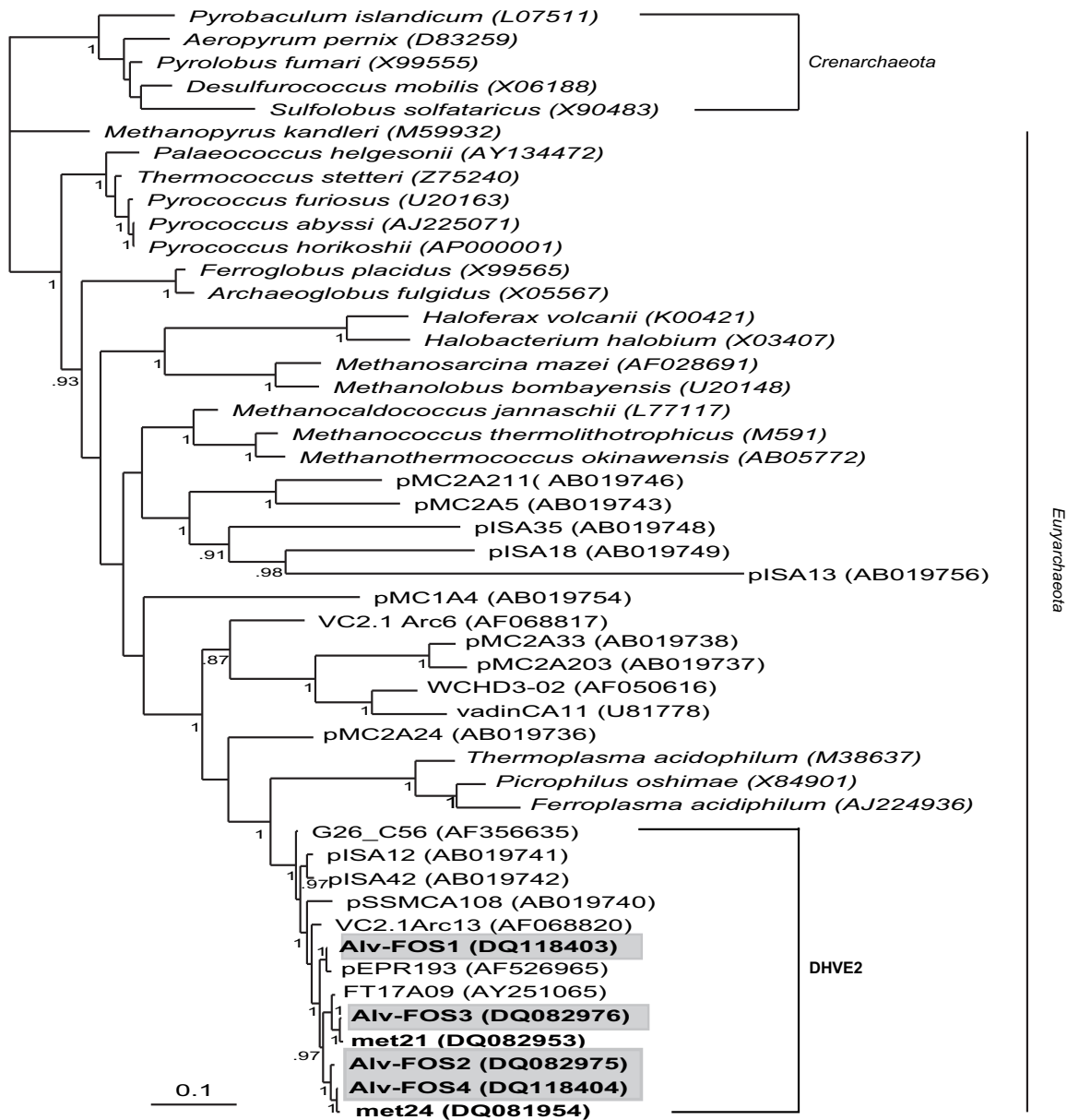


Fig 1. Bayesian phylogenetic tree of DHVE2 16S rRNA gene sequences retrieved from *A. pompejana* tubes. The 16S rDNA gene sequences derived from the PCR-based libraries (37) are in boldface and those identified in the fosmid library are in boldface and shaded in grey. Accession numbers of sequences retrieved from GenBank are given between brackets. 947 unambiguously aligned positions were used in the analysis. Posterior probabilities > 0.8 are indicated at nodes; the scale bar represents the estimated number of nucleotide substitutions per site.

Table 1. Predicted RNA and protein coding genes, and remarkable features in the euryarchaeotal fosmid Alv-FOS1. ORFs are numbered 1-41 following the sense 5' to 3' in Fig. 2. D, direction of transcription. Shaded ORF numbers indicate the co-linear region.

ORF	Nucleotide range	D	Protein size (aa)	Putative function (COG accession number if available)	Most similar homologue (BLAST e-value <e-05)	Comments	% G+C
1	<1..1453	-	484	Hypothetical conserved protein	<i>Haloarcula marismortui</i> AAIV45993 (3e-17)	1 predicted transmembrane helix and 1 predicted signal peptide	51.03
2	1450-2448	-	332	ABC transporter involved in lipoprotein release (COG4591)	<i>Clostridium perfringens</i> BAB81246 (1e-10)	5 predicted transmembrane helices. Present in <i>Bacteria</i> and in <i>Archaea</i> . Missing in <i>Thermoplasmatales</i> . Poorly conserved, phylogeny not resolved	42.84
3	2494-3147	+	217	ABC transporter (COG1136)	<i>Clostridium thermocellum</i> ZP_00313596 (2e-19)	Present in <i>Archaea</i> and in <i>Bacteria</i> . Missing in <i>Thermoplasmatales</i> . Phylogenetic analysis supported an archaeal affinity with a low bootstrap proportion	47.09
4	3137-3856	+	239	Unknown	None	4 predicted transmembrane helices and 1 predicted signal peptide	51.09
5	3932-7234	+	1100	Archaeal DNA polymerase II, large subunit (DP2) (COG1933)	<i>Thermoplasma volcanium</i> NP_110554 (0.0)	Only present in <i>Euryarchaeota</i> . ORF4 and ORF5 form an operon-like structure shared only with <i>Thermoplasmatales</i> except <i>P. torridus</i>	53.69
6	7231-7998	+	255	GTPase (COG1100)	<i>Ferroplasma acidarmanus</i> ZP_00306591 (5e-68)	Present in <i>Archaea</i> and in <i>Eukaryota</i>	49.87
7	8000-8539	+	179	Orotate phosphoribosyltransferase (COG0461)	<i>Thermoplasma acidophilum</i> NP_393535 (4e-33)	Distributed in the three domains	48.33
8	8556-9725	+	389	Hypothetical conserved protein	<i>Archaeoglobus fulgidus</i> NP_068866 (4e-17)	12 predicted transmembrane helices. Present in <i>Archaea</i> and in <i>Bacteria</i> . Protein containing a major facilitator superfamily PFAM domain	54.27
9	9740-9955	-	71	Unknown	None	1 predicted transmembrane helix	43.52
10	10165-10383	+	72	Unknown	None		43.38
11	10380-10628	+	82	Unknown	None		48.19
12	10618-11412	+	264	NH(3)-dependent NAD ⁺ synthetase (COG0171H)	<i>Archaeoglobus fulgidus</i> NP_069833 (6e-51)	Present in <i>Archaea</i> and in <i>Bacteria</i> .	52.7
13	11414-11698	-	94	Hypothetical conserved protein (COG2412)	<i>Pyrococcus abyssi</i> CAB49898 (6e-19)	Only present in <i>Euryarchaeota</i> and in <i>Sulfolobales</i> . ORF12 and ORF13 form an operon-like structure also shared with <i>Thermoplasmatales</i> and <i>Methanosarcinales</i>	43.51
14	11712-13760	-	682	DNA replication licensing factor MCM related protein (COG1241)	<i>Ferroplasma acidarmanus</i> ZP_00306098 (e-148)	Present in <i>Archaea</i> and in <i>Eukaryota</i> .	49.59
15	13803-14462	-	219	Unknown	None	4 predicted transmembrane helices	48.33
16	14494-16839	-	781	DNA polymerase, family B (COG0417)	<i>Picrophilus torridus</i> YP_022906 (0.0)	Distributed in the three domains	50.26
17	17032-17988	+	318	Fe-S oxidoreductase (COG0731)	<i>Pyrococcus horikoshii</i> NP_142898 (e-114)	Only present in <i>Euryarchaeota</i> and in <i>Bacteria</i> . In <i>Thermoplasmatales</i> , only present in <i>T. acidophilum</i>	53.71
18	18024-18392	+	122	Unknown	None	2 predicted transmembrane helices	46.61
19	18423-18701	+	92	Predicted transcriptional regulator containing an	<i>Pyrococcus furiosus</i> NP_578339	1 predicted transmembrane helix. Only present in	48.03

20	18885-19733	+	282	HTH domain fused to a Zn ribbon (COG3357) SAM-dependent methyltransferase (COG0500)	(1e-13) <i>Archaeoglobus fulgidus</i> NP_069435 (2e-17)	<i>Archaea</i> . Missing in <i>Thermoplasmales</i> Present in <i>Archaea</i> and in few <i>Bacteria</i>	41.58
21	19738-22305	-	855	Valyl tRNA synthetase (COG0525)	<i>Methanocaldococcus jannaschii</i> NP_248001 (0.0)	Distributed in the three domains	51.48
22	22346-23404	-	352	Geranylgeranyl pyrophosphate synthase (COG0142)	<i>Thermoplasma volcanium</i> NP_110781 (7e-67)	Distributed in the three domains	50.99
23	23512-24675	+	387	Pantothenate metabolism flavoprotein (COG0452)	<i>Archaeoglobus fulgidus</i> NP_070473 (3e-51)	Present in <i>Archaea</i> and in <i>Bacteria</i> . Phylogenetic affiliation within the <i>Euryarchaeota</i> unclear	50.43
24	24659-25474	+	271	Predicted archaeal kinase (COG1829)	<i>Methanosarcina mazei</i> NP_634306 (1e-23)	Only present in <i>Archaea</i> but missing in <i>Thermoplasmales</i> ; no close relatives	54.17
25	25547-26596	+	349	Unknown	None	1 predicted transmembrane helix and 1 predicted signal peptide	47.71
26	26605-26982	+	125	Unknown	None	4 predicted transmembrane helices	42.59
27	26992-28389	+	465	Archaeal DNA polymerase II, small subunit (DPI) (COG1311)	<i>Ferroplasma acidarmanus</i> ZP_00307183 (7e-94)	Only present in <i>Euryarchaeota</i> . Affiliated to the <i>Thermoplasmales</i>	48.43
28	28386-29369	-	327	Archaeal chlorohydrolyase family protein (COG0402)	<i>Picrophilus torridus</i> YP_023489 (1e-48)	Present in prokaryotes. Affiliated to the <i>Thermoplasmales</i> (low bootstrap proportion)	51.73
29	29396-29986	-	196	Uracyl DNA glycosylase (COG1573)	<i>Pyrococcus abyssi</i> CAB49606 (2e-58)	Present in prokaryotes. Affiliated to the <i>Thermococcales</i>	50.25
30	30036-30659	+	207	Thiamine phosphate pyrophosphorylase (COG0352)	<i>Pyrococcus furiosus</i> NP_579063 (8e-43)	Present in <i>Euryarchaeota</i> , in <i>Bacteria</i> and in few <i>Eukaryota</i> . Affiliated to the <i>Thermococcales</i>	54.17
31	30616-31044	-	142	Unknown	None	4 predicted transmembrane helices. 52 overlapping nucleotides with ORF18	45.92
32	31041-32126	-	361	Unknown	None	1 predicted transmembrane helix and 1 predicted signal peptide	44.75
rRNA	32210-33682			16S rRNA	None		60.83
33	33736-33912	-	58	Unknown	None		46.33
34	33941-34102	+	53	Unknown	None		43.83
35	34172-35359	+	395	Aspartate/tyrosine/aromatic aminotransferase (COG0436)	<i>Pyrococcus abyssi</i> CAB49738 (1-e113)	1 predicted transmembrane helix	49.58
36	35449-36132	+	227	SSU ribosomal protein S3AE (COG1890)	<i>Thermococcus kodakaraensis</i> BAD85443 (1-e43)	Distributed in the three domains. Affiliated to the <i>Thermococcales</i> Only present in <i>Archaea</i> and in <i>Eukaryota</i> . Affiliated to the <i>Thermoplasmales</i>	44.44
tRNA	36147-36257			tRNA ^{Met}		anticodon CAT	60.36
37	36261-36911	-	216	Uracil phosphoribosyltransferase (COG0035)	<i>Pyrococcus furiosus</i> NP_578970 (2e-66)	Distributed in the three domains. Affiliated to the <i>Thermococcales</i> ; missing in <i>Thermoplasmales</i>	49.77
38	36904-37505	-	133	Hypothetical conserved protein (COG1833)	<i>Methanocaldococcus jannaschii</i> NP_247597 (1e-17)	Present in <i>Archaea</i> and <i>Bacteria</i> . Missing in <i>Thermoplasmales</i> . Phylogeny not resolved	45.77
39	37380-38681	+	433	Metal-dependent phosphohydrolase, HD superfamily (COG1078)	<i>Thermococcus kodakaraensis</i> BAD84729 (5e-88)	Distributed in the three domains. Phylogenetic position within the <i>Euryarchaeota</i> not resolved	48.54

40	38694-39437	+	247	Hypothetical conserved protein	<i>Methanococcolides burtonii</i> ZP_00349116 (2e-29)	2 predicted transmembrane helices. Only present in few methanogens. Few bacterial and eukaryal homologues. Phylogeny not resolved.	45.77
41	39428-40474	-	348	L-alanine-DL-glutamate epimerase and related enzymes of enolase superfamily (COG4948)	<i>Picrophilus torridus</i> YP_022799 (9e-61)	Present in <i>Archaea</i> and <i>Bacteria</i> . Very poorly conserved protein; unclear phylogeny	49.19

Table 2. Predicted RNA, protein coding genes, and remarkable features in the euryarchaeotal fosmid Alv-FOS4. ORFs are numbered 1-44 following the sense 5' to 3' in Fig. 2. D, direction of transcription. Shaded ORF numbers indicate the co-linear region.

ORF	Nucleotide range	D	Protein size (aa)	Putative function (COG accession if available)	Most similar homologue (BLAST e-value <e-05)	Comments	% G+C
1	40-195	+	51	Unknown	None		47.44
2	232-966	+	244	Predicted archaeal kinase (COG1608)	<i>Thermoplasma acidophilum</i> NP_393581 (1e-37)	Distributed in the three domains. Operon-like structure formed by ORFs 2-3-4 shared with <i>Thermoplasmatales</i> only.	50.34
3	963-1976	+	337	Isopentenyl-diphosphate delta-isomerase, FMN-dependent (COG1304)	<i>Thermoplasma acidophilum</i> NP_393580 (4e-73)	Present in <i>Archaea</i> and <i>Bacteria</i>	54.14
4	1967-2584	+	205	Conserved protein containing ZPR1 zinc-finger domain (COG1779)	<i>Methanocaldococcus jannaschii</i> NP_247509 (7e-41)	Only present in <i>Archaea</i> and <i>Eukaryota</i> . Poorly resolved phylogeny	49.19
5	2548-3225	-	225	Sugar fermentation stimulation protein (COG1489)	<i>Methanococcoides burtonii</i> ZP_00349063 (1e-48)	Only present in prokaryotes	52.8
6	3194-3817	-	207	Hypothetical conserved protein	<i>Thermococcus kodakaraensis</i> BAD84931 (3e-08)	6 predicted transmembrane helices. Only present in few archaea. No homologs within <i>Thermoplasmatales</i>	48.72
7	3872-6862	-	996	Hypothetical conserved protein	<i>Methanococcus maripaludis</i> NP_987932 (2e-42)	2 predicted transmembrane helices and 1 predicted signal peptide. Composite protein containing a peptidase C1A, papain domain (IPR000668) and a fibronectin, type III domain (IPR003961)	49.08
8	6920-8251	-	443	Phosphoribosylamine-glycine ligase (COG0151)	<i>Methanothermobacter thermoautotrophicus</i> AAB85920 (e-117)	Distributed in the three domains. ORF8 and ORF9 are arranged in tandem. This organization is only present in <i>Thermoplasmatales</i> . Affiliated to the <i>Thermoplasmatales</i>	53.98
9	8244-8753	-	169	Predicted transcriptional regulator containing the HTH domain (COG1386)	<i>Archaeoglobus fulgidus</i> NP_070409 (6e-32)	Present in <i>Archaea</i> and <i>Bacteria</i>	50.78
10	8860-10314	+	484	Permeases of the major facilitator superfamily (COG0477)	<i>Pyrobaculum aerophilum</i> NP_560682 (e-114)	14 predicted transmembrane helices. Present in <i>Archaea</i> and <i>Bacteria</i>	53.81
11	10373-11530	+	385	Pantothenate metabolism flavoprotein (COG0452)	<i>Methanocaldococcus jannaschii</i> NP_247908 (7e-62)	Present in <i>Archaea</i> and <i>Bacteria</i> . Phylogenetic affiliation within the <i>Euryarchaeota</i> unclear	51.9
12	11514-12329	+	271	Predicted archaeal kinase (COG1829)	<i>Methanosarcina acetivorans</i> NP_616218 (4e-21)	Only present in <i>Archaea</i> but missing in <i>Thermoplasmatales</i> ; no close relatives	53.8
13	12402-13451	+	349	Unknown	None	1 predicted transmembrane helix and 1 predicted signal peptide	42
14	13460-13837	+	125	Unknown	None	4 predicted transmembrane helices	36.24
15	13847-15247	+	466	DNA polymerase II, small subunit (DPI) (COG1311)	<i>Picrophilus torridus</i> YP_023247 (2e-96)	Only present in <i>Euryarchaeota</i> . Affiliated with the <i>Thermoplasmatales</i>	46.25
16	15244-16230	-	328	Archaeal chlorohydrolyase family protein (COG0402)	<i>Picrophilus torridus</i> YP_023489 (8e-46)	Present in <i>Archaea</i> and <i>Bacteria</i> . Affiliated to the <i>Thermoplasmatales</i> (low bootstrap proportion)	53.5
17	16257-16847	-	196	Uracyl-DNA glycosylase (COG1573)	<i>Thermococcus kodakaraensis</i> BAD86332 (e-59)	Present in <i>Archaea</i> and <i>Bacteria</i> . Affiliated to the <i>Thermococcales</i>	51.95
18	16896-17522	+	208	Thiamine phosphate pyrophosphorylase (COG0352)	<i>Pyrococcus furiosus</i> NP_579063 (5e-48)	Present in <i>Euryarchaeota</i> , in <i>Bacteria</i> , and in few <i>Eukaryota</i> . Affiliated to the <i>Thermococcales</i>	54.23
19	17470-17898	-	142	Unknown	None	4 predicted transmembrane helices. 52 overlapping nucleotides with ORF18	46.39

20	17895-18980	-	361	Unknown	None	1 predicted transmembrane helix (a signal peptide was predicted in its homologue in FOS1)	37.75
rRNA	19062-20535			16S rRNA	None		60.79
21	20591-20818	-	75	Unknown	None		44.3
22	20819-22009	+	396	Aspartate/tyrosine/aromatic aminotransferase (COG0436)	<i>Pyrococcus abyssi</i> CAB49738 (e-113)	Distributed in the three domains. Affiliated to the <i>Thermococcales</i>	49.45
23	22044-22721	+	225	SSU ribosomal protein S3AE (COG1890)	<i>Thermococcus kodakaraensis</i> BAD85443 (1-e44)	Only present in <i>Archaea</i> and <i>Eukaryota</i> . Affiliated to the <i>Thermoplasmales</i>	44.69
tRNA ^{Met}	22736-22846			tRNA ^{Met}	None		60.36
24	22847-23497	-	216	Uracyl phosphoribosyltransferase (COG0035)	<i>Pyrococcus furiosus</i> NP_578970 (2e-72)	Distributed in the three domains. Affiliated to the <i>Thermococcales</i> ; missing in <i>Thermoplasmales</i>	50.08
25	23494-23892	-	132	Hypothetical conserved protein (COG1833)	<i>Methanothermobacter thermautotrophicus</i> AAB85146 (4e-17)	Present in <i>Archaea</i> and <i>Bacteria</i> . Missing in <i>Thermoplasmales</i> . Phylogeny not resolved	48.37
26	23967-25268	+	433	Metal-dependent phosphohydrolase, HD superfamily (COG1078)	<i>Thermococcus kodakaraensis</i> BAD84729 (1e-83)	Distributed in the three domains. Phylogenetic position within the <i>Euryarchaeota</i> not resolved	50.61
27	25272-26021	+	249	Hypothetical conserved protein	<i>Methanococoides burtonii</i> ZP_00349116 (7e-37)	2 predicted transmembrane helices. Only present in few methanogens. Few bacterial and eukaryal homologues. Phylogeny not resolved.	42.93
28	26018-27058	-	346	L-alanine-DL-glutamate epimerase and related enzymes of enolase superfamily (COG4948)	<i>Picrophilus torridus</i> YP_022799 (2e-59)	Only present in <i>Archaea</i> and <i>Bacteria</i> . Very poorly conserved protein, phylogeny unclear	52.74
29	27260-27529	+	89	Unknown	None	1 predicted transmembrane helix	50.74
30	27584-28564	+	326	Diphthamide synthase subunit DPH2 (COG1736)	<i>Methanococoides burtonii</i> ZP_00148013 (7e-66)	Present in <i>Archaea</i> and <i>Eukaryota</i> .	50.97
31	28603-30312	+	569	Unknown	None	2 predicted transmembrane helices and 1 predicted signal peptide	40.35
32	30286-31440	-	384	Aspartate aminotransferase (COG0436)	<i>Thermoplasma volcanium</i> NP_111524 (2e-94)	Distributed in the three domains	44.24
33	31738-32007	+	89	Nucleoid DNA binding protein (COG0776)	<i>Thermoplasma acidophilum</i> NP_393571 (4e-21)	Present in various <i>Bacteria</i> . Exclusively present in <i>Thermoplasmales</i> within the <i>Archaea</i>	49.26
34	32016-33104	-	362	DNA topoisomerase VI, subunit A (COG1697)	<i>Archaeoglobus fulgidus</i> NP_069773 (e-125)	Present in various <i>Archaea</i> and <i>Eukaryota</i> , but absent in <i>Thermoplasmales</i>	47.66
35	33094-34989	-	631	DNA topoisomerase VI, subunit B (COG1839)	<i>Methanococoides burtonii</i> ZP_00148378 (e-162)	Present in various <i>Archaea</i> and <i>Eukaryota</i> , but absent in <i>Thermoplasmales</i>	52.06
36	35185-35892	+	235	Hypothetical conserved protein	<i>Thermoplasma volcanium</i> NP_111729 (7e-20)	Only present in <i>Euryarchaeota</i>	44.77
37	35882-36208	+	108	Unknown	None		37.31
38	36205-37248	+	347	Deacetylase (COG0123)	<i>Archaeoglobus fulgidus</i> NP_068969 (2e-41)	Present in various <i>Archaea</i> , <i>Eukaryota</i> and <i>Bacteria</i> , but absent in <i>Thermoplasmales</i>	43.77
39	37255-37551	+	98	Unknown	None		37.04
40	37562-37996	+	144	Hypothetical conserved protein	<i>Geobacter metallireducens</i> ZP_00300493 (8e-05)	Very poorly conserved protein, only one bacterial homolog with e-value<8e-05	36.32
41	38122-38409	+	95	Unknown	None		39.58
42	38394-39227	+	277	Unknown	None		38.13
43	39217-40098	+	293	ATPase RecA-superfamily (COG0467)	<i>Archaeoglobus fulgidus</i> NP_069609 (8e-30)	Present in <i>Archaea</i> and <i>Bacteria</i>	43.54
44	40100-40885		261	Unknown	None		40.46

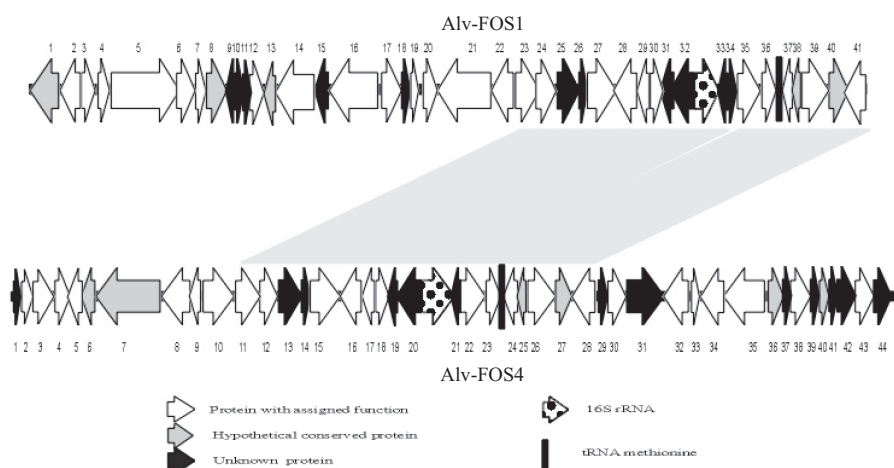


Fig 2. Schematic representation of the euryarchaeal clones Alv-FOS1 and Alv-FOS4. Open reading frames (ORFs) categorized in proteins with assigned functions (COGs), conserved hypothetical proteins, unknown proteins and rRNAs are differentiated as indicated in the inset. The tRNA genes are indicated by black bars. ORF numbers of the FOS1 and FOS4 fragments refer to Tables 1 and 2, respectively. Shading between the clones indicates homologous regions with the percentage of identities of deduced amino acid sequences.

A significant number of ORFs (17) were predicted to represent core components of information processing systems, e.g., involved in DNA replication/conformation/repair and in translation/transcription (47). The predicted genes belonging to this class encoded a MCM homolog (Alv-FOS1 ORF14), a DNA polymerase of the family B (DNA polB) (Alv-FOS1 ORF16), a small (DP1) (Alv-FOS1 ORF27 and Alv-FOS4 ORF 15) and large (DP2) (Alv-FOS1 ORF5) subunits of a DNA polymerase II (DNA pol II or family D DNA polymerase). The putative family B DNA polymerase exhibited the six conserved motifs indicative of the family B DNA polymerases and the tree motifs 3'-5' exonuclease motifs, suggesting that the protein exhibits both activities (62). All putative DNA polymerases predicted in both fosmids had *Thermoplasmatales* homologs as closest relatives (Table 1 and Table 2). Whereas all *Archaea* studied to date contain one or more DNA polymerases B (10), DNA polymerases belonging to the family D were only found to exist in *Euryarchaeota* (24, 58). Interestingly, the genes that were predicted to encode the small and large subunits of the family D DNA polymerase were separated. This situation is similar in all euryarchaeotal genomes sequenced to date, except in that of *Thermococcales* and of *Haloferax volcanii*, where the two genes were contiguous (38, 54). A GTPase (ORF6) adjacent to the DP2 subunit was predicted in Alv-FOS1. Comparative genome analyses revealed that this gene tandem arrangement was only encountered in the genomes of *Thermoplasmatales* except *P. torridus*. Similarly, the gene asso-

ciation formed by a hypothetical conserved protein and the MCM homolog (Alv-FOS1 ORFs 12-13) was specific to *Thermoplasmatales* and *Methanosarcinales*. Other gene clusters (ORFs 2-3-4 and ORFs 8-9) specific to *Thermoplasmatales* and the DHVE2 lineage were identified in Alv-FOS4 (Table 2).

The predicted proteins involved in the DNA conformation were all identified in Alv-FOS4. They consisted in a histone-like DNA binding protein (ORF33), two subunits of a DNA topoisomerase VI (ORF34 and ORF35) and a histone-deacetylase (ORF38). Interestingly, DNA topoisomerases VI are present in all archaea whose genomes have been sequenced except in *Thermoplasmatales* (21).

The predicted proteins involved in the transcription or the translation included protein S3AE which is a constituent protein of the small ribosomal subunit (Alv-FOS1 ORF36, Alv-FOS4 ORF23), the valyl-tRNA synthetase (Alv-FOS1 ORF21), the subunit DPH2 of the diptamide synthase (Alv-FOS4 ORF30), and two distinct putative transcriptional regulators containing the helix-turn-helix (HTH) domain (Alv-FOS1 ORF19 and Alv-FOS4 ORF9). The diptamide synthase enzyme catalyses the synthesis of the diptamide, a posttranslationally modified histidine residue present in the translation elongation factor 2 (35). The two putative transcriptional regulators have no apparent sequence similarity to each other; but both displayed a helix-turn-helix (HTH) domain which is a common denominator in basal and specific transcription factors for the three domains of life (3).

A large co-linear region (~16.6 kb) between Alv-FOS1 and Alv-FOS4 contained 18 ORFs, the 16S rRNA- and the Met-tRNA-encoding genes (Fig. 2). The only difference between both gene clusters corresponded to a small protein of unknown function (ORF34 of Alv-FOS1) absent in Alv-FOS4. Five other proteins of unknown functions were also encoded by the syntenic region. Most of them contained strong predicted transmembrane helices, suggesting that their products were membrane-anchored. The co-linear section included genes encoding putative proteins involved in the metabolism of coenzymes and cofactors (Alv-FOS1 ORF23 and Alv-FOS4 ORF11; Alv-FOS1 ORF30 and Alv-FOS4 ORF18), nucleotides (Alv-FOS1 ORF37 and Alv-FOS4 ORF23), amino acids (Alv-FOS1 ORF35 and Alv-FOS4 ORF 22) and nucleic acids (Alv-FOS1 ORF39 and Alv-FOS4 ORF28) (Table 1 and Table 2). The syntenic portion also contained an uracyl DNA glycosylase, an enzyme involved in DNA repair (50). The similarities of the predicted protein sequences within the co-linear region ranged from 58% (unknown protein encoded by Alv-FOS1 ORF26 and Alv-FOS4 ORF14) to 96% (metal-dependent phosphohydrolase). No synteny identical to that observed between Alv-FOS1 and Alv-FOS4 could be identified in other genomes.

Phylogenetic analysis of protein-coding genes and detection of horizontal gene

transfers.

Until now, the sisterhood of the DHVE2 lineage with the *Thermoplasmatales* was only inferred from phylogenetic analyses based on the 16S rRNA genes. To clarify the branching position of the lineage and to identify possible horizontal gene transfer (HGT) events, phylogenetic analyses were carried out using a set of well-conserved protein-coding genes predicted in Alv-FOS1 and/or Alv-FOS4.

The putative family B DNA polymerase identified in Alv-FOS1 exhibited the highest similarities with homologs from *Thermoplasmatales* (51-53% identity). This was confirmed by the Bayesian phylogenetic analysis of these sequences, which showed the emergence of the Alv-FOS1 DNA polymerase in a very well supported group (PP=1) with homologs from the *Thermoplasmatales* (Fig. 3A). Within this group, the Alv-FOS1 DNA polymerase branched with those of *Thermoplasma* spp. with a relative low posterior probability (PP=0.67). These results supported the sisterhood of Alv-FOS1 with the *Thermoplasmatales* deduced from the 16S rRNA data. It is interesting to note that this Bayesian phylogenetic tree of family B DNA polymerases failed to retrieve a clear division between the *Crenarchaeota* and the *Euryarchaeota*. This atypical phylogeny has been previously reported and explained by ancient duplication events of the genes in crenarchaeotes (10, 18).

Among the predicted proteins identified in the co-linear region and having homologs in databases, three yielded tree topologies similar of those obtained with the 16S rRNA genes and the Alv-FOS1 DNA polymerase. These proteins comprised a well conserved small subunit of the DNA polymerase II (Alv-FOS1 ORF27 and Alv-FOS4 ORF15), an archaeal protein of the chlorohydrolase family (Alv-FOS1 ORF28 and Alv-FOS4 ORF16), and the small subunit ribosomal protein S3AE (Alv-FOS1 ORF36 and Alv-FOS4 ORF23) (data not shown). According to the BLAST predictions, the latter protein had a homolog from *Thermococcus kodakaraensis* as closest relative (Table 1 and 2), suggesting a possible HGT event between Alv-FOS1 and Alv-FOS4 and *Thermococcales*. However, both phylogenetic analyses (neighbor-joining and Bayesian analyses) supported the affiliation of DHVE2 members close to the *Thermoplasmatales* (data not shown). This exemplified the difficulties to infer evolutionary relationships using BLAST rather than phylogenetic analyses (31).

DNA topoisomerases VI have been detected in all the archaeal genomes sequenced to date except in *Thermoplasmatales*. The phylogeny of the A subunit of topoisomerase VI is highly consistent with that based on 16S rRNA (21). Bayesian phylogenetic analysis showed that the A subunit of Alv-FOS4 topoisomerase emerged at a position occupied by the *Thermoplasmatales* in phylogenies reconstructed using 16S rRNA genes (Fig. 3B). It has been suggested that DNA topoisomerases VI were present in the last common ancestor of all archaea (21). If the DHVE2 lineage shared a common origin with the *Thermoplasmatales* (as supported by above phylogenetic analyses), this could suggest that the *Thermoplasmatales* have lost the genes en-

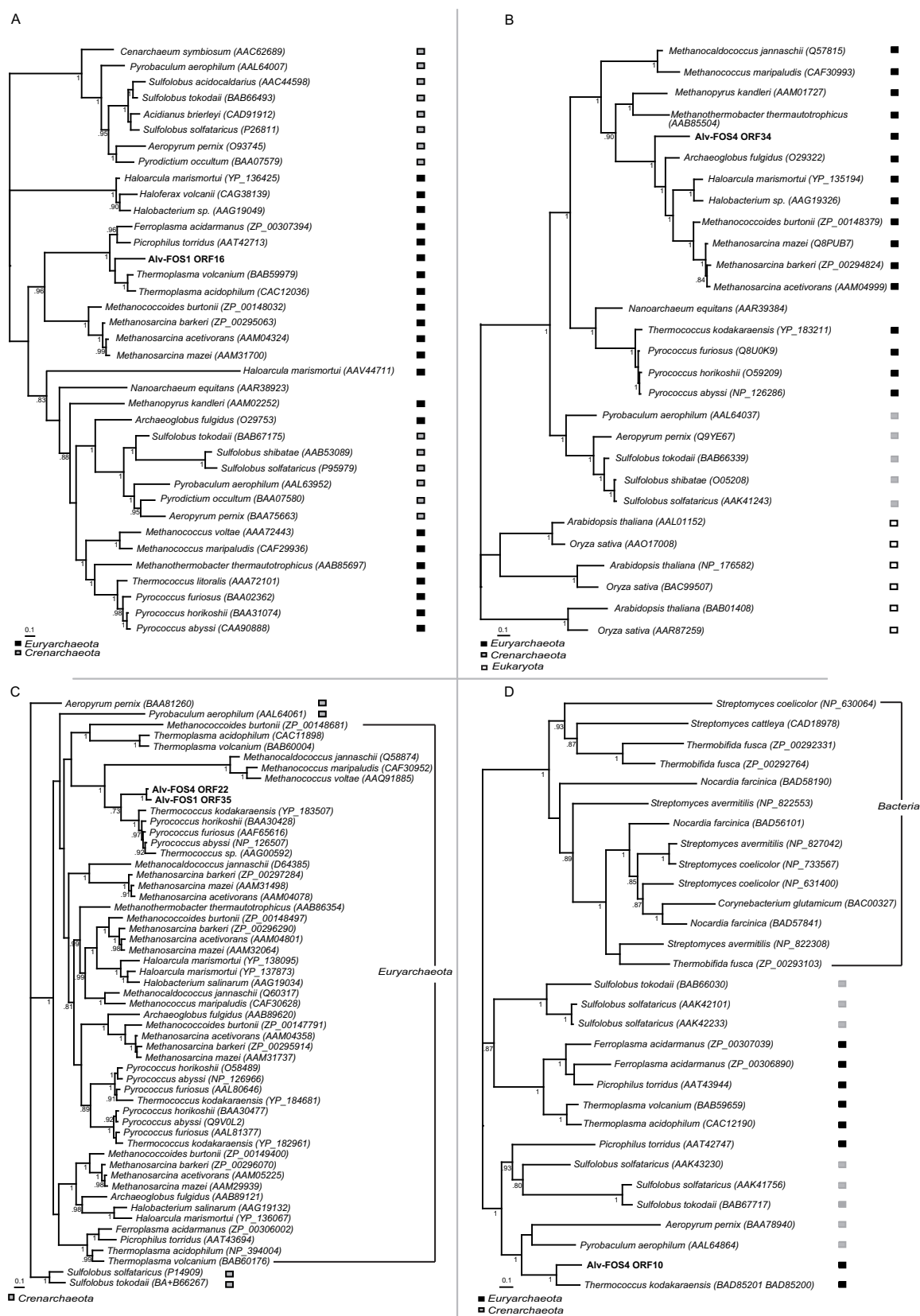


Fig 3. Phylogenetic trees of the (A) DNA polymerases B (330 amino acids), (B) subunits A of DNA topoisomerases VI (273 amino acids), (C) aspartate/tyrosine/aromatic aminotransferases (260 amino acids), and (D) permeases of the major facilitator superfamily (399 amino acids). Posterior probabilities > 0.8 are indicated at nodes; the scale bar represents the estimated number of amino acid substitutions per site.

coding DNA topoisomerases VI after their divergence from the DHVE2 lineage. Interestingly, our Bayesian phylogenetic analysis confirmed the position of the hyperthermophilic archaeal symbiont *Nanoarchaeum equitans* as sister group of *Thermococcales* within the *Euryarchaeota* (9) with a strong support (PP=1).

Unexpected branching positions were obtained for the uracyl DNA glycosylases, uracyl phosphoribosyltransferases, thiamine phosphate pyrophosphorylases and aspartate/tyrosine/aromatic aminotransferases predicted in both fragments. On the basis of BLAST and neighbor-joining analyses, these four genes were most closely related to their *Thermococcales* homologs. Bayesian phylogenetic analyses performed on the latter gene supported the proximity to the *Thermococcales* but failed to retrieve an unambiguous phylogenetic profile (Fig. 3C). This was probably due to the existence of many paralogues of this gene in most euryarchaeotal species. However, although supported by a relatively low posterior probability (PP=0.73), the branching position of Alv-FOS1 and Alv-FOS4 close to the *Thermococcales* remained consistent with the BLAST and neighbor-joining analysis. Our results suggest that the four genes have been likely acquired from *Thermococcales* by HGT events. An additional case of HGT from *Thermococcales* concerned the predicted permease of the major facilitator (FOS4 ORF10). This protein is present in genomes of most crenarchaeotal species sequenced to date and has only a few homologs within the *Euryarchaeota* (only predicted in the *Thermoplasmatales* and in *T. kodakaraensis*). Bayesian phylogenetic analysis supported with high posterior probabilities the grouping of *Thermoplasmatales* homologs with that of *Sulfolobales* (Fig. 3D), confirming other likely cases of HGT between the *Thermoplasmatales* and the *Crenarchaeota* (48, 61). Our phylogenetic analysis showed that the FOS4 permease gene clustered with that of *T. kodakaraensis*. As this gene is absent in other *Thermococcales*, this could suggest a recent case of HGT between Alv-FOS4 and *T. kodakaraensis*.

Expression and activity of Alv-FOS1 family B DNA polymerase.

To determine some biochemical properties of the Alv-FOS1 family B DNA polymerase (ORF16), the gene was amplified by PCR and cloned into the prokaryotic expression vector pAHR5. The DNA polymerase was produced in recombinant form in *E. coli* cells after IPTG induction. The protein was purified to near homogeneity and the SDS/PAGE analysis revealed a single band of ~90 kDa (data not shown). This value was in accordance with the molecular mass of 92.8 kDa derived from the amino acid sequence. The Alv-FOS1 DNA polymerase activity was tested in the incubation buffers designed for DNA polymerases B and D purified from *P. abyssi* (17, 24). The incubation buffer of DNA polymerase B [commercialized as Isis DNA polymerase (Q-biogene)] displayed an alkaline pH and a low Mg^{2+} concentration. The incubation buffer of the DNA polymerase D (27) was more acidic and had a high Mg^{2+} concentration. The DNA polymerase activity of the Alv-FOS1 DNA polymerase was detectable in both in-

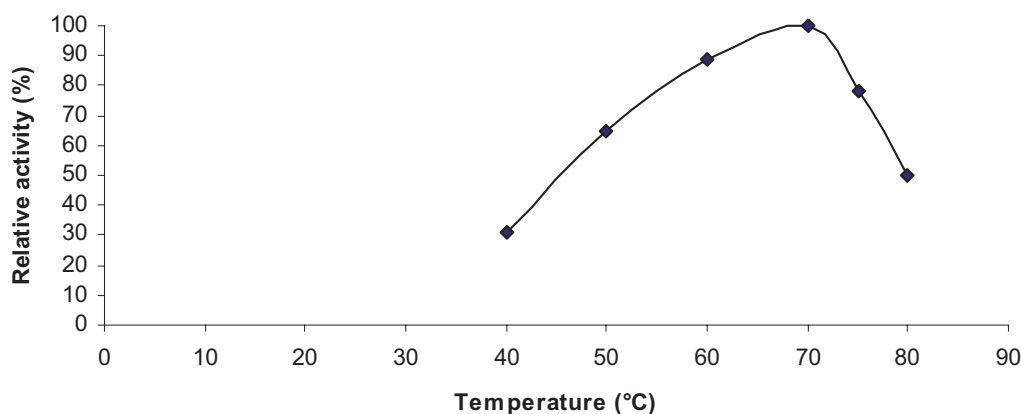


Fig 4. Effect of the temperature on Alv-FOS1 DNA polymerase activity. FOS1 DNA polymerase (1.4 pmol) was assayed under standard conditions for 20 min at the indicated temperature. DNA polymerase activity was determined by measuring the incorporation of [³H]dTTP into activated calf thymus DNA.

cubation buffers (data not shown) but was higher in that supplied with Isis DNA polymerase. Under these conditions, the DNA polymerase activity was found to be 0.3 U/ml.

Temperature optimum and thermostability.

The optimal temperature for the activity of Alv-FOS1 DNA polymerase and its thermostability were compared with those of representative DNA polymerases (Klenow fragment of *E. coli* DNA polymerase I, Taq and Isis DNA polymerases). The effect of elevated temperatures onto incorporation of nucleotides was limited by the thermal stability of the calf thymus DNA. The Alv-FOS1 DNA polymerase activity was optimal at 70°C (Fig. 4). At lower temperatures, incorporation of nucleotides decreased slowly and 30% of the optimal activity was observed at 40°C. At temperatures higher than 70°C, dNTP incorporation rapidly decreased, with approximately 50% of residual activity at 80°C. The behaviour of Alv-FOS1 DNA polymerase activity towards temperature exhibited nearly similar trends to that of several thermophiles and hyperthermophiles (42). However, the optimal temperature of Alv-FOS1 DNA polymerase could not be accurately determined since activated DNA was not stable above 75°C. To better evaluate its ability to resist to elevated temperatures, we compared its thermal stability with that of DNA polymerases from microorganisms of different thermal classes.

To test their thermostability, the Klenow fragment of *E. coli* DNA polymerase I, the Alv-FOS1, Taq and Isis DNA polymerases B were pre-incubated for 30 min at temperatures ranging from 4°C to 90°C and their ability to incorporate dNTPs into the DNA substrate was tested at 65°C for 20 min. The Alv-FOS1 DNA polymerase was found to be resistant (for 30 min) to pre-incubation temperatures up to 60°C (Fig. 5). Its thermostability decreased dramatically at higher temperatures and the residual activities, after pre-incubation at 70°C and

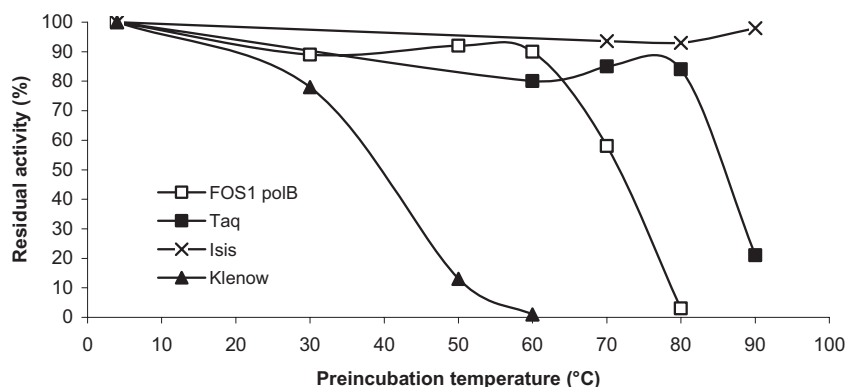


Fig 5. Thermostability of Alv-FOS1 DNA polymerase compared with that of *T. aquaticus* and *P. abyssi* and the Klenow fragment of *E. coli* DNA polymerase I. The residual DNA polymerase activity was measured for 20 min at 65°C after a 30-min pre-incubation of the enzymes at increasing temperatures.

80°C, were about 58% and 3%, respectively. The Alv-FOS1 DNA polymerase was much more thermostable than that of *E. coli* for which no residual activity was observed after a 30-min pre-incubation at 60°C. However, contrary to DNA polymerases purified from *P. abyssi* and *T. aquaticus*, Alv-FOS1 DNA polymerase was almost completely inactivated after a 30-min pre-incubation at 80°C. Its thermostability was comparable to that of the *T. acidophilum* DNA polymerase (25). Despite the similarities between the thermophilicity and the thermostability of their DNA polymerases, it is not possible to assume that the temperature range and optimal temperature for growth of Alv-FOS1 are very close to that of *T. acidophilum* (45–63°C and ~60°C, respectively) (13). However, the physical properties of Alv-FOS1 enzyme as determined here suggest that Alv-FOS1 is a thermophilic rather than a hyperthermophilic organism.

4 CONCLUSIONS

With molecular ecological studies that rely on the use of primers specific for the archaeal 16S rRNA gene or specific probes, the diversity and abundance of DHVE2 members have been well documented (28–29, 40–41, 46, 56). Based on their phylogenetic diversity and ecological distribution, members of DHVE2 lineage can be considered as likely candidates to represent a group having a significant impact in high-temperature hydrothermal habitats. Since they escaped cultivation despite valuable efforts (41), their phenotypic and metabolic features were unknown.

We initiated this study in order to provide access to genomic information for this group

of archaea. Some general properties can be deduced from the genomic fragments : (i) the separation of the 16S rRNA gene from 23S and 5S rRNA genes, (ii) a compact organization with short intergenic sequences, (iii) a G+C content of the rRNA and tRNA genes significantly higher than that of the genome fragment as a whole, and (iv) phylogenetic analyses with deduced protein sequences generally confirmed the affiliation of DHVE2 members with the *Thermoplasmatales* as predicted by 16S rRNA phylogeny. The identification of individual genes and gene clusters specific to both lineages provides further insight into the relationship of the DHVE2 lineage to the *Thermoplasmatales*. In contrast with these findings, the genome fragments analysed here also encoded proteins that have not been predicted in genomes of *Thermoplasmatales*, suggesting that this group has lost genes after their divergence from the DHVE2 lineage (e.g. the topoisomerase VI subunits). Based on 16S rRNA sequences, the two genomic fragments were closely related with only 2% sequence divergence. Although a co-linear region was identified, the genome fragments were not fully syntenic. The DNA regions upstream this common section were different, indicating genome rearrangements. Several other environmental genomic studies have revealed widespread microheterogeneity at the species level (6, 51), emphasizing the possible impact of these genomic variations on the functional diversity within very closely related organisms.

The G+C content of the rRNA and tRNA genes of both genome fragments was significantly higher than that of the genome fragment as a whole. This feature is common to all thermophiles and hyperthermophiles sequenced so far. On the basis of the high G+C content of their 16S rRNA genes, a thermophilic lifestyle had been speculated for members of the DHVE2 lineage. The biochemical characterization of the DNA polymerase encoded by Alv-FOS1 demonstrated that this protein has temperature requirements similar to that of its thermophilic homologs. Therefore, our expression study substantiated the prediction that Alv-FOS1 (and most likely its relatives of the DHVE2 lineage) are thermophiles.

By providing new insight into the evolution, mode of life and diversity of members of the DHVE2 lineage, our study illustrates the potential of the metagenomic approach. Our analysis identified several genes indicative of metabolic pathways. Unfortunately, these informations were not sufficient to predict precise metabolic traits that could suggest useful strategies to elaborate culture media. Better knowledge of the environmental distribution and abundance of DHVE2 members could provide strategies for capturing new genomic sequences of this particular phylogenetic group and/or direct selective isolation attempts.

Acknowledgment

We are grateful to Nadine Le Bris and Françoise Gaill, chief scientist and project leader of the Phare cruise for having inviting us on board. Didier Flament and Gaël Erauso

are gratefully acknowledged for valuable discussions and advices. We thank the captain and crews of 'L'Atalante' and the pilots and support crew of the ROV 'Victor'. We greatly acknowledge Emmanuelle Morin and Patrick Durand (Irisa-Inria, Rennes) for their help in the use of the Genostar platform. This work was financed by the CNRS program "Geomex" and by the CNRS-Ministère de la Recherche program "Séquençage à Grande Échelle". H. M. was supported by a grant from the Ministère de la Recherche.

Bibliographie

- [1] **Allen, E. E., and J. F. Banfield.** 2005. Community genomics in microbial ecology and evolution. *Nature Rev. Microbiol.* **3** : 489-498.
- [2] **Altschul, S., T. Madden, A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. Lipman.** 1997. Gapped BLAST and PSI-BLAST : a new generation of protein database search programs. *Nucl. Acids Res.* **25** : 3389-3402.
- [3] **Aravind, L., V. Anantharaman, S. Balaji, M. M. Babu, and L. M. Iyer.** 2005. The many faces of the helix-turn-helix domain : transcription regulation and beyond. *FEMS Microbiol. Rev.* **29** : 231-262.
- [4] **Béjà, O., L. Aravind, E. V. Koonin, M. T. Suzuki, A. Hadd, L. P. Nguyen, S. B. Jovanovich, C. M. Gates, R. A. Feldman, J. L. Spudich, E. N. Spudich, and E. F. DeLong.** 2000. Bacterial rhodopsin : evidence for a new type of phototrophy in the sea. *Science* **289** : 1902-1906.
- [5] **Béjà, O., E. N. Spudich, J. L. Spudich, M. Leclerc, and E. F. DeLong.** 2001. Proteorhodopsin phototrophy in the ocean. *Nature* **411** : 786-789.
- [6] **Béjà, O., E. V. Koonin, L. Aravind, L. T. Taylor, H. Seitz, J. L. Stein, D. C. Bensen, R. A. Feldman, R. V. Swanson, and E. F. DeLong.** 2002a. Comparative genomic analysis of archaeal genotypic variants in a single population and in two different oceanic provinces. *Appl. Environ. Microbiol.* **68** : 335-345.
- [7] **Béjà, O., M. T. Suzuki, J. F. Heidelberg, W.C. Nelson, C. M. Preston, T. Hamada, J. A. Eisen, C. M. Fraser, and E. F. DeLong.** 2002b. Unsuspected diversity among marine aerobic anoxygenic phototrophs. *Nature* **415** : 630-633.
- [8] **Bradford, M. M.** 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72** : 248-254.
- [9] **Brochier, C., S. Gribaldo, Y. Zivanovic, F. Confalonieri, and P. Forterre.** 2005. *Nanoarchaeota* : representatives of a novel archaeal phylum or a fast-evolving euryarchaeal lineage related to *Thermococcales*. *Genome Biol.* **6** : R42.
- [10] **Cann, I. K., and Y. Ishino.** 1999. Archaeal DNA replication : identifying the pieces to solve a puzzle. *Genetics* **152** : 1249-1267.
- [11] **Dalgaard, J., and R. Garrett.** 1993. Archaeal hyperthermophilic genes, p. 535-563. In M. Kates, D. J. Kushner, and A. T. Matheson (ed.), *The biochemistry of Archaea (Archaeobacteria)*. Elsevier, vol 26, Amsterdam, The Netherlands.
- [12] **Daniel, R.** 2005. The metagenomics of soil. *Nature Rev. Microbiol.* **3** : 470-478.
- [13] **Darland, G., T. D. Brock, W. Samsonoff, and S. F. Conti.** 1970. A thermophilic, acidophilic mycoplasma isolated from a refuse pile. *Science* **170** : 1416-1418.
- [14] **DeLong, E. F.** 2005. Microbial community genomics in the ocean. *Nature Rev.* **3** : 459-469.
- [15] **De Moerlooze, L., I. Struman, A. Renard, and J. A. Martial.** 1992. Stabilization of T7-promoter-based pARHS expression vectors using the parB locus. *Gene* **119** : 91-93.
- [16] **Desbruyères, D., P. Chevaldonné, A. M. Alayse, D. Jollivet, F. H. Lallier, J.-T. C., F. Zal, P. M. Sarradin, R. Cosson, J. C. Caprais, C. Arndt, J. O'Brien, J. Guezennec, S. Hourdez, R. Riso, F. Gaill, L. Laubier, and A. Toulmond.** 1998. Biology and ecology of the "Pompeii worm" (*Alvinella pompejana* Desbruyères and Laubier), a normal dweller of an extreme deep-sea environment : a synthesis of current knowledge and recent developments. *Deep-Sea Res. II* **45** : 383-422.
- [17] **Dietrich, J., P. Schmitt, M. Zieger, B. Preve, J. L. Rolland, H. Chaabihi, and Y. Gueguen.** 2002. PCR performance of the highly thermostable proof-reading B-type DNA polymerase from *Pyrococcus abyssi*. *FEMS Microbiol. Lett.* **217** : 89-94.
- [18] **Edgell, D. R., S-B. Malik, and W. F. Doolittle.** 1998. Evidence of independent gene duplications during the evolution of archaeal and eukaryotic family B DNA polymerases. *Mol. Biol. Evol.* **15** : 1207-1217.
- [19] **Edwards, R. A., and F. Rohwer.** 2005. Viral metagenomics. *Nature Rev. Microbiol.* **3** : 504-509.

- [20] **Futterer, O., A. Angelov, H. Liesegang, G. Gottschalk, C. Schleper, B. Schepers, C. Dock, G. Antranikian, and W. Liebl.** 2004. Genome sequence of *Picrophilus torridus* and its implications for life around pH 0. *Proc. Natl. Acad. Sci. USA* **101** : 9091-9096.
- [21] **Gadelle, D., J. Filee, C. Buhler, and P. Forterre.** 2003. Phylogenomics of type II DNA topoisomerases. *Bioessays* **25** : 232-242.
- [22] **Galtier, N., N. Tourasse, and M. Gouy.** 1999. A nonhyperthermophilic common ancestor to extant life forms. *Science* **283** : 220-221.
- [23] **Golyshina, O. V., T. A. Piparova, G. I. Karavaiko, T. F. Kondrat'eva, E. R. B. Moore, W. R. Abraham, H. Lünsdorf, K. N. Timmis, M. M. Yakimov, and P. N. Golyshin.** 2000. *Ferroplasma acidiphilum* gen. nov., sp. nov., an acidophilic, autotrophic, ferrous-oxidizing, cell-wall-lacking, mesophilic member of the *Ferropasmaceae* fam. nov., comprising a distinct lineage of the *Archaea*. *Int. J. Syst. Evol. Microbiol.* **50** : 997-1006.
- [24] **Gueguen, Y., J. L. Rolland, O. Lecompte, P. Azam, G. Le Romancer, D. Flament, J. P. Raffin, and J. Dietrich.** 2001. Characterization of two DNA polymerases from the hyperthermophilic euryarchaeon *Pyrococcus abyssi*. *Eur. J. Biochem.* **268** : 5961-5969.
- [25] **Hamal, A., P. Forterre, and C. Elie.** 1990. Purification and characterization of a DNA polymerase from the archaeobacterium *Thermoplasma acidophilum*. *Eur. J. Biochem.* **190** : 517-521.
- [26] **Handelsman, J.** 2004. Metagenomics : application of genomics to uncultured microorganisms. *Microbiol. Mol. Biol. Rev.* **68** : 669-685.
- [27] **Henneke, G., D. Flament, U. Hübscher, J. Quérellou, and J. P. Raffin.** 2005. The hyperthermophilic *Euryarchaeota* *Pyrococcus abyssi* likely requires the two DNA polymerases D and B for DNA replication. *J. Mol. Biol.* **350** : 53-64.
- [28] **Hoek, J., A. Banta, F. Hubler, and A.-L. Reysenbach.** 2003. Microbial diversity of a sulphide spire located in the Edmond deep-sea hydrothermal vent field on the Central Indian Ridge. *Geobiology* **1** : 119-127.
- [29] **Huber, J. A., D. A. Butterfield, and J. A. Baross.** 2002. Temporal changes in archaeal diversity and chemistry in a mid-ocean ridge seafloor habitat. *Appl. Environ. Microbiol.* **68** : 1585-1594.
- [30] **Huelsenbeck, J. P., and F. Ronquist.** 2001. MrBayes : Bayesian inference of phylogenetic trees. *Bioinformatics* **17** : 754-755.
- [31] **Koski, L. B., and G. B. Golding.** 2001. The closest BLAST hit is often not the nearest neighbor. *J. Mol. Evol.* **52** : 540-542.
- [32] **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227** : 680-685.
- [33] **López-García, P., C. Brochier, D. Moreira, and F. Rodríguez-Valera.** 2004. Comparative analysis of a genome fragment of an uncultivated mesopelagic cre-narchaeote reveals multiple horizontal gene transfers. *Environ. Microbiol.* **6** : 19-34.
- [34] **Lorenz, P., and J. Eck.** 2005. Metagenomics and industrial applications. *nature Rev. Microbiol.* **3** : 510-516.
- [35] **Moehring, J. M., and T. J. Moehring.** 1988. The post-translational trimethylation of diphthamide studied in vitro. *J. Biol. Chem.* **263** : 3840-3844.
- [36] **Moreira, D., F. Rodríguez-Valera, and P. López-García.** 2004. Analysis of a genome fragment of a deep-sea uncultivated Group II euryarchaeote containing 16S rDNA, a spectinomycin-like operon and several energy metabolism genes. *Environ. Microbiol.* **6** : 959-969.
- [37] **Moussard, H., D. Moreira, M. A. Cambon-Bonavita, P. López-García, and C. Jeanthon.** Uncultured *Archaea* in a hydrothermal microbial assemblage : Phylogenetic diversity and characterization of a genome fragment from an euryarchaeote. *FEMS Microbiol. Ecol.* **57** : 452-469.
- [38] **Myllykallio, H., P. Lopez, P. López-García, R. Heilig, W. Saurin, Y. Zivanovic, H. Philippe, and P. Forterre.** 2000. Bacterial mode of replication with eukaryotic-like machinery in a hyperthermophilic archaeon. *Science* **288** : 2212-2215.
- [39] **Nercessian, O., N. Bienvenu, D. Moreira, D. Prieur, and C. Jeanthon.** 2005. Diversity of functional genes of methanogens, methanotrophs, and sulfate reducers in deep-sea hydrothermal environments. *Environ. Microbiol.* **5** : 118-132.
- [40] **Nercessian, O., M. Prokof'eva, A. Lebedinski, S. L'Haridon, C. Cary, D. Prieur, and C. Jeanthon.** 2004. Design of 16S rRNA-targeted oligonucleotide probes for detecting cultured and uncultured archaeal lineages in high-temperature environments. *Environ. Microbiol.* **6** : 170-182.
- [41] **Nercessian, O., A. L. Reysenbach, D. Prieur, and C. Jeanthon.** 2003. Archaeal diversity associated with in situ samplers deployed on hydrothermal vents on the East Pacific Rise (13 degrees N). *Environ. Microbiol.* **5** : 492-502.

- [42] **Perler, F. B., S. Kumar, and H. Kong.** 1996. Thermostable DNA polymerases. *Adv. Protein Chem.* **48** : 377-435.
- [43] **Philippe, H.** 1993. MUST, a computer package of management utilities for sequences and trees. *Nucleic Acids Res.* **21** : 5264-5272.
- [44] **Ree, H. K., and R. A. Zimmermann.** 1990. Organization and expression of the 16S, 23S and 5S ribosomal RNA genes from the archaeobacterium *Thermoplasma acidophilum*. *Nucleic Acids Res.* **18** : 4471-4478.
- [45] **Reysenbach, A. L.** 2001. *Thermoplasmatales* ord. nov., p. 335. In D. R. Boone, R. W. Castenholz, and G. Garrity (ed.), *Bergey's manual of systematic bacteriology*, Springer-Verlag, vol. 1, New-York, N.Y.
- [46] **Reysenbach, A. L., K. Longnecker, and J. Kirshtein.** 2000. Novel bacterial and archaeal lineages from an in situ growth chamber deployed at a Mid-Atlantic Ridge hydrothermal vent. *Appl. Environ. Microbiol.* **66** : 3798-3806.
- [47] **Rivera, M. C., R. Javi, J. E. Moore, and J. A. Lake.** 1998. Genomic evidence for two functionally distinct gene classes. *Proc. Natl. Acad. Sci. USA* **95** : 6239-6244.
- [48] **Ruepp, A., W. Graml, M.-L. Santos-Martinez, K. K. Koretke, C. Volker, H. W. Mewes, D. Frishman, S. Stocker, A. N. Lupas, and W. Baumeister.** 2000. The genome sequence of the thermoacidophilic scavenger *Thermoplasma acidophilum*. *Nature* **407** : 508-513.
- [49] **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. *Molecular cloning : a laboratory manual*. Cold Spring Harbor Laboratory Press, New York.
- [50] **Sartori, A. A., S. Fitz-Gibbon, H. Yang, J. H. Miller, and J. Jiricny.** 2002. A novel uracil-DNA glycosylase with broad substrate specificity and an unusual active site. *EMBO J.* **21** : 3182-3191.
- [51] **Schleper, C., E. F. DeLong, C. M. Preston, R. A. Feldman, K. Y. Wu, and R. V. Swanson.** 1998. Genomic analysis reveals chromosomal variation in natural populations of the uncultured psychrophilic archaeon *Cenarchaeum symbiosum*. *J. Bacteriol.* **180** : 5003-5009.
- [52] **Schleper, C., G. Jurgens, and M. Jonuscheit.** 2005. Genomic studies of uncultivated archaea. *Nature Rev. Microbiol.* **3** : 479-488.
- [53] **Schleper, C., R. V. Swanson, E. J. Mathur, and E. F. DeLong.** 1997. Characterization of a DNA polymerase from the uncultivated psychrophilic archaeon *Cenarchaeum symbiosum*. *J. Bacteriol.* **179** : 7803-7811.
- [54] **Shen, Y., X.-F. Tang, and I. Matsui.** 2003. Subunit interaction and regulation of activity through terminal domains of the family D DNA polymerase from *Pyrococcus horikoshii*. *J. Biol. Chem.* **278** : 21247-21257.
- [55] **Slesarev, A. I., K. V. Mezhevaya, K. S. Makarova, N. N. polushin, O. V. Shscherbinina, V. V. Shakhova, G. I. Belova, L. Aravind, D. A. Natale, I. B. Rogozin, R. L. Tatusov, Y. I. Wolf, K. O. Stetter, A. G. Malykh, E. V. Koonin, and S. A. Kozyavkin.** 2002. The complete genome of hyperthermophile *Methanopyrus kandleri* AV19 and monophyly of archaeal methanogens. *Proc. Natl. Acad. Sci. USA* **99** : 4644-4649.
- [56] **Takai, K., and K. Horikoshi.** 1999. Genetic diversity of *Archaea* in deep-sea hydrothermal vent environments. *Genetics* **152** : 1285-1297.
- [57] **Takai, K., T. Komatsu, F. Inagaki, and K. Horikoshi.** 2001. Distribution of archaea in a black smoker chimney structure. *Appl. Environ. Microbiol.* **67** : 3618-3629.
- [58] **Tang, X. F., Y. Shen, E. Matsui, and I. Matsui.** 2004. Domain topology of the DNA polymerase D complex from a hyperthermophilic archaeon *Pyrococcus horikoshii*. *Biochemistry* **43** : 11818-11827.
- [59] **Tatusov, R. L., D. A. Natale, I. V. Garkavtsev, T. A. Tatusova, U. T. Shankavaram, B. S. Rao, B. Kiryutin, M. Y. Galperin, N. D. Fedorova, and E. V. Koonin.** 2001. The COG database : new developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res.* **29** : 22-28.
- [60] **Thompson, J. D., D. G. Higgins, and T. J. Gibson.** 1994. *CLUSTAL W* : improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22** : 4673-4680.
- [61] **Tyson, G.W., J. Chapman, P. Hugenholtz, E. E. Allen, R. J. Ram, P. M. Richardson, V.V. Solovyev, E. M. Rubin, D. S. Rokhsar, and J. F. Banfield.** (2004). Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **428** : 37-43.
- [62] **Zhao, Y., D. Jeruzalmi, I. Moarefi, L. Leighton, R. Lasken, and J. Kuriyan.** (1999). Crystal structure of an archaeal DNA polymerase. *Structure* **7** : 1189-1199.

Nous disposons de résultats complémentaires pour cette étude. Ces résultats complémentaires sont présentés dans les paragraphes suivants.

1 Prédiction automatique des gènes codant des protéines

Différents programmes, utilisant des méthodes fondées sur les modèles de chaînes de Markov ou sur les modèles des chaînes de Markov cachées (Hidden Markov Models ou HMM), ont été testés pour prédire les gènes de Alv-FOS1 et Alv-FOS4. Ces programmes sont les suivants :

- GeneMark.hmm (http://opal.biology.gatech.edu/GeneMark/gmhmm2_prok.cgi),
- FGENESB (<http://www.softberry.com>),
- et celui proposé par le module GenoAnnot de la plateforme Genostar (<http://www.genostar.org>).

Les paramètres par défaut de ces trois programmes ont été presque exclusivement choisis pour prédire les gènes. Le résultat des prédictions est présenté dans la Tableau 1.

	Genostar	GeneMark	FGENESB
Alv-FOS1	66	42	52
Alv-FOS4	82	46	48

Tableau 1 – Nombre de gènes prédits chez Alv-FOS1 et Alv-FOS4 par les différents programmes utilisés.

Nous avons préféré poursuivre l'étude en utilisant les résultats du programme qui prédisait le plus grand nombre de gènes, à savoir le programme proposé par Genostar.

2 Recherche des signaux de régulation des gènes codant des protéines

Définir les positions des CDS (coding sequence) implique de trouver un codon d'initiation et un triplet de terminaison (codon stop) de la traduction. Il n'y a pas d'ambiguïté possible dans le choix du codon stop, puisque ces codons ne codent pas d'acides aminés. Cependant, le choix du codon d'initiation est problématique puisque les codons ATG, GTG et TTG servent à la fois à initier la traduction mais également à coder l'acide aminé méthionine, valine et leucine respectivement.

Afin de préciser les positions des sites bordant les CDS (positions prédites par Genostar), ce qui revient entre autre à identifier le "vrai" codon d'initiation, nous avons fait une recherche manuelle des signaux de régulation de l'expression des gènes, à savoir ceux de la région promotrice de la transcription, ainsi que les sites de fixation du ribosome (RBS).

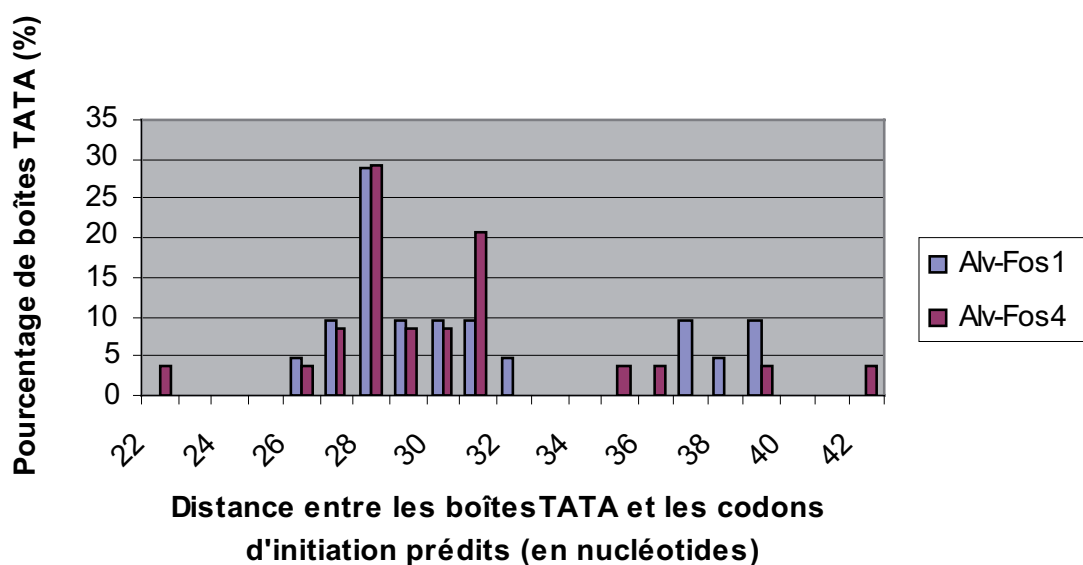


Fig. 1 – Distribution des distances entre les boîtes TATA et les codons d'initiation de la traduction prédits.

2.1 Recherche des régions promotrices de la transcription : boîtes TATA et BRE

Les plus proches représentants cultivés de Alv-FOS1 et Alv-FOS4 sont affiliés aux *Euryarchaeota*. D'après la figure 7 du chapitre 1 de l'introduction de ce manuscrit de thèse, la séquence consensus de la boîte TATA des *Euryarchaeota* (consensus des halophiles, méthanogènes et *Pyrococcus*) est la suivante : YTWWW où Y=(C ou T) et W=(A ou T). Cette séquence est généralement localisée de -25 à -32 nucléotides en amont du +1 de la transcription.

La recherche manuelle de la boîte TATA a été effectuée sur l'ensemble des gènes prédits par Genostar. Environ la moitié des boîtes TATA identifiées chez Alv-FOS1 et Alv-FOS4, correspondait à la séquence consensus recherchée. La plupart des autres boîtes TATA ne s'éloignait de la séquence consensus que d'un seul nucléotide (T du consensus remplacé par un A). Une certaine flexibilité de la distance entre la boîte TATA et le codon initiateur présumé de la traduction a été observée (Fig. 1). Cependant la majorité des boîtes TATA était localisée entre les positions -25 et -32. Une étude récente portant sur l'étude de la position des motifs nucléiques conservés de *T. acidophilum* et *T. volcanium*, a montré que pour ces génomes, les boîtes TATA étaient centrées aux positions -26 / -27 (Torarinnsson *et al.*, 2005).

Des séquences riches en A et G pouvant correspondre au BRE ont été identifiées en amont (-3 à -6 nucléotides) de la plupart des boîtes TATA identifiées (voir §5.2.1 du chapitre 1 de l'introduction).

2.2 Recherche des sites de fixation du ribosome

Un seul promoteur de la traduction a été recherché, il s'agit du site de fixation du ribosome (RBS). Le ribosome se fixe sur le RBS au moment de la phase d'initiation de la traduction. Ce site est placé à l'extrémité 5' de l'ARNm et est localisé 8 à 10 nucléotides en amont du site de démarrage de la traduction. Il est caractérisé par un motif Shine-Dalgarno, séquence de 6 nucléotides complémentaire à l'extrémité 3' de l'ARNr 16S. Ce motif est sujet à peu de variation chez les procaryotes, reflet de la conservation de l'ARNr 16S.

L'algorithme de RBSfinder utilise cette propriété essentielle des gènes pour localiser le RBS. Le programme travaille sur les 50 nucléotides précédant le codon initiateur de la traduction par recherche du consensus du RBS (AGGAGG). Les RBS correspondant au consensus étaient souvent prédits à plus de 20 nucléotides en amont du codon initiateur. Comme les RBS de *M. jannaschii*, l'organisme cultivé le plus proche de Alv-Fos 1 et Alv-Fos4 sur la base de l'ARNr 16S, sont localisés pour la plupart aux positions -3 à -7 du codon d'initiation de la traduction (Hayes & Borodovsky, 1998), nous avons préféré rechercher manuellement les RBS. Pour cela, nous avons utilisé le consensus du RBS de *M. jannaschii* DRDGD (D = A ou G ou T et R = A ou G).

La plupart des RBS identifiés étaient localisés conformément aux résultats de Hayes & Borodovsky (1998).

2.3 Usage des codons d'initiation de la traduction

Le codon ATG est le codon d'initiation de la traduction le plus utilisé chez Alv-FOS1 et Alv-FOS4, constituant respectivement 75 et 77% des codons d'initiation prédits. Chez Alv-FOS1, les codons GTG et TTG sont ensuite équitablement utilisés; tandis que chez Alv-FOS4, ils se répartissent en 16% de GTG et 7% de TTG.

3 Comparaison de la séquence de l'ADN polB d'Alv-FOS1 avec les autres ADN polB connues

La séquence protéique de l'ADN polB de Alv-FOS1 a été alignée avec celles de toutes les *Euryarchaeota* disponibles. Cette séquence a pu être très facilement alignée avec celles des autres *Thermoplasmatales* disponibles. La figure 2 illustre cet alignement multiple, où seules les ADN polB de Alv-FOS1, des *Thermoplasmatales* et de *P. furiosus* y sont représentées. Les six motifs conservés des ADN polymérase de la famille B ont pu être localisés dans la séquence déduite de l'ADN polB de Alv-FOS1 (Braithwaite & Ito, 1993 ; Zhao *et al.*, 1999). Les trois motifs 3'-5' exonucléases typiquement identifiés dans les ADN polB ont également été localisés sur cette sé-

quence.

D'une manière générale, les motifs polymérase et exonucléase des ADN polB d'*Archaea* sont très bien conservés, tandis que les régions séparant ces motifs le sont beaucoup moins (Schleper *et al.*, 1997b). Il est de ce fait important de noter, que l'ADN polB d'Alv-FOS1 s'alignait très bien avec ses homologues *Thermoplasmatales*, sans l'introduction d'importants gaps ou insertions.

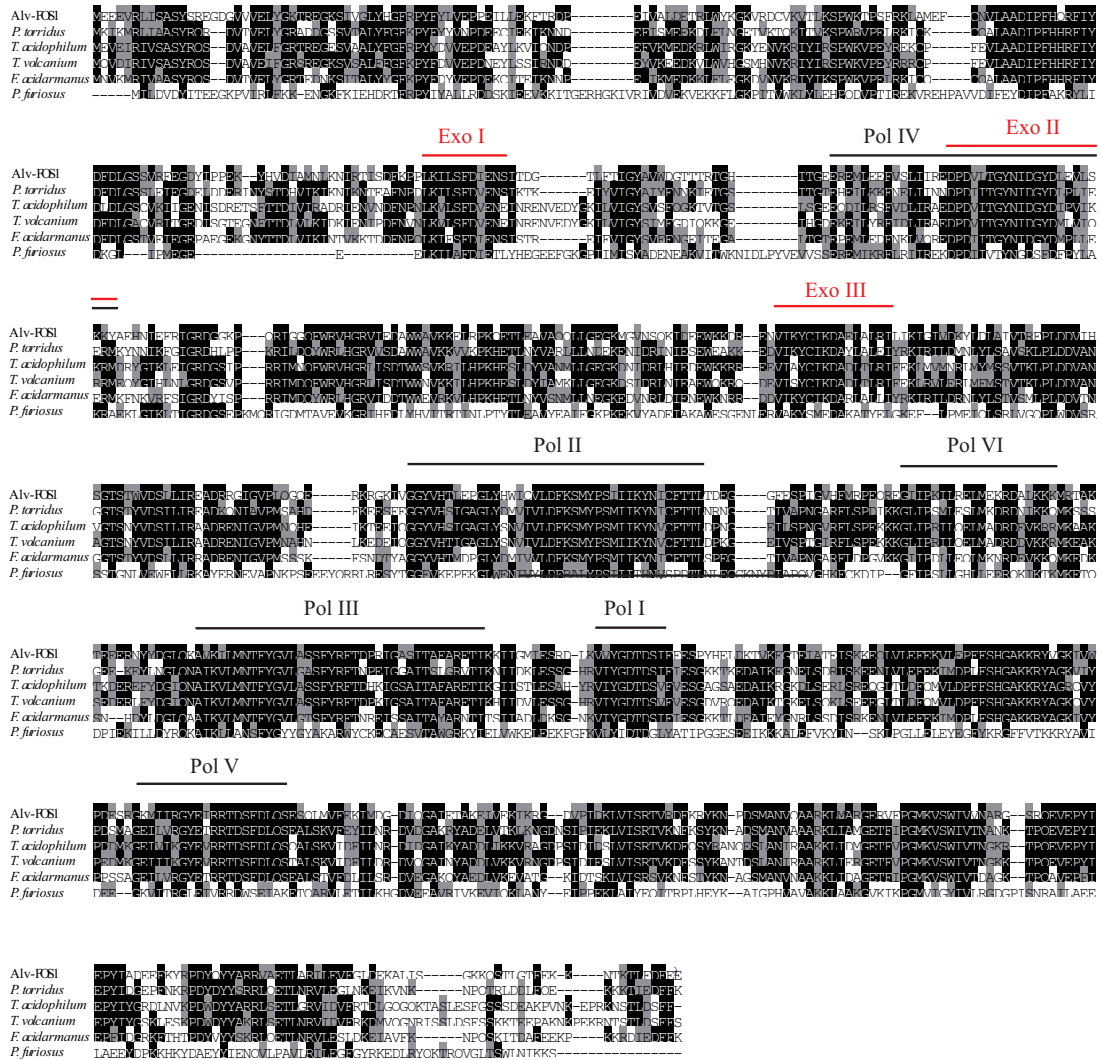


Fig. 2 – Alignement de séquences d'ADN polB de différentes *Euryarchaeota*. Les séquences ont été alignées avec ClustalW. Les régions Pol I à Pol VI représentent les six domaines polymérase conservées des ADN polymérases de la famille B. Les trois domaines exonucléases (3'-5' exonucléase) sont notés Exo I à III. Les écritures blanches sur fond noir représentent des blocs d'acides aminés identiques. Les écritures noires sur fond gris représentent des blocs d'acides aminés similaires. Les écritures noires sur fond blanc représentent les acides aminés qui ne sont pas conservés.

Discussion générale

Au sein de l'écosystème hydrothermal marin profond se développe une faune riche et abondante. Cet environnement est particulièrement dynamique. Il se caractérise par une variabilité spatiale et temporelle très importante, qui est la conséquence du mélange du fluide hydrothermal avec l'eau de mer environnante. Ce mélange entre l'eau de mer et le fluide hydrothermal crée d'abrupts gradients physico-chimiques qui définissent une multitude de micro-niches propices au développement de communautés microbiennes très diversifiées.

Les méthodes d'écologie microbienne, dépendantes ou non de la mise en culture des organismes, ont révélé une extraordinaire diversité microbienne au sein de cet écosystème. Cette diversité s'exprime tant au point de vue morphologique que physiologique, génétique et métabolique. La détection de nouvelles lignées génétiques pour lesquelles aucun représentant n'est cultivé à ce jour, est courante. Le but de ce travail de thèse était de rechercher des informations sur les micro-organismes incultivés de l'écosystème hydrothermal. Les communautés microbiennes occupant les habitats exposés aux fluides hydrothermaux étaient plus particulièrement visées.

Pour répondre à cet objectif, notre choix s'est porté sur l'utilisation de deux approches moléculaires indépendantes de la mise en culture des organismes. La première, décrite dans l'étude 1, a consisté à inventorier la diversité à l'aide d'un marqueur moléculaire phylogénétique (le gène codant pour l'ARNr 16S) et d'un marqueur fonctionnel (gène *aclB*). Cette étude avait pour but d'identifier d'abord les membres de la communauté microbienne d'un biofilm prélevé en milieu hydrothermal puis d'étudier leur mode de fixation du carbone. La deuxième (études 2 et 3) a quant à elle consisté à analyser des fragments de génomes de micro-organismes incultivés par l'approche métagénomique.

Les expériences conduites au cours de ce travail de recherche ont permis de dégager plusieurs points qui sont discutés dans les paragraphes suivants.

1 Importance écologique des ϵ -*Proteobacteria* dans l'écosystème hydrothermal marin profond

1.1 Diversité phylogénétique et abondance des ϵ -*Proteobacteria* dans l'écosystème hydrothermal marin profond

Dans un premier temps (étude 1), ce travail de thèse a permis de confirmer l'importance écologique des ϵ -*Proteobacteria* dans l'écosystème hydrothermal marin profond. L'analyse phylogénétique des séquences d'ARNr 16S du biofilm que nous avons collecté sur le site hydrothermal 13°N EPR, a révélé que tous les phylotypes détectés dans cet inventaire moléculaire étaient affi-

liés aux ε -*Proteobacteria*. Cette étude ainsi que bons nombres d'inventaires moléculaires réalisés cette dernière décennie (Campbell *et al.*, 2006, pour revue), ont révélé l'omniprésence, l'abondance et la diversité phylogénétique des ε -*Proteobacteria* au sein de l'écosystème hydrothermal marin profond. Des séquences environnementales d'ARNr 16S affiliées aux ε -*Proteobacteria* ont en effet été détectées dans toutes les niches écologiques proposées par Karl (1995). Elles ont ainsi été identifiées dans des tapis microbiens prélevés sur divers substrats exposés au fluide hydrothermal (roches, cheminées, sédiments, surface d'animaux, modules de colonisation), mais ont été également détectées parmi la microflore présente sous forme libre ou associée aux particules entraînées par le fluide hydrothermal et dans la microflore symbiotique des invertébrés hydrothermaux (Campbell *et al.*, 2006, pour revue).

Dans ces inventaires moléculaires, on a constaté que les séquences environnementales affiliées aux ε -*Proteobacteria* étaient souvent les séquences les plus représentées dans les banques de clones d'ADNr 16S. L'hypothèse selon laquelle les ε -*Proteobacteria* étaient des micro-organismes largement distribués et avaient une fonction écologique de première importance dans l'environnement hydrothermal marin profond a donc été émise. Cette hypothèse n'était toutefois soutenue que par des résultats d'expériences basées sur l'amplification par PCR. Afin de nous affranchir des biais liés à l'utilisation de cette technique et de révéler l'abondance réelle de ces organismes dans l'environnement, nous avons choisi d'utiliser la technique d'hybridation *in situ*. Ces dix dernières années, des sondes nucléiques pour hybridation *in situ* avaient été dessinées afin de cibler spécifiquement les divers groupes incultivés d' ε -*Proteobacteria* identifiés dans des inventaires moléculaires réalisés sur des échantillons hydrothermaux ou non (Polz & Cavanaugh, 1995 ; Cary *et al.*, 1997 ; Snaidr *et al.*, 1997 ; Longnecker & Reysenbach, 2001 ; Engel *et al.*, 2003). Cependant ces sondes avaient une spécificité très étroite : elles ne couvraient que les groupes d' ε -*Proteobacteria* incultivés qui avaient été détectés dans les études que nous venons de citer. Nous avons donc mis au point une sonde nucléique pour hybridation *in situ* (sonde EPSI682), ciblant plus largement les ε -*Proteobacteria*. Cette sonde couplée à l'utilisation de la sonde EUB338 spécifique des *Bacteria* nous a permis de confirmer dans notre échantillon de biofilm, la prédominance des ε -*Proteobacteria*. Elle pourra être dorénavant utilisée pour quantifier et estimer l'importance écologique des ε -*Proteobacteria* dans l'écosystème hydrothermal, mais également dans d'autres environnements où elles sont également fréquemment détectées, comme les sources d'eau riches en sulfures, les gisements de pétrole, les eaux souterraines contaminées ou non par des hydrocarbures et les grottes calcaires (Campbell *et al.*, 2006, pour revue).

Seules les techniques moléculaires indépendantes des méthodes de culture ont pendant longtemps permis de mettre en évidence les ε -*Proteobacteria* dans l'écosystème hydrothermal marin profond. Ce n'est que récemment, que les premières espèces d' ε -*Proteobacteria* provenant de cet environnement, ont pu être isolées et caractérisées (Fig. 1). Ces souches ont été isolées et carac-

Isolat/association phylogénétique	Lieu d'isolement	Température de croissance	Métabolisme du carbone	Donneur d'électron	Accepteur d'électron	Réduction nitrate ou soufre en:	Ref.
Ordre Nautiliales, Famille Nautiliaceae							
<i>Nautilia lithotrophica</i>	Tube d' <i>Alvinella pompejana</i> , 13°N EPR	53°C	Mixotrophe	H ₂ , formate	Sulphite, soufre élémentaire	H ₂ S	Miroshnichenko <i>et al.</i> , 2002
<i>Nautilia</i> sp. str. Am-H	Tube d' <i>Alvinella pompejana</i> , 13°N EPR	45°C	Mixotrophe	H ₂ , formate	Soufre élémentaire	H ₂ S	Campbell <i>et al.</i> , 2001
<i>Caminiibacter hydrogeniphilus</i>	Tube d' <i>Alvinella pompejana</i> , 13°N EPR	60°C	Mixotrophe	H ₂ , composés organiques complexes	Nitrate, soufre élémentaire	H ₂ S/NH ₃	Alain <i>et al.</i> , 2002
<i>Caminiibacter profundus</i>	Vent cap, Rainbow Field, MAR	55 °C	Autotrophe	H ₂	Nitrate, oxygène (microaérobiose) soufre élémentaire	H ₂ S/NH ₃	Miroshnichenko <i>et al.</i> , 2004
<i>Caminiibacter mediatlanticus</i>	Cheminée, Rainbow Field, MAR	55 °C	Autotrophe	H ₂	Nitrate, soufre élémentaire	H ₂ S/NH ₃	Voordeckers <i>et al.</i> , 2005
<i>Lebetimonas acidiphila</i>	Colonisateur <i>in situ</i> TOTO, MA	50 °C	Autotrophe	H ₂	Soufre élémentaire	H ₂ S	Takai <i>et al.</i> , 2005
Ordre incertain, Famille Hydrogenimonaceae							
<i>Hydrogenimonas thermophila</i>	Cheminée, Kairei Field, CIR	55 °C	Autotrophe	H ₂	Nitrate, oxygène (microaérobiose), soufre élémentaire	H ₂ S/NH ₃	Takai <i>et al.</i> , 2004
Ordre incertain, Famille Nitratiruptoraceae							
<i>Nitratiruptor tergaricus</i>	Cheminée, Iheya North Field, OT	55 °C	Autotrophe	H ₂	Nitrate, oxygène (microaérobiose), soufre élémentaire	H ₂ S/N ₂	Nakagawa <i>et al.</i> , 2005
Ordre incertain, Famille Thioreductoraceae							
<i>Thioreductor micantisoli</i>	Sédiment, Iheya North Field, OT	32 °C	Autotrophe	H ₂	Nitrate, soufre élémentaire	H ₂ S/NH ₃	Nakagawa <i>et al.</i> , 2005
Ordre Campylobacterales, Famille Campylobacteraceae							
<i>Sulfurospirillum</i> sp. str. Am-N	<i>Alvinella pompejana</i> , 13°N EPR	41 °C	Hétérotrophe	Formate, fumarate	Soufre élémentaire	H ₂ S	Campbell <i>et al.</i> , 2001
<i>Arcobacter</i> sp. str. FWKO B	Eau de production, Coleville oil field	30 °C	Autotrophe	H ₂ , formate sulfure	Nitrate, oxygène (microaérobiose), soufre élémentaire	H ₂ S/NO ₂ ⁻	Gevertz <i>et al.</i> , 2000
Ordre incertain, Famille Thiovolgaceae							
<i>Sulfurovum lithotrophicum</i>	Sédiment, Iheya North Field, OT	30 °C	Autotrophe	Soufre élémentaire thiosulphate	Nitrate, oxygène (microaérobiose)	N ₂	Inagaki <i>et al.</i> , 2003
<i>Nitratifactor salsuginis</i>	Ceminée, Iheya North Field, OT	37 °C	Autotrophe	H ₂	Nitrate, oxygène (microaérobiose)	N ₂	Nakagawa <i>et al.</i> , 2005
<i>Sulfurimonas autotrophica</i>	Sédiment, Hatoma Knoll, OT	25°C	Autotrophe	Soufre élémentaire thiosulphate	Oxygène (microaérobiose)		Inagaki <i>et al.</i> , 2003
<i>Sulfuricurvum kujiense</i>	Eau souterraine, Japan oil storage cavity	25°C	Autotrophe	H ₂ , sulfures thiosulphate, soufre élémentaire	Nitrate, oxygène (microaérobiose)	NO ₂ ⁻	Kodama <i>et al.</i> , 2004
<i>Thiomicrospira</i> sp. str. CVO	Eau de production, Coleville oil field	30 °C	Mixotrophe	Sulfures soufre élémentaire	Oxygène (microaérobiose), nitrate, nitrite	N ₂ , N ₂ O	Gevertz <i>et al.</i> , 2000

Abréviations des lieux: CIR, Central Indian Ridge; EPR, East Pacific Rise; MA, Mariana Volcanic Arc; MAR, Mid-Atlantic Ridge; OT, Okinawa Trough; TOTO, TOTO caldera deep-sea hydrothermal field. Abréviations chimiques: H₂S, hydrogen sulphide; NH₃, ammonia; N₂O, nitrous oxide; NO₂⁻, nitric oxide.

Fig. 1 – Caractéristiques physiologiques des ϵ -*Proteobacteria* isolées d'habitats hydrothermaux marins profonds et de quelques autres environnements sélectionnés (d'après Campbell *et al.*, 2006).

térisées pendant que j'effectuais ce travail de thèse. Les espèces cultivées incluent des membres de la famille des *Nautiliaceae*, des "*Hydrogenimonaceae*", des "*Nitratiruptoraceae*", des "*Thioreductoraceae*", des *Campylobacteraceae*, et des "*Thiovolgaceae*". Les ϵ -*Proteobacteria* demeurent cependant encore très peu représentées dans les collections de culture. Il n'y a pas de difficulté technique pour cultiver ces ϵ -*Proteobacteria*. Leur découverte si tardive peut venir du fait que peu de gens s'étaient intéressés à la culture des communautés microbiennes mésophiles et thermophiles modérées avant Takai et ses collaborateurs (2003). Un effort de culture reste donc à produire. On peut notamment optimiser des stratégie de culture reproduisant les conditions *in situ* rencontrées par ces organismes, comme cela a été fait avec la mise au point d'un réacteur spécifique pour la culture de la souche "*Candidatus Arcobacter sulfidicus*" (Taylor *et al.*, 1999).

L'ensemble de ces données ajoutées à celles issues de ce travail de thèse, nous permettent de dire que l'écosystème hydrothermal marin profond peut être considéré comme l'un des plus grands réservoirs d'*ε-Proteobacteria* connu sur Terre. Au sein de cet écosystème, les *ε-Proteobacteria* sont considérées comme des colonisateurs primaires et des producteurs primaires de premier ordre. Ces deux derniers points seront discutés dans les paragraphes qui suivent.

1.2 Les *ε-Proteobacteria* : des colonisateurs primaires de l'écosystème hydrothermal marin profond

Les résultats issus de nos 5 jours d'expérience de colonisation (étude 1), suggèrent qu'une surface vierge se trouvant à l'interface entre le fluide hydrothermal et l'eau de mer environnante, peut être très rapidement colonisée par des *ε-Proteobacteria*. Ces dernières peuvent donc être considérées comme des colonisateurs primaires de cet environnement hydrothermal. Des expériences de colonisation menées précédemment sur divers sites hydrothermaux à l'aide de colonisateur *in situ* ou autre engins collecteurs de micro-organismes, avaient elles aussi abouti à la même conclusion (Taylor *et al.*, 1999 ; Reysenbach *et al.*, 2000 ; Corre *et al.*, 2001 ; Lopez-Garcia *et al.*, 2003a ; Alain *et al.*, 2004 ; Higashi *et al.*, 2004 ; Nakagawa *et al.*, 2005b). L'apparition d'un biofilm très dense visible au bout de 3 jours sur notre colonisateur *in situ*, laisse sous-entendre que les *ε-Proteobacteria*, seuls micro-organismes détectés dans ce biofilm, sont particulièrement bien adaptées à la colonisation de supports/surfaces vierges exposés aux fluides hydrothermaux.

Nous sommes en droit de nous demander pourquoi les *ε-Proteobacteria* sont-elles plus adaptées que d'autres micro-organismes pour coloniser cet environnement ? Qu'est-ce qui fait des *ε-Proteobacteria* des espèces pionnières dans les processus de colonisation ?

Des mesures physico-chimiques réalisées à l'endroit où notre colonisateur *in situ* a été déposé (mesures réalisées quelques jours avant notre expérience de colonisation par Alain et collaborateurs (2004)) ont indiqué que cette zone était caractérisée par une très forte instabilité (variations de pH, de température et de concentration en H₂S). Ces conditions environnementales instables sont typiques des zones de mélange se trouvant à l'interface oxygène-anoxique, entre le fluide hydrothermal et l'eau de mer. Ces habitats très instables pourraient paraître hostiles à l'établissement de communautés microbiennes. Les conditions *in situ* extrêmes comme des concentrations en sulfure d'hydrogène qui peuvent être très élevées, doivent certainement contribuer à une sélection drastique des micro-organismes pouvant s'y développer.

Plusieurs auteurs ont mis en avant la flexibilité nutritionnelle des *ε-Proteobacteria* pour expliquer leur facilité d'adaptation aux conditions physico-chimiques fluctuantes de l'environnement hydrothermal (Nakagawa *et al.*, 2005b ; Campbell *et al.*, 2006). Cette versatilité nutritionnelle des *ε-Proteobacteria* leur permettrait ainsi de coloniser plus facilement que d'autres micro-organismes cet environnement très instable. Les premières souches d'*ε-Proteobacteria* qui ont été récem-

ment isolées d'environnements hydrothermaux, ont en effet montré qu'elles pouvaient utiliser une gamme très variée de donneurs d'électrons (l'hydrogène, le formate, le fumarate, les sulfures...) et d'accepteurs d'électrons (les sulfures, les sulfites, le soufre élémentaire, les nitrates...). De plus, la gamme de température dans laquelle ces souches sont capables de se développer (si l'on considère les *ε-Proteobacteria* mésophiles et les thermophiles modérées) est relativement large, allant de 20°C à 60°C, ce qui leur procure un avantage supplémentaire pour s'adapter à cet environnement instable.

Il est probable que le type d'expérience que nous avons mené ne mette en évidence que les micro-organismes capables d'attachement et de colonisation rapide. Les *ε-Proteobacteria* identifiées dans notre expérience de colonisation ont la capacité de produire rapidement de la biomasse (dès le 3^{me} jour de l'expérience, le biofilm était déjà bien épais). Le temps de génération de ces *ε-Proteobacteria* est certainement très rapide dans les conditions *in situ* auxquelles ont été réalisées l'expérience, c'est-à-dire au niveau d'une interface oxygène-anoxique, à des températures modérées allant de 11°C à 45°C et où de fortes concentrations en H₂S sont disponibles. L'aptitude des *ε-Proteobacteria* à adhérer à un support vierge via la formation d'un biofilm (étude 1, Alain *et al.*, 2004) procure certainement à ces organismes un avantage pour la colonisation de cet environnement dynamique et extrême. Il a en effet été suggéré que la formation de biofilm pouvait être un avantage pour certains micro-organismes : ces derniers pourraient être protégés de conditions environnementales non favorables, et facilement bénéficier d'échanges de nutriments et de matériels génétiques (La Paglia & Hartzell, 1997 ; Davey & O'Toole, 2000).

Il est intéressant de noter que lors d'une expérience de colonisation menée par Alain et ses collaborateurs (2004), expérience réalisée quelques jours avant la notre et sur le même site hydrothermal, les micro-organismes identifiés dans cette étude n'étaient pas les mêmes que ceux que nous avons identifiés dans la notre. L'originalité des séquences d'*ε-Proteobacteria* que nous avons obtenues dans notre étude est discutée dans la paragraphe suivant.

1.3 Détection d'*ε-Proteobacteria* affiliées au genre *Arcobacter* dans l'écosystème hydrothermal marin profond

L'inventaire des ARNr 16S ainsi que les analyses de FISH (étude 1) ont révélé que la communauté microbienne du biofilm était peu diversifiée. Tous les clones environnementaux analysés étaient affiliés à des *ε-Proteobacteria*, confirmant leur abondance dans l'écosystème hydrothermal. Parmi ces *ε-Proteobacteria*, la majorité des phylotypes que nous avons obtenus, était affiliée à des représentants incultivés du genre *Arcobacter*. Étonnamment, de précédentes études menées sur différents sites hydrothermaux et à l'aide de divers supports de colonisation (Reysenbach *et al.*, 2000 ; Corre *et al.*, 2001 ; Lopez-Garcia *et al.*, 2003a ; Alain *et al.*, 2004 ; Higashi *et al.*, 2004 ; Nakagawa *et al.*, 2005c) avaient pourtant montré que les *ε-Proteobacteria* du groupe F

(groupe qui inclut deux souches caractérisées : *Sulfurovum lithotrophicum* et *Nitratifactor salsuginis*) et du groupe B (incluant les souches du genre *Sulfurimonas*) représentaient les phylotypes les plus fréquemment détectés dans les banques de clones d'ADNr 16S.

Les micro-organismes du genre *Arcobacter* figurent parmi les principaux groupes de bactéries responsables d'entérites, de mammites et de bactériémies chez les humains et les animaux (<http://www.bacterio.cict.fr/bacdico/aa/arcobacter.html>). Les espèces animales, notamment les volailles, ainsi que l'eau, seraient les principales sources de contamination. Outre les espèces d'*Arcobacter* isolées du milieu médical, quelques séquences environnementales *Arcobacter* ont également été mises en évidence, notamment dans de l'eau de mer, des boues activées, de l'eau d'égouts, de l'eau de rivière et dans de l'eau de boisson non chlorée.

Malgré les attentions portées envers les souches pathogènes, les *Arcobacter* ont été peu caractérisés, tant au niveau de leur métabolisme, que de leur taxonomie et de leur écologie. Leur mode de vie et le type d'habitat que ces bactéries peuvent coloniser n'ont pas été étudiés en détail (Campbell *et al.*, 2006). C'est d'ailleurs essentiellement dans le cadre d'études médicales, et à l'aide de tests bactériologiques, que ces bactéries du genre *Arcobacter* ont été identifiées. Les espèces qui ont été caractérisées ont des optima de croissance à température ambiante et semblent préférer des conditions de microaérophilie pour se développer. Au sein de l'écosystème hydrothermal marin profond, aucune souche du genre *Arcobacter* n'a été à notre connaissance isolée et caractérisée à ce jour. De plus, très peu d'inventaires moléculaires réalisés dans cet écosystème, relatent la présence de séquences environnementales affiliées à ce genre.

Parmi les souches du genre *Arcobacter* à avoir été caractérisées, la souche "*Candidatus Arcobacter sulfidicus*" (Wirsen *et al.*, 2002) requiert notre attention. Cette souche marine n'a pas été isolée de l'environnement hydrothermal marin profond. Cependant, sa morphologie ainsi que les filaments de soufre qu'elle excrète, sont identiques à ceux de micro-organismes hydrothermaux qui ont été observés dans des flocculats (Taylor & Wirsen, 1997 ; Taylor *et al.*, 1999). La souche "*Candidatus Arcobacter sulfidicus*" est une bactérie mobile, chimiolithoautotrophe, mésophile. C'est une bactérie sulfo-oxydante, microaérophile obligatoire (Wirsen *et al.*, 2002). Difficile à cultiver, elle nécessite une interface géochimique bien définie entre des concentrations en oxygène dissous et en H₂S. Un réacteur spécifique qui délivre un gradient opposé en concentrations d'O₂ et en H₂S, a d'ailleurs été spécialement mis au point pour sa culture (Wirsen *et al.*, 2002). La particularité de cette souche vis-à-vis des autres bactéries sulfo-oxydantes connues ce jour, vient du fait qu'elle est capable de supporter des concentrations en H₂S bien plus élevées (1 à 2 mM) que ces dernières. De plus, elle est peut se développer à des concentrations en O₂ très basses (1 à 10 μM).

Lorsque les conditions environnementales lui conviennent, la souche "*Candidatus Arcobacter*

sulfidicus” excrète des filaments de soufre (0.5-2.0 x 20-500 μm) (Wirsen *et al.*, 2002). La formation de ces filaments de soufre semble être un mécanisme d’adaptation conçu pour coloniser efficacement des surfaces qui sont caractérisées par des concentrations très élevées en H_2S et qui sont exposées aux flux de fluides puissants (Taylor & Wirsen, 1997 ; Sievert *et al.*, 1999 ; Taylor *et al.*, 1999 ; Wirsen *et al.*, 2002). Le processus de formation de ces filaments de soufre vient d’être étudié (Sievert *et al.*, 2006). Les hypothèses suivantes portant sur l’intérêt écologique de ce mécanisme d’adaptation, ont été émises suite à cette étude :

- la formation des filaments de soufre permettrait de convertir rapidement l’ H_2S en soufre élémentaire, fonctionnant ainsi comme un mécanisme de détoxification du milieu qui permet aux bactéries de supporter des concentrations très élevées en H_2S ;
- les filaments de soufre permettraient aux organismes qui les excrètent, de se maintenir dans des micro-niches qui conviennent à leur survie et à leur développement, notamment au sein d’une interface en oxygène/ H_2S qui leur est optimale. Si on se place à l’échelle d’une population de “*Candidatus Arcobacter sulfidicus*”, la structure physique fabriquée par ces micro-organismes à partir d’un produit de dégradation de leur métabolisme (le soufre), leur permettrait d’adhérer et de coloniser efficacement toute surface qui leur conviendrait. Dans les environnements dynamiques sans cesse perturbés par des mouvements de fluides, ces micro-organismes capables de former des filaments de soufre seraient ainsi plus adaptés que d’autres bactéries sulfo-oxydantes incapables de former ces structures, faisant d’eux des colonisateurs primaires.

Les différences observées dans notre expérience et dans celle menée par Alain et ses collaborateurs (2004), sont certainement dues au fait que la paroi de la cheminée hydrothermale sur laquelle a été déposé notre colonisateur *in situ*, a été fragilisée par le bras télé-opéré du robot Victor 6000 (voir le matériel et méthode de l’étude 1). De ce fait, les fluides hydrothermaux s’échappant de la paroi étaient plus intenses dans notre expérience, ce qui a favorisé un enrichissement de l’habitat en sulfures. Seuls des micro-organismes capables de supporter des concentrations très élevées en H_2S mais également capables de se développer dans un environnement dynamique perturbé par le flux des fluides (via la formation d’un biofilm), pouvaient dans un premier temps coloniser notre support. Les conditions *in situ* dans lesquelles notre biofilm s’est formé, ainsi que la présence de particules de soufre dans le biofilm, suggèrent que les *Arcobacter* détectées dans nos expériences, ont très probablement un métabolisme proche de la souche “*Candidatus Arcobacter sulfidicus*”, c’est-à-dire un métabolisme lié à l’oxydation des composés soufrés réduits.

Les conditions *in situ* qui prévalent dans l’écosystème hydrothermal, notamment au niveau de la zone de mélange entre le fluide hydrothermal et l’eau de mer, ont sélectionné de manière très drastique les micro-organismes susceptibles de coloniser notre support vierge. La communauté microbienne du biofilm était en effet très peu diversifiée. On peut supposer que cette communauté microbienne aurait pu évoluer en se diversifiant, si l’expérience avait duré plus longtemps. La

création de micro-niches anaérobies au sein du biofilm, aurait en effet pu convenir par la suite à la colonisation de nouveaux micro-organismes réduisant par exemple le soufre élémentaire détecté dans le biofilm, bouclant ainsi le cycle du soufre.

Après avoir identifié les micro-organismes du biofilm, nous avons utilisé une approche fonctionnelle pour tenter d'obtenir des réponses concernant l'écophysio­logie de ces micro-organismes. Cette approche est décrite dans la partie suivante.

1.4 Etude fonctionnelle du mode de fixation du carbone des ϵ -*Proteobacteria* détectées dans le biofilm

L'analyse phylogénétique de la diversité microbienne d'un échantillon environnemental (via un marqueur phylogénétique comme l'ARNr 16S par ex.) combinée à une analyse fonctionnelle visant un ou plusieurs gènes codant pour des enzymes spécifiques de voies métaboliques, constitue une méthode d'investigation d'écologie microbienne très prometteuse. Cette technique a pour intérêt de cibler les micro-organismes ayant le même trait métabolique dans une communauté microbienne complexe. Il a été récemment démontré que les ϵ -*Proteobacteria* fixaient leur carbone via le cycle rTCA (Campbell *et al.*, 2003 ; Campbell & Cary, 2004 ; Hugler *et al.*, 2005 ; Takai *et al.*, 2005a). Notons que les γ -*Proteobacteria* qui occupent pourtant les mêmes niches écologiques que les ϵ -*Proteobacteria*, fixent leur carbone via le cycle de Calvin (Cavanaugh *et al.*, 1981 ; Felbeck *et al.*, 1981 ; Cavanaugh, 1983 ; Karl *et al.*, 1988 ; Nelson *et al.*, 1989 ; Ruby *et al.*, 1981). Afin d'établir une relation entre l'identité génétique des ϵ -*Proteobacteria* détectées dans le biofilm et leur mode de fixation du carbone, nous avons donc choisi de rechercher dans le biofilm la présence des gènes *aclB* et *oorA* qui codent pour des enzymes clés du cycle rTCA (voir les résultats complémentaires de l'étude 1).

Les résultats de ces recherches nous ont permis de mettre en évidence la présence de gènes *aclB* dans le biofilm. Cependant, aucun produit d'amplification du gène *oorA* n'a pu être obtenu à partir de l'ADN extrait du biofilm. L'analyse de la diversité des gènes *aclB* a révélé que toutes les séquences étaient affiliées à des ϵ -*Proteobacteria*. Étonnamment, aucune de ces séquences n'était apparentée aux *Arcobacter*. Ces micro-organismes étaient pourtant dominants dans notre expérience de quantification par hybridation *in situ*. Tous les gènes *aclB* étaient en fait affiliés aux *Nautiliaceae*. Aucune séquence affiliée aux *Nautiliaceae* n'avait pourtant pu être mise en évidence dans les banques de clones d'ADNr 16S.

Les deux approches moléculaires qui ont été utilisées pour étudier la diversité et la fonction des micro-organismes présents dans le biofilm, ont fourni des résultats contradictoires mais des informations complémentaires. En effet, si les *Nautiliaceae* n'ont pas été détectées dans les inventaires d'ARNr 16S, c'est probablement parce que ces micro-organismes sont minoritaires au sein de la communauté microbienne du biofilm. De plus, comme nous l'avons évoqué dans l'étude 1,

tout fait penser que les couples d'amorces que nous avons utilisés pour amplifier les gènes *aclB* et *oorA* ne couvraient pas l'ensemble des ε -*Proteobacteria*. Cela expliquerait donc pourquoi nous n'avons pas détecté de gènes *aclB* et *oorA* affiliés aux *Arcobacter*. L'accumulation progressive de nouvelles séquences de gènes clés intervenant dans le cycle rTCA (via le biais d'études métagénomiques, par exemple) permettra à l'avenir de redessiner de nouvelles amorces à couverture plus large.

Le métabolisme du carbone et celui lié à l'utilisation de sources d'énergies, ont été récemment testés chez plusieurs isolats d' ε -*Proteobacteria* provenant de l'environnement hydrothermal marin profond (Takai *et al.*, 2005a). Des tests enzymatiques (portant sur des enzymes du cycle de Calvin, du cycle rTCA, et des enzymes intervenant dans l'utilisation de l'hydrogène et des sulfites comme source d'énergie) réalisés en parallèle avec des tests génétiques (via l'utilisation de marqueurs fonctionnels ciblant les cycles du carbone, mais aussi d'un marqueur fonctionnel ciblant l'opéron *hyn* qui intervient dans l'utilisation de l'hydrogène comme source d'énergie) ont permis d'aller plus loin dans la caractérisation fonctionnelle des ε -*Proteobacteria*. Cette étude montre l'utilité d'utiliser des outils moléculaires pour élucider la fonction et le rôle écophysiologique des ε -*Proteobacteria* hydrothermales. Notons, qu'à l'heure actuelle, deux ε -*Proteobacteria* isolées en milieu hydrothermal marin profond, ont été sélectionnées pour un séquençage complet de leur génome. Les souches choisies sont *Caminibacter mediatlanticus* et *Nautilia* sp AmH. La souche "*Candidatus Arcobacter sulfidicus*" a elle aussi été sélectionnée pour séquençage de son génome (source <http://www.genomesonline.org/>). L'accès à ces génomes ouvrira certainement de nouvelles portes pour la compréhension des ε -*Proteobacteria*

Pour aller plus loin dans l'étude des micro-organismes incultivés de l'écosystème hydrothermal, nous avons opté pour une autre approche qui n'avait jamais été expérimentée dans le laboratoire où j'ai effectué ma thèse. Cette approche nous a permis d'accéder à des fragments de génomes appartenant à des micro-organismes incultivés, et ce, sans avoir à les cultiver.

2 L'approche métagénomique pour l'étude des communautés microbiennes hydrothermales

Avec l'objectif de rechercher des informations sur les lignées incultivées de l'écosystème hydrothermal, ce travail de thèse a été l'objet du développement méthodologique de l'approche métagénomique, au sein du laboratoire. Le but de ce travail était d'identifier parmi les gènes mis en évidence, des gènes pouvant fournir des informations sur la physiologie et le métabolisme des micro-organismes incultivés. Pour accéder à cette information génomique, nous avons choisi de construire des banques à l'aide de vecteurs de clonage du type cosmide et fosmide, qui acceptent des inserts de grande taille, allant de 35 à 45 kpb.

2.1 Difficulté de construire des banques métagénomiques à grands inserts avec des échantillons hydrothermaux

Avant de parvenir à construire la banque métagénomique présentée dans les études 2 et 3 de ce manuscrit de thèse, plusieurs essais d'extraction d'ADN génomique de grande taille ont été réalisés sur des échantillons de sédiments hydrothermaux. Ces échantillons d'ADN génomique ont été clonés dans des vecteurs de type cosmide ou des fosmide. Cependant, des problèmes d'ordre technique nous ont empêché de parvenir à nos fins. Ces problèmes sont certainement dus à une biomasse microbienne insuffisante dans nos échantillons (il est recommandé d'avoir au moins 2.5 μg d'ADN génomique de haut poids moléculaire pour construire une banque métagénomique dans un vecteur de type fosmide), et à la difficulté d'obtenir de l'ADN génomique non contaminé par des inhibiteurs de réactions enzymatiques qui peuvent être co-extraits avec les acides nucléiques. Le rendement de grands fragments génomiques à partir de ces échantillons de sédiments s'est révélé trop faible pour pouvoir générer des banques génomiques avec des vecteurs de type cosmide ou fosmide. Les extractions d'ADN génomique ont été réalisées à l'aide du kit UltraClean DNA (Laboratoires MoBio, Carlsbad, CA). Une autre méthode d'extraction de l'ADN génomique, comme celle basée sur la technique d'électrophorèse à deux étapes (Quaiser *et al.*, 2002), aurait été probablement plus adaptée pour purifier l'ADN de haut poids moléculaire contenu dans ces sédiments hydrothermaux.

2.2 Motivations du choix de tubes d'*Alvinellidae* comme support de biomasse microbienne pour l'élaboration d'une banque métagénomique

N'ayant pas réussi à construire des banques métagénomiques à partir de l'ADN extrait des sédiments hydrothermaux, nous avons choisi d'extraire l'ADN génomique d'une communauté microbienne associée à des tubes d'*Alvinella pompejana*. Les tubes d'*Alvinellidae* constituent un support privilégié pour l'étude des communautés microbiennes thermophiles. Les *Alvinellidae* colonisent la partie la plus chaude des édifices hydrothermaux de la dorsale du Pacifique Oriental. Pour se protéger des fluides hydrothermaux, ces annélides polychètes sécrètent des tubes. Ces tubes composés de matière organique et minérale, sont directement exposés aux fluides hydrothermaux. Ils permettent à ces métazoaires d'être localisés dans des environnements où les températures peuvent atteindre 80 °C (Gaill & Hunt, 1991). La fraction minérale du tube est composée d'environ 12 à 25 % de soufre élémentaire, le restant étant composé de sulfures, de phosphates et de carbonates. Environ 30 à 40 % de la fraction organique du tube correspondent à des protéines et 7% correspondent à des sucres (hexoses) qui sont le plus souvent des poly- ou des oligo-saccharides liés aux protéines (Desbruyères *et al.*, 1998). Ces tubes de part leur composition organique et leur interface constituée par l'environnement immédiat du polychète, offrent un support qui permet de collecter une biomasse microbienne suffisante pour être extraite.

L'inventaire moléculaire que nous avons réalisé sur l'ADN extrait de la communauté microbienne associée à ces tubes, a révélé plus de la moitié des séquences d'*Archaea* étaient affiliées à des lignées incultivées (étude 2). Cet ADN a donc été préférentiellement utilisé pour une analyse visant à étudier l'information génomique des organismes incultivés. Plus de 5000 clones ayant des inserts d'environ 40 kpb ont été obtenus dans la banque métagénomique construite à partir de cet ADN. Cette banque a été criblée par PCR afin d'identifier les clones porteurs de séquences d'ARNr 16S et 18S. Les articles 2 et 3 de ce manuscrit de thèse exposent l'analyse de la séquence génomique de trois clones (Alv-FOS1, Alv-FOS4 et Alv-FOS5) porteurs de séquences d'ARNr 16S d'*Archaea* affiliés aux DHVE.

2.3 Identification de gènes fournissant des informations sur la physiologie et le métabolisme des micro-organismes incultivés

Notre connaissance des procaryotes est essentiellement fondée sur les expérimentations réalisées à partir des micro-organismes cultivés, qui représentent finalement une part infime de la diversité microbienne sur Terre. L'analyse du génome de tous les organismes d'une niche écologique, autrement dit l'analyse d'un métagénome donné, peut permettre de détecter des gènes informationnels et l'existence de voies métaboliques chez des organismes incultivés. La plupart des annotations disponibles à l'heure actuelle proviennent de l'analyse de programmes informatiques qui prédisent les fonctions des nouvelles séquences de gènes, sur la base de leur similitude avec des gènes de fonction connue. Cependant, cette annotation des gènes prédits n'est pas toujours évidente. A titre d'exemple, une annotation fonctionnelle précise n'a pu être attribuée qu'à seulement la moitié des gènes prédits dans la séquence génomique du fosmide Alv-FOS5 (article 2). De plus, seules ces prédictions ne peuvent être réellement validées que si elles sont expérimentalement testées.

2.4 Améliorer la mise au point des amorces

Les amorces PCR qui ciblent les ARNr 16S des *Archaea* ou des *Bacteria*, ont été mises au point à partir de séquences répertoriées dans les banques de données. Leur utilisation permet de mettre régulièrement en évidence de nouvelles lignées phylogénétiques. Toutefois, les séquences d'ARNr 16S de ces lignées n'hybrident pas toujours avec les amorces universelles couramment utilisées lors des inventaires moléculaires. La réévaluation de ces amorces s'avère donc l'un des points critiques. L'article 2 de cette thèse illustre ce point. Le fosmide Alv-FOS5 analysé dans cette étude, forme une nouvelle lignée de micro-organismes affiliés au groupe II des DHVE. Lors du criblage par PCR de notre banque métagénomique, le gène de l'ARNr 16S de ce fosmide a été identifié de manière inattendue. Il a en effet été mis en évidence à l'aide d'amorces qui ciblaient le gène de l'ARNr 18S des eucaryotes et non pas à l'aide de celles ciblant spécifiquement l'ARNr 16S des *Archaea*. L'explication tient dans le fait que la séquence de ce gène présente plusieurs

mésappariements - cinq - avec l'une des deux amorces spécifiques du domaine des *Archaea* que nous avons utilisées. Nous avons eu la chance de découvrir, avec des amorces spécifiques des *Eucarya*, cette nouvelle lignée phylogénétique au sein des DHVE. Cependant, nous aurions pu potentiellement détecter ce nouvel organisme par un séquençage total ou partiel de notre banque métagénomique. Etant donné qu'a priori, tout fragment d'ADN génomique d'un métagénome donné peut être cloné puis séquencé, les données issues de la métagénomique laissent présager la détection de nouvelles séquences qui, s'ajoutant à celles déjà répertoriées, vont permettre d'affiner la spécificité et la couverture phylogénétique des sondes et des amorces PCR.

2.5 Proposition du mode de vie thermophile pour les membres d'une lignée d'organismes incultivés endémiques des sources hydrothermales marines profondes

L'étude 3 a été l'occasion par le biais de l'approche métagénomique, de nous fournir une hypothèse testable expérimentalement pour nous éclairer sur la physiologie d'une lignée d'organismes incultivés fréquemment identifiée dans l'écosystème hydrothermal. Le criblage de la banque métagénomique construite dans le cadre de cette thèse a révélé que quatre fosmidés contenaient une séquence d'ARNr 16S affiliés aux DHVE2, deux d'entre eux ont été entièrement séquencés, les fosmidés Alv-FOS1 et Alv-FOS4. La lignée des DHVE2 a été identifiée la première fois par Takai et Horikoshi (1999). Depuis, des micro-organismes incultivés affiliés aux DHVE2 ont été identifiés dans tous les système hydrothermaux marins profonds étudiés à ce jour (voir par exemple, Takai & Horikoshi, 1999 ; Reysenbach *et al.*, 2000 ; Nercessian *et al.*, 2003 ; Hoek *et al.*, 2003 ; Nakagawa *et al.*, 2006 ; Takai *et al.*, 2001a). Ils sont d'ailleurs abondamment représentés dans les banques de clones réalisées sur les échantillons hydrothermaux (Fig. 2).

Le pourcentage en guanine et cytosine contenu dans leurs ARNr 16S (> 60%, valeur propre aux micro-organismes thermophiles), ainsi que leur identification dans des échantillons contenant également des hyperthermophiles de l'ordre des *Thermococcales*, laissaient présager que les DHVE2 étaient des organismes thermophiles, certainement hétérotrophes, ayant une croissance stimulée par le soufre (Reysenbach & Shock, 2002 ; Takai *et al.*, 2001a). Cependant, malgré de nombreux efforts entrepris pour cultiver ces micro-organismes, leur présence n'a pas été détectée dans les cultures d'enrichissement réalisées à 70 °C, 80 °C et 90 °C (Nercessian *et al.*, 2003).

A l'instar de toutes les séquences affiliées aux DHVE2 répertoriées dans les banques de données, le contenu en G+C des séquences d'ARNr 16S contenues dans les fosmidés Alv-FOS1 et Alv-FOS4, était supérieur à > 60%. L'ADN polymérase recombinante issue du clonage d'un gène présent dans l'un de ces fosmidés, a confirmé de part ses propriétés de thermostabilité et d'optimum d'activité, que les DHVE2 présentaient un mode de vie thermophile proche de celui de *Thermoplasma acidophila*. Les micro-organismes de l'ordre des *Thermoplasmatales* sont

Lieu	Lat/Long	% de clones dans la banque	Exemple de clone	N° d'accession Genbank	Type d'échantillon	Ref
Axial Volcano (Juan de Fuca Ridge), North Pacific Ocean	46°N, 130°W	6.7	33-P127A99	AF355837	fluide hydrothermal	Huber <i>et al.</i> , 2002
Axial Volcano (Juan de Fuca Ridge), North Pacific Ocean	46°N, 130°W	6.1	33-P23A98	AF355840	fluide hydrothermal	Huber <i>et al.</i> , 2002
Guaymas Basin (Sea of Cortez), Pacific Ocean	27°N, 110°W	4.9	G26_C73	AF356637	cheminée hydrothermale	Longnecker, 2001
East Pacific Rise 9°N	9°50'N, 104°18'N	1.5	clone CH8_7a	AY672495	cheminée hydrothermale	Kormas <i>et al.</i> , 2006
East Pacific Rise 13°N, Pacific Ocean	12°48'N, 104°W	ND*	pEPR122	AF526963	échantillon hydrothermal	Nercessian <i>et al.</i> , 2003 Nercessian <i>et al.</i> , 2004
Myojin Knoll (Izu-Ogasawara Arc), North western Pacific Ocean	32°06'N, 139°52'E	33.0	pMC2A24	AB019736	cheminée hydrothermale	Takai et Horikoshi, 1999
Myojin Knoll (Izu-Ogasawara Arc), North western Pacific Ocean	28°34'N, 140°38'E	4.8	pSSMCA108	AB019740	cheminée hydrothermale	Takai et Horikoshi, 1999
Iheya Basin (Middle Okinawa Trough), Northwestern Pacific Ocean	27°32'.N, 126°58'E	29.8	pISA12	AB019741	sédiments hydrothermaux	Takai et Horikoshi, 1999
Pacmanus (Manus Basin), Western Pacific Ocean	03°43'S, 151°40'E	6.8-22.9	pPACMA-M	AB052983	cheminée hydrothermale	Takai <i>et al.</i> , 2001
Snake Pit (Mid-Atlantic Ridge), Atlantic Ocean	23°22' N, 44°57'W	4.0	VC2.1 Arc6	AF068817	échantillon hydrothermal	Reysenbach <i>et al.</i> , 2000
Edmond Vent Field (Central Indian Ridge), Indian Ocean	23°S, 69°E	93.0	FT17A03	AY251064	cheminée hydrothermale	Hoek <i>et al.</i> , 2003
TOTO caldera (Mariana Volcanic Arc), Western Pacific Ocean	12°42'N, 143°32'E	35.7	TOTO-A6-12	AB167486	cheminée hydrothermale	Nagakawa <i>et al.</i> , 2006
Cascadia Margin-eastern Pacific-Deep Marine Sediments, Eastern Pacific Ocean	44°34'N, 125°04'W	ND*	ODP1251A15.24	AB177273	sédiments marins profonds (123-304 mbsf)	Inagaki <i>et al.</i> , 2006

* non déterminé

Fig. 2 – Etudes publiées relatant la présence de DHVE2 dans les banques de clones d'*Archaea*. (D'après Reysenbach *et al.*, 2006.)

d'ailleurs les représentants cultivés les plus proches des DHVE2.

Etant donné que les inventaires moléculaires avaient montré que les DHVE2 étaient présentes et abondamment distribuées dans l'écosystème hydrothermal marin profond, on pouvait envisager de les cultiver à partir d'échantillons hydrothermaux. Très récemment, le premier représentant cultivé de la lignée des DHVE2 a été cultivé à partir d'un milieu permettant d'enrichir les organismes thermoacidophiles hétérotrophes (Reysenbach *et al.*, 2006) (Fig. 3). Cette souche appelée "*Aciduliprofundum boonei*" est une *Euryarchaeota* anaérobie hétérotrophe, thermoacidophile, réduisant le soufre ou le fer. Ce membre de la lignée des DHVE2 peut se développer dans une gamme de température allant de 55 °C à 75 °C, et une gamme de pH allant de 3.3 à 5.8. Son optimum de croissance est observé à 70 °C, dans des pH allant de 4.2 à 4.8. Elle a été isolée sur le site hydrothermal Mariner, du Bassin de Lau.

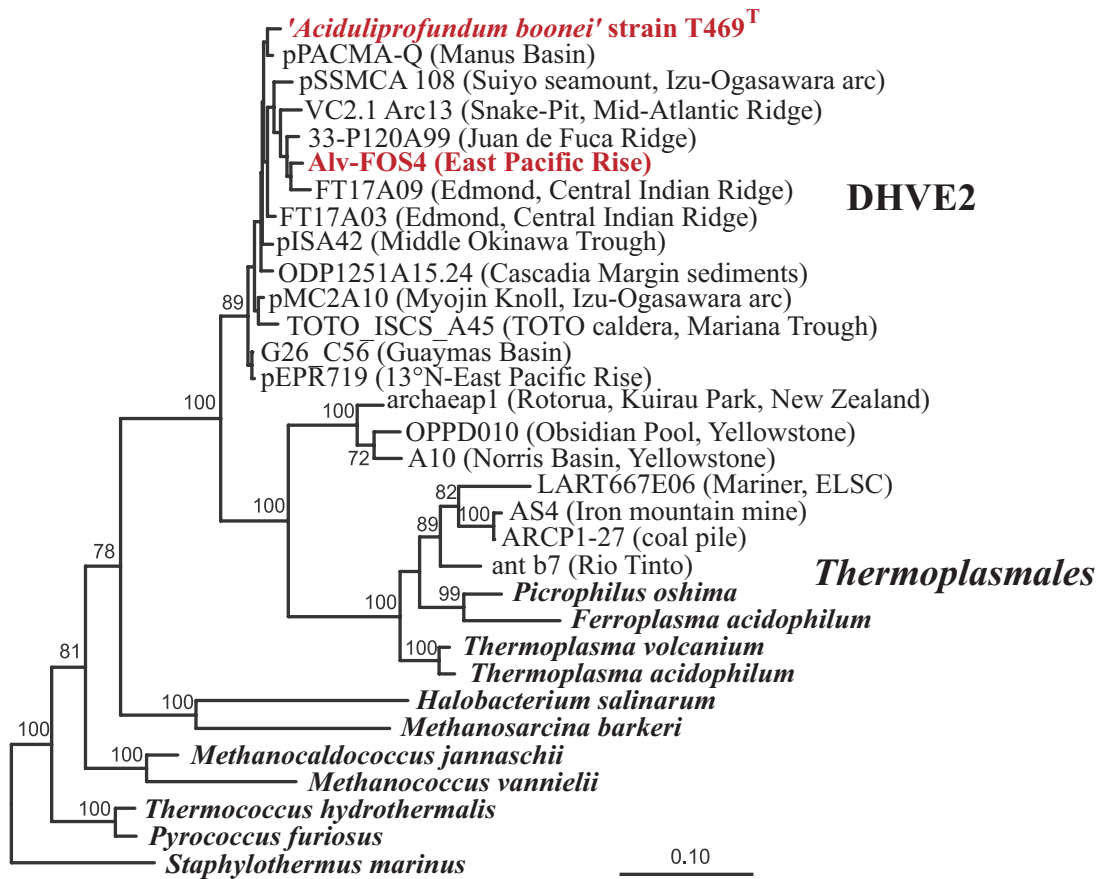


Fig. 3 – Relations phylogénétiques entre les séquences d'ARNr 16S de la souche T469, de représentants des DHVE2 dont Alv-FOS4 et d'*Euryarchaeota*. (D'après Reysenbach *et al.*, 2006) (voir cette référence pour les détails de la méthode phylogénétique utilisée)

Conclusion et perspectives

La démarche de cette thèse était d'étudier la diversité des micro-organismes rencontrés dans l'écosystème hydrothermal marin profond. Les techniques moléculaires qui ont été employées pour ce travail de thèse, ont contribué à étudier ces micro-organismes sans le biais de sélection lié à leur mise en culture. La première partie de cette thèse a permis d'étendre nos connaissances actuelles sur la diversité et l'abondance des ε -*Proteobacteria* au sein de cet écosystème. La détection de bactéries affiliées à des *Arcobacter* incultivés et phylogénétiquement éloignés de la souche "*Candidatus Arcobacter sulfidicus*" a révélé qu'un nouveau groupe d' ε -*Proteobacteria* pouvait être impliqué dans la formation de biofilms composés principalement de filaments de soufre que ces micro-organismes sécrèteraient. Cette étude participe à démontrer qu'il reste encore de nombreux clades qui n'ont pas de représentants cultivés parmi les ε -*Proteobacteria* détectées dans l'écosystème hydrothermal marin profond.

La banque métagénomique construite lors de ce travail de thèse, pourrait maintenant être exploitée de sorte à apporter des éléments de réponse pour la culture de ces ε -*Proteobacteria*. On pourrait en effet cribler cette banque pour la présence d'ARNr 16S bactériens, dans l'espoir d'identifier parmi les clones porteurs de ces gènes, des clones affiliés à des ε -*Proteobacteria* incultivées. Tout porte à croire que nous pourrions identifier parmi les quelques 5000 clones de notre banque métagénomique, des clones porteurs d'ARNr 16S affiliés à des ε -*Proteobacteria*. Tout d'abord, les résultats complémentaires présentés à la suite de l'étude 2, montrent en effet que l'ADN ayant servi à construire notre banque génomique contient entre autres, de nombreux gènes d'ARNr 16S bactériens affiliés aux ε -*Proteobacteria*. De plus, parmi les 6 clones de la banque métagénomique qui portaient un gène codant pour un ARNr 16S affilié aux *Archaea*, quatre appartenaient à des DHVE2, et un autre appartenait à une *Thermococcale*. Or, il a été suggéré que les DHVE2, les *Thermococcales* et les ε -*Proteobacteria* occupaient la même niche écologique au sein de l'écosystème hydrothermal marin profond : les ε -*Proteobacteria* qui pour la plupart oxydent les composés réduits du soufre ou l'hydrogène, contribueraient en effet à abaisser localement le pH, créant ainsi une microniche anaérobie acidophile qui conviendrait aux DHVE2 (Reysenbach *et al.*, 2006).

Le criblage de la banque métagénomique pourrait également être orienté à la recherche de gènes fonctionnels impliqués dans des métabolismes importants. Nous pourrions par exemple tenter d'étendre nos connaissances actuelles concernant la fixation autotrophe du carbone dans l'écosystème hydrothermal marin profond. L'utilisation de marqueurs fonctionnels tels que ceux impliqués dans l'utilisation du cycle rTCA, permettraient ainsi d'approfondir nos connaissances sur les micro-organismes utilisant potentiellement cette voie pour fixer leur carbone. Mais d'autres gènes pourraient potentiellement être utilisés pour cribler cette banque métagénomique : citons par exemple ceux impliqués dans la voie de la réduction du sulfate (gène de la sulfite réductase qui catalyse la réduction du sulfite en sulfure d'hydrogène), de la méthanogénèse (gène *mcrA*),

dans le cycle de l'azote (gènes *nif*, *nor*, *nar*, *nos* et *amo*).

L'étude d'une communauté microbienne associées à des tubes d'*Alvinellideae* nous a permis d'approfondir nos connaissances sur le compartiment microbien du pôle chaud de l'écosystème hydrothermal. Les compartiments plus froids de cet écosystème mériteraient également d'être soumis à des études de ce genre. L'étude métagénomique de la communauté microbienne associée à des tubes du ver *Riftia pachyptila* permettrait par exemple d'accéder à un compartiment plus froid. Tout comme les tubes d'*A. pompejana*, ces tubes forment un support idéal pour la colonisation de communautés microbiennes très diversifiées (Lopez-Garcia *et al.*, 2002). La communauté microbienne associée à ces tubes diffère significativement de celle observée sur les tubes d'*A. pompejana*. Cela tient en partie à l'environnement physico-chimique des tubes d'*R. pachyptila* : ces tubes sont exposés à des températures plus froides et à un environnement moins réduit que ceux d'*A. pompejana* qui sont d'avantage exposés aux fluides hydrothermaux (Lopez-Garcia *et al.*, 2002).

Les tapis microbiens recouvrant les roches, les cheminées, les sédiments ou les surfaces des animaux (Karl, 1995) pourraient également constituer un échantillon de choix pour une nouvelle étude métagénomique. D'un point de vue méthodologique, les biofilms offrent une biomasse microbienne importante, permettant d'extraire de l'ADN génomique en quantité suffisante pour pouvoir construire une banque métagénomique. D'autre part, les biofilms semblent être un mode de vie favorisé des micro-organismes des systèmes hydrothermaux (Huber *et al.*, 2002, 2003).

Annexes

ANNEXE A

BANQUES DE DONNÉES DE SÉQUENCES
NUCLÉIQUES ET PROTÉIQUES

1 Banques de données de séquences nucléiques

1.1 Banques nucléiques généralistes

Les séquences nucléiques et leurs annotations sont accessibles sur trois principales banques généralistes internationales :

- la banque européenne EMBL diffusée par l'EBI (Européan Bioinformatics Insitute) (<http://www.ebi.ac.uk/embl/>)
- la banque européenne EMBL diffusée par l'EBI (Européan Bioinformatics Insitute) (<http://www.ebi.ac.uk/embl/>)
- la banque américaine GenBank soutenue par le NCBI (National Center for Biotechnology Information) (<http://www.ncbi.nlm.nih.gov/>)
- et la banque japonaise DDBJ (DNA Data Bank of Japan) du centre NIG (National Institute of Genetics) (<http://www.ddbj.nig.ac.jp/>) Depuis 1987, ces trois banques ont adopté un système une convention communes qui leur permet d'échanger systématiquement tous leurs fichiers grâce à une collaboration internationale appelée Internationnal Nucleotide Sequence Database (INSD : <http://www.ncbi.nlm.nih.gov/projects/collab/>).

1.2 Banques nucléiques spécialisées

Il existe également d'autres banques de données de séquences nucléiques telles que la banque RDP (Ribosomal Database Project, <http://rdp.cme.msu.edu/>) qui regroupe des séquences d'ARNr 16S annotées et alignées (Cole et al., 2003), et la banque dédiée aux ARNt (<http://www.uni-bayreuth.de/departments/biochemie/sprinzi/trna/>) (Sprinzi and Vassilenko, 2005).

2 Banques de données des séquences protéiques

2.1 Banques généralistes

Les principales banques de données de séquences protéiques sont les banques TrEMBL (<http://www.ebi.ac.uk/trembl/>) et GENPEPT, qui correspondent en fait aux versions protéiques des deux banques nucléiques EMBL et GenBank. Il y a également la banque de données SWISS PROT (<http://www.expasy.org/spot/>) maintenue et distribuée par la SIB (Swiss Insitute of Bioinformatics) et l'EBI, et la banque de données NBRF-PIR (Protein Information Ressource) (<http://www-nbrf.georgetown.edu/pir/>) diffusée conjointement par le NBRF (National Bio-medical Research Foundation) aux États-Unis, le MIPS (Martinsried Institute for Protein Sequences) en Allemagne, et la JIPID (Japan International Protein Information Database) au Japon. La banque SWISS PROT fournit une remarquable qualité des données, dans la mesure où les séquences sont annotées et expertisées non pas par les auteurs qui les soumettent, mais par une équipe d'annotateurs. Certaines banques centralisent les séquences protéiques et leurs

annotations, c'est le cas de la banque UNIPROT (<http://www.ebi.ac.uk/uniprot>) qui joint les informations contenues dans Swiss-Prot, TrEMBL et PIR.

2.2 Banques protéiques spécialisées

A titre d'exemple, voici quelques types de banques protéiques spécialisées qui ont été mises en place. Tout d'abord, on peut citer les banques protéiques d'alignement, c'est le cas des banques PFAM (<http://www.sanger.ac.uk/Software/Pfam/>), BLOCKS (<http://blocks.fhcrc.org/blocks/>) et CDD (http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd_help.shtml). Les banques de motifs protéiques sont les suivantes : InterPro (<http://www.ebi.ac.uk/interpro/>), PROSITE (<http://www.expasy.org/prosite/>) et PRINTS (<http://www.biochem.ucl.ac.uk/bsm/dbbrowser/PRINTS/printsman.html>). Il existe également des banques d'enzymes comme la banque ENZYME (<http://www.expasy.org/txt/enzuser.txt>).

**COMPOSÉS CHIMIQUES DES
ENVIRONNEMENTS HYDROTHERMAUX
MARINS PROFONDS**

Concentration et origine des composés chimiques identifiés dans les environnements hydrothermaux marins profonds (d'après Kelley *et al.*, 2002)

Chemical species	Vent fluids conc/kg SW	Seawater conc/kg	Sources in vents	Biological significance
CO ₂	3.9–215 mmol	2.3 mmol	Magma degassing, water/rock reactions, fermentation	Carbon source for chemoautotrophs and methanogens
CH ₄	0.05–4.5 mmol	0.3 nmol	Magma degassing, reduction of CO ₂ , methanogenesis	Aerobic and anaerobic microbial oxidation
H ₂	0.1–50 mmol	0.3 nmol	Magma degassing, water/rock reactions, microbial fermentation	Methanogenesis; aerobic and anaerobic oxidation by microorganisms
S (as H ₂ S)	3–110 mmol	Not detected 28 mmol (as SO ₄ ²⁻)	Water/rock reactions, chemical reduction to anhydrate, microbial reduction of SO ₄ ²⁻	H ₂ S is the primary energy source for aerobic chemoautotrophs including symbionts; oxidized sulfur species reduced by high diversity of microorganisms; present in some amino acids and Fe-S clusters as protein cores
N (as NH ₃ or NH ₄ ⁺)	< 0.01–1 mmol ^a	< 0.01 mmol 30 μ mol (as NO ₃ ⁻)	Organic-N in buried sediments, possibly N ₂ -fixation by microorganisms and/or chemical reduction of N ₂ ^b	A source of nitrogen for subsurface microorganisms; oxidation to NO ₃ ⁻ or NO ₂ ⁻ by nitrifying bacteria
P (as PO ₄ ²⁻)	0.5 μ mol	2.5 μ mol	Water/rock reactions (P ₂ O ₅ in basalts), SW PO ₄ ²⁻	All organisms require P for nucleic acids, energy reactions and fatty acids; animals and aerobic microbes use SW-PO ₄ ²⁻ and/or detrital organic PO ₄ ²⁻ compounds
Fe	0.009–18 mmol as Fe(II)	< 1 nmol	Water/rock reactions (12% iron in crust), major component of sulfides as pyrite	Energy source for Fe(II) oxidizing bacteria. Fe(III) as electron acceptor for specific groups of vent microbes. Trace element required by all organisms
Mn	0.1–4.5 mmol as Mn(III)	< 1 nmol	Water/rock reactions	Energy source for Mn(III) oxidizing bacteria. Trace element required by all organisms
Si	2.7–23 mmol	0.16 mmol	Water/rock reactions, SW entrainment, major component of outer layers of sulfide deposits	Some evidence for microbially mediated Si precipitation in sulfides

Chemical species	Vent fluids conc/kg SW	Seawater conc/kg	Sources in vents	Biological significance
Zn	2–100 μmol	0.01 μmol	Water/rock reactions, major component of sulfides as pyrite	Trace element required for key enzymes including alkaline phosphatase and RNA polymerase
Cu	0.02–44 μmol	0.007 μmol	Water/rock reactions, major component of sulfides as chalcopyrite	Trace element required for enzymes including oxidative enzymes involved in electron transfer; component of haemocyanin in arthropods; toxic in μmol concentrations
Co	20–200 nmol	0.03 nmol	Water/rock reactions	Trace element required by all organisms usually involving enzymes requiring B ₁₂ coenzymes (contain Co)
Cd	1–180 nmol	< 1 nmol	Water/rock reactions	Toxic to all organisms in nmol levels; binds to S, N centers of proteins and DNA
Pb	9–359 nmol	0.01 nmol	Water/rock reactions	Toxic to all organisms in nmol levels
Mo	1–33 nmol	< 1 nmol	Water/rock reactions	Trace element required for specific enzymes including nitrogenase
W ^c	< 0.01–2.1 μmol	< 1 nmol	Water/rock reactions	Trace element required for specific enzymes by hyperthermophilic archaea
Ni ^c	< 0.01–53 μmol	< 1 nmol	Water/rock reactions	Important component in coenzymes (F-430) of methanogens and enzymes of hyperthermophilic archaea

^aChemical data from Butterfield et al. 1995; Elderfield & Schultz 1996; Lilley et al. 1993; Von Damm 1990, 1995.

^bThere are experimental data for the reduction of N₂ to NH₃ under hydrothermal vent conditions (Brandes et al. 1998).

^cData from high temperature fluids from nine different sites at Endeavour on the Juan de Fuca Ridge (Baross & Adams, unpublished).

PUBLICATION DE L'ARTICLE :
“*Thermodesulfatator indicus* GEN.
NOV., SP. NOV., A NOVEL THERMOPHILIC
CHEMOLITHOAUTOTROPHIC
SULFATE-REDUCING BACTERIUM
ISOLATED FROM THE CENTRAL INDIAN
RIDGE”

H. Moussard, S. L'Haridon, B. J. Tindall, A. Banta, P. Schumann, E. Stackebrandt,
A.-L. Reysenbach, C. Jeanthon

Article publié en 2004, dans International Journal of Systematic and Evolutionary Microbiology

Thermodesulfatator indicus gen. nov., sp. nov., a novel thermophilic chemolithoautotrophic sulfate-reducing bacterium isolated from the Central Indian Ridge

H. Moussard,¹ S. L'Haridon,¹ B. J. Tindall,² A. Banta,³ P. Schumann,²
E. Stackebrandt,² A.-L. Reysenbach³ and C. Jeanthon¹

Correspondence

H. Moussard
moussard@univ-brest.fr

A.-L. Reysenbach
reysenbacha@pdx.edu

¹UMR 6539, Centre National de la Recherche Scientifique & Université de Bretagne Occidentale, Institut Universitaire Européen de la Mer, 29280 Plouzané, France

²DSMZ – German Collection of Microorganisms and Cell Cultures, Mascheroder Weg 1b, D-38124 Braunschweig, Germany

³Department of Biology, Portland State University, Portland, OR 97201, USA

A thermophilic, marine, anaerobic, chemolithoautotrophic, sulfate-reducing bacterium, strain CIR29812^T, was isolated from a deep-sea hydrothermal vent site at the Kaiei vent field on the Central Indian Ridge. Cells were Gram-negative motile rods that did not form spores. The temperature range for growth was 55–80 °C, with an optimum at 70 °C. The NaCl concentration range for growth was 10–35 g l⁻¹, with an optimum at 25 g l⁻¹. The pH range for growth was 6–6.7, with an optimum at approximately pH 6.25. H₂ and CO₂ were the only electron donor and carbon source found to support growth of the strain. However, several organic compounds were stimulatory for growth. Sulfate was used as electron acceptor, whereas elemental sulfur, thiosulfate, sulfite, cystine, nitrate and fumarate were not. No fermentative growth was observed with malate, pyruvate or lactate. The phenotypic characteristics of strain CIR29812^T were similar to those of *Thermodesulfobacterium hydrogeniphilum*, a recently described thermophilic, chemolithoautotrophic sulfate-reducer. However, phylogenetic analyses of the 16S rRNA gene sequences showed that the new isolate was distantly related to members of the family *Thermodesulfobacteriaceae* (similarity values of less than 90%). The chemotaxonomic data (fatty acids and polar lipids composition) also indicated that strain CIR29812^T could be distinguished from *Thermodesulfobacterium commune*, the type species of the type genus of the family *Thermodesulfobacteriaceae*. Finally, the G + C content of the genomic DNA of strain CIR29812^T (46.0 mol%) was not in the range of values obtained for members of this family. On the basis of phenotypic, chemotaxonomic and genomic features, it is proposed that strain CIR29812^T represents a novel species of a new genus, *Thermodesulfatator*, of which *Thermodesulfatator indicus* is the type species. The type strain is CIR29812^T (= DSM 15286^T = JCM 11887^T).

In the latest edition of *Bergey's Manual of Systematic Bacteriology*, the class *Thermodesulfobacteria* (Garrity & Holt, 2001) contained two species: *Thermodesulfobacterium commune* and *Thermodesulfobacterium mobile* (recently renamed *Thermodesulfobacterium thermophilum*) (Judicial

Commission of the International Committee on Systematics of Prokaryotes, 2003). In the past few years, two novel species of the genus *Thermodesulfobacterium*, *Thermodesulfobacterium hveragerdense* (Sonne-Hansen & Ahring, 1999) and *Thermodesulfobacterium hydrogeniphilum* (Jeanthon *et al.*, 2002), have been described and classified in this genus. All *Thermodesulfobacterium* species are anaerobic, thermophilic, non-spore-forming, deeply branching, sulfate-reducing bacteria. With the exception of the marine chemolithoautotrophic organism *T. hydrogeniphilum*, all *Thermodesulfobacterium* spp. are chemo-organotrophs that thrive in terrestrial and subterrestrial environments (Zeikus *et al.*, 1983; Rozanova & Khudyakova, 1974;

Published online ahead of print on 1 August 2003 as DOI 10.1099/ij.s.0.02669-0.

The GenBank accession number for the 16S rRNA gene sequence of *Thermodesulfatator indicus* CIR29812^T is AF393376.

Details for the fatty acids of strain CIR29812^T are available in IJSEM Online.

Sonne-Hansen & Ahring, 1999). Recently, a new strictly chemolithoautotrophic, iron-reducing species, '*Geothermobacterium ferrireducens*', was isolated from hot springs at Yellowstone National Park (USA). This organism represents the only member of the family *Thermodesulfobacteriaceae* that is unable to reduce sulfate (Kashefi *et al.*, 2002).

We describe here the isolation and characterization of another novel thermophilic, strictly chemolithoautotrophic, sulfate-reducing bacterium. The new isolate was obtained from a sample of an active black smoker collected at a depth of 2420 m at the Kairei vent field (25°19' S, 70°02' E) on the Central Indian Ridge (Van Dover *et al.*, 2001) in April 2001. The chimney fragment was collected by the ROV *Jason* and was placed in an isolated container for the trip to the surface. Subsamples of the chimney fragment were ground in a mortar and the slurry was stored under an atmosphere of nitrogen at 4 °C until used as an inoculum.

Initial enrichments were done using the following medium that contained (l⁻¹ distilled water): 29 g NaCl; 7 g MgSO₄·7H₂O; 4 g NaOH; 0.5 g KCl; 2 g Na₂S₂O₃·5H₂O; 1.66 g MgCl₂·6H₂O; 0.4 g CaCl₂·2H₂O; 0.2 g NH₄Cl; 0.3 g K₂HPO₄·3H₂O; and 10 ml of a trace element stock solution according to Boone *et al.* (1989) (http://methanogens.pdx.edu/OCM_media.html). The medium was prepared with anoxic water and, prior to autoclaving, the pH was adjusted to pH 6 at room temperature with sulfuric acid. The medium was dispensed under a CO₂ atmosphere into Bellco tubes and capped with butyl-rubber stoppers. After inoculation with the sulfide slurry [10% (v/v) inoculum], the tubes were pressurized with H₂ (100%; 138 kPa) and incubated at 70 °C without shaking.

After 4 days, cultures of small motile rods producing sulfide were observed. Enrichments that produced sulfide were subsequently transferred to a sulfate-reducing bacteria (SRB) medium that consisted of (l⁻¹ distilled water): 20 g NaCl; 4 g Na₂SO₄; 3 g MgCl₂·6H₂O; 0.2 g KH₂PO₄; 0.5 g KCl; 0.25 g NH₄Cl; 3.46 g PIPES; 0.15 g CaCl₂·2H₂O; 1 mg resazurin; 2 mg sodium tungstate; 0.5 mg sodium selenate; 1 ml vitamin mixture (Widdel & Bak, 1992); 1 ml thiamin solution (Widdel & Bak, 1992); and 0.05 mg vitamin B₁₂. The pH of the medium was adjusted to pH 6.7 at room temperature using 5 M HCl. After autoclaving under N₂ (100%), the pH had decreased to 6.5. Medium (10 ml) was dispensed anaerobically in 50 ml vials sealed with butyl-rubber stoppers and reduced with 0.1 ml of a 10% (w/v) Na₂S·9H₂O sterile solution; H₂/CO₂ (80:20; 200 kPa) was used as the gas phase. Cultures were incubated at 70 °C with shaking (150 r.p.m.). The pH of the medium in uninoculated vials checked at room temperature after incubation at 70 °C decreased from 6.5 to 6.3.

One pure culture, strain CIR29812^T, was obtained by using shake dilution tubes (Widdel & Bak, 1992) of SRB solidified medium, where agar was replaced by 0.7% (w/v) Phytigel (Sigma). After 6 days incubation at 70 °C, smooth, brown, spindle-shaped colonies of approximately 1 mm in

diameter were transferred into SRB medium and checked for purity microscopically. Furthermore, the purity of the isolate was checked at 55 and 70 °C. SRB medium supplemented with 2 g Difco yeast extract l⁻¹, 2 g tryptone l⁻¹ and 10 mM glucose with air in the headspace was used to check for aerobic contaminants. The latter medium prepared anaerobically with N₂ (100%; 200 kPa) or H₂ (100%; 100 kPa) as the gas phase was used to detect anaerobic contaminants. The presence of possible autotrophic contaminants was checked in SRB medium where sulfate was omitted but where 2 g Difco yeast extract l⁻¹ and 2 mM acetate were added. Stock cultures of strain CIR29812^T were stored in SRB medium at 4 °C. However, frequent transfers (twice per month) with 10% (v/v) of inoculum in freshly prepared culture medium were found optimal to ensure re-growth. Alternatively, the isolate was stored in liquid nitrogen in the same medium containing 5% (w/v) DMSO.

Cells of strain CIR29812^T were small rods, approximately 0.8–1 µm in length and 0.4–0.5 µm in width, with a single polar flagellum (Fig. 1a, b). Cells occurred singly, in pairs or in chains of three cells, and elongated during the stationary phase of growth. Occasionally, visible creamy aggregates that corresponded to large clumps of cells could be observed in the liquid medium. No spores were produced.

Unless otherwise stated, growth experiments were performed in duplicate in SRB medium supplemented with

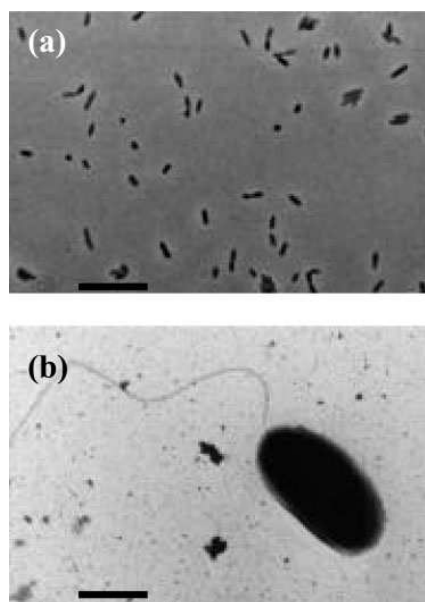


Fig. 1. (a) Phase-contrast micrograph of strain CIR29812^T; bar, 5 µm. (b) Electron micrograph of negatively stained cell (method as described by Jeanthon *et al.*, 2002); bar, 500 nm.

0.5 g tryptone l⁻¹ and 2 mM acetate. Growth was monitored by measuring the increase in optical density at 600 nm with a Spectronic 401 spectrophotometer (Bioblock). The temperature range for growth was determined without agitation with 20 g NaCl l⁻¹ at pH 6.5. The NaCl range was obtained at 70 °C and pH 6.5 under agitation (150 r.p.m.). To determine the pH range for growth, SRB medium was buffered with 20 mM MES (pH adjusted to 6) or 20 mM PIPES (pH adjusted to 6.7 and 7.2). After autoclaving, these pH values decreased to pH 5.9 (with MES as buffer), pH 6.5 and pH 7 (with PIPES as buffer). The pH ranges from 5.9 to 6.75 (with MES) and 6 to 7 (with PIPES) were obtained by the addition of varying concentrations of NaHCO₃. The pH of the media was checked at room temperature after overnight incubation of uninoculated tubes under H₂/CO₂ at 70 °C.

Under these conditions, strain CIR29812^T grew between 55 and 80 °C, with an optimum at 70 °C. No growth was observed at 50 or 82 °C. Growth occurred between 10 and 35 g NaCl l⁻¹, with a growth optimum at 25 g NaCl l⁻¹. No growth was detected after 96 h in media containing 5 and 40 g NaCl l⁻¹. Growth occurred between pH 6 and 6.7 in PIPES buffered medium, with an optimum at approximately pH 6.25. Growth occurred in MES buffered medium from pH 6 to 6.25. Under optimal growth conditions with shaking (150 r.p.m.), the doubling time of strain CIR29812^T was around 2 h (maximal OD₆₄₀ 0.11).

The new isolate was a strict anaerobe and was transferred at least six times under strict chemolithoautotrophic conditions, using H₂ as the electron donor and sulfate as the electron acceptor. Hydrogen sulfide was produced during growth. Elemental sulfur (1%), thiosulfate (10 mM), cystine (1%), nitrate (5 mM), fumarate (10 mM) and sulfite (2 mM) were not used as electron acceptors. CO₂ was the sole carbon source used by strain CIR29812^T. In the presence of H₂/CO₂ and sulfate, growth was stimulated by acetate (2 mM), methanol (0.5%), monomethylamine (0.2%), glutamate (5 mM), peptone (0.1%), fumarate (15 mM), tryptone (0.1%), isobutyrate (5 mM), 3-CH₃ butyrate (5 mM), ethanol (10 mM) and propanol (5 mM). In the presence of H₂/CO₂ and sulfate, growth was not affected by isovalerate (5 mM), glucose (5 mM), fructose (5 mM) or succinate (10 mM), whereas acetate (15 mM), propionate (10 mM), butyrate (10 mM), 2-CH₃ butyrate (5 mM) and yeast extract (0.2%) were slightly inhibitory. Growth was completely inhibited by lactate (15 mM), caprylate (2.5 mM), caproate (5 mM), caprate (2.5 mM), formate (15 mM), malate (10 mM), valerate (5 mM), pyruvate (10 mM) and heptanoate (5 mM). In sulfate-free medium, no fermentative growth was observed with malate, pyruvate or lactate. The strain preferentially used ammonium (5 mM) as the nitrogen source but peptone (0.5%), nitrate (5 mM) and tryptone (0.1%) also supported growth.

Unlike the control culture of *Desulfovibrio fructosovorans*

DSM 3604^T (Ollivier *et al.*, 1988), strain CIR29812^T did not contain desulfovirdin (Postgate, 1959).

Sensitivity to antibiotics (at 25, 50, 100 and 200 µg ml⁻¹) was tested at 70 °C. Strain CIR29812^T was resistant to penicillin and kanamycin (200 µg ml⁻¹) and streptomycin (100 µg ml⁻¹), but was inhibited by tetracycline (50 µg ml⁻¹), ampicillin, chloramphenicol and rifampicin (all at 25 µg ml⁻¹).

Respiratory lipoquinones and polar lipids were extracted from 100 mg of freeze-dried cell material using the two-stage method described by Tindall (1990a, b). Respiratory quinones were extracted using methanol/hexane (Tindall, 1990a, b) and the polar lipids were extracted by adjusting the remaining methanol/0.3% aqueous NaCl phase (containing the cell debris) to give a chloroform/methanol/0.3% aqueous NaCl mixture (1:2:0.8, by vol.). The extraction solvent was stirred overnight and the cell debris pelleted by centrifugation. Respiratory lipoquinones were separated into their different classes (menaquinones and ubiquinones) by TLC on silica gel (Macherey-Nagel art. no. 805 023), using hexane/*tert*-butylmethylether (9:1, v/v) as solvent. UV-absorbing bands corresponding to menaquinones or ubiquinones were removed from the plate and further analysed by HPLC. This step was carried out on an LDC Analytical (Thermo Separation Products) HPLC apparatus fitted with a reverse phase column (2 mm × 125 mm, 3 µm, RP18; Macherey-Nagel) using methanol/heptane as the eluant. Respiratory lipoquinones were detected at 269 nm.

Examination of the respiratory lipoquinone composition of strain CIR29812^T indicated that menaquinones were the sole respiratory quinones present. The major component was a menaquinone with seven isoprenologues, i.e. menaquinone 7 (MK-7). MK-7 had also been identified as the major menaquinone in *T. commune* and *T. thermophilum* (Collins & Widdel, 1986).

Polar lipids were recovered into the chloroform phase by adjusting the chloroform/methanol/0.3% aqueous NaCl mixture to a ratio of 1:1:0.9 (by vol.). They were separated by two-dimensional silica-gel TLC (Macherey-Nagel art. no. 818 135). The first direction was developed in chloroform/methanol/water (65:25:4, by vol.) and the second was developed in chloroform/methanol/acetic acid/water (80:12:15:4, by vol.). Total lipid material and specific functional groups were detected using dodecamolybdo-phosphoric acid (total lipids), Zinzadze reagent (phosphate), ninhydrin (free amino groups), periodate-Schiff (α -glycols), Dragendorff (quaternary nitrogen) and anisaldehyde-sulfuric acid (glycolipids).

The polar lipids of strain CIR29812^T were predominantly phospholipids (Fig. 2). The two major lipids were identified initially on the basis of their R_F values and staining behaviour as phosphatidylinositol and phosphatidylethanolamine. In addition, one of the minor phospholipids was identified as phosphatidylglycerol. Additional phospholipids

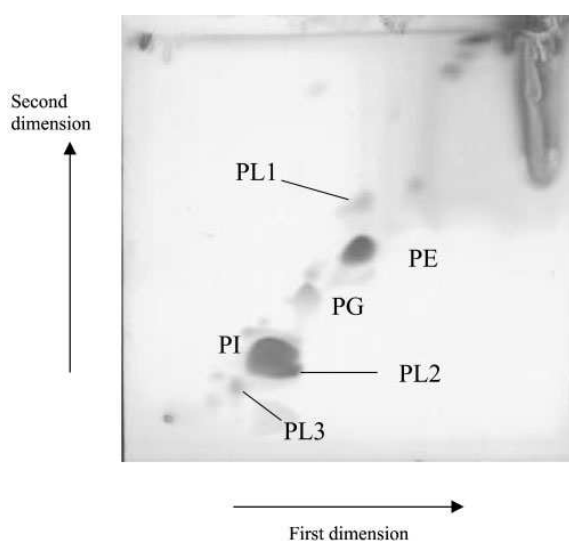


Fig. 2. Two-dimensional thin-layer chromatogram of the polar lipids of strain CIR29812^T. All polar lipids were stained with 5% ethanolic molybdophosphoric acid. Solvents: chloroform/methanol/water (65:25:4, by vol.), first direction; chloroform/methanol/acetic acid/water (80:12:15:4, by vol.), second direction. PE, phosphatidylethanolamine; PG, phosphatidylglycerol; PI, phosphatidylinositol; PL1, PL2 and PL3, phospholipids of unknown structure.

(PL1, PL2, PL3) were present in small amounts and could not be identified unambiguously. Confirmation of the head-group structures was made by ESI-MS/MS studies, details of which will be reported elsewhere. The lipid composition of *T. commune* was similar. However, a third phospholipid identified in *T. commune* was not present in strain CIR29812^T. This phospholipid had the R_F value and staining behaviour of the phosphatidyl aminopentatetrol, which has also been reported in *Hydrogenobacter thermophilus* (Yoshino *et al.*, 2001), the head-group structure having originally been described in the methanogenic members of the *Archaea* (Ferrante *et al.*, 1987, 1988).

Fatty acids were analysed as the methyl ester derivatives prepared from 10 mg of dry cell material. Cells were subjected to differential hydrolysis in order to detect ester-linked and non-ester-linked (amide-bound) fatty acids (B. J. Tindall, unpublished). Fatty acid methyl esters were analysed by GC using a 0.2 $\mu\text{m} \times 25$ m non-polar capillary column and flame-ionization detection. The run conditions were injection and detector port temperature 300 °C, inlet pressure 60 kPa, split ratio 50:1, injection volume 1 μl , with a temperature programme from 130 to 310 °C at a rate of 4 °C min^{-1} .

The fatty acids of strain CIR29812^T comprised both saturated and unsaturated straight-chain, as well as hydroxylated, fatty acids. The major fatty acids of strain CIR29812^T

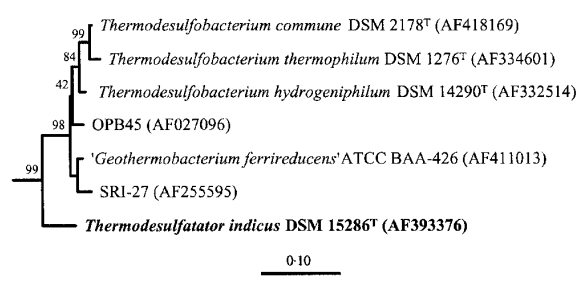


Fig. 3. Phylogenetic relationships of *Thermodesulfator indicus* (strain CIR29812^T) and other members of the family *Thermodesulfobacteriaceae*, produced by maximum-likelihood analysis. The 16S rRNA gene sequence of strain CIR29812^T was aligned with other 16S rRNA gene sequences from the Ribosomal Database Project (Maidak *et al.*, 2001) and GenBank. Environmental sequences AF027096 (OPB45) and AF411013 (SRI27) have been retrieved from hot springs in Yellowstone National Park (USA) and in Iceland, respectively (Hugenholtz *et al.*, 1998; Skirnisdottir *et al.*, 2000). Bar, expected number of changes per sequence position. The numbers at the branch nodes are bootstrap values based on 100 bootstrap resamplings. Only bacteria belonging to the family *Thermodesulfobacteriaceae* are shown. The tree was generated with *Bacillus subtilis* (GenBank accession no. K00637), *Hellobacterium chlorum* (M11212), *Escherichia coli* (J01695), *Flexibacter flexilis* (M62794), *Thermus thermophilus* (X07998), *Deinococcus radiodurans* (M21413), *Thermotoga maritima* (M21774), *Thermosipho melanesiensis* (Z70248), *Persephonella marina* (AF188332) and *Aquifex pyrophilus* (M83548), and *Methanocaldococcus jannaschii* (M59126) as the outgroup.

consisted of $C_{18:0}$ (42.7–50.9%) and $C_{18:1}$ (19.2–23.6%) (Table I, available from IJSEM Online). By comparison, Langworthy *et al.* (1983) reported the presence of iso-, anteiso- and straight-chain fatty acids in *T. commune*, a pattern which could be confirmed in this study. Although Langworthy *et al.* (1983) examined the fatty acid composition of the lipid fraction, a re-examination of the fatty acid composition from whole cells confirmed these results, but also indicated the presence of hydroxyl fatty acids (data not shown). We assume that the hydroxyl fatty acids are bound to the cell, perhaps in the form of lipopolysaccharide-bound fatty acids.

For the determination of the G+C content, DNA was isolated after disruption of cells using a French pressure cell (Thermo Spectronic) and purified by hydroxyapatite chromatography (Cashion *et al.*, 1977). The DNA was hydrolysed with P1 nuclease and the nucleotides dephosphorylated with bovine alkaline phosphatase (Mesbah *et al.*, 1989). The G+C content of the DNA of strain CIR29812^T determined by the HPLC method described by Tamaoka & Komagata (1984) was 46 mol%.

A total of 1496 nt from the 16S rRNA gene were sequenced as described previously (Götz *et al.*, 2002). The sequence was

reconfirmed using the Thermo Sequenase TM Primer Cycle Sequencing Kit (Amersham) and the reactions were run on a LI-COR automatic sequencer (model 4200) using the LI-COR BASE IMAGEIR software (Science Tec) for analysis.

Distance and maximum-likelihood analyses (De Soete 1983; Olsen *et al.*, 1994) (1301 nt were used) revealed that strain CIR29812^T clustered with all other members of the family *Thermodesulfobacteriaceae* and was most closely related to *T. hydrogeniphilum* (10.3 % distant) (Fig. 3).

The metabolic and physiological properties of strain CIR29812^T are very similar to those of *T. hydrogeniphilum* SL6^T. Contrary to other members of the family *Thermodesulfobacteriaceae*, both organisms are thermophilic, chemolithoautotrophic, sulfate-reducing bacteria, which are non-fermenting, unable to reduce thiosulfate or sulfite, and require NaCl for growth (Table 1). However, their optimal temperature and NaCl range for growth and their resistance to streptomycin and penicillin represent phenotypic characteristics that distinguish these sulfate-reducing chemolithoautotrophs from one another. In addition, the 16S rRNA gene sequences of strain CIR29812^T and *T. hydrogeniphilum* SL6^T are very different (10.3 % distance). Moreover, when analysed using the same method (HPLC), a 15 % difference discriminates the G + C content of their DNA. Lastly, the major fatty acids present in strain CIR29812^T differed from those of *T. commune*, the type species of the type genus of the family

Thermodesulfobacteriaceae. They were more similar to the fatty acid patterns reported for members of the *Aquifex-Hydrogenobacter* group (Stöhr *et al.*, 2001), although the longer-chain components found in the latter group were not found in strain CIR29812^T.

Based on a combination of 16S rRNA, chemotaxonomic and physiological data, we propose that strain CIR29812^T be placed into a new genus within the family *Thermodesulfobacteriaceae*, for which we propose the name *Thermodesulfatator*, as a new species, *Thermodesulfatator indicus*, which is the sole and type species of this genus.

Description of *Thermodesulfatator* gen. nov.

Thermodesulfatator (Ther.mo.de.sul.fa.ta'tor. Gr. masc. n. *thermos* heat; N.L. n. *desulfatator* sulfate-reducer; N.L. masc. n. *Thermodesulfatator* thermophilic sulfate-reducer).

Thermophilic. Strictly anaerobic. Marine. Cells are Gram-negative, rod-shaped (0.8–1 µm long and 0.4–0.5 µm wide) and motile by means of a single polar flagellum. They occur singly, in pairs, in chains of three cells and may form cell aggregates in stationary-phase cultures. Do not form spores. Chemolithoautotrophs growing exclusively with hydrogen as the sole electron donor and sulfate as the sole electron acceptor. 16S rRNA gene sequence comparison differentiates *Thermodesulfatator* from the other genera of the family *Thermodesulfobacteriaceae*.

Table 1. Differentiating characteristics of cultivated members of the family *Thermodesulfobacteriaceae*

Data were obtained from Zeikus *et al.* (1983), Rozanova & Pivovarova (1988), Henry *et al.* (1994), Sonne-Hansen & Ahring (1999), Jeanthon *et al.* (2002) and Kashefi *et al.* (2002). Electron donors were tested with CO₂ as the carbon source. ND, Not determined. Species: 1, strain CIR29812^T; 2, *T. hydrogeniphilum*; 3, *G. ferrireducens*; 4, *T. commune*, *T. thermophilum* and *T. hveragerdense*.

Characteristic	1	2	3	4
G + C content (mol%)	46	28 (31.5)*	ND	31–40
NaCl range (g l ⁻¹)	10–35	5–55	0–7.5	<5
Optimal salinity (g l ⁻¹)	25	30	0–0.5	0–1
Temperature range (°C)	55–80	50–80	65–100	45–85
Optimal temperature (°C)	70	75	85–90	65–74
pH range for growth	6–6.7	6.3–6.8	ND	6–8 for <i>T. commune</i> ; ND for the other species
Electron donors:				
H ₂	+‡	+‡	+‡	–
Pyruvate	–	–	–	+
Lactate	–	–	–	+
Pyruvate fermentation	–	–	–	+
Electron acceptors:				
Sulfate	+	+	–	+
Thiosulfate	–	–	–	+
Antibiotic resistance:				
Streptomycin (200 µg ml ⁻¹)	+	–	+	ND
Penicillin (200 µg ml ⁻¹)	–	+	+	ND

*The value in parentheses was obtained in this study by HPLC.

‡Autotrophic growth.

The type species is *Thermodesulfatator indicus*.

Description of *Thermodesulfatator indicus* sp. nov.

Thermodesulfatator indicus (in.di'cus. L. adj. *indicus* referring to the Indian Ocean, from where the strain was isolated).

Gram-negative rods (0.8–1 µm long by 0.4–0.5 µm wide), motile by means of a single polar flagellum. Cells occur singly, in pairs or in chains of three cells in early cultures. Growth occurs between 55 and 80 °C (optimum 70 °C), pH 6 and 6.7 (optimum at about pH 6.25) and in the presence of 10 and 35 g NaCl l⁻¹ (optimum 25 g l⁻¹). Anaerobic. Strictly chemolithoautotrophic using sulfate as electron acceptor and H₂ as electron donor. No fermentative metabolism. With H₂/CO₂ and sulfate, growth is stimulated by methanol, monomethylamine, glutamate, peptone, fumarate, tryptone, isobutyrate, 3-CH₃ butyrate, ethanol, propanol and low amounts of acetate. Unable to use sulfur, cystine, thiosulfate, sulfite, fumarate and nitrate as electron acceptor. Ammonium is the preferred nitrogen source. Sensitive to ampicillin, chloramphenicol and rifampicin (25 µg ml⁻¹). Resistant to tetracycline and streptomycin (100 µg ml⁻¹), penicillin and kanamycin (200 µg ml⁻¹). The major lipoquinone is MK-7. Predominant polar lipids are phosphatidylethanolamine and phosphatidylinositol. Small amounts of phosphatidylglycerol and three unknown phospholipids (PL1, PL2, PL3) are detected. Fatty acid profile is mainly composed of C_{18:0} and C_{18:1}.

The type strain (CIR29812^T = DSM 15286^T = JCM 11887^T) was isolated from an active hydrothermal sulfide chimney deposit at the Kairei vent field on the Central Indian Ridge. The G + C content of its DNA is 46.0 mol%.

Acknowledgements

We are grateful to Bernard Ollivier (IRD, Marseille) for the gift of *Desulfovibrio fructosovorans* DSM 3604^T. Our thanks go also to Manfred Nimtz and Andrea Abrahamik (GBF, Braunschweig) for running and help in the interpretation of the preliminary ESI-MS/MS experiments. We also acknowledge Adelaide Nieguitsila for her help with phenotypic characterization. The work performed at Braunschweig and Plouzané was supported by an INTAS grant (99-1250). The work performed at Plouzané was also supported by a CNRS/Rhône-Poulenc grant, a PRIR (THERMOSH) and the programme 'Souchothèque de Bretagne' from the Conseil Régional de Bretagne. H. M. was supported by a grant from the Ministère de la Recherche. The work was also supported by grants from the National Science Foundation (NSF-OCE9712358 and NSF-OCE0083134).

References

Boone, D. R., Johnson, R. L. & Lui, Y. (1989). Diffusion of the interspecies electron carriers H₂ and formate in methanogenic ecosystems and its implications in the measurement of K_m for H₂ or formate uptake. *Appl Environ Microbiol* **55**, 1735–1741.

Cashion, P., Holder-Franklin, M. A., McCully, J. & Franklin, M. (1977). A rapid method for the base ratio determination of bacterial DNA. *Anal Biochem* **81**, 461–466.

Collins, M. D. & Widdel, F. (1986). Respiratory quinones of sulfate-reducing and sulphur-reducing bacteria: a systematic investigation. *Syst Appl Microbiol* **8**, 8–18.

DeSoete, G. (1983). A least squares algorithm for fitting additive trees to proximity data. *Psychometrika* **48**, 621–626.

Ferrante, G., Ekiel, I. & Sprott, G. D. (1987). Structures of diether lipids of *Methanospirillum hungatei* containing novel head groups N,N-dimethylamino- and N,N,N-trimethylaminopentane-tetrol. *Biochim Biophys Acta* **921**, 281–291.

Ferrante, G., Ekiel, I., Patel, G. B. & Sprott, D. (1988). Structure of the major polar lipids isolated from the aceticlastic methanogen, *Methanotheroxiphilium concilii* GP6. *Biochim Biophys Acta* **963**, 162–172.

Garrity, G. M. & Holt, J. G. (2001). Phylum BIII. *Thermodesulfobacteria* phyl. nov. In *Bergey's Manual of Systematic Bacteriology*, 2nd edn, vol. 1, pp. 389–393. Edited by D. R. Boone, R. W. Castenholz & G. M. Garrity. New York: Springer.

Götz, D., Banta, A., Beveridge, T. J., Rushdi, A. I., Simoneit, B. R. T. & Reysenbach, A.-L. (2002). *Persephonella marina* gen. nov., sp. nov. and *Persephonella guaymasensis* sp. nov., two novel, thermophilic, hydrogen-oxidizing microaerophiles from deep-sea hydrothermal vents. *Int J Syst Evol Microbiol* **52**, 1349–1359.

Henry, E. A., Devereux, R., Maki, J. S., Gilmour, C. C., Woese, C. R., Mandelco, L., Schauder, R., Remsen, C. C. & Mitchell, R. (1994). Characterization of a new thermophilic sulfate-reducing bacterium *Thermodesulfobacterium yellowstonii*, gen. nov. and sp. nov.: its phylogenetic relationship to *Thermodesulfobacterium commune* and their origins deep within the bacterial domain. *Arch Microbiol* **161**, 62–69.

Hugenholtz, P., Pitulle, C., Herschberger, K. L. & Pace, N. R. (1998). Novel division level bacterial diversity in a Yellowstone hot spring. *J Bacteriol* **180**, 366–376.

Jeanthon, C., L'Haridon, S., Cueff, V., Banta, A., Reysenbach, A.-L. & Prieur, D. (2002). *Thermodesulfobacterium hydrogeniphilum* sp. nov., a thermophilic, chemolithoautotrophic, sulfate-reducing bacterium isolated from a deep-sea hydrothermal vent at Guaymas Basin, and emendation of the genus *Thermodesulfobacterium*. *Int J Syst Evol Microbiol* **52**, 765–772.

Judicial Commission of the International Committee on Systematics of Prokaryotes (2003). Valid publication of the genus name *Thermodesulfobacterium* and the species names *Thermodesulfobacterium commune* (Zeikus *et al.* 1983) and *Thermodesulfobacterium thermophilum* (ex *Desulfovibrio thermophilus* Rozanova and Khudyakova 1974). Opinion 71. *Int J Syst Evol Microbiol* **53**, 927.

Kashefi, K., Holmes, D. E., Reysenbach, A. L. & Lovley, D. R. (2002). Use of Fe(III) as an electron acceptor to recover previously uncultured hyperthermophiles: isolation and characterization of *Geothermobacterium ferrireducens* gen. nov., sp. nov. *Appl Environ Microbiol* **68**, 1735–1742.

Langworthy, T. A., Hölzer, G., Zeikus, J. G. & Tornabene, T. G. (1983). Iso- and anteiso-branched glycerol diethers of the thermophilic anaerobe *Thermodesulfotobacterium commune*. *Syst Appl Microbiol* **4**, 3–17.

Maidak, B. L., Cole, J. R., Lilburn, T. G. & 7 other authors (2001). The RDP-II (Ribosomal Database Project). *Nucleic Acids Res* **29**, 173–174.

Mesbah, M., Premachandran, U. & Whitman, W. B. (1989). Precise measurement of the G + C content of deoxyribonucleic acid by high-performance liquid chromatography. *Int J Syst Bacteriol* **39**, 159–167.

Ollivier, B., Cord-Ruwisch, R., Hatchikian, E. C. & Garcia, J. L. (1988). Characterization of *Desulfovibrio fructosovorans* sp. nov. *Arch Microbiol* **149**, 447–450.

- Olsen, G. J., Matsuda, H., Hagstrom, R. & Overbeek, R. (1994). FASTDNAML: a tool for construction of phylogenetic trees of DNA sequences using maximum likelihood. *Comput Appl Biosci* **10**, 41–48.
- Postgate, J. (1959). A diagnostic reaction of *Desulphovibrio desulphuricans*. *Nature* **14**, 481–482.
- Rozanova, E. P. & Khudyakova, A. I. (1974). A new non-spore-forming thermophilic sulfate-reducing organism, *Desulfovibrio thermophilus* nov. sp. *Microbiology* (English translation of *Mikrobiologiya*) **43**, 908–912.
- Rozanova, E. P. & Pivovarova, T. A. (1988). Reclassification of *Desulfovibrio thermophilus* (Rozanova & Khudyakova 1974). *Microbiology* (English translation of *Mikrobiologiya*) **57**, 102–106.
- Skirnisdottir, S., Hreggvidsson, G. O., Hjorleifsdottir, S., Marteinsson, V. T., Petursdottir, S. K., Holst, O. & Kristjansson, J. K. (2000). Influence of sulfide and temperature on species composition and community structure of hot spring microbial mats. *Appl Environ Microbiol* **66**, 2835–2841.
- Sonne-Hansen, J. & Ahring, B. K. (1999). *Thermodesulfobacterium hveragerdense* sp. nov., and *Thermodesulfovibrio islandicus* sp. nov., two thermophilic sulfate reducing bacteria isolated from a Icelandic hot spring. *Syst Appl Microbiol* **22**, 559–564.
- Stöhr, R., Waberski, A., Völker, H., Tindall, B. J. & Thomm, M. (2001). *Hydrogenothermus marinus* gen. nov., sp. nov., a novel thermophilic hydrogen-oxidizing bacterium, recognition of *Calderobacterium hydrogenophilum* as a member of the genus *Hydrogenobacter* and proposal of the reclassification of *Hydrogenobacter acidophilus* as *Hydrogenobacterium acidophilum* gen. nov., comb. nov., in the phylum 'Hydrogenobacter/Aquifex'. *Int J Syst Evol Microbiol* **51**, 1853–1862.
- Tamaoka, J. & Komagata, K. (1984). Determination of DNA base composition by reverse-phase high-performance liquid chromatography. *FEMS Microbiol Lett* **25**, 125–128.
- Tindall, B. J. (1990a). A comparative study of the lipid composition of *Halobacterium saccharovorum* from various sources. *Syst Appl Microbiol* **13**, 128–130.
- Tindall, B. J. (1990b). Lipid composition of *Halobacterium lacusprofundi*. *FEMS Microbiol Lett* **66**, 199–202.
- Van Dover, C. L., Humphris, S. E., Fornari, D. & 24 other authors (2001). Biogeography and ecological setting of Indian Ocean hydrothermal vents. *Science* **294**, 818–823.
- Widdel, F. & Bak, F. (1992). Gram-negative mesophilic sulfate-reducing bacteria. In *The Prokaryotes*, 2nd edn, pp. 3352–3378. Edited by A. Balows, H. G. Trüper, M. Dworkin, W. Harder & K. H. Schleifer. New York: Springer.
- Yoshino, J., Sugiyama, Y., Sakuda, S., Kodama, T., Nagasawa, H., Ishii, M. & Igarashi, Y. (2001). Chemical structure of a novel aminophospholipid from *Hydrogenobacter thermophilus* strain TK-6. *J Bacteriol* **183**, 6302–6304.
- Zeikus, J. G., Dawson, M. A., Thompson, T. E., Ingvorsen, K. & Hatchikian, E. C. (1983). Microbial ecology of volcanic sulphidogenesis: isolation and characterization of *Thermodesulfobacterium commune* gen. nov. and sp. nov. *J Gen Microbiol* **129**, 1159–1169.

Bibliographie

- Alain K.** (2003) Approches culturales et moléculaires des assemblages microbiens associés aux polychètes hydrothermaux de la famille *Alvinellidae*. *Thèse de l'Université de Bretagne Occidentale*
- Alain K., Olagnon M., Desbruyères D., Pagé A., Barbier G., Juniper S.K. et al.** (2002a) Phylogenetic characterization of the bacterial assemblage associated with mucous secretions of the hydrothermal vent polychaete *Paralvinella palmiformis*. *FEMS Microbiol Ecol* **42** : 463–476
- Alain K., Querellou J., Lesongeur F., Pignet P., Crassous P., Raguenes G. et al.** (2002b) *Caminibacter hydrogeniphilus* gen. nov., sp. nov., a novel thermophilic, hydrogen-oxidizing bacterium isolated from an East Pacific Rise hydrothermal vent. *Int J Syst Evol Microbiol* **52** : 1317–1323
- Alain K., Zbinden M., Le Bris N., Lesongeur F., Querellou J., Gaill F. et al.** (2004) Early steps in microbial colonization processes at deep-sea hydrothermal vents. *Environ Microbiol* **6** : 227–241
- Alayse-Danet A.M., Desbruyères D. & Gaill F.** (1987) The possible nutritional or detoxification role of the epibiotic bacteria of Alvinellid polychaetes : review of the current data. *Symbiosis* **4** : 51–62
- Allen E.E. & Banfield J.F.** (2005) Community genomics in microbial ecology and evolution. *Nat Rev Microbiol* **3** : 489–498
- Altschul S.F., Gish W., Miller W., Myers E.W. & Lipman D.J.** (1990) Basic local alignment search tool. *J Mol Biol* **215** : 403–410
- Altschul S.F. & Lipman D.J.** (1990) Protein database searches for multiple alignments. *Proc Natl Acad Sci USA* **87** : 5509–5513
- Amann R.I., Ludwig W. & Schleifer K.H.** (1995) Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. *Microbiol Rev* **59** : 143–169
- Arp A.J. & Childress J.J.** (1983) Sulfide binding by the blood of the hydrothermal vent tubeworm *Riftia pachytila*. *Science* **219** : 295–297
- Arrigo K.R.** (2005) Marine microorganisms and global nutrient cycles. *Nature* **437** : 349–55
- Baker B.J. & Banfield J.F.** (2003) Microbial communities in acid mine drainage. *FEMS Microbiol Ecol* **44** : 139–152
- Béjà O.** (2004) To BAC or not to BAC : marine ecogenomics. *Curr Opin Biotechnol* **15** : 187–190
- Béjà O., Aravind L., Koonin E.V., Suzuki M.T., Hadd A. & Nguyen L.P.** (2000a) Bacterial rhodopsin : evidence for a new type of phototrophy in the sea. *Science* **289** : 1902–1906
- Béjà O., Koonin E.V., Aravind L., Taylor L.T., Seitz H., Stein J.L. et al.** (2002a) Comparative genomic analysis of archaeal genotypic variants in a single population and in two different oceanic provinces. *Appl Environ Microbiol* **68** : 335–345
- Béjà O., Marcelino T.S., Eugene V.K., Aravind L., Hadd A., Nguyen L.P. et al.** (2000b) Construction and analysis of bacterial artificial chromosome libraries from a marine microbial assemblage. *Environ Microbiol* **2** : 516–529
- Béjà O., Spudich E.N., Spudich J.L., Leclerc M. & DeLong E.F.** (2001) Proteorhodopsin phototrophy in the ocean. *Nature* **411** : 786–789
- Béjà O., Suzuki M.T., Heidelberg J.F., Nelson W.C., Preston C.M., Hamada T. et al.** (2002b) Unsuspected diversity among marine aerobic anoxygenic phototrophs. *Nature* **415** : 630–633

- Bell P.J., Sunna A., Gibbs M.D., Curach N.C., Nevainainen H. & Bergquist P.L.** (2002) Prospecting for novel lipase genes using PCR. *Microbiology* **148** : 2283–2291
- Bell S.D. & Jackson S.P.** (1998) Transcription and translation in *Archaea* : a mosaic of eukaryal and bacterial features. *Trends Microbiol* **6** : 222–8
- Bintrim S.B., Donohue T.J., Handelsman J., Roberts G.P. & Goodman R.M.** (1997) Molecular phylogeny of *Archaea* from soil. *Proc Natl Acad Sci USA* **94** : 277–82
- Bock E. & Wagner M.** (2001) Oxidation of inorganic nitrogen compounds as an energy source. In *Prokaryotes - an evolving electronic resource for the microbiological community*, Dworkin M., Falkow S., Rosenberg E., Schleifer K.-H. & Stackebrandt E. (eds). (<http://link.springerlink-ny.com/link/service/books/10125>) release3.7, 3rd ed. Springer-Verlag, New York
- Boetius A., Ravensschlag K., Schubert C.J., Rickert D., Widdel F., Gieseke A. et al.** (2000) A marine microbial consortium apparently mediating anaerobic oxidation of methane. *Nature* **407** : 623–626
- Boetius A. & Suess E.** (2004) Hydrate ridge : a natural laboratory for the study of microbial life fuelled by methane from near-surface gas hydrates. *Source Chem Geol* **205** : 291–310
- Bond P.L., Smriga S.P. & Banfield J.F.** (2000) Phylogeny of microorganisms populating a thick, subaerial, predominantly lithotrophic biofilm at an extreme acid mine drainage site. *Appl Environ Microbiol* **66** : 3842–3849
- Brady S.F., Chao C.J. & Clardy J.** (2004) Long-chain n-acyltyrosine synthases from environmental DNA. *Appl Environ Microbiol* **70** : 6865–6870
- Braithwaite D.K. & Ito J.** (1993) Compilation, alignment, and phylogenetic relationships of DNA polymerases. *Nucleic Acids Res* **21** : 787–802
- Breitbart M., Hewson I., Felts B., Mahaffy J.M., Nulton J., Salamon P. et al.** (2003) Metagenomic analyses of an uncultured viral community from human feces. *J Bacteriol* **185** : 6220–6223
- Breitbart M., Salamon P., Andresen B., Mahaffy J.M., Segall A.M., Mead D. et al.** (2002) Genomic analysis of uncultured marine viral communities. *Proc Natl Acad Sci USA* **99** : 14250–14255
- Brinkmann A.** (2002) Regulation of transcription in hyperthermophilic *Archaea*. Ph.D. thesis, Wageningen University, the Netherlands
- Brofft J.E., McArthur J.V. & Shimkets L.J.** (2002) Recovery of novel bacterial diversity from a forested wetland impacted by reject coal. *Environ Microbiol* **4** : 764–9
- Bruns A., Nubel U., Cypionka H. & Overmann J.** (2003) Effect of signal compounds and incubation conditions on the culturability of freshwater bacterioplankton. *Appl Environ Microbiol* **69** : 1980–1989
- Buckley D.H., Graber J.R. & Schmidt T.M.** (1998) Phylogenetic analysis of nonthermophilic members of the kingdom *Crenarchaeota* and their diversity and abundance in soils. *Appl Environ Microbiol* **64** : 4333–4339
- Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D., Sutton G.G. et al.** (1996) Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science* **273** : 1058–1073
- Burggraf S., Huber H. & Stetter K.O.** (1997) Reclassification of the crenarchael orders and families in accordance with 16S rRNA sequence data. *Int J Syst Bacteriol* **47** : 657–60
- Button D.K., Schut F., Quang P., Martin R. & Robertson B.R.** (1993) Viability and isolation of marine bacteria by dilution culture : theory, procedures, and initial results. *Appl Environ Microbiol* **59** : 881–891
- Byrd J.J., Xu H.S. & Colwell R.R.** (1991) Viable but non culturable bacteria in drinking water. *Appl Environ Microbiol* **57** : 875–878
- Campbell B.J. & Cary S.C.** (2001) Characterization of a novel spirochete associated with the hydrothermal vent polychaete annelid, *Alvinella pompejana*. *Appl Environ Microbiol* **67** : 110–117
- Campbell B.J. & Cary S.C.** (2004) Abundance of reverse tricarboxylic acid cycle genes in free-living microorganisms at deep-sea hydrothermal vents. *Appl Environ Microbiol* **70** : 6282–6289
- Campbell B.J., Engel A.S., Porter M.L. & Takai K.** (2006) The versatile ϵ -Proteobacteria : key players in sulphidic habitats. *Nat Rev Microbiol* **4** : 458–68
- Campbell B.J., Jeanthon C., Kostka J.E., Luther G. W. r. & Cary S.C.** (2001) Growth and phylogenetic properties of novel bacteria belonging to the epsilon subdivision of the *Proteobacteria* enriched from *Alvinella pompejana* and deep-sea hydrothermal vents. *Appl Environ Microbiol* **67** : 4566–4572
- Campbell B.J., Stein J.L. & Cary S.C.** (2003) Evidence of chemolithoautotrophy in the bacterial community associated with *Alvinella pompejana*, a hydrothermal vent polychaete. *Appl Environ Microbiol* **69** : 5070–5078

- Cane D.E., Walsh C.T. & Khosla C.** (1998) Biochemistry - harnessing the biosynthetic code : combinations, permutations, and mutations. *Science* **282** : 63-68
- Cary S.C., Cottrell M.T., Stein J.L., Camacho F. & Desbruyeres D.** (1997) Molecular identification and localization of filamentous symbiotic bacteria associated with the hydrothermal vent annelid *Alvinella pompejana*. *Appl Environ Microbiol* **63** : 1124-1130
- Cary S.C., Shank T. & Stein J.** (1998) Worms bask in extreme temperatures. *Nature* **391** : 545-546
- Cary S.C., Warren W., Anderson E. & Giovannoni S.J.** (1993) Identification and localization of bacterial endosymbionts in hydrothermal vent taxa with species-specific PCR amplification and *in situ* hybridization techniques. *Mar Mol Biol Biotechnol* **2** : 51-62
- Cavanaugh C.** (1983) Symbiotic chemoautotrophic bacteria in marine invertebrates from sulphide-rich habitats. *Nature* **302** : 58-61
- Cavanaugh C.M., Jones M.L., Jannasch H.W. & Waterbury J.B.** (1981) Prokaryotic cells in the hydrothermal vent tube worm *Riftia pachyptila* : possible chemoautotrophic symbionts. *Science* **213** : 340-342
- Charlou J.L., Donval J.P., Fouquet Y., Jean-Baptiste P. & Holm N.** (2002) Geochemistry of high H₂ and CH₄ vent fluids issuing from ultramafic rocks at the Rainbow hydrothermal field (36°14'N, MAR). *Chem Geol* **191** : 345-359
- Chevaldonné P., Desbruyères D. & Childress J.J.** (1992) Some like it hot... and some even hotter. *Nature* **359** : 593-594
- Collins J. & Hohn B.** (1978) Cosmids : a type of plasmid gene-cloning vector that is packageable in vitro in bacteriophage lambda heads. *Proc Natl Acad Sci USA* **75** : 4242-4246
- Colwell R.R.** (2000) Viable but non culturable bacteria : a survival strategy. *J Infect Chemother* **6** : 121-125
- Conte M.H.** (1998) The oceanic flux program : twenty years of particle flux measurements in the deep Sargasso Sea. *Oceanus* **40** : 15-19
- Corliss J.B. & Ballard R.D.** (1977) Oasis of life in the cold abyss. *Natl Geogr* **152** : 441-453
- Corliss J.B., Dymond J., Gordon L.I., Edmond J.M., Von Herzen R.P., Ballard R.D. et al.** (1979) Submarine thermal springs on the Galapagos rift. *Science* **203** : 1073-1083
- Corre E.** (2000) Approches moléculaires de la diversité microbienne de deux environnements extrêmes : les sources hydrothermales profondes et les réservoirs pétroliers. *Thèse de l'Université de Bretagne Occidentale*
- Corre E., Reysenbach A.L. & Prieur D.** (2001) ϵ -Proteobacterial diversity from a deep-sea hydrothermal vent on the Mid-Atlantic Ridge. *FEMS Microbiol Lett* **205** : 329-335
- Cottrell M.T., Moore J.A. & Kirchman D.L.** (1999) Chitinases from uncultured marine microorganisms. *Appl Environ Microbiol* **65** : 2553-2557
- Courtois S., Cappellano C.M., Ball M., Francou F.X., Normand P., Helynck G. et al.** (2003) Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products. *Appl Environ Microbiol* **69** : 49-55
- Courtois S., Frostegard A., Goransson P., Depret G., Jeannin P. & Simonet P.** (2001) Quantification of bacterial subgroups in soil : comparison of DNA extracted directly from soil or from cells previously released by density gradient centrifugation. *Environ Microbiol* **3** : 431-439
- Curtis T.P., Sloan W.T. & Scannell J.W.** (2002) Estimating prokaryotic diversity and its limits. *Proc Natl Acad Sci USA* **99** : 10494-10499
- Dalgaard J. & Garrett R.** (1993) Archaeal hyperthermophilic genes. In *The biochemistry of Archaea (Archaeobacteria)*, Kates M., Kushner D.J. & Matheson A.T. (eds). Amsterdam, The Netherlands : Elsevier vol.26 pp. 535-563. Elsevier, Amsterdam, the Netherlands
- Damm K.L.V., Oosting S.E., Kozłowski R., Buttermore L.G., Colodner D.C., Edmonds H.N. et al.** (1995) Evolution of East Pacific Rise hydrothermal vent fluids following a volcanic eruption. *Nature* **375** : 47-50
- Daniel R.** (2004) The soil metagenome - a rich resource for the discovery of novel natural products. *Curr Opin Biotechnol* **15** : 199-204
- Daniel R.** (2005) The metagenomics of soil. *Nat Rev Microbiol* **3** : 470-478
- Dauga C., Dore J. & Sghir A.** (2005) Expanding the known diversity and environmental distribution of cultured and uncultured *Bacteria*. *Med Sci (Paris)* **21** : 290-296
- Davey M.E. & O'Toole G.A.** (2000) Microbial biofilms : from ecology to molecular genetics. *Microbiol Mol Biol Rev* **64** : 847-867
- Dayhoff M.O.** (1965) Computer aids to protein sequence determination. *J Theor Biol* **8** : 97-112

- de la Torre J.R., Christianson L.M., Bèjà O., Suzuki M.T., Karl D.M., Heidelberg J. et al.** (2003) Proteorhodopsin genes are distributed among divergent marine bacterial taxa. *Proc Natl Acad Sci USA* **100** : 12830–12835
- DeLong E.F.** (1992) Archaea in coastal marine environments. *Proc Natl Acad Sci USA* **89** : 5685–5689
- DeLong E.F.** (2005) Microbial community genomics in the ocean. *Nat Rev Microbiol* **6** : 159–169
- DeLong E.F., Wickham G.S. & Pace N.R.** (1989) Phylogenetic stains : ribosomal RNA-based probes for the identification of single cells. *Science* **243** : 1360–1363
- Desbruyères D., Biscoito M., Caprais J.C., Colaco A., Comtet T., Crassous P. et al.** (2001) Variations in deep-sea hydrothermal vent communities on the Mid-Atlantic Ridge near the Azores plateau. *Deep-Sea Res* **48** : 1325–1346
- Desbruyères D., Chevalloné P., Alayse A.M., Jollivet D., Lallier F.H., C J.T. et al.** (1998) Biology and ecology of the "pompeii worm" (*Alvinella pompejana* Desbruyères and Laubier), a normal dweller of an extreme deep-sea environment : a synthesis of current knowledge and recent developments. *Deep-Sea Res II* **45** : 383–422
- Desbruyères D., Gaill F., Laubier L. & Fouquet Y.** (1985) Polychaetous annelids from hydrothermal vent ecosystems : an ecological overview. *Biol Soc Wash Bull* **6** : 103–106
- Desbruyères D., Gaill F., Laubrier L., Prieur D. & Rau G.** (1983) Unusual nutrition of the "pompeii worm" *Alvinella pompejana* (polychaetous annelid) from a hydrothermal vent environment : SEM, TEM, super(13)C, and super(15)N evidence. *Mar Biol* **75** : 201–205
- Di Meo-Savoie C.A., Luther G. W. r. & Cary S.C.** (2004) Physicochemical characterization of the microhabitat of *Alvinella pompejana*, a hydrothermal vent annelid. *Geochim Cosmochim Acta* **68** : 2055–2066
- Diez B., Pedros-Alio C. & Massana R.** (2001) Study of genetic diversity of eukaryotic picoplankton in different oceanic regions by small-subunit rRNA gene cloning and sequencing. *Appl Environ Microbiol* **67** : 2932–2941
- Ding M. & Yelton D.B.** (1993) Cloning and analysis of the *leuB* gene of *Leptospira interrogans* serovar pomona. *J Gen Microbiol* **139** : 1093–1103
- Distel D.L., Lane D.J., Olsen G.J., Giovannoni S.J., Pace B., Pace N.R. et al.** (1988) Sulfur-oxidizing bacterial endosymbionts : analysis of phylogeny and specificity by 16S rRNA sequences. *J Bacteriol* **170** : 2506–2510
- Dopson M., Baker-Austin C., Hind A., Bowman J.P. & Bond P.L.** (2004) Characterization of *Ferroplasma* isolates and *Ferroplasma acidarmanus* sp. nov., extreme acidophiles from acid mine drainage and industrial bioleaching environments. *Appl Environ Microbiol* **70** : 2079–2088
- Dumont M.G. & Murrell J.C.** (2005) Stable isotope probing - linking microbial identity to function. *Nat Rev Microbiol* **3** : 499–504
- Edmonds H.N., Michael P.J., Baker E.T., Connelly D.P., Snow J.E., Langmuir C.H. et al.** (2003) Discovery of abundant hydrothermal venting on the ultraslow-spreading Gakkel Ridge in the Arctic Ocean. *Nature* **421** : 252–256
- Eilers H., Pernthaler J., Gloeckner F.O. & Amann R.** (2000) Culturability and *in situ* abundance of pelagic bacteria from the North sea. *Appl Environ Microbiol* **66** : 3044–3051
- Engel A.S., Lee N., Porter M.L., Stern L.A., Bennett P.C. & Wagner M.** (2003) Filamentous "epsilonproteobacteria" dominate microbial mats from sulfidic cave springs. *Appl Environ Microbiol* **69** : 5503–5511
- Entcheva P., Liebl W., Johann A., Hartsch T. & Streit W.** (2001) Direct cloning from enrichment cultures, a reliable strategy for isolation of complete operons and genes from microbial consortia. *Appl Environ Microbiol* **77** : 89–99
- Erkel C., Kemnitz D., Kube M., Ricke P., Chin K.J., Dedysh S. et al.** (2005) Retrieval of first genome data for rice cluster I methanogens by a combination of cultivation and molecular techniques. *FEMS Microbiol Ecol* **53** : 187–204
- Ermler U., Grabarse W., Shima S., Goubeaud M. & Thauer R.K.** (1997) Crystal structure of methyl-coenzyme M reductase : the key enzyme of biological methane formation. *Science* **278** : 1457–1462
- Evans M.C., Buchanan C.W. & Arnon B.B.** (1966) A new ferredoxin-dependant carbon reduction cycle in a photosynthetic bacterium. *Proc Natl Acad Sci USA* **55** : 928–934
- Felbeck H., Childress J.J. & Somero G.N.** (1981) Calvin-Benson cycle and sulfide oxydation enzymes in animals from sulfide rich habitats. *Nature* **293** : 291–293
- Felbeck H. & Jarchow J.** (1998) Carbon release from purified chemoautotrophic bacterial symbionts of the hydrothermal vent tubeworm *Riftia pachyptila*. *Physiol Zool* **71** : 294–302

- Ferreira R.G., Soncini F.C. & Viale A.M.** (1993) Cloning, characterization, and functional expression in *Escherichia coli* of chaperonin (*groESL*) genes from the phototrophic sulfur bacterium *Chromatium vinosum*. *J Bacteriol* **175** : 1514–1523
- Finster K., Coates J.D., Liesack W. & Pfennig N.** (1997) *Desulfuromonas thiophila* sp. nov., a new obligately sulfur-reducing bacterium from anoxic freshwater sediment. *Int J Syst Bacteriol* **47** : 754–8
- Fisher C.R., Childress J.J. & Sanders N.K.** (1988) The role of vestimentiferan hemoglobin in providing an environment suitable for chemoautotrophic sulfide-oxidizing endosymbionts. *Symbiosis* **5** : 229–246
- Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F., Kerlavage A.R. et al.** (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269** : 496–512
- Flores J.F., Fisher C.R., Carney S.L., Green B.N., Freytag J.K., Schaeffer S.W. et al.** (2005) Sulfide binding is mediated by zinc ions discovered in the crystal structure of a hydrothermal vent tubeworm hemoglobin. *Proc Natl Acad Sci USA* **102** : 2713–2718
- Fouquet Y., Auclair G., Cambon P. & Etoubleau J.** (1988) Geological setting and mineralogical and geochemical investigations on sulfide deposits near 13°N on the East Pacific Rise. *Marine Geology* **84** : 145–178
- Fuchs G., Stupperich E. & Eden G.** (1980) Autotrophic CO₂ fixation in *Chlorobium limicola* - evidence for the operation of a reductive tricarboxylic acid cycle in growing cells. *Arch Microbiol* **128** : 64–71
- Fuhrman J.A., McCallum K. & Davis A.A.** (1992) Novel major archaeobacterial group from marine plankton. *Nature* **356** : 148–149
- Fujiwara S.** (2002) Extremophiles : Developments of their special functions and potential resources. *J Biosci Bioeng* **94** : 518–525
- Gabor E.M.** (2004) Harvesting novel biocatalyst from the metagenome. Ph.D. thesis, Unniv. Grohingen
- Gabor E.M., de Vries E.J. & Janssen D.B.** (2004) Construction, characterization, and use of small-insert gene banks of DNA isolated from soil and enrichment cultures for the recovery of novel amidases. *Environ Microbiol* **6** : 948–958
- Gaill F. & Hunt S.** (1991) The biology of Annelid worms from high temperature hydrothermal vent regions. *Rev Aquat Sci* **4** : 107–137
- Galperin M.Y.** (2004) Metagenomics : from acid mine to shining sea. *Environ Microbiol* **6** : 543–545
- Galperin M.Y.** (2005) The molecular biology database collection : 2005 update. *Nucleic Acids Res* **33** : D5–D24
- Galtier N. & Lobry J.R.** (1997) Relationships between genomic G+C content, RNA secondary structures, and optimal growth temperature in prokaryotes. *J Mol Evol* **44** : 632–6
- Galtier N., Tourasse N. & Gouy M.** (1999) A nonhyperthermophilic common ancestor to extant life forms. *Science* **283** : 220–221
- Garrity G.M. & Holt J.G.** (2001) The road map to the manual. In *Bergey's manual of systematic bacteriology*, Boone D.R., Castenholz R.W., & Garrity G.M. (eds). New-Ork : Springer-Verlag, 2nd edn, vol. 1, pp. 119–166
- Gebruck A.V., Southward E.C., Kennedy H. & Southward A.J.** (2000) Food source behaviour and distribution of hydrothermal vent shrimps at the Mid-Atlantic Ridge. *Journal of Marine Biological Association UK* **80** : 485–499
- Gillespie D.E., Brady S.F., Bettermann A.D., Cianciotto N.P., Liles M.R., Rondon M.R. et al.** (2002) Isolation of antibiotics turbomycin A and B from a metagenomic library of soil microbial DNA. *Appl Environ Microbiol* **68** : 4301–4306
- Giovannoni S.J., Britschgi T.B., Moyer C.L. & Field K.G.** (1990) Genetic diversity in sargasso sea bacterioplankton. *Nature* **345** : 60–63
- Giovannoni S.J., DeLong E.F., Olsen G.J. & Pace N.R.** (1988) Phylogenetic group-specific oligodeoxynucleotide probes for identification of single microbial cells. *J Bacteriol* **170** : 720–726
- Goffeau A., Barrell B.G., Bussey H., Davis R.W., Dion B., Feldmann H. et al.** (1996) Life with 6000 genes. *Science* **274** : 546, 563–7
- Gupta R., Beg K.Q. & Lorenz P.** (2002) Bacterial alkaline, proteases : molecular approaches and industrial applications. *Appl Microbiol Biotechnol* **59** : 15–32
- Haddad A., Camacho F., Durand P. & Cary S.C.** (1995) Phylogenetic characterization of the epibiotic bacteria associated with the hydrothermal vent polychaete *Alvinella pompejana*. *Appl Environ Microbiol* **61** : 1679–1687
- Hallam S.J., Girguis P.R., Preston C.M., Richardson P.M. & DeLong E.F.** (2003) Identification of methyl coenzyme M reductase A (*mcrA*) genes associated with methane-oxidizing *Archaea*. *Appl Environ Microbiol* **69** : 5483–5491

- Hallam S.J., Putnam N., Preston C.M., Detter J.C., Rokhsar D., Richardson P.M. *et al.* (2004) Reverse methanogenesis : testing the hypothesis with environmental genomics. *Science* **305** : 1457–1462
- Handelsman J. (2004) Metagenomics : Application of genomics to uncultured microorganisms. *Microbiol Mol Biol Rev* **68** : 669–685
- Handelsman J. (2005) Sorting out metagenomes. *Nat Biotechnol* **23** : 38–9
- Handelsman J., Liles M., Mann D., Riesenfeld C. & Goodman R.M. (2002) Cloning the metagenome : culture-independent access to the diversity and functions of the uncultivated microbial world. *Methods in Microbiology - Functional Microbial Genomics Academic Press* **33** : 241–255
- Handelsman J., Rondon M.R., Brady S., Clardy J. & Goodman R.M. (1998) Molecular biology provides access to the chemistry of unknown soil microbes : a new frontier for natural products. *Chem Biol* **5** : 245–249
- Harmsen H., Prieur D. & Jeanthon C. (1997) Distribution of microorganisms in deep-sea hydrothermal vent chimneys investigated by whole-cell hybridization and enrichment culture of thermophilic subpopulations. *Appl Environ Microbiol* **63** : 2876–2883
- Haymon R.M., Fornari D.J., Von Damm K.L., Lilley M.D., Perfit M.R., Edmond J.M. *et al.* (1993) Volcanic eruption of the mid-ocean ridge along the East Pacific Rise crest at 9 degree 45-52'N : Direct submersible observations of seafloor phenomena associated with an eruption event in April, 1991. *Earth and Planet Sci Lett* **119** : 85–101
- Henne A., Daniel R., Schmitz R.A. & Gottschalk G. (1999) Construction of environmental DNA libraries in *Escherichia coli* and screening for the presence of genes conferring utilization of 4-hydroxybutyrate. *Appl Environ Microbiol* **65** : 3901–3907
- Henne A., Schmitz R.A., Bomeke M., Gottschalk G. & Daniel R. (2000) Screening of environmental DNA libraries for the presence of genes conferring lipolytic activity on *Escherichia coli*. *Appl Environ Microbiol* **66** : 3113–3116
- Higashi Y., Sunamura M., Kitamura K., Nakamura K.i., Kurusu Y., Ishibashi J.i. *et al.* (2004) Microbial diversity in hydrothermal surface to subsurface environments of Suiyo Seamount, Izu-Bonin Arc, using a catheter-type *in situ* growth chamber. *FEMS Microbiol Ecol* **47** : 327–336
- Hinrichs K.U., Hayes J.M., Sylva S.P., Brewer P.G. & DeLong E.F. (1999) Methane-consuming archaeobacteria in marine sediments. *Nature* **398** : 802–805
- Hoehler T.M., Alperin M.J. & Albert D.B. (1994) Field and laboratory studies of methane oxydation in an anoxic marine sediment : evidence for a methanogen-sulfate-reducer consortium. *Global Biogeochem Cycles* **8** : 451–463
- Hoek J., Banta A., Hubler F. & Reysenbach A.L. (2003) Microbial diversity of a sulphide spire located in the Edmond deep-sea hydrothermal vent field on the Central Indian Ridge. *Geobiology* **1** : 119–127
- Hohn M.J., Hedlund B.P. & Huber H. (2002) Detection of 16S rDNA sequences representing the novel phylum “Nanoarchaeota” : indication for a wide distribution in high temperature biotopes. *Syst Appl Microbiol* **25** : 551–554
- Huber H. & Stetter K.O. (1998) Hyperthermophiles and their possible potential in biotechnology. *J Biotechnol* **64** : 39–52
- Huber J.A., Butterfield D.A. & Baross J.A. (2002) Temporal changes in archaeal diversity and chemistry in a mid-ocean ridge seafloor habitat. *Appl Environ Microbiol* **68** : 1585–1594
- Huber J.A., Butterfield D.A. & Baross J.A. (2003) Bacterial diversity in a seafloor habitat following a deep-sea volcanic eruption. *FEMS Microbiol Ecol* **43** : 393–409
- Huber R. & Stetter K.O. (1992) The order *Thermotogales*. In *The Prokaryotes*, Balows A., Trüper H.G., Dworkin M., Harder W. & Schleifer K.H. (eds). Heidelberg, New York (1992) pp. 3809–3815. Springer, Berlin
- Hugenholtz P. (2002) Exploring prokaryotic diversity in the genomic era. *Genome Biol* **3** : 1–8
- Hughes D.S., Felbeck H. & Stein J.L. (1997) A histidine protein kinase homolog from the endosymbiont of the hydrothermal vent tubeworm *Riftia pachyptila*. *Appl Environ Microbiol* **63** : 3494–3498
- Hughes J.B., Hellmann J.J., Ricketts T.H. & Bohannan B.J. (2001) Counting the uncountable : statistical approaches to estimating microbial diversity. *Appl Environ Microbiol* **67** : 4399–4406
- Hugler M., Huber H., Stetter K.O. & Fuchs G. (2003) Autotrophic CO₂ fixation pathway in archaea (*Crenarchaeota*). *Arch Microbiol* **179** : 160–173
- Hugler M., Wirsén C.O., Fuchs G., Taylor C.D. & Sievert S.M. (2005) Evidence for autotrophic CO₂ fixation via the reductive tricarboxylic acid cycle by members of the ϵ subdivision of *Proteobacteria*. *J Bacteriol* **187** : 3020–3027
- Hungate R.E. (1950) The anaerobic mesophilic cellulolytic bacteria. *Bacteriol Rev* **14** : 1–49

- Inagaki F., Takai K., Kobayashi H., Nealson K.H. & Horikoshi K.** (2003) *Sulfurimonas autotrophica* gen. nov., sp. nov., a novel sulfur-oxidizing ϵ -*Proteobacterium* isolated from hydrothermal sediments in the Mid-Okinawa Trough. *Int J Syst Evol Microbiol* **53** : 1801–1805
- Inagaki F., Takai K., Kobayashi H., Nealson K.H. & Horikoshi K.** (2006) Biogeographical distribution and diversity of microbes in methane hydrate-bearing deep marine sediments on the Pacific Ocean Margin. *Proc Natl Acad Sci USA* **103** : 2815–2820
- Inagaki F., Takai K., Nealson K.H. & Horikoshi K.** (2004) *Sulfurovum lithotrophicum* gen. nov., sp. nov., a novel sulfur-oxidizing chemolithoautotroph within the *Epsilon-Proteobacteria* isolated from Okinawa Trough hydrothermal sediments. *Int J Syst Evol Microbiol* **54** : 1477–1482
- Jannasch H.W.** (1985) The chemosynthetic support of life and the microbial diversity at deep-sea hydrothermal vents. *Proc R Soc Lond B* **225** : 277–297
- Jannasch H.W.** (1995) Microbial interactions with hydrothermal fluids. In *Seafloor hydrothermal systems : physical, chemical, biological, and geological interactions*, Humphris S.E., Zierenberg R.A., Mullineaux L.S. & Thomson, R.E. (eds). Washington : American Geophysical Union. pp. 273–296
- Jannasch H.W. & Jones G.E.** (1959) Bacterial populations in sea water as determined by different methods of enumeration. *Limnol Oceanogr* **4** : 128–139
- Jannasch H.W. & Wirsen C.O.** (1979) Chemosynthetic primary production at East Pacific sea floor spreading centers. *BioScience* **29** : 592–598
- Janssen P.H., Yates P.S., Grinton B.E., Taylor P.M. & Sait M.** (2002) Improved culturability of soil bacteria and isolation in pure culture of novel members of the divisions *Acidobacteria*, *Actinobacteria*, *Proteobacteria*, and *Verrucomicrobia*. *Appl Environ Microbiol* **68** : 2391–2396
- Jeanthon C.** (2000) Molecular ecology of hydrothermal vent microbial communities. *Antonie Van Leeuwenhoek* **77** : 117–133
- Jeanthon C., L'Haridon S., Cuffe V., Banta A., Reysenbach A.L. & Prieur D.** (2002) *Thermodesulfobacterium hydrogenophilum* sp. nov., a thermophilic, chemolithoautotrophic, sulfate-reducing bacterium isolated from a deep-sea hydrothermal vent at Guaymas Basin, and emendation of the genus *Thermodesulfobacterium*. *Int J Syst Evol Microbiol* **52** : 765–72
- Jeanthon C., L'Haridon S., Reysenbach A.L., Corre E., Vernet M., Messner P. et al.** (1999) *Methanococcus vulcanius* sp. nov., a novel hyperthermophilic methanogen isolated from East Pacific Rise, and identification of *Methanococcus* sp. DSM 4213T as *Methanococcus fervens* sp. nov. *Int J Syst Bacteriol* **49** : 583–9
- Jeanthon C., L'Haridon S., Reysenbach A.L., Vernet M., Messner P., Sleytr U.B. et al.** (1998) *Methanococcus infernus* sp. nov., a novel hyperthermophilic lithotrophic methanogen isolated from a deep-sea hydrothermal vent. *Int J Syst Bacteriol* **48 Pt 3** : 913–9
- Judd A.G., Hovland M., Dimitrov L.I., Gil S.G. & V J.** (2002) The geological methane budget at continental margins and its influence on climate change. *Geofluids* **2** : 109–126
- Jurgens G., Glockner F., Amann R., Saano A., Montonen L., Likolammi M. et al.** (2000) Identification of novel *Archaea* in bacterioplankton of a boreal forest lake by phylogenetic analysis and fluorescent *in situ* hybridization. *FEMS Microbiol Ecol* **34** : 45–56
- Jurgens G., Lindstrom K. & Saano A.** (1997) Novel group within the kingdom *Crenarchaeota* from boreal forest soil. *Appl Environ Microbiol* **63** : 803–805
- Jurgens K., Pernthaler J., Schalla S. & Amann R.** (1999) Morphological and compositional changes in a planktonic bacterial community in response to enhanced protozoan grazing. *Appl Environ Microbiol* **65** : 1241–1250
- Kaeberlein T., Lewis K. & Epstein S.S.** (2002) Isolating "uncultivable" microorganisms in pure culture in a simulated natural environment. *Science* **296** : 1127–1129
- Karl D.M.** (1995) Ecology of free-living, hydrothermal vent microbial communities. In *The microbiology of deep-sea hydrothermal vents*. Karl D.M. (eds). Boca Raton : CRC Press Inc., 35–124
- Karl D.M., Taylor G.T., Novitsky J.A., Jannasch H.W., Wirsen C.O., Pace N.R. et al.** (1988) A microbiological study of Guaymas Basin high temperature hydrothermal vents. *Deep -Sea Res* **35** : 777–791
- Karner M.B., DeLong E.F. & Karl D.M.** (2001) Archaeal dominance in the mesopelagic zone of the Pacific Ocean. *Nature* **409** : 507–10
- Keller M. & Zengler K.** (2004) Tapping into microbial diversity. *Nat Rev Microbiol* **2** : 141–150
- Kelley D.S., Baross J.A. & Delaney J.R.** (2002) Volcanoes, fluids, and life at Mid-Ocean ridge spreading centers. *Annual Review of Earth and Planetary Sciences* **30** : 385–491

- Kelley D.S., Karson J.A., Blackman D.K., Fruh-Green G.L., Butterfield D.A., Lilley M.D. et al.** (2001) An off-axis hydrothermal vent field near the Mid-Atlantic Ridge at 30 degrees N. *Nature* **412** : 145–149
- Kelley D.S., Karson J.A., Fruh-Green G.L., Yoerger D.R., Shank T.M., Butterfield D.A. et al.** (2005) A serpentinite-hosted ecosystem : the Lost City hydrothermal field. *Science* **307** : 1428–1434
- Keough B.P., Schmidt T.M. & Hicks R.E.** (2003) Archaeal nucleic acids in picoplankton from great lakes on three continents. *Microb Ecol* **46** : 238–248
- Kim U.J., Shizuya H., de Jong P.J., Birren B. & Simon M.I.** (1992) Stable propagation of cosmid sized human DNA inserts in an F factor based vector. *Nucleic Acids Res* **20** : 1083–1085
- Kluyver A.J. & Van Niel C.B.** (1936) Prospects for a natural system of classification of *Bacteria*. zbl. bakt., ii abt. 94 :369-403. *Int J Syst Bacteriol* **30** : 225–420
- Knietsch A., Bowien S., Whited G., Gottschalk G. & Daniel R.** (2003a) Identification and characterization of coenzyme B12-dependent glycerol dehydratase- and diol dehydratase-encoding genes from metagenomic DNA libraries derived from enrichment cultures. *Appl Environ Microbiol* **69** : 3048–3060
- Knietsch A., Waschowitz T., Bowien S., Henne A. & Daniel R.** (2003b) Construction and screening of metagenomic libraries derived from enrichment cultures : generation of a gene bank for genes conferring alcohol oxidoreductase activity on *Escherichia coli*. *Appl Environ Microbiol* **69** : 1408–1416
- Knietsch A., Waschowitz T., Bowien S., Henne A. & Daniel R.** (2003c) Metagenomes of complex microbial consortia derived from different soils as sources for novel genes conferring formation of carbonyls from short chain polyols on *Escherichia coli*. *J Mol Microbiol* **5** : 46–56
- Knittel K., Losekann T., Boetius A., Kort R. & Amann R.** (2005) Diversity and distribution of methanotrophic archaea at cold seeps. *Appl Environ Microbiol* **71** : 467–479
- Könneke M., Bernhard A.E., de la Torre J.R., Walker C.B., Waterbury J.B. & Stahl D.A.** (2005) Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature* **437** : 543–546
- Kormas K.A., Tivey M.K., Von Damm K. & Teske A.** (2006) Bacterial and archaeal phylotypes associated with distinct mineralogical layers of a white smoker spire from a deep-sea hydrothermal vent site (9°N, East Pacific Rise). *Environ Microbiol* **8** : 909–920
- Krüger M., Meyerdierks A., Glockner F.O., Amann R., Widdel F., Kube M. et al.** (2003) A conspicuous nickel protein in microbial mats that oxidize methane anaerobically. *Nature* **426** : 878–881
- Kuypers M.M., Sliemers A.O., Lavik G., Schmid M., Jorgensen B.B., Kuenen J.G. et al.** (2003) Anaerobic ammonium oxidation by anammox bacteria in the Black Sea. *Nature* **422** : 608–11
- Kvenvolden K.A.** (1988) Methane hydrates and global climate. *Global Biogeochem Cycles* **2** : 221–229
- Kvenvolden K.A.** (1999) Potential effects of gas hydrate on human welfare. *Proc Natl Acad Sci USA* **96** : 3420–3426
- La Paglia C. & Hartzell P.L.** (1997) Stress-induced production of biofilm in the hyperthermophilic *Archaeoglobus fulgidus*. *Appl Environ Microbiol* **63** : 3158–3163
- Laue B.E. & Nelson D.C.** (1994) Characterization of the gene encoding the autotrophic ATP sulfurylase from the bacterial endosymbiont of the hydrothermal vent tubeworm *Riftia pachyptila*. *J Bacteriol* **176** : 3723–3929
- Le Bris N., Sarradin P.M. & Caprais J.C.** (2003) Contrasted sulphide chemistries in the environment of 13°N EPR vent fauna. *Deep Sea Research Part I : Oceanographic Research Papers* **50** : 737–747
- Le Bris N., Zbinden M. & Gaill F.** (2005) Processes controlling the physico-chemical micro-environments associated with Pompeii worms. *Deep Sea Res I* **52** : 1071–1083
- L'Haridon S., Cilia V., Messner P., Raguene G., Gambacorta A., Sleytr U.B. et al.** (1998) *Desulfurobacterium thermolithotrophum* gen. nov., sp. nov., a novel autotrophic, sulphur-reducing bacterium isolated from a deep-sea hydrothermal vent. *Int J Syst Bacteriol* **48** : 701–11
- Liles M.R., Manske B.F., Bintrim S.B., Handelsman J. & Goodman R.M.** (2003) A census of rRNA genes and linked genomic sequences within a soil metagenomic library. *Appl Environ Microbiol* **69** : 2684–2691
- Lonsdale P.** (1977) Clustering of suspension-feeding macrobenthos near abyssal hydrothermal vents at oceanic spreading centers. *Deep-Sea Res* **24** : 857–863
- Longnecker K.** (2001) *MSc Oregon State University, 2001, Oceanography*
- Longnecker K. & Reysenbachach A.L.** (2001) Expansion of the geographic distribution of a novel lineage of ϵ -*Proteobacteria* to a hydrothermal vent site on the Southern East Pacific Rise. *FEMS Microbiol Ecol* **35** : 287–293

- Lopez-Garcia P., Brochier C., Moreira D. & Rodriguez-Valera F.** (2004) Comparative analysis of a genome fragment of an uncultivated mesopelagic crenarchaeote reveals multiple horizontal gene transfers. *Environ Microbiol* **6** : 19–34
- Lopez-Garcia P., Duperron S., Philippot P., Foriel J., Susini J. & Moreira D.** (2003a) Bacterial diversity in hydrothermal sediment and epsilonproteobacterial dominance in experimental microcolonizers at the Mid-Atlantic Ridge. *Environ Microbiol* **5** : 961–976
- Lopez-Garcia P., Gaill F. & Moreira D.** (2002) Wide bacterial diversity associated with tubes of the vent worm *Riftia pachyptila*. *Environ Microbiol* **4** : 204–215
- Lopez-Garcia P., Philippe H., Gail F. & Moreira D.** (2003b) Autochthonous eukaryotic diversity in hydrothermal sediment and experimental microcolonizers at the Mid-Atlantic Ridge. *Proc Natl Acad Sci USA* **100** : 697–702
- Lorenz P. & Eck J.** (2005) Metagenomics and industrial applications. *Nat Rev Microbiol* **3** : 510–516
- Lorenz P., Liebeton K., Niehaus F. & Eck J.** (2002) Screening for novel enzymes for biocatalytic processes : accessing the metagenome as a resource of novel functional sequence space. *Curr Opin Biotechnol* **13** : 572–577
- Lorenz P. & Schleper C.** (2002) Metagenome - a challenging source of enzyme discovery. *J Mol Catal B-Enzym* **19-20** : 13–19
- Lowe T.M. & Eddy S.R.** (1997) tRNAscan-SE : a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* **25** : 955–964
- Ludwig W. & Klenk H.P.** (2001) Overview : A phylogenetic backbone and taxonomic framework for procaryotic systematics. In *Bergey's Manual of Systematic Bacteriology*, Boone D.R., Castenholz R.W. & Garrity G.M. (eds). Berlin : Springer-Verlag. 2nd edn, vol.1, pp. 49-65
- MacGregor B.J., Moser D.P., Alm E.W., Neelson K.H. & Stahl D.A.** (1997) *Crenarchaeota* in Lake Michigan sediment. *Appl Environ Microbiol* **63** : 1178–1181
- MacNeil I.A., Tiong C.L., Minor C., August P.R., Grossman T.H., Loiacono K.A. et al.** (2001) Expression and isolation of antimicrobial small molecules from soil DNA libraries. *J Mol Microbiol Biotechnol* **3** : 301–308
- Madigan M.T.** (1996) Brock : Biology of Microorganisms. Madigan M.T., Martinko J.M. & Parker J. (eds). Prentice Hall College Div ; 8th Package edition
- Madsen E.L.** (2005) Identifying microorganisms responsible for ecologically significant biogeochemical processes. *Nat Rev Microbiol* **3** : 439–446
- Majernik A., Gottschalk G. & Daniel R.** (2001) Screening of environmental DNA libraries for the presence of genes conferring (Na⁺)(Li⁺)/H⁺ antiporter activity on *Escherichia coli* : characterization of the recovered genes and the corresponding gene products. *J Bacteriol* **183** : 6645–6653
- Man D., Wang W., Sabehi G., Aravind L., Post A.F., Massana R. et al.** (2003) Diversification and spectral tuning in marine proteorhodopsins. *Embo J* **22** : 1725–1731
- Manefield M., Whiteley A.S., Griffiths R.I. & Bailey M.J.** (2002a) RNA stable isotope probing, a novel means of linking microbial community function to phylogeny. *Appl Environ Microbiol* **68** : 5367–73
- Manefield M., Whiteley A.S., Ostle N., Ineson P. & Bailey M.J.** (2002b) Technical considerations for RNA-based stable isotope probing : an approach to associating microbial diversity with microbial community function. *Rapid Commun Mass Spectrom* **16** : 2179–83
- Marteinsson V., Birrien J.L., Kristjansson J.K. & Prieur D.** (1995) First isolation of thermophilic aerobic non-sporulating heterotrophic bacteria from deep-sea hydrothermal vents. *FEMS Microbiology Ecology* **18** : 163–174
- Marteinsson V.T., Birrien J.L., Jeanthon C. & Prieur D.** (1996) Numerical taxonomic study of thermophilic *Bacillus* isolated from three geographically separated deep-sea hydrothermal. *FEMS Microbiol Ecol* **21** : 255–266
- Médigue C., Bocs S., Labarre L., Mathe C. & Vallenet D.** (2002) Bio-informatique (1) : L'annotation *in silico* des séquences génomiques. *Ms Med Sci* **18** : 237–250
- Meyerdierks A., Kube M., Lombardot T., Knittel K., Bauer M., Glockner F.O. et al.** (2005) Insights into the genomes of archaea mediating the anaerobic oxidation of methane. *Environ Microbiol* **7** : 1937–1951
- Millikan D.S., Felbeck H. & Stein J.L.** (1999) Identification and characterization of a flagellin gene from the endosymbiont of the hydrothermal vent tubeworm *Riftia pachyptila*. *Appl Environ Microbiol* **65** : 3129–3133
- Miroshnichenko M.L., Kostrikina N.A., Chernyh N.A., Pimenov N.V., Tourova T.P., Antipov A.N. et al.** (2003) *Caldithrix abyssi* gen. nov., sp. nov., a nitrate-reducing, thermophilic, anaerobic bacterium isolated from a Mid-Atlantic Ridge hydrothermal vent, represents a novel bacterial lineage. *Int J Syst Evol Microbiol* **53** : 323–9
- Miroshnichenko M.L., Kostrikina N.A., L'Haridon S., Jeanthon C., Hippe H., Stackebrandt E. et al.** (2002) *Nautilia lithotrophica* gen. nov., sp. nov., a thermophilic sulfur-reducing *epsilon*-proteobacterium isolated from a deep-sea hydrothermal vent. *Int J Syst Evol Microbiol* **52** : 1299–1304

- Miroshnichenko M.L., L'Haridon S., Schumann P., Spring S., Bonch-Osmolovskaya E.A., Jeanthon C. *et al.* (2004) *Caminibacter profundus* sp. nov., a novel thermophile of *Nautiliales* ord. nov. within the class "*Epsilon-proteobacteria*", isolated from a deep-sea hydrothermal vent. *Int J Syst Evol Microbiol* **54** : 41–45
- Moreira D., Rodriguez-Valera F. & Lopez-Garcia P. (2004) Analysis of a genome fragment of a deep-sea uncultivated group II euryarchaeote containing 16S rDNA, a spectinomycin-like operon and several energy metabolism genes. *Environ Microbiol* **6** : 959–969
- Moyer C.L., Dobbs F.C. & Karl D.M. (1994) Estimation of diversity and community structure through restriction fragment length polymorphism distribution analysis of bacterial 16S rRNA genes from a microbial mat at an active, hydrothermal vent system, Loihi Seamount, Hawaii. *Appl Environ Microbiol* **60** : 871–9
- Moyer C.L., Dobbs F.C. & Karl D.M. (1995) Phylogenetic diversity of the bacterial community from a microbial mat at an active, hydrothermal vent system, Loihi Seamount, Hawaii. *Appl Environ Microbiol* **61** : 1555–1562
- Moyer C.L., Tiedje J.M., Dobbs F.C. & Karl D.M. (1998) Diversity of deep-sea hydrothermal vent Archaea from Loihi Seamount, Hawaii. *Deep Sea Research Part II : Topical Studies in Oceanography* **45** : 303–317
- Nakagawa S., Inagaki F., Takai K., Horikoshi K. & Sako Y. (2005a) *Thioreductor micantisoli* gen. nov., sp. nov., a novel mesophilic, sulfur-reducing chemolithoautotroph within the *ε-Proteobacteria* isolated from hydrothermal sediments in the Mid-Okinawa Trough. *Int J Syst Evol Microbiol* **55** : 599–605
- Nakagawa S., Takai K., Inagaki F., Chiba H., Ishibashi J., Kataoka S. *et al.* (2005b) Variability in microbial community and venting chemistry in a sediment-hosted backarc hydrothermal system : Impacts of seafloor phase-separation. *FEMS Microbiol Ecol* **54** : 141–55
- Nakagawa S., Takai K., Inagaki F., Hirayama H., Nunoura T., Horikoshi K. *et al.* (2005c) Distribution, phylogenetic diversity and physiological characteristics of *epsilon-Proteobacteria* in a deep-sea hydrothermal field. *Environ Microbiol* **7** : 1619–1632
- Nakagawa S., Takai K., Inagaki F., Horikoshi K. & Sako Y. (2005d) *Nitratiruptor tergarucus* gen. nov., sp. nov. and *Nitratifactor salsuginis* gen. nov., sp. nov., nitrate-reducing chemolithoautotrophs of the *ε-Proteobacteria* isolated from a deep-sea hydrothermal system in the Mid-Okinawa Trough. *Int J Syst Evol Microbiol* **55** : 925–933
- Nakagawa T., Takai K., Suzuki Y., Hirayama H., Konno U., Tsunogai U. *et al.* (2006) Geomicrobiological exploration and characterization of a novel deep-sea hydrothermal system at the TOTO caldera in the Mariana Volcanic Arc. *Environ Microbiol* **8** : 37–49
- Nelson D.C., Wirsén C.O. & Jannasch H.W. (1989) Characterization of large, autotrophic *Beggiatoa* spp. abundant at hydrothermal vents of the Guaymas Basin. *Appl Environ Microbiol* **55** : 2909–2917
- Nelson K.E. (2003) The future of microbial genomics. *Environ Microbiol* **5** : 1223–1225
- Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J., Haft D.H. *et al.* (1999) Evidence for lateral gene transfer between *Archaea* and *Bacteria* from genome sequence of *Thermotoga maritima*. *Nature* **399** : 323–329
- Nercessian O. (2003) Etude de la diversité spécifique et fonctionnelle des communautés microbiennes associées à des sources hydrothermales marines profondes par des approches moléculaires. Ph.D. thesis, Thèse de l'Université de Bretagne Occidentale
- Nercessian O., Bienvenu N., Moreira D., Prieur D. & Jeanthon C. (2005a) Diversity of functional genes of methanogens, methanotrophs and sulfate reducers in deep-sea hydrothermal environments. *Environ Microbiol* **7** : 118–32
- Nercessian O., Fouquet Y., Pierre C., Prieur D. & Jeanthon C. (2005b) Diversity of *Bacteria* and *Archaea* associated with a carbonate-rich metalliferous sediment sample from the Rainbow vent field on the Mid-Atlantic Ridge. *Environ Microbiol* **7** : 698–714
- Nercessian O., Prokofeva M., Lebedinski A., L'Haridon S., Cary C., Prieur D. *et al.* (2004) Design of 16S rRNA-targeted oligonucleotide probes for detecting cultured and uncultured archaeal lineages in high-temperature environments. *Environ Microbiol* **6** : 170–182
- Nercessian O., Reysenbach A.L., Prieur D. & Jeanthon C. (2003) Archaeal diversity associated with *in situ* samplers deployed on hydrothermal vents on the East Pacific Rise (13 degrees N). *Environ Microbiol* **5** : 492–502
- Newman J.R. & Fuqua C. (1999) Broad-host-range expression vectors that carry the l-arabinose-inducible *Escherichia coli* *araBAD* promoter and the *araC* regulator. *Gene* **227** : 197–203
- Noonan J.P., Hofreiter M., Smith D., Priest J.R., Rohland N., Rabeder G. *et al.* (2005) Genomic sequencing of Pleistocene cave bears. *Science* **309** : 597–9

- Ochsenreiter T., Selezi D., Quaiser A., Bonch-Osmolovskaya L. & Schleper C. (2003) Diversity and abundance of *Crenarchaeota* in terrestrial habitats studied by 16S RNA surveys and real time PCR. *Environ Microbiol* **5** : 787–797
- Ogram A., Sayler G.S. & Barkay T. (1987) The extraction and purification of microbial DNA from sediments. *J Microbiol Methods* **7** : 57–66
- Olsen G.J., Lane D.J., Giovannoni S.J., Pace N.R. & Stahl D.A. (1986) Microbial ecology and evolution : a ribosomal RNA approach. *Annu Rev Microbiol* **40** : 337–365
- Orphan V.J., House C.H., Hinrichs K.U., McKeegan K.D. & DeLong E.F. (2002) Multiple archaeal groups mediate methane oxidation in anoxic cold seep sediments. *Proc Natl Acad Sci USA* **99** : 7663–7668
- Pace N.R. (1997) A molecular view of microbial diversity and the biosphere. *Science* **276** : 734–740
- Pace N.R., Stahl D.A., Lane D.J. & Olsen G.J. (1985) Analyzing natural microbial populations by rRNA sequences. *ASM news* **51** : 4–12
- Pearson A., McNichol A.P., Benitez-Nelson B.C., Hayes J.M. & Eglinton T.I. (2001) Origins of lipid biomarkers in Santa Monica Basin surface sediment : a case study using compound-specific D¹⁴C analysis. *Geochim Cosmochim Acta* **65** : 3123–3137
- Pfeifer B.A., Admiraal S.J., Gramajo H., Cane D.E. & Khosla C. (2001) Biosynthesis of complex polyketides in a metabolically engineered strain of *E. coli*. *Science* **291** : 1790–1792
- Philippe H. & Germot A. (2000) Phylogeny of eukaryotes based on ribosomal RNA : long-branch attraction and models of sequence evolution. *Mol Biol Evol* **17** : 830–4
- Piel J., Hui D., Wen G., Butzke D., Platzer M., Fusetani N. *et al.* (2004) Antitumor polyketide biosynthesis by an uncultivated bacterial symbiont of the marine sponge *Theonella swinhoei*. *Proc Natl Acad Sci USA* **101** : 16222–16227
- Polz M.F. & Cavanaugh C.M. (1995) Dominance of one bacterial phylotype at a Mid-Atlantic Ridge hydrothermal vent site. *Proc Natl Acad Sci USA* **92** : 7232–7236
- Preston C.M., Wu K.Y., Molinski T.F. & DeLong E.F. (1996) A psychrophilic crenarchaeon inhabits a marine sponge : *Cenarchaeum symbiosum* gen. nov., sp. nov. *Proc Natl Acad Sci USA* **93** : 6241–6
- Prieur D., Chamroux S., Durand P., Erauso G., Fera P., Jeanthon C. *et al.* (1990) Metabolic diversity in epibiotic microflora associated with the Pompeii worms *Alvinella pompejana* and *A. caudata* (polychaete : Annelida) from deep-sea hydrothermal vents. *Mar Biol* **106** : 361–367
- Prieur D., Erauso G. & Jeanthon C. (1995) Hyperthermophilic life at deep-sea hydrothermal vents. *Planet Space Sci* **43** : 115–121
- Prieur D. & Jeanthon C. (1987) Preliminary study of heterotrophic bacteria isolated from two deep-sea hydrothermal vent invertebrates : *Alvinella pompejana* (polychaete) and *Bathymodiolus thermophilus* (bivalve). *Symbiosis* **4** : 87–98
- Pruzzo C., del Mar Lleo M., Signoretto C., Zampini M., Pane L., Colwell R.R. *et al.* (2003) Persistence of adhesive properties in *Vibrio cholerae* after long-term exposure to sea water. *Environ Microbiol* **5** : 850–858
- Quaiser A., Ochsenreiter T., Klenk H.P., Kletzin A., Treusch A.H., Meurer G. *et al.* (2002) First insight into the genome of an uncultivated crenarchaeote from soil. *Environ Microbiol* **4** : 603–611
- Quaiser A., Ochsenreiter T., Lanz C., Schuster S.C., Treusch A.H., Eck J. *et al.* (2003) *Acidobacteria* form a coherent but highly diverse group within the bacterial domain : evidence from environmental genomics. *Mol Microbiol* **50** : 563–575
- Radajewski S., Webster G., Reay D.S., Morris S.A., Ineson P., Nedwell D.B. *et al.* (2002) Identification of active methylophilic populations in an acidic forest soil by stable-isotope probing. *Microbiology* **148** : 2331–2342
- Rappe M.S. & Giovannoni S.J. (2003) The uncultured microbial majority. *Annu Rev Microbiol* **57** : 369–394
- Rees H.C., Grant S., Jones B., Grant W.D. & Heaphy S. (2003) Detecting cellulase and esterase enzyme activities encoded by novel genes present in environmental DNA libraries. *Extremophiles* **7** : 415–421
- Reysenbach A.L., Liu Y., Banta A.B., Beveridge T.J., Kirshtein J.D., Schouten S. *et al.* (2006) A ubiquitous thermoacidophilic archaeon from deep-sea hydrothermal vents. *Nature* **442** : 444–447
- Reysenbach A.L., Longnecker K. & Kirshtein J. (2000) Novel bacterial and archaeal lineages from an *in situ* growth chamber deployed at a Mid-Atlantic Ridge hydrothermal vent. *Appl Environ Microbiol* **66** : 3798–3806
- Reysenbach A.L. & Shock E. (2002) Merging genomes with geochemistry in hydrothermal ecosystems. *Science* **296** : 1077–1082

- Richardson T.H., Tan X., Frey G., Callen W., Cabell M., Lam D. et al.** (2002) A novel, high performance enzyme for starch liquefaction. discovery and optimization of a low ph, thermostable alpha-amylase. *J Biol Chem* **277** : 26501–26507
- Riesenfeld C.S., Schloss P.D. & Handelsman J.** (2004) Metagenomics : genomic analysis of microbial communities. *Annu Rev Genet* **38** : 525–552
- Robinson J.J. & Cavanaugh C.M.** (1995) Expression of form I and form II ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubsico) in chemoautotrophic symbioses : implications for the interpretation of stable carbon isotope ratios. *Limnol Oceanogr* **40** : 1496–1502
- Robinson J.J., Stein J.L. & Cavanaugh C.M.** (1998) Cloning and sequencing of a form II ribulose-1,5-biphosphate carboxylase/oxygenase from the bacterial symbiont of the hydrothermal vent tubeworm *Riftia pachyptila*. *J Bacteriol* **180** : 1596–1599
- Rondon M.R., August P.R., Bettermann A.D., Brady S.F., Grossman T.H., Liles M.R. et al.** (2000) Cloning the soil metagenome : a strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Appl Environ Microbiol* **66** : 2541–2547
- Rondon M.R., Goodman R.M. & Handelsman J.** (1999a) The Earth's bounty : assessing and accessing soil microbial diversity. *Trends Biotechnol* **17** : 403–409
- Rondon M.R., Raffel S.J., Goodman R.M. & Handelsman J.** (1999b) Toward functional genomics in bacteria : Analysis of gene expression in *Escherichia coli* from a bacterial artificial chromosome library of *Bacillus cereus*. *Proc Natl Acad Sci USA* **96** : 6451–6455
- Ruby E.G., Wirsen C.O. & Jannasch H.W.** (1981) Chemolithotrophic sulfur-oxidizing bacteria from the Galapagos Rift hydrothermal vents. *Appl Environ Microbiol* **42** : 317–324
- Rutherford K., Parkhill J., Crook J., Horsnell T., Rice P., Rajandream M.A. et al.** (2000) Artemis : sequence visualization and annotation. *Bioinformatics* **16** : 944–945
- Sabehi G., Massana R., Bielawski J.P., Rosenberg M., DeLong E.F. & B ej a O.** (2003) Novel proteorhodopsin variants from the Mediterranean and Red seas. *Environ Microbiol* **5** : 842–8499
- Sambrook J., Fritsch E.F. & Maniatis T.** (1989) Molecular Cloning : a Laboratory Manual. Cold Spring Harbor Laboratory Press, 2nd ed., Plainview, New York
- Sandaa R.A., Enger O. & Torsvik V.** (1999) Abundance and diversity of *Archaea* in heavy-metal-contaminated soils. *Appl Environ Microbiol* **65** : 3293–3297
- Sandler S.J., Hugenholtz P., Schleper C., DeLong E.F., Pace N.R. & Clark A.J.** (1999) Diversity of *radA* genes from cultured and uncultured archaea : comparative analysis of putative RadA proteins and their use as a phylogenetic marker. *J Bacteriol* **181** : 907–915
- Sanger F., Air G.M., Barrell B.G., Brown N.L., Coulson A.R., Fiddes C.A. et al.** (1977) Nucleotide sequence of bacteriophage phi X174 DNA. *Nature* **265** : 687–695
- Santangelo T.J. & Reeve J.N.** (2006) Archaeal RNA polymerase is sensitive to intrinsic termination directed by transcribed and remote sequences. *J Mol Biol* **355** : 196–210
- Schleper C.** (2004) Population genomics of soil microbial communities. *104th General meeting of the American Society for Microbiology, New Orleans, Louisiana*
- Schleper C., DeLong E.F., Preston C.M., Feldman R.A., Wu K.Y. & Swanson R.V.** (1998) Genomic analysis reveals chromosomal variation in natural populations of the uncultured psychrophilic archaeon *Cenarchaeum symbiosum*. *J Bacteriol* **180** : 5003–5009
- Schleper C., Holben W. & Klenk H.P.** (1997a) Recovery of crenarchaeotal ribosomal DNA sequences from freshwater-lake sediments. *Appl Environ Microbiol* **63** : 321–323
- Schleper C., Jurgens G. & Jonscheit M.** (2005) Genomic studies of uncultivated *Archaea*. *Nat Rev Microbiol* **3** : 479–488
- Schleper C., Swanson R.V., Mathur E.J. & DeLong E.F.** (1997b) Characterization of a DNA polymerase from the uncultivated psychrophilic archaeon *Cenarchaeum symbiosum*. *J Bacteriol* **179** : 7803–7811
- Schloss P.D. & Handelsman J.** (2003) Biotechnological prospects from metagenomics. *Curr Opin Biotechnol* **14** : 303–310
- Schmid M., Twachtmann U., Klein M., Strous M., Juretschko S., Jetten M. et al.** (2000) Molecular evidence for genus level diversity of *Bacteria* capable of catalyzing anaerobic ammonium oxidation. *Syst Appl Microbiol* **23** : 93–106
- Schmidt T.M., DeLong E.F. & Pace N.R.** (1991) Analysis of a marine picoplankton community by 16s rRNA gene cloning and sequencing. *J bacteriol* **173** : 4371–4378
- Schrenk M.O., Kelley D.S., Delaney J.R. & Baross J.A.** (2003) Incidence and diversity of microorganisms within the walls of an active deep-sea sulfide chimney. *Appl Environ Microbiol* **69** : 3580–3592

- Schwecke T., Aparicio J.F., Molnar I., Konig A., Khaw L.E., Haydock S.F. *et al.* (1995) The biosynthetic gene cluster for the polyketide immunosuppressant rapamycin. *Proc Natl Acad Sci USA* **92** : 7839–7843
- Segonzac M. (1992) Les peuplements associés à l'hydrothermalisme océanique du Snake Pit (dorsale medio-Atlantique; 23°N, 3480 m) : composition et microdistribution de la mégafaune. *CR Acad Sci paris* **314** : 593–600
- Shizuya H., Birren B., Kim U., Mancino V., Slepak T., Tachiiri Y. *et al.* (1992) Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proc Natl Acad Sci USA* **89** : 8794–8797
- Sievert S., Wieringa E., Wirsén C. & Taylor C. (2006) Growth and mechanism of filamentous-sulfur formation by *Candidatus Arcobacter sulfidicus* in opposing oxygen-sulfide gradients. *Environ Microbiol* : doi :10.1111/j.1462-2920.2006.01156.x
- Sievert S.M., Brinkhoff T., Muyzer G., Ziebis W. & Kuever J. (1999) Spatial heterogeneity of bacterial populations along an environmental gradient at a shallow submarine hydrothermal vent near Milos Island (Greece). *Appl Environ Microbiol* **65** : 3834–2842
- Sievert S.M. & Kuever J. (2000) *Desulfacinum hydrothermale* sp. nov., a thermophilic, sulfate-reducing bacterium from geothermally heated sediments near Milos Island (Greece). *Int J Syst Evol Microbiol* **50** : 1239–46
- Simon H.M., Dodsworth J.A. & Goodman R.M. (2000) *Crenarchaeota* colonize terrestrial plant roots. *Environ Microbiol* **2** : 495–505
- Slobodkin A., Campbell B., Cary S.C., Bonch-Osmolovskaya E. & Jeannot C. (2001) Evidence for the presence of thermophilic Fe(III)-reducing microorganisms in deep-sea hydrothermal vents at 13 degrees N (East Pacific Rise). *FEMS Microbiol Ecol* **36** : 235–243
- Snaird J., Amann R., Huber I., Ludwig W. & Schleifer K.H. (1997) Phylogenetic analysis and *in situ* identification of bacteria in activated sludge. *Appl Environ Microbiol* **63** : 2884–2896
- Somerville C.C., Knight I.T., Straube W.L. & Colwell R.R. (1989) Simple, rapid method for direct isolation of nucleic acids from aquatic environments. *Appl Environ Microbiol* **55** : 548–554
- Stach J.E.M. & Burns R.G. (2002) Enrichment versus biofilm culture : a functional and phylogenetic comparison of polycyclic aromatic hydrocarbon-degrading microbial communities. *Environ Microbiol* **4** : 169–182
- Steffan R.J., Goksoyr J., Bej A.K. & Atlas R.M. (1988) Recovery of DNA from soils and sediments. *Appl Environ Microbiol* **54** : 2908–2915
- Stein J.L., Marsh T.L., Wu K.Y., Shizuya H. & DeLong E.F. (1996) Characterization of uncultivated prokaryotes : isolation and analysis of a 40-kilobase-pair genome fragment from a planktonic marine archaeon. *J Bacteriol* **178** : 591–599
- Steinberg D.K., Carlson C.A., Bates N.R., Johnson R.J., Michaels A.F. & Knap A.F. (2001) Overview of the U.S. JGOFS Bermuda Atlantic Time-series Study (BATS) : A decade-scale look at ocean biology and biogeochemistry. *Deep-Sea Research II* **48** : 1405–1447
- Straub K.L. & Buchholz-Cleven B.E. (2001) *Geobacter bremensis* sp. nov. and *Geobacter pelophilus* sp. nov., two dissimilatory ferric-iron-reducing bacteria. *Int J Syst Evol Microbiol* **51** : 1805–8
- Strous M. & Jetten M.S.M. (2004) Anaerobic oxidation of methane and ammonium. *Annual Review of Microbiology* **58** : 99–117
- Strous M., Pelletier E., Mangenot S., Rattei T., Lehner A., Taylor M.W. *et al.* (2006) Deciphering the evolution and metabolism of an anammox bacterium from a community genome. *Nature* **440** : 790–4
- Takai K., Campbell B.J., Cary S.C., Suzuki M., Oida H., Nunoura T. *et al.* (2005a) Enzymatic and genetic characterization of carbon and energy metabolisms by deep-sea hydrothermal chemolithoautotrophic isolates of *Epsilonproteobacteria*. *Appl Environ Microbiol* **71** : 7310–7320
- Takai K., Gamo T., Tsunogai U., Nakayama N., Hirayama H., Nealson K.H. *et al.* (2004a) Geochemical and microbiological evidence for a hydrogen-based, hyperthermophilic subsurface lithoautotrophic microbial ecosystem (HyperSLiME) beneath an active deep-sea hydrothermal field. *Extremophiles* **8** : 269–282
- Takai K., Hirayama H., Nakagawa T., Suzuki Y., Nealson K.H. & Horikoshi K. (2005b) *Lebetimonas acidiphila* gen. nov., sp. nov., a novel thermophilic, acidophilic, hydrogen-oxidizing chemolithoautotroph within the 'Epsilonproteobacteria', isolated from a deep-sea hydrothermal fumarole in the Mariana Arc. *Int J Syst Evol Microbiol* **55** : 183–189
- Takai K. & Horikoshi K. (1999) Genetic diversity of *Archaea* in deep-sea hydrothermal vent environments. *Genetics* **152** : 1285–1297

- Takai K., Inagaki F., Nakagawa S., Hirayama H., Nunoura T., Sako Y. et al.** (2003) Isolation and phylogenetic diversity of members of previously uncultivated ϵ -*Proteobacteria* in deep-sea hydrothermal fields. *FEMS Microbiol Lett* **218** : 167–174
- Takai K., Komatsu T., Inagaki F. & Horikoshi K.** (2001a) Distribution of *Archaea* in a black smoker chimney structure. *Appl Environ Microbiol* **67** : 3618–3629
- Takai K., Moser D.P., DeFlaun M., Onstott T.C. & Fredrickson J.K.** (2001b) Archaeal diversity in waters from deep south african gold mines. *Appl Environ Microbiol* **67** : 5750–5760
- Takai K., Nealson K.H. & Horikoshi K.** (2004b) *Hydrogenimonas thermophila* gen. nov., sp. nov., a novel thermophilic, hydrogen-oxidizing chemolithoautotroph within the ϵ -*Proteobacteria*, isolated from a black smoker in a Central Indian Ridge hydrothermal field. *Int J Syst Evol Microbiol* **54** : 25–32
- Tang L., Shah S., Chung L., Carney J., Katz L., Khosla C. et al.** (2000) Cloning and heterologous expression of the epothilone gene cluster. *Science* **287** : 640–642
- Taylor C.D. & Wirsen C.O.** (1997) Microbiology and ecology of filamentous sulfur formation. *Science* **277** : 1483–1485
- Taylor C.D., Wirsen C.O. & Gaill F.** (1999) Rapid microbial production of filamentous sulfur mats at hydrothermal vents. *Appl Environ Microbiol* **65** : 2253–2255
- Tebbe C.C. & Vahjen W.** (1993) Interference of humic acids and DNA extracted directly from soil in detection and transformation of recombinant DNA from bacteria and a yeast. *Appl Environ Microbiol* **59** : 2657–2665
- Teeling H., Meyerdierks A., Bauer M., Amann R. & Glockner F.O.** (2004a) Application of tetranucleotide frequencies for the assignment of genomic fragments. *Environ Microbiol* **6** : 938–947
- Teeling H., Waldmann J., Lombardot T., Bauer M. & Glockner F.O.** (2004b) TETRA : a web-service and a stand-alone program for the analysis and comparison of tetranucleotide usage patterns in DNA sequences. *BMC Bioinformatics* **5** : 163
- Teske A., Hinrichs K.U., Edgcomb V., de Vera Gomez A., Kysela D., Sylva S.P. et al.** (2002) Microbial diversity of hydrothermal sediments in the Guaymas Basin : evidence for anaerobic methanotrophic communities. *Appl Environ Microbiol* **68** : 1994–2007
- Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D., Peterson S. et al.** (2001) Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*. *Science* **293** : 498–506
- Torarinsson E., Klenk H.P. & Garrett R.A.** (2005) Divergent transcriptional and translational signals in *Archaea*. *Environ Microbiol* **7** : 47–54
- Torsvik V. & Ovreas L.** (2002) Microbial diversity and function in soil : from genes to ecosystems. *Curr Opin Microbiol* **5** : 240–245
- Treusch A.H., Kletzin A., Raddatz G., Ochsenreiter T., Quaiser A., Meurer G. et al.** (2004) Characterization of large-insert DNA libraries from soil for environmental genomic studies of *Archaea*. *Environ Microbiol* **6** : 970–980
- Tringe S.G. & Rubin E.M.** (2005) Metagenomics : DNA sequencing of environmental samples. *Nature Rev Genet* **6** : 805–814
- Tringe S.G., von Mering C., Kobayashi A., Salamov A.A., Chen K., Chang H.W. et al.** (2005) Comparative metagenomics of microbial communities. *Science* **308** : 554–557
- Tsai Y.L. & Olson B.H.** (1992) Detection of low numbers of bacterial cells in soils and sediments by polymerase chain reaction. *Appl Environ Microbiol* **58** : 754–757
- Tyson G.W. & Banfield J.F.** (2005) Cultivating the uncultivated : a community genomics perspective. *Trends Microbiol* **13** : 411–415
- Tyson G.W., Chapman J., Hugenholtz P., Allen E.E., Ram R.J., Richardson P.M. et al.** (2004) Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **428** : 37–43
- Van Dover C.L. & Fry B.** (1989) Stable isotopic compositions of hydrothermal vent organisms. *Mar Biol* **102** : 257–263
- Van Dover C.L., Humphris S.E., Fornari D., Cavanaugh C.M., Collier R., Goffredi S.K. et al.** (2001) Biogeography and ecological setting of Indian Ocean hydrothermal vents. *Science* **294** : 818–23
- Van Dover C.L., Szuts E.Z., Chamberlain S.C. & Cann J.R.** (1989) A novel eye in 'eyeless' shrimp from hydrothermal vents of the Mid-Atlantic Ridge. *Nature* **337** : 458–460
- Venter J.C., Remington K., Heidelberg J.F., Halpern A.L., Rusch D., Eisen J.A. et al.** (2004) Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304** : 66–74

- Vicente M.F., Basilio A., Cabello A. & Pelaez F. (2003) Microbial natural products as a source of antifungals. *Clin Microbiol Infect* **9** : 15–32
- Voget S., Leggewie C., Uesbeck A., Raasch C., Jaeger K.E. & Streit W.R. (2003) Prospecting for novel biocatalysts in a soil metagenome. *Appl Environ Microbiol* **69** : 6235–6242
- Von Damm K.L. (2001) Lost City found. *Nature* **412** : 127–128
- von Wintzingerode F., Gobel U.B. & Stackebrandt E. (1997) Determination of microbial diversity in environmental samples : pitfalls of PCR-based rRNA analysis. *FEMS Microbiol Rev* **21** : 213–229
- Voordeckers J.W., Starovoytov V. & Vetriani C. (2005) *Caminibacter mediatlanticus* sp. nov., a thermophilic, chemolithoautotrophic, nitrate-ammonifying bacterium isolated from a deep-sea hydrothermal vent on the Mid-Atlantic Ridge. *Int J Syst Evol Microbiol* **55** : 773–9
- Waksman S.A. & Curtis R.E. (1916) The actinomycetes of soil. *Soil Sci* **1** : 99–134
- Wang G.Y., Graziani E., Waters B., Pan W., Li X., McDermott J. *et al.* (2000) Novel natural products from soil DNA libraries in a streptomycete host. *Org Lett* **2** : 2401–2404
- Washio T., Sasayama J. & Tomita M. (1998) Analysis of complete genomes suggests that many prokaryotes do not rely on hairpin formation in transcription termination. *Nucleic Acids Res* **26** : 5456–63
- Waters C.M. & Bassler B.L. (2005) Quorum sensing : cell-to-cell communication in *Bacteria*. *Annual Review of Cell and Developmental Biology* **21** : 319–346
- Werner E., Holder A.A. & Hoheisel J.D. (1997) Growth and storage of YAC clones in Hogness freezing medium. *Nucleic Acids Res* **25** : 1467–1468
- Whitehead N.A., Barnard A.M., Slater H., Simpson N.J. & Salmond G.P. (2001) Quorum-sensing in gram-negative bacteria. *FEMS Microbiol Rev* **25** : 365–404
- Whitman W.B., Boone D.R. & Koga Y. (2001) *Methanocaldococaceae*. In *Bergey's Manual of Systematic Bacteriology*, Boone D.R., Castenholz R.W. & Garrity G. (eds), vol. 1, pp 242–246. 2nd edn. Springer-Verlag, New York
- Whitman W.B., Coleman D.C. & Wiebe W.J. (1998) Prokaryotes : The unseen majority. *Proc Natl Acad Sci USA* **95** : 6578–6583
- Wirsen C.O., Jannasch H.W. & Molyneaux S.J. (1993) Chemosynthetic microbial activity at Mid-Atlantic Ridge hydrothermal vent sites. *J Geophys* **98** : 9693–9703
- Wirsen C.O., Sievert S.M., Cavanaugh C.M., Molyneaux S.J., Ahmad A., Taylor L.T. *et al.* (2002) Characterization of an autotrophic sulfide-oxidizing marine *Archaeobacter* sp. that produces filamentous sulfur. *Appl Environ Microbiol* **68** : 316–325
- Woese C.R. (1987) Bacterial evolution. *Microbiol Rev* **51** : 221–271
- Woese C.R. & Fox G.E. (1977) Phylogenetic structure of the prokaryotic domain : the primary kingdoms. *Proc Natl Acad Sci USA* **74** : 5088–5090
- Woese C.R., Kandler O. & Wheelis M.L. (1990) Towards a natural system of organisms : proposal for the domains *Archaea*, *Bacteria*, and *Eucarya*. *Proc Natl Acad Sci USA* **87** : 4576–4579
- Yin B., Crowley D., Sparovek G., De Melo W.J. & Borneman J. (2000) Bacterial functional redundancy along a soil reclamation gradient. *Appl Environ Microbiol* **66** : 4361–5
- Yurkov V.V., Krieger S., Stackebrandt E. & Beatty J.T. (1999) *Citromicrobium bathyomarinum*, a novel aerobic bacterium isolated from deep-sea hydrothermal vent plume waters that contains photosynthetic pigment-protein complexes. *J Bacteriol* **181** : 4517–4525
- Zbinden M. & Cambon-Bonavita M.A. (2003) Occurrence of *Deferribacterales* and *Entomoplasmatales* in the deep-sea *Alvinocarid* shrimp *Rimicaris exoculata* gut. *FEMS Microbiol Ecol* **46** : 23–30
- Zbinden M., Le Bris N., Compere P., Martinez I., Guyot F. & Gaill F. (2003) Mineralogical gradients associated with alvinellids at deep-sea hydrothermal vents. *Deep Sea Research Part I : Oceanographic Research Papers* **50** : 269–280
- Zehnder A.J. & Brock T.D. (1979) Methane formation and methane oxidation by methanogenic *Bacteria*. *J Bacteriol* **137** : 430–432
- Zengler K., Toledo G., Rappe M., Elkins J., Mathur E.J., Short J.M. *et al.* (2002) Cultivating the uncultured. *Proc Natl Acad Sci USA* **99** : 15681–15686
- Zhao Y., Jeruzalmi D., Moarefi I., Leighton L., Lasken R. & Kuriyan J. (1999) Crystal structure of an archaeobacterial DNA polymerase. *Structure* **7** : 1189–99

Zillig W. & Reysenbach A.L. (2001) *Thermococcales*. In *Bergey's Manual of Systematic Bacteriology*, Boone D.R., Castenholz R.W. & Garrity G. (eds), vol. 1. pp 341-348. 2nd edn. Springer-Verlag, New York

Zuckerkandl E. & Pauling L. (1965) Molecules as documents of evolutionary history. *J Theor Biol* **8** : 357-366

Résumé

Analyses moléculaires de la diversité et des fonctions de micro-organismes incultivés des sources hydrothermales profondes

Nos connaissances de la diversité des communautés microbiennes marines ont longtemps été restreintes aux informations précieuses, mais incomplètes, issues de l'utilisation de méthodes de culture.

Lors de cette étude, des techniques moléculaires (PCR, clonage, séquençage, hybridation, construction d'une banque métagénomique) et des marqueurs génétiques [ARNr 16S, gènes codants pour des enzymes spécifiques de la voie inverse du cycle tricarboxylique (*aclB*, *oorA*)] ont été utilisés pour contourner les limites inhérentes aux méthodes de culture et accéder à la diversité spécifique ainsi qu'aux fonctions des micro-organismes incultivés des sources hydrothermales profondes.

Ce travail de recherche a permis :

- de confirmer l'importance écologique des ε -*Proteobacteria* dans l'écosystème hydrothermal profond, notamment dans les processus de colonisation de surfaces vierges exposées aux fluides hydrothermaux,
- de mettre au point une sonde oligonucléotidique spécifique des ε -*Proteobacteria* permettant de les quantifier et de les observer dans un échantillon environnemental,
- de montrer la présence de bactéries *Arcobacter-like* vraisemblablement sulfo-oxydantes et qui seraient impliquées dans la formation d'un biofilm filamenteux contenant essentiellement du soufre,
- et d'obtenir via l'approche métagénomique, les premiers fragments génomiques d'*Archaea* incultivées hydrothermales.

Grâce à l'utilisation combinée de différentes approches moléculaires, ce travail a permis d'apporter un nouvel éclairage sur la diversité des communautés microbiennes de l'écosystème hydrothermal marin profond.

Mots clefs : analyse moléculaire de la diversité microbienne, *Archaea*, *Bacteria*, ε -*Proteobacteria*, DHVE, ARNr 16S, *aclB*, *oorA*, métagénomique, génomique environnementale, sources hydrothermales marines, FISH, biofilm.